Edge Hill University

The impact of arts for health on the health, wellbeing and quality of life of older adults in care homes: systematic review and national survey.

Amy Rachel Curtis

A thesis submitted in partial fulfilment of the requirements of Edge Hill University for the degree of Doctor of Philosophy.

August 2019

Acknowledgements

This thesis is dedicated first and foremost to my Mum, with a big thank you for all the love and support. Without a doubt it was much easier doing this with you than without and I know how proud you'd be that I've (finally) finished. Secondly to my beautiful princesses Violet and Autumn, for keeping me grounded and providing me with PhD breaks by giving me alternative stresses to think about. To Josh for coming along at just the right time, for all the coffee and pancakes and for putting up with me when the world ended three times a week. And to all my extended family and friends who I am not going to list for fear of inadvertently forgetting someone.

I would like to thank my director of studies, Professor Brenda Roe for both her academic and invaluable personal support. Without her encouragement, support and guidance this would not have been as enjoyable a process as it was. To my supervisory team Professor Mary O'Brien and Dr Lucy Gibson for their help and guidance. I promise I will never ask you to double check a data extraction or quality appraisal form again. To Olly, Greg and Emmie for the support both in and out of the office during this process together.

Finally, I would like to thank all the care homes who took part in the survey and those who participated in the public involvement for the creation of the study.

Abstract

Older adults residing in care homes are vulnerable to physical and psychological stressors that can impact upon their health, wellbeing and quality of life. Despite increased interest in the potential for the use of arts for health there has been no evaluation of current empirical evidence assessing its impact exclusively within the care home population. Furthermore, there has been no mapping of arts for health activities delivered within care homes and associated benefits.

Study 1 consisted of a systematic review that evaluated published empirical research focused on assessing the impact of arts for health activities for older adults residing in care homes. Databases were searched from inception with continual updates until August 2018. A total of 71 studies were eligible for inclusion in the review and these underwent data extraction and quality appraisal with a subsequent descriptive narrative synthesis of all included studies. Studies were classified to form an arts typology which included music, performing arts, literary spoken and written word, multisensory activities, and applied arts and crafts. Following this, a further descriptive narrative synthesis of studies was conducted according to arts type.

Study 2 was a national survey conducted between 2017 and 2018 and evaluated the delivery and potential impact of arts for health activities in care homes. A total of 184 care home managers responded to the survey, with ten follow-up telephone interviews carried out with a self-selecting sub-sample of those completing the survey.

Findings showed existing empirical research evaluating arts for health within the care home population is of varying quality and focused mainly upon the evaluation of music activities. There was evidence of benefit for measures of psychological wellbeing, agitation, cognition,

socialisation and improving the caring process with less support for measures of quality of life or verbally disruptive behaviours. Low baseline levels were reported for depression and behavioural disturbances and high baseline measures of quality of life were reported by both studies, this may have contributed to a lack of observable improvements for these measures. Therefore, future research should focus on the potential of arts for health in the maintenance of health, wellbeing and quality of life. Quantitative findings alone did not show arts activities to be more beneficial than other forms of social activities delivered within care homes. However, qualitative evidence showed findings unique to arts activities which would be difficult to capture quantitatively, such as creativity and self-expression which reinforces the need for further research to utilise mixed methods.

Key Words: health, wellbeing, quality of life, arts for health, older adults, care home, systematic review, national survey, narrative synthesis.

Contents

List of included figures

List of included tables

Appendix

Chapter 1- Introduction and background

1.1 Introduction

This thesis evaluated the impact of arts for health activities on the health, wellbeing and quality of life of older people in care homes. Two studies are included; Study 1 a systematic review of existing empirical evidence evaluating the impact of arts for health activities within the care home population. Study 2 is a national survey of care homes in England assessing the current provision of arts for health activities, with follow-up interviews conducted with care home managers to explore the benefits associated with such activities.

There are eight chapters, Chapter 1 introduces the care home population and an overview of arts for health. Chapter 2 outlines the methods for the systematic review, with the results of an overall high-level synthesis presented in Chapter 3 and a narrative synthesis by arts type in Chapter 4. The methods for the national survey are found in Chapter 5 and the results in Chapter 6. A discussion section synthesising the results from Study 1 and Study 2 is within Chapter 7, along with limitations of both studies. Recommendations for practice, policy and conclusions are shown in Chapter 8.

This chapter (Chapter 1) provides background on the current ageing population and care home demographic. The chapter also discusses the development of arts for health, both policy and evidence, and its role in maintaining the health of older people.

1.2 Ageing Population

There is no universally adopted definition of being 'old', or the age at which individuals begin to be classed as 'older adults'. The United Nations defines older adults as those aged 60 and over (United Nations Population Fund, 2013 p3). Within the developed world, the term 'older adult' is often accepted as those aged 65 and over, parallel to current retirement age (ONS, 2018). However, in the developing world where there is a lower life expectancy, or where

chronological age is not used as a benchmark, lower definitions of 'old age' are more appropriate, often including those aged 55 and over (Yasamy et al 2013).

The proportion of older adults worldwide is increasing at a faster rate than any other age group. The World Health Organisation (WHO) predicts that by the year 2050 the proportion of adults aged 60 and over will have reached 2 billion, accounting for 22% of the total global population, doubling from the present-day figure of 11% (WHO, 2014; Office for National Statistics (ONS), 2018). Furthermore, life expectancy is anticipated to increase with the percentage of older adults over the age of 80 the fastest growing within the older age bracket. Consequently, the number of individuals who will live to see their 80s and 90s is now greater than ever before (WHO 2002; WHO 2014).

The increasing overall age is also reflected within figures for the UK where there are currently 12 million adults over the age of 65 (ONS, 2018). The increased number of older adults will change the dynamic of the UK population such that for the first time by the year 2020 there will be more adults over the age of 65 than children under the age of 18 (Allen, 2008; ONS, 2015a). In line with worldwide population trends the most dramatically increased prevalence within the UK has been shown to be in the 'oldest old'; those aged 85 and over (Allen, 2008; Age UK, 2015; Mortimer & Green, 2016 ONS, 2018).

A number of factors are responsible for the increasingly ageing population, including reduced mortality rates among infants in developing countries, reduced mortality rates among older adults in developed countries (from conditions such as cardiovascular disease and certain cancers) and an overall decline in fertility levels (Ham et al. 2011; WHO, 2015). However, whilst there has been an increase in life expectancy, the rate at which we remain healthy has not increased with the same pace (Allen, 2008; WHO, 2015 Thomson, 2015; Age UK 2015;

Mortimer & Green, 2015; APPG, 2017). Worldwide, health care systems within their current format will struggle to cope with the demand that will be placed on them because of the inevitable increase of age-related problems (WHO, 2015). With this in mind, there is a current drive to encourage active ageing and to maintain physical and psychological health for as long as possible.

As the population ages, the number of older adults requiring assistance with daily living in developed countries will quadruple therefore a greater number of individuals will live within nursing, care homes and assisted living facilities (WHO, 2014; ONS, 2018). Aside from the increase in life expectancy and chronic health conditions, the role of the family and familial structure has also contributed to the increasing numbers of older adults who require assistance with daily living within developed countries. These changes include a decrease in the number of family members staying at home to become carers, increased working age of relatives who would traditionally have taken up the caring role and an increase in geographical distances amongst families (Ham et al. 2011; WHO, 2012; Vella-Burrows, 2015). In developing countries, the proportion of older adults is increasing at a sharper rate than developed countries and there are greater numbers of older people reliant upon care (WHO, 2015). Despite this, older adults who live within these areas of the world are often cared for by their family members and less likely to live within care homes and other long-term care facilities (United Nations Population Fund, 2012; de la Cuesta and Roe, 2015).

Whilst a large number of older adults within the UK will continue to live within their own homes, there are currently 352,000 individuals within England and Wales who are living in nursing or care home facilities and 80% of these are over the age of 65 (ONS, 2014a). 'Care homes' is a term used to refer to long term facilities including nursing homes, residential

homes, dual registered homes (which offer both nursing and social care) and aged care homes and from now on the term 'care home' will be used to refer to all these facilities. Care homes are designed to offer a place where individuals can receive assistance with activities of daily living and receive nursing support and social care if required. The level and type of support offered varies between care homes and based upon need and the type of care home. Care home facilities are operated by local councils, voluntary agencies, or are privately owned and within the UK they must all be registered with the Care Quality Commission (CQC) (Age UK, 2013; NICE 2013).

Disability-free life expectancy has not decreased to a significant level and the prevalence of several chronic and progressive healthcare problems such as Chronic Obstructive Pulmonary Disease (COPD), arthritis, type 2 diabetes and dementia has increased in recent years meaning there are greater numbers of older adults with long term health conditions (Ham et al. 2011; APPG, 2017). Further complicating this, older people often have combinations of different illnesses and disabilities (Flood and Philips, 2007; O'Neill, 2011). This combination of illness and disability can be a factor in an older person moving in to long-term care (Age UK 2015) and consequently there is a high proportion of disability within the care home population. Around 70% of older people within care homes will have dementia and a number of those who do not have a formal dementia diagnosis will present with some level of cognitive impairment (Alzheimer's Society 2014; Thraves, 2016).

Currently long-term care facilities provide the levels of physical care required for older adults, but there is now recognition that there needs to be a more holistic approach to long-term care that maintains not only physical health but also nurtures the psychological and social functioning of older people (New Economics Foundation, 2008; NICE 2013; WHO, 2014; WHO,

2015; ONS, 2018). The traditional medical model is not effective in the management of long-term conditions within older adults, which are often progressive or chronic in nature and therefore unlikely to be impacted upon by medication alone (O'Neill, 2011; Clift & Camic, 2015). This view has been echoed by research carried out within care homes which has found that whilst the physical needs of residents are almost always met, psychosocial needs are often missed which can result in mood disturbances and anxiety (deGuzman et al., 2009; Hancock et al., 2009; Cohen-Mansfield et al. 2015; Ferreira, Dias & Fernandes 2016).

1.3 Inequalities in social care access

Within the UK there is a divide with respect of access to health services. Those within the most deprived areas of the UK have reduced life expectancy and show a greater number of years spent in poor health. Although deprivation is found within different regions of England, the majority of these are located within the North of England, suggesting a 'North-South' divide (Public Health England, 2017). These inequalities also exist within the older adult population, where the gap between income of those within the greatest 20% is more than double that of individuals within the lowest 20% (Centre for Ageing Better, 2017). Equalities are also shown for those from Black and ethnic backgrounds, who show increased morbidity and decreased access to health services (Public Health England, 2017), a trend which continues into older adulthood (Centre for Ageing Better, 2017).

The ongoing economic climate within the UK has impacted upon many vulnerable groups, including older adults. Direct cuts to pensions and social care were found to impact on mortality rates in older-adults, particularly those who were over the age of 85 years (REF, 2016). The impact of austerity is also apparent in those requiring long-term care, the most recent Directors of Adult Social Services Budget (ADASS, 2019) concluded that the current system of social care was failing those who needed it most. Furthermore, The Kings Fund

(2018) showed the gap between older adults who requested access to adult social care and those who received it was increasingly widening.

Funding for adult social care has decreased by 5% since 2010 (ADASS, 2019), meaning local authorities are left with an increasing population of those in need and less money dedicated to meeting such needs. Consequently, spending on social care by local authorities decreased by 1.4 billion between the years 2010 and 2018 (Age UK, 2019). In addition, many care services delivered at home or within the community are being cut, or reduced, and the cost to pay privately for services such as 24 in-home care for older adults is so high that individuals are left with no choice but to move into residential care homes (Human Rights Watch, 2019). Access to social care for older adults is also affected by geographical location, whilst recommendations for social care spending are set by the government, individual local authorities are responsible for their own decisions regarding budgets and spending, meaning variation in the level of support offered (The Kings Fund, 2019).

Currently within England help with meeting the cost of care home fees is means-tested and the monetary threshold for receiving any support for care home fees is met when an individual has up to £23,250 in 'financial assets' (Human Rights Watch, 2019). Only 21% of older adults within the UK have most of their social care funded by their local authority (Age UK, 2019), with the rest relying on using their own savings and financial assets, or that of family members (Human Rights Watch, 2019; The Kings Fund, 2019).

With continuing economic uncertainty there has also been in decrease in the number of care homes operating within the UK. Although commissioned by local authorities, most care homes are owned by private companies. As funding available to councils decreases and their availability to pay for services decreases, care homes struggle to cope with the costs of

dealing with increasingly complex health needs (Age UK, 2019; The Kings Fund, 2019). There are also large number of vacant positions for care home staff members attributed to low pay, lack of job security (such as zero hours' contracts) and a lack of training to effectively cope with complex behaviours (The Kings Fund, 2019). The staff turnover within the care home sector is also high (around 30%), highlighting a difficulty in recruiting and, more importantly, retaining staff members (ADASS, 2019; Age UK, 2019).

1.4 Quality of Life, Wellbeing and Health

Definitions of health, wellbeing and quality of life are closely related and overlap in terms of qualities and measures. As such, they are likely to impact and moderate one another. The World Health Organisation (WHO) recognises the need to have an overall sense of feeling healthy, not just in relation to physical health. The WHO definition of health (which has remained unchanged since 1946) is:

> *'A state of complete physical, mental and social wellbeing and not merely the absence of disease or infirmity' (WHO, 2006, p1).*

Specifically, within the UK, the British Medical Association (BMA) has highlighted that the future of healthcare needs to encompass the spiritual, social and environmental needs of patients in addition to receiving treatment for physical illness (British Medical Association, 2011).

Definitions of quality of life vary amongst disciplines and the term is often used synonymously with wellbeing (CDC, 2013). Within a health context, quality of life includes measurements that overlap with wellbeing definitions: physical and psychological health, level of independence, social relationships, environment and religious and/or personal beliefs (WHO

Division of Mental Health and Prevention of Substances Abuse, 1997; British Medical Association, 2011).

The concept of wellbeing may be thought of both on a societal and personal level. Within the UK, there is increasing interest in the wellbeing of the nation as a measure of progress (Allen, 2008; ONS 2018) and 10 domains have been highlighted by the Office for National Statistics which effectively measure the wellbeing of society. These domains are: personal or subjective wellbeing (our happiness and psychological health); our relationships with others; health (encompassing physical; social and psychological aspects); how we spend our time ('what we do'); where we live (geographical location and housing type); personal finances, education and skills; the economy as a whole; how much confidence we have in government and the natural environment (ONS, 2014b). Within a health context, wellbeing is often explored on a personal level, with subjective wellbeing arguably focused upon the most alongside social functioning (relationships with others) and overall health.

In 2008, the New Economics Foundation (NEF) outlined two elements of personal wellbeing: feeling good and functioning well. The first element therefore focuses on our feelings of happiness, contentment, enjoyment, curiosity and engagement. The second incorporates how well we function within the world, our experiences and how we form and maintain relationships. Using this information NEF then devised 'Five Ways to Wellbeing', which identifies the factors that are important to maintaining this wellbeing. This included: connecting, being active, taking notice, keeping learning and giving. Each of these factors is explained as enhancing an individual's sense of feeling good which in turn moderates our functioning and this cycle continues within a feedback loop each reinforcing the other. The

‘Five Ways to Wellbeing’ has since been utilised by the NHS, mental health charities and government agencies as a way of promoting and explaining wellbeing and healthy living.

The most recent measure of national wellbeing within the UK (Office for National Statistics, 2018) showed that adults between the ages of 65 and 79 reported the highest levels of personal wellbeing. In spite of this, three of the measures of wellbeing (life satisfaction, feelings that what one does are worthwhile and happiness) began to fall as age increased past the age of 75, with the biggest decrease being shown for the ‘feelings that what one does is worthwhile’ domain.

Despite the population of older adults showing a steady increase, ageism is an increasingly prevalent problem (O’Neill, 2011; WHO, 2014). O’Neill (2011) highlights the two ways in which ageism occurs within healthcare, the first of which is by restricted access to healthcare, for example older people are less likely to receive chemotherapy (Age UK, 2015) and are less frequently referred to specialist mental health services (Anderson et al 2009; Thraves, 2016). The second is a failure to incorporate gerontological principles into healthcare; demonstrated by a lack of specialist knowledge and tailoring of older person’s healthcare (ONeill, 2009; Anderson et al, 2009; Age UK, 2015). Societally there is more of a focus upon the negative aspects of ageing and despite the ability for older people to effectively contribute to society, views of old age include that older people are frail, dependent on others and society and are unable to learn new skills and concepts (WHO, 2015). Such misconceptions can reinforce the stereotypes often focused upon by older people with a focus on lost abilities and lack of opportunity and can therefore be contributory factors for mood disorders (Basting, 2009), with poor mental health being shown to be associated with decreased levels of wellbeing (Royal Society for Public Health, 2013).

Worldwide around 20% of older adults are thought to have a mental health condition (WHO, 2017), however a recent report on the care home population in the UK showed prevalence within this demographic to be 60% (Thraves, 2016). Depression within the older adult population is complex and often differs to that experienced by younger adults mainly being mediated by loss (such as loss of relationships, position in society or health) and defective adaption to such losses (Flood and Philips, 2007; Myskja and Nord, 2008; Basting, 2009; DeGuzman et al, 2009; Age UK 2015). It is predicted that many older people will experience depressive symptoms in addition to those with a formal diagnosis and many older people fail to receive help from mental health services or wait longer than the rest of the population (Allen, 2008; Anderson, 2009; Thraves 2016). Given that the move to care homes for many older people can occur quickly (NICE, 2015), and in relation to complex medical needs and social changes, it is not surprising that there is a high level of depression within the care home population (Xiu-Ying et al. 2012; Atkins et al. 2013; Ferreira, Dias & Fernandes, 2015; Age UK, 2016; Thraves, 2016).

Whilst there is an increasing focus upon the psychological aspects of health, the impact of living with a long-term physical health condition should not be overlooked. Many older adults are living with some form of long-term condition and physical disabilities and poor physical health is correlated with poor psychological health in older adults (Flood and Philips, 2007; Anderson, 2009; Atkins et al. 2013; Thraves, 2016). Despite this, half of the older people within the UK who live with a long-term condition do not see this as a barrier to living their lives (Mortimer & Green, 2016). Similarly, psychological health impacts upon physical health, people who are happier or fulfilled in their everyday lives are less likely to present with health problems (Cameron et al. 2013). Furthermore, studies have demonstrated how older people

are perceived by their caregivers as having a reduced quality of life due to their physical disabilities and dependencies; whilst older people themselves rate their mood and psychological health as having the biggest influences on their quality of life (Hoe et al. 2006; Hoe et al. 2009).

Another factor which can impact upon the quality of life, health and wellbeing in older people is feeling lonely. Loneliness is prevalent within the older population and is a strong predictor for reduced quality of life and wellbeing (ONS 2015; ONS, 2018) and poor health outcomes (Age UK 2015). A survey for Age UK found that 1 million older people felt lonely 'always' or 'often' (Davidson and Rossall, 2014). In the same way in which depression in older adults is influenced by change in circumstances, loneliness occurs due to changes such as loss of friends and family, reduction in physical functioning and reduction in income (Social Care Institute for Excellence, 2012). Increased loneliness is also a predictor of early admissions to residential care facilities and an increase in frequency of GP visits and can therefore increase the strain on the National Health Service and social care services (Age UK, 2015). Conversely, the Marmot Review (2010) showed that those who participated in social activities on a frequent basis lengthened their life expectancy.

Frequent contact with friends and family is a predictor for increased wellbeing and reduced loneliness in older adults (World Health Organisation, 2015). Interestingly older people who reside in care homes, where there are many other people around, are more likely to report loneliness and become socially isolated than those who live within their own homes (Cowl and Gaugler, 2014; Davidson and Rossall, 2013). A reason for this may be that older people within care homes actually have less contact with their friends and family than those who do not, which can lead to them feeling socially isolated and low in mood (Allen, 2008).

Experiencing ‘positive relationships’ is important for wellbeing (New Economics Foundation, 2008) and therefore encouraging fostering of good quality relationships and links with family and friends as opposed to simply having other people around may exert a greater influence on the wellbeing of long-term care residents (World Health Organisation, 2015).

Older adults within the care home population need to be safeguarded effectively in order to conserve their quality of life and wellbeing (NICE, 2015) and as such this is a measure of quality for care home facilities. The Adult Social Care Outcomes Framework (ASCOF) consists of four domains and is used by the Department of Health to measure quality in care homes and identify areas for future improvement. The first domain reported within the ASCOF is the need to ‘ensure quality of life for people with care and support needs’ (Department of Health, 2014). The domain recognises, among other things, the requirements of people in care to receive social care related quality of life (1A), maintain a sense of control over their care needs (1B) and maintain levels of social contact (1L).

Further to this, the National Institute for Health and Care Excellence (NICE) has discussed the ways in which quality of life can be maintained within the care home population in relation to domain one of the ASCOF. They outline the requirement for older people living within a care home to have the opportunity to take part in ‘meaningful activities’ that can help maintain the health of residents and also decrease the cost to health systems by reducing the number of aggressive behaviours towards staff and residents (NICE, 2015). One such way in which older people may partake in ‘meaningful activities’ is through engaging with arts based activities.

1.5 Arts for Health

Broadly speaking, art forms fit within one of the following categories: performing arts, visual arts, combined arts and recreational activities (Arts Council England, 2015). The All-Party Parliamentary Group on Arts, Health and Wellbeing (2017) further specify 'arts' to include visual and performing arts, crafts, dance, film, literature, music and singing. In addition to this, the arts can be engaged with on either an active or passive level. Active participation may involve activities such as painting, creative writing or being involved in a performance such as singing or acting. Passive participation may be experienced by observing others or the artistic creations of others, such as attending a production, concert, exhibition or gallery (Staricoff, 2004; Arts Council England, 2015). The arts can also be experienced individually or with others and there is evidence of benefit when participating within the arts in both these ways (Cameron et al. 2013; Arts Council England, 2015). A review of the evidence by Arts Council England has shown the arts to be beneficial to the economy, society and education, with these benefits evidenced across a range of different demographic backgrounds (Mowlah et al., 2014). The review also highlighted the potential for the arts to positively impact upon health and wellbeing.

In 2013, the Scottish Government carried out a survey that explored the role of cultural engagement on the health and life satisfaction of the public (Scottish Government Social Research, 2013). The survey found that people who had engaged with a cultural institution (library, museum, cinema or live music event) within the last year were almost 60% more likely to report good health in comparison to those who had not. In addition, those who participated in cultural and creative activities (such as dance, reading and arts) within the last 12 months were 38% more likely to report good health in comparison to those who had not participated in such activities.

Further positive impact from experience of the arts was demonstrated in a recent study where participants were asked to report their levels of cultural activity and to answer questions that measured their perceived levels of subjective wellbeing (Fujiwara and MacKerron, 2015). It was found that engaging in any arts and culture activity improved both levels of happiness and relaxation. Interestingly, experiencing the arts passively was ranked as being most effective for both happiness (attending the theatre, dance or concert) and relaxation (attending a museum, exhibition or library).

Attitudes towards the arts in health have shifted from viewing arts-based activities as an optional 'add on' activity to a beneficial and worthwhile resource for patients (Department of Health and Arts Council England, 2007). There is an emergence of arts for health principles within medical guidelines and recommendations. The British Medical Association recognises the need for creative participation within its publication surrounding the psychological and social needs of patients who are receiving treatment in a hospital environment. It recommends that patients are given the option to take part in 'creative and recreational activities' that should be personalised to both residents and staff (British Medical Association, 2011) and The Department of Health classes arts as being "integral to health and health services" (Department of Health and Arts Council England, 2007). Additionally, the 2011 Government report 'No Health without Mental Health' outlines the ability to partake in meaningful activities to be a factor in maintaining positive mental health (Department of Health, 2011).

The idea that the arts can be beneficial to health and wellbeing has existed for centuries. However as global and UK-based policies and strategies are becoming focused upon the importance of health care to encompass the social and psychological wellbeing of

people, interest in the potential of arts for health is on the increase (Clift, 2012; Clift and Camic, 2015; Crossick and Kaszynska, 2016; Public Health England, 2016). This increased interest is evidenced not only within academia and research publications, but also by the establishment of a number of Arts for Health organisations worldwide (Clift and Camic, 2015).

There is much discussion within the literature as to the distinction between arts *activities* and art *therapy*. Whereas art therapy is focused upon treatment outcomes and therefore delivered by trained therapists, arts for health is not intended to be a form of therapy but may, by its process, have therapeutic benefits (Stickley, 2012; Beard, 2012; Crossick and Kaszynska, 2016). Arts Council England (2015) recognises the benefits not only to health, but also to the impact on care provision and environment and therefore defines 'arts for health' as:

> *"…arts based activities that aim to improve individual community health and healthcare delivery, and which enhances the healthcare environment by providing artwork or performances" (p5).*

The various ways in which the arts can be beneficial to medicine have been explored in several studies. Reviews of the medical literature have shown different art forms to be beneficial in the areas of cancer care, cardiovascular care, post-operative care, mental health, paediatrics and obstetrics mainly by reducing the anxiety and depression of patients during treatment (Staricoff, 2004; British Medical Association 2011; Staricoff and Clift, 2011). There are also benefits to health staff and students in taking part in the arts; there is evidence that taking part in medical humanities allows medical students to gain a better understanding of arguments and context (Downie 1999; Downie, 2016). Medical staff who took part in the arts also showed a greater appreciation of the complexities of humans when treating patients (Anderson and Schiedermayer, 2003) and allows for thinking of societal changes rather than

just medical ones (Downie, 2016). The arts can also aid communication between medical staff and patients, by using visual images to explain complex medical procedures patients have a better understanding of their conditions and treatment (Department of Health, 2007).

Specifically focused on older people, the arts may offer the ability to take part in activities and further learning, which may consequently improve mood, functioning and health. NICE guidelines for the care of those living in care homes and the management of dementia both promote the idea of engaging with activities that are 'worthwhile', meaning they can facilitate maintenance of wellbeing and quality of life (NICE, 2013; NICE; 2015). The WHO's active ageing policy framework (World Health Organisation, 2015) also recommends that older people can 'participate' by having access to 'learning opportunities throughout the life-course'.

In keeping with the inclusive nature of arts for health, the focus should not be upon the production of *meaningful* art in older people, but on the creative process and the benefits that may be achieved by it (Cowl and Gaugler, 2014).

1.6 Evidence of Arts for Health Benefit in Older Adults

Given the prevalence of dementia and cognitive impairment within the older population, it is not surprising that many studies have focused upon the benefits of arts within this subgroup of older adults. Studies that focus on participants with dementia often use clinical outcome measures as a benchmark for success, with improvements being noted for behaviour, mood, psychosocial functioning, cognition and relationships with carers (O'Connor, 2009a; O'Connor, 2009b, Vink, Bruinsma and Sholten, 2011; Kolanowski, 2011). A systematic review that included 112 studies reported the most commonly explored outcome measures were

improvement in agitation, behaviour, cognition, depression and mood (Cowl and Gaugler, 2014) with evidence of benefit being concluded for all aside from cognition.

Another systematic review carried out by Beard (2012) also showed there to be improvements in behavioural manifestations of dementia, however the review also highlighted the importance of findings obtained from studies focused on arts-based music activities, as opposed to those which provided structured therapies. One benefit frequently reported by such studies was an improvement in socialisation, an important factor in ensuring older people do not become isolated from others. This has been reported by a number of studies where participants have taken part in group-based arts interventions (DeGuzman et al 2009; Cowl and Gaugler, 2014; Young, Camic and Tischler, 2015). There is evidence that engagement with places of culture, when experienced as part of a group, promotes discussion and interaction with other members of the group, including staff members, highlighting the importance of shared experiences (MacPherson, 2009; Roe et al. 2014). Also, of interest is evidence that changing the environment of a ward within a nursing home with art or other sensory stimulation can promote interaction through experience of a novel change (Scott, Masser and Pachana, 2014). Artwork produced by older adults (either individually or as part of a group) and subsequently displayed can also promote discussions amongst peers, with caregivers and family members, allowing for an outlet for self-expression and communication (DeGuzman et al. 2009; Stallings, 2010; Greer, Fleuriet and Cantu, 2012**;** Roe et al. 2014). Improvements in socialisation need not be as obvious as through discussion and conversation, those with advanced dementia have been shown to improve their social contact through engaging with music, demonstrated through prolonged eye contact and head turning (Clair 1996; Clair and Bernstein, 1990; McDermott, Ridder and Orelle, 2014).

Arts activities that facilitate discussion through encouraging memories (though not in the form of traditional structured reminiscence therapy which attempts to promote memory recall without the inclusion of creativity) can also provide an opportunity for social interaction through common interests and sharing of pleasant memories (MacPherson et al., 2009). The idea that the arts can facilitate socialisation and improve mood but not focus on traditional reminiscence therapy has been shown by use of the Time*Slips* programme. The emphasis on the programme is to move away from the standard requirement of older people with dementia, cognitive impairment and memory loss where the emphasis lies on simply being required to recall memories from their past. Instead, the programme facilitates creative writing processes by letting older people each contribute lines to a story which is then told to the group. Whilst there is an emphasis on creativity, the programme also encourages socialisation through discussion of topics, which participants can relate to, and ideas they may have, without the need for them to focus upon recall of specific memories (Basting, 2003; Basting, 2009).

Another way in which arts activities may benefit older adults is through the encouragement of a sense of empowerment and independence (MacPherson et al. 2009; Beard 2012). Over three quarters of older people feel that their skills are not appreciated or utilised (Age UK, 2015) and as previously discussed, older people show a rapid decline with age in feeling that they partake in activities that are worthwhile (ONS, 2015b). Participation in arts based activities can encourage older adults to develop a sense of control over their wellbeing and health (Cohen, 2006; Vella-Burrows, 2015). This is highlighted in a study in which older adults living in a residential facility were taught traditional Filipino arts (which was culturally relevant to the group). Prior to participating in the programme, levels of depression amongst

participants was high and they described a number of stressors such as reduced contact with family, physical disabilities and struggling to adapt to a residential environment. Following the course, participants showed an increased sense of self-esteem and subsequent increased mood. Furthermore, it was identified how participants developed a sense of resilience to their physical challenges, by recognising and nurturing their mental abilities (DeGuzman et al. 2009). Older people have also highlighted how the respect and dignity shown to them when partaking in arts activities and indeed being offered to take part in activities that are not solely focused on being 'old'. This allowed them to feel they were treated equally with the care staff and staff leading the interventions, which in turn can positively influence their sense of self (Basting, 2009).

An aspect of the World Health Organisation's healthy ageing strategy is the opportunity to continue learning throughout the lifespan (World Health Organisation, 2015). Creativity and aesthetic appreciation remains throughout our lives and the opportunity to be creative, and learn new skills by doing so, can be beneficial to the health, wellbeing and quality of life of older people (Cowl and Gaugler, 2014; Age UK 2015; Clift and Camic, 2015; Vella-Burrows, 2015). Furthermore, in older people with cognitive or physical impairment, producing art and learning new skills can serve as a reminder of abilities that are still present rather than focusing on what has been lost (DeGuzman et al, 2009; Beard, 2012; Ullan et al. 2013). A systematic literature review, which included 17 studies, emphasised the learning potential for arts interventions in those with dementia (Young, Camic and Tischler, 2015). Singing was shown to enhance learning by allowing those taking part to learn new songs and actions, particularly in a group setting. Visual arts promoted the learning process either by taking part and learning new methods, or by viewing art and therefore enhancing intellectual stimulation.

McDermott, Ridder and Orelle (2014) encouraged older people from care homes and day hospitals to take part in a group music intervention. Follow-up focus group interviews found that learning how to play an instrument was something that many of the group had wished to do in the past but had not had the opportunity and they valued being given the opportunity to learn. In addition, the participants also stated that they enjoyed the opportunity to learn something with others, including staff members, and this had made the experience more enjoyable. Conversely, a study by Ullan (2013) showed that whilst there was a high level of satisfaction and pride following completion of artwork, there was an increase in vocalisations amongst participants with dementia who were not given enough time to finish their projects, this highlights the importance of planning to minimise the risk of ill-effects.

There is an importance not only in establishing the benefits of the arts for health, but also in determining factors for best practice in carrying out these activities in order to maximise positive outcomes. Of significance are the findings from studies that highlight the role of matching activities and interventions to the personal preferences of participating. Gerdner (2000) compared two groups taking part in music therapy, one who had the music carefully matched to their preferences and another who had the music chosen for them. Whilst agitation was found to reduce for both groups, greater effects were observed for the group where the music was matched to the participants' preferences. A subsequent study by Ragenskog and colleagues (2001) found no benefit for listening to classical music (which was not chosen by participants), but a reduction in agitation for individualised music. A study by Nair et al (2011) carried out music therapy with participants who listened to Baroque music and found that behavioural disturbances actually increased following the music programme, with a number of factors such as volume, speed and type of music reported as possible

aspects which may need to be matched to participants to avoid adverse effects. Similarly, Kolanowski (2011) showed an increase in agitation and decrease in mood when activities were not matched to the preferences of those taking part in comparison to a group who did have theirs matched. This demonstrated the importance of enjoyment and pleasure when taking part in the arts and a need for activities to be of the interest to those taking part in order for them to be beneficial. However, matching the activities to the individual participants may be difficult in settings such as care homes where activities are often carried out within a group setting (McDermott, Ridder and Orelle, 2014).

Whilst there are many identifiable benefits for older people taking part in the arts, a number of reviews and studies are in agreement that these effects are often short lived and do not last for long beyond the programme duration (Sung and Chang, 2006; MacPherson et al, 2009; Vink, Bruinsma and Sholten, 2011; Beard 2012; Cowl and Gaugler, 2014). It has also been noted that withdrawal of activities results in lower mood and expressions of pleasure which suggests that residents become used to the stimulation provided by taking part in arts activities (Myskja and Nord, 2008; Kolanowski, 2011). Further reinforcement of this is provided by qualitative interviews carried out with family members who had taken part in a music therapy programme. They indicated that whilst there was evidence of behavioural change, these changes did not extend much beyond the programme duration and the benefits were very much rooted in the 'here and now' (McDermott, Orelle and Ridder, 2014). A further study by Myskja and Nord (2008) evaluated depression scores in participants who had previously had a music therapy programme in their home which had subsequently ended 11 weeks before the study. Levels of depression were measured at baseline (when participants had not taken part in music activity for 11 weeks) and then six weeks following the

programme being reinstated. Depression scores were significantly lower following the programme indicating that there needs to be consistency in carrying out activities in order to maximise their benefits.

1.7 Limitations of the current literature

Whilst there is clearly a growing body of evidence for the benefits of arts for health activities, there are limitations to current literature and completed reviews. There is a dominance of clinical outcome measures within the research relating to arts and health and the qualitative evidence that does exist is not robust, lacking adequate reporting of analysis and results. There is therefore a need for more qualitative evidence to accompany the body of quantitative evidence (Beard, 2012; Cowl and Gaugler, 2014; Young, Camic and Tischler, 2015; Crossick and Kaszynska, 2016). Such evidence can offer important contextual and person-centred information to further complement the findings of quantitative studies and offer insight into perceived benefits. This is evidenced by research exploring the role of dance for people with Parkinson's disease, which returned statistically non-significant results for clinical measures of improvement (such as posture and movement), yet anecdotally reported a number of improvements noted by care staff, family members, course leaders and the participants (Houston and McGill, 2016).

A number of existing systematic reviews that have explored the role of arts for health in older people have highlighted a number of flaws in existing evidence. Studies generally consist of small sample sizes, measures that have not been validated, poor reporting and analysis of qualitative results and an over-reliance upon clinical outcome measures as a reflection of success (Beard, 2012; Cowl & Gaugler, 2014; McDermott, Orrelle & Ridder, 2014). A large proportion of research to date is focused upon the medium of music (Beard, 2012) and as music therapy is a well-established field most arts for health systematic reviews (Sung and

Chang, 2006; Vink, Bruinsma and Sholten, 2011) focus on this alone. In a similar vein, many of the systematic reviews to date (O'Connor, 2009a; O'Connor, 2009b Vink, Bruinsma and Scholten, 2011; Cowl and Gaugler, 2014) have used terms related to 'art therapy' which may exclude studies which explore arts activities. Studies that focus upon such 'activities' may report different outcomes to those focused upon a traditional therapy method of delivery (Beard, 2012). Additionally, whilst there were a number of methodological limitations reported for the studies included within systematic reviews, there were hardly any which used quality appraisal tools in order to establish the quality of evidence of included studies.

As discussed, there are challenges that are faced by older people who live within residential or institutional care, such as care homes. Whilst there has been reviews which have included participants who live within care homes and residential facilities, there has been no systematic review which focuses solely on the benefits of arts for health within the care home population alone. A large proportion of studies have focused specifically on individuals with dementia; whilst this is obviously expected within the older adult population it would be beneficial to provide insight into arts for health for the care home population in general. Additionally, there is a need to determine whether there is evidence of different activities being beneficial to different health conditions or different subgroups of older adults.

There are a number of studies providing evidence for arts for health programmes, however the majority of these are stand-alone research projects and there is little information about what arts activities are currently being carried out within the UK.

1.8 The research project

Within the national inquiry into arts for health (APPG, 2017), three factors were determined as the main aims of carrying out research into the potential benefits of arts for health. These

were: to increase knowledge, to provide accountability for funding and to aid reflective practice. Therefore, research into the area can contribute towards the development and facilitation of arts for health activities for older adults.

Current evidence includes individuals from care homes, but there has been no systematic review focused upon the impact of arts for health exclusively within the care home population. Therefore, Study 1 is a systematic review of pre-existing empirical studies evaluating the impact of arts for health on the health, wellbeing and quality of life of older people residing in care homes.

The objectives for Study 1 were:

1. To identify published empirical studies focused upon arts for health and the impact to older adults with respect of health, wellbeing and quality of life.

2. To create a typology of arts for health activities and their perceived benefits.

Secondly, there has been no mapping of current activities being carried out within UK care homes and their perceived benefits. Study 2 focused upon this and comprised an electronic survey completed by a random stratified sample of care home managers within England. The questionnaire asked questions based around current arts activities being carried out within homes, how they are undertaken and whom they are led by. Follow-up interviews were conducted with a self-selecting sample of participating care home managers that completed the survey to provide contextual information to the questionnaire responses.

The objectives of Study 2 were:

1. To conduct an online survey of care home managers in England to determine what (if any) activities were delivered and map their prevalence.

2. To carry out telephone interviews with a self-selecting sample of care home managers to add further contextual information.

3. To examine the characteristics surrounding delivery of such activities to gather information including frequency and method of delivery, funding and associated benefits.

The final objective was to synthesise the results from both Study 1 and Study 2 and form a typology of arts activities and their perceived benefits and make recommendations and implications for policy and practice.

1.9 Summary

The population of older adults is increasing and there is a recognised need for maintaining health, wellbeing and quality of life in later life. Older adults who reside in care homes are particularly vulnerable to a number of psychological and physical health concerns which can impact upon their daily functioning. Current legislation for older people in care homes promotes the ability for individuals to take part in meaningful and worthwhile activates in order to maintain health and wellbeing. There is an increased interest in arts for health and benefit has been evidenced in a range of populations, therefore there is a potential for its use with older adults who live within care home facilities.

As yet there has been no systematic review assessing the impact of arts for health within older adults who reside in care homes. Nor has there been a national survey of what arts activities are currently delivered within England or the benefits associated with such activities. This thesis therefore consists of two studies, a systematic review (Study 1) and a national electronic survey and follow-up interviews with care home managers (Study 2).

Chapters 2-4 focus on Study 1, with an overview of the systematic review methodology and methods outlined in Chapter 2. Following this an overall high-level descriptive narrative synthesis of all included studies is presented in Chapter 3. Finally, Chapter 4 consists of a descriptive narrative synthesis by arts type for included studies. Chapter 5 then provides an overview of the methodology and methods for the national survey and telephone interviews and the results of these are presented in Chapter 6. The results from both studies are synthesised within the discussion (Chapter 7) and overall conclusions, recommendations for policy and practice are in the final chapter (Chapter 8).

Chapter 2- Study 1: Systematic review and descriptive narrative synthesis. Methodology and methods.

2.1 Introduction

The previous chapter (Chapter 1) introduced arts for health and provided an overview of its potential benefits to older people who live in care homes. This chapter outlines the methodology and methods to Study 1, a systematic review of existing literature on arts for health in care homes. The first half of the chapter will provide an overview and context of the importance of systematic reviews, particularly within the field of health. The second half of the chapter outlines the methods for the systematic review in Study 1 with focus on searching for literature, selection of potentially eligible studies, synthesis of result and appraisal of evidence. The search outcome and an overall high-level synthesis of results is presented in the following chapter (Chapter 3).

2.2 Overall Personal Methodology

The philosophical underpinnings of this research are based in that of critical realism. This epistemological approach recognises that there is a real world which exists independent of our own theories, perceptions and constructions (Maxwell, 2011; Jessop, 2019). Furthermore, this reality involves biological, psychological and social foundations which are interconnected and form the basis of life experiences (Kjorstad & Solem, 2017). This approach directly contrasts with that of constructivist ontologies which state that we construct our own realities, meaning there is no other reality than our own (Maxwell, 2011).

Such a perspective also acknowledges that facts and knowledge of the world are socially constructed meaning they are historically situated (Collier, 1994; Belfrage & Hauf, 2016). However, this also means that critical realist research can be used not only to enhance our understanding of society, but to implement emancipatory change (Belfrage & Hauf, 2016;

Kjorstad & Solem, 2017) within a physical reality (Maxwell, 2011). This directly relates to my time working within the NHS and allows me to recognise that different social realities, structures and political events may impact on my findings and recommendations. Furthermore, it allows me to be mindful that whilst events can be observed, there may be underpinning unobservable factors which may be responsible for the results.

Critical realism lends to several research methods (Kjorstad & Solem, 2017), with the focus of being as objective as possible in order to minimise bias (Collier, 1994). This factors into both studies but is particularly related to the systematic review, which subscribes to the methods of transparency. Furthermore, the nature of this study is to explain the current understanding of the impact of arts for health for older adults rather than 'predict' or 'hypothesise' in order to develop understanding and make recommendations on that basis.

2.3 Systematic Review Methodology

2.3.1 Conducting a systematic review

Within the field of health there is an interest in evidenced based practice and the potential for unbiased and reliable research to inform and shape healthcare (Cochrane, 1972; Sackett et al. 2000; Egger, Davey Smith and Altman, 2005; Centre for Reviews and Dissemination, 2009). However, the volume of health research is vast and ever increasing, making it difficult for professionals to keep up-to-date with the latest developments and know which recommendations to follow (Petticrew and Roberts, 2006; Pope, Mays and Popay, 2007; Khan et al., 2011). As there are such a large number of publications covering a wide range of topics with the health field, it would be nearly impossible for individuals to read and summarise all of the literature available to them in order to reach conclusions and recommendations (Egger, Davey Smith and O'Rourke, 2008; Khan et al., 2011). Single studies may often not be enough to conclusively provide the benefits of an intervention or

treatment, or may lead to over-inflation of the importance of findings (Petticrew and Roberts, 2006). Different studies exploring the same intervention may also greatly vary in their findings making it difficult to know which evidence provides a true reflection of an intervention's effectiveness (Glasziou et al. 2001).

Previously, 'traditional literature reviews' were carried out in order to summarise and provide evidence of an intervention's efficacy. Such reviews were typically carried out by industry experts and involved presenting a series of research studies in support of a certain treatment or intervention. However, these reviews have several problems. Firstly, they leave themselves open to bias as authors could be highly selective in choosing literature and the way in which such literature is sourced is not reported. Secondly, they are likely to be focused upon the outcome of the study and may fail to take into account the methodological quality of the evidence. Taking these points into account they therefore tend to lack critical appraisal of their included evidence and are prone to biases (Pope, Mays and Popay, 2007; Egger, Davey Smith and O'Rourke, 2008).

In 1972 Archie Cochrane published a textbook in which he suggested medical research was missing robust and reliable summaries of evidence from a number of different studies (specifically randomised controlled trials) in order to provide evidence of an intervention's efficacy (Cochrane, 1972). Following this, interest in systematic reviews began to develop and in 1993 the Cochrane Collaboration was formed, an organisation which provides current and accurate information on health care interventions which may be accessed through the Cochrane Collaboration library (Webb and Roe, 2007; Higgins and Green, 2011).

Once seen as a method for quantitative evidence, qualitative research is increasingly being recognised as a valuable source of data for systematic reviews, providing information not

able to be captured by quantitative methods (Pope, Mays and Popay, 2007; Webb and Roe 2007). Webb and Roe (2007, p255-257) mapped the timeline of the systematic review process and methods for synthesis of quantitative or qualitative evidence and mixed methods quantitative and qualitative evidence in integrative reviews. The Cochrane Qualitative Research Methods Group identifies that evidence from qualitative studies is important for health research by complementing quantitative findings and providing answers to questions other than simply reporting the efficacy of treatments or interventions (Noyes et al. 2011).

The aim of carrying out a systematic review is to identify and summarise all the relevant evidence related to a research question, appraise this evidence and then synthesise the findings from individual studies in order to form overall conclusions and/or recommendations (Pope, Mays and Popay, 2007; Webb and Roe 2007; Liberati et al. 2009). Systematic reviews therefore offer summarisation of evidence that is less of a discussion as found in traditional literature reviews and more focused upon critically reviewing and appraising the available research and findings (Petticrew and Roberts, 2006). By synthesising the data from a range of studies, systematic reviews can provide evidence for robustness of studies and add statistical power to small scale studies (Glasziou et al. 2001). In contrast to this, by summarising available evidence, systematic reviews can also help identify the gaps in research study or standards and recommend the need for further study (Egger, Davey Smith and O'Rourke, 2008).

The Cochrane Collaboration indicates that the characteristics of a systematic review are as follows:

1. A clearly outlined set of objectives which highlights inclusion and exclusion criteria for eligibility of studies.
2. An explicit methodology which is reproducible.
3. The carrying out of a systematic search using a pre-determined search strategy in order to identify all potentially relevant studies.
4. Assessment of the risk of bias for included studies in order to determine the methodological qualities of included studies.
5. A systematic presentation with clear synthesis of characteristics and findings of the studies which are included within the review (Higgins and Green, 2011).

A review of systematic reviews (that had been published prior to 2005) was carried out in order to update previous guidelines on how they are conducted and reported. The review resulted in the identification of a number of flaws which were present in their methodology and reporting. This led to the development of the PRISMA statement, which identifies a systematic review as consisting of the following steps:

1. Clearly outlining the research question and objectives of the review
2. A search for literature, which must be systematic and identify relevant studies of inclusion
3. Subsequent extraction of data from identified studies
4. An assessment of the confidence which may be placed in the findings (quality appraisal)
5. Synthesis of the data from included sections.

A 27 item checklist was also devised from exploration of the systematic reviews within the paper and this offers detailed information as to what is required at each step of the undertaking and reporting of systematic reviews (Liberati et al. 2009; Moher et al. 2009).

2.3.2 Defining the research question

The first step in carrying out a systematic review is to create the research question which the review will address. Such questions need to be broad enough to ensure that a variety of literature is sourced and allow identification of any factors which may impact upon the efficacy of a treatment or outcome. The research question should be defined prior to carrying out the systematic review to maintain a level of objectivity to the review and allow for critical appraisal (Khan et al, 2011). It is therefore strongly recommended that following identification of the research question and study aims, reviews are registered with an appropriate body (such as the International Prospective Register of Systematic Reviews (PROSPERO), the database for health reviews currently being undertaken or completed- http://www.crd.york.ac.uk/PROSPERO), in order to provide access to the stud aims and protocol (Liberati et al. 2009; Moher et al. 2009).

The focus of systematic reviews is often upon the effectiveness of a specific intervention and this will form the basis of the research question. Typically, such questions are devised using the acronym 'PICOS', which refers to the population, intervention, comparison (if one is used, often in health research this is usual care), outcome and study design (if only a certain study design is being included, such as randomised controlled trials) (Robinson, Saldanha and McKoy, 2011; Liberati et al. 2011).

Identification of the population of interest allows the systematic review to remain relevant to a group of people but also allows for some level of specification, for example looking at a

certain age group, people with a certain condition or a specific gender. The identification of the intervention is normally straightforward as this forms the basis of the review. If the research question calls for a specific comparison, such as another treatment (i.e comparison of two different medications) then this can also be acknowledged along with the outcome which is being explored (Glasziou et al. 2001; Khan et al., 2011; Higgins, and Green, 2011).

The effective framing of an answerable and relevant research question and identification of the 'PICOS' for the review can be beneficial for the location and selection of studies which forms the second step in conducting a systematic review (Pope, Mays and Popay, 2007).

2.3.3 Identification of studies

When carrying out literature searches in order to determine potentially relevant studies it is important to adopt a thorough approach so that the maximum number of appropriate studies are found (Khan et al., 2011). With the ever increasing scope of the internet, there is now a great deal of research available and one way of accessing this information is through using electronic databases, which allow advanced searching to identify studies of interest. It is important that different electronic databases are utilised as they are likely to return different results based on their subject area and their indexing terms (Pope, Mays and Popay, 2007). Furthermore, there is differentiation in the geographical location of included journals from each database (i.e. more American journals for MEDLINE, more from Europe in EMBASE) further illustrating the need for comprehensive searching across a number of databases (Roe, 2007). Aspects of the research question and PICOS may be combined with Boolean logic in order to form these search strings (Glasziou, 2001; Roe, 2007; Khan et al., 2011).

For example, the research question of this study is: 'According to the current literature, what are the benefits of arts for health on the health, wellbeing and quality of life in older people

in care homes'. The individual components of the question may then be used to produce a search string using the population (older people who live within care homes), intervention (arts for health) and outcome (health, wellbeing and quality of life). As the research question does not specify a comparison or study type then these are not used. The search may then comprise of the components such as: (older people) AND (care homes) AND (arts for health) AND (wellbeing). Strings can also be formed using different aspects of the intervention, in this case using different arts forms, for example replacing 'arts for health' with 'music' or 'drama' etc. In order to allow for replication of a review, all strings should be reported within the methods section of the write up (Roe, 2007).

Electronic databases, however, only represent a fraction of the hundreds of research articles that are available for any given subject area (Glasziou et al, 2001) and it is therefore necessary to source relevant literature from other areas. Both identification of relevant studies and prior systematic reviews related to the topic area can allow for hand-searching of referencing lists which may identify studies not obtained from the database search alone. In addition, using online search engines such as Google, can return links to so called 'grey literature' as can searching of conference proceedings, abstracts, dissertation theses, charity organisations and research council websites. Sourcing of such information can reduce the risk of 'publication bias' in selected studies such as greater likelihood of statistically significant results being published or those from more 'prestigious' universities and organisations (Roe, 2007; Pope, Mays and Popay, 2008; Khan et al., 2011).

The focus of a systematic review should be to ensure that all *relevant* information is found (Petticrew and Roberts, 2006; Webb and Roe, 2007) and, with this in mind, it is important to specify relevant inclusion and exclusion criteria. These criteria are often a checklist comprised

of a combination of the PICOS and research question, which determines whether the study addresses the aim of the current review.

The premise of inclusion exclusion forms is to provide readers with evidence of a systematic way for study selection, however whilst this minimises the risk of bias, it does not eliminate it completely. In order to further address the possibility of bias, it is best practice to ensure that potentially relevant studies are assessed by two researchers for eligibility and then cross checked to form an agreement, if there is still no agreement then a third reviewer may be called upon in order to determine inclusion or not (Pope, Mays and Popay, 2007; Roe, 2007). Every stage of the literature search should be reported, including the numbers of papers screened and excluded at each stage. The PRISMA statement provides a flow chart which outlines a transparent way to report on how literature is sourced and selected for inclusion (figure 2.1).

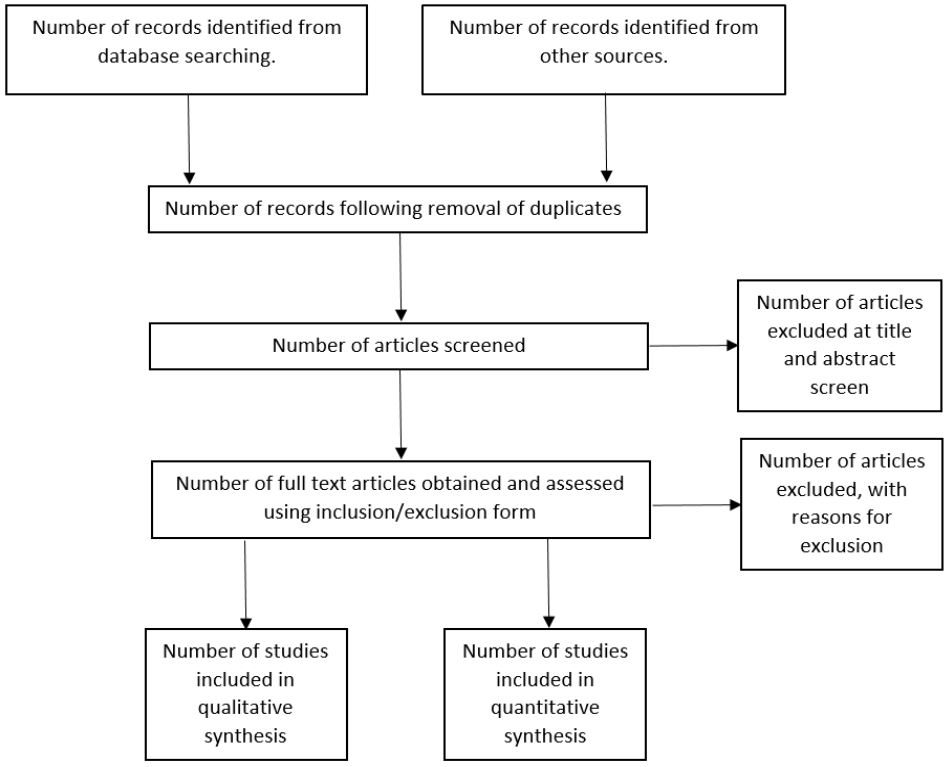


Figure 2.1. PRISMA flow chart showing how to report the selection and inclusion of literature (Liberti et al. 2011; Moher et al. 2011).

2.3.4 Data Extraction

Data extraction is the procedure of ensuring all data which is relevant to the research question and the PICOS is identified and recorded (Pope, Mays and Popay, 2007). A guide as to what data will be extracted should be provided within the registered protocol prior to conducting the review, however as knowledge of a subject increases during the extraction process specific aspects of data which are extracted may change and this should be amended within the protocol if necessary (Centre for Reviews and Dissemination, 2009).

Data extraction errors within reviews are common (Eden et al. 2011) and it is therefore recommended that data extraction be carried out by two reviewers to ensure that no omissions occur, with any disagreements being discussed and resolved with a third reviewer if necessary (Roe 2007). If time constraints are a problem, as a minimum requirement, data may be extracted by one member of the review team and cross checked by another, however this may increase the chances of errors (Centre for Reviews and Dissemination, 2009; Eden et al. 2011).

It is useful to devise a data extraction sheet which can be piloted on a number of the studies in order to ensure it captures all relevant information; use of a standardised sheet for all included studies can lessen the risk of bias and increase the reliability of results (Centre for Reviews and Dissemination, 2009; Moher et al. 2009). Such forms should include not only collection of data related to the PICOS, but should also record an identifier for the study and full reference of the paper. The name of the person carrying out the data extraction should also be recorded, along with any revisions made (Higgins and Green, 2011).

Sometimes, not all of the necessary data is reported within the results section of a study. If this is the case, the authors of such a study may be contacted for further information or

clarification; if this occurs then it should be explicitly stated in the write up (Liberati et al. 2009). In the result of two papers being identified which refer to the same study, these should be tagged together and only one set of data be extracted in reference to the whole study to avoid duplications (Higgins and Green, 2011). Data extraction may be carried out simultaneously with quality appraisal in order to further reduce the risk of mistakes occurring as the reviewer is already involved with the data (Centre for Reviews and Dissemination, 2009).

2.3.5 Quality Appraisal

The premise of quality appraisal is to assess the levels of bias of each study based upon scoring the design of the study, how the study was conducted and how the analysis of findings was carried out (Khan et al, 2011). Quality appraisal should not simply involve listing what is perceived to be wrong with a study, but appraising what aspects of the study may be open to bias based upon the evidence from the paper (Petticrew and Roberts, 2006). Inclusion of studies that demonstrate poor quality on any of these factors may lead to acceptance of results which are biased and therefore do not represent best available evidence (Higgins et al. 2011). Items 12 and 19 of the PRISMA statement refer to the quality of studies, stating that risks and sources of bias, and their assessment of, should be clearly reported within the review (Liberati et al. 2009; Moher et al. 2009).

Within research design, there exists a hierarchy of evidence, upon which systematic reviews are viewed as offering the highest level of evidence. Randomised controlled trials are seen as offering the highest standard of primary research due to their low risk of bias and observational studies reflecting a high risk of bias (Pope, Mays and Popay, 2007; Guyatt et al. 2011). For some systematic reviews quality appraisal will form the basis of inclusion (such as reviews of randomised controlled trials). Appraisal of study design may be important for

subjects where there is variation of research designs as it may help to explain discrepancies in results across studies (Centre for Reviews and Dissemination, 2009). However, it is important to recognise the variation of study designs amongst different topics and to place caution in over-reliance upon randomised controlled trials, which may not always effectively answer the research question or may not be ethically appropriate for the intervention (Khan et al., 2011). Additionally, randomised controlled trials may offer evidence of benefit in the short term, but are not able to provide longitudinal data or effects of the intervention past the point of study (Shaw et al. 2009).

Selection of participants included within the study and the allocation of study groups may leave studies open to biases and this should be considered for all studies, not just randomised controlled trials (Higgins and Green, 2011). Failure to address such an issue can result in difference between the groups which can then produce unreliable results (Khan et al., 2011). It is desirable for participants to be blinded within their treatment groups and for researchers to also be blinded, however, this is not always possible and so blinding of those carrying out the outcome assessment may be more appropriate (Centre for Reviews and Dissemination, 2009).

The way in which data are analysed may also impact upon whether or not a study is susceptible to bias. Studies of all designs in which participants withdraw or are 'lost to follow-up' can lead to findings being reported which are incorrect (Petticrew and Roberts, 2006). If intention-to-treat analysis (whereby data from *all* participants who took part within the intervention phase of the study) is not carried out within the analysis then the results are open to bias (Khan et al, 2011). At the very least, there needs to be an adequate account of drop outs from the study and their reasons behind doing so (Higgins and Green, 2011).

There are many quality appraisal tools used to assess studies, or one may be created which is specific to the study (Webb and Roe, 2007; Shaw et al. 2009). Quantitative and qualitative tools differ as the research process is not the same for each of the methods. One example of a quantitative quality appraisal tool was developed by the Grades of Recommendation, Assessment, Development and Evaluation (GRADE) Working Group to address the risk of bias. The GRADE approach recognises six areas which may pose a risk of bias; selection (related allocation sequence generation), performance bias (allocation concealment), detection bias (blinding of research personnel), attrition bias (omission of results), reporting bias (not reporting items stated in the protocol and vice versa) and 'other biases' (any other clear biases which would impact upon the credibility of the study). Assessment is twofold with a description of the biases firstly being recorded and then a corresponding risk of bias score of 'high', 'low', 'very low' or 'unclear' being assigned to each item) (Higgins and Green, 2011; Scheunemann et al. 2013).

The appraisal of qualitative studies has formed the subject of debate as some feel to do so takes away from the interpretive nature of the process of qualitative research (Shaw et al. 2009). As such, whilst the factors assessing quality of quantitative research are, to some extent, agreed upon there is not an agreed definition of how to appraise the quality of qualitative evidence (Centre for Reviews and Dissemination, 2009). The focus on qualitative appraisal largely lies upon the appropriateness of methodology chosen, how this was implemented and how the authors reached the conclusions and themes presented. A number of individual checklists exist for appraising qualitative evidence and these were the subject of a review by Walsh and Downe (2006). The review explored items used on seven existing qualitative appraisal tools (Murphy et al. 1998; Popay and Rogers, 1998; CASP, 1999; Mays

and Pope, 2000, Yardley, 2000; Cesario et al 2002; Spencer et al. 2003) in order to assess common features of the scales and synthesise these to produce a quality appraisal tool for qualitative research. The resulting checklist consists of 8 factors with a series of prompts to make judgements of quality of studies consisting of: the purpose of the research, the study design, the sampling strategy, size and characteristics, how the data was analysed, how results were interpreted, whether there was evidence of reflexivity and whether ethical considerations were adhered to.

2.3.6 Synthesis of results.

The end result of a systematic review is to synthesise all of the available evidence so as to form conclusions and recommendations based upon cumulative findings from a number of studies. With homogenous data collected from studies with the same methods and outcome measures the desired way to do this is by meta-analysis, the pooling together of statistical data so as to form a weighted average which represents an overall figure (Egger, Davey Smith and Altman, 2005; Higgins and Green, 2011).

Carrying out a meta-analysis relies upon having data that was all obtained in a similar manner and measured the same across studies and is therefore used for quantitative data. If carrying out a review which involves a number of different study designs, or has varying outcomes then it is more appropriate to carry out a synthesis of results (Ryan, 2013).

One such way of synthesising the data is through narrative synthesis. Unlike statistical methods of combining data, narrative synthesis uses words and texts in order to synthesise the findings of the review and forms more of a story about the data (Popay et al. 2006). Narrative syntheses may pose a problem as they leave themselves open to bias, such as over-emphasising the results of a study or being selective when reporting on studies (Higgins and

Green, 2011). For this reason, it is important to ensure that the methods used for narrative synthesis are as systematic as those for the rest of the review.

There have been four steps identified when carrying out narrative synthesis (Popay et al. 2006; Ryan 2013). The first step recommends that the reviewers become familiar with the data and begin to form a theory as to who the intervention is beneficial to and why this is the case.

Secondly, a preliminary synthesis may be undertaken which can be done in a number of ways such as: describing each individual study making sure to include exactly the same information for each study and in the same order; grouping the studies (i.e. by participant characteristic, intervention type, outcome measure or study design; using Tables to visually identify patterns within the data or transforming the data via thematic or content analysis.

The third step involves an exploration of the relationships which have emerged from the data in accordance with the heterogeneity amongst studies. When considering differences within studies the focus should be on different effects in participants (i.e. whether different effects were observed for a certain age group or a certain sex) and when considering relationships between, studies should explore whether different effects were observed for different study designs, different populations or intervention types. How this is carried out is up to the reviewer, but should be conducted within a systematic way, such as forming a graph or visual plot, carrying out subgroup analysis or developing conceptual models (mapping or conceptual triangulation).

Finally, how robust the evidence is should be considered in accordance with the quality of included evidence as determined by the quality appraisal carried out earlier on in the review

process. A further way in which the synthesis of qualitative data may be examined for robustness is the use of the Confidence in the Evidence from Reviews of Qualitative Research (CERQual) method (Glenton et al. 2013; Lewin et al. 2015). The CERQual approach is used where overall themes have been identified from a qualitative synthesis and examines the methodological limitations, relevance, coherence and adequacy of the results in order to give an assessment of the confidence that can be placed in them. CERQual therefore works where data has been analysed via methods such as thematic analysis during synthesis in order to identify overall themes.

Synthesis for studies which include both quantitative and qualitative data may be carried out in one of two ways; either a multi-level synthesis can be carried out, in which quantitative and qualitative data are synthesised separately and then combined in a third level to form one overall conclusion, or parallel synthesis may be carried out in order to explore two separate syntheses which can then be used to provide context to one another (Higgins and Green, 2011).

2.4 Systematic Review Methods

2.4.1 Research Aim

Despite recognition of emerging benefits to older people, to date, there has been no systematic review focused exclusively on evaluating the evidence of arts for health activities for older people who reside in care homes. The aim of this systematic review was to locate, analyse, appraise and synthesise existing empirical studies which have evaluated the impact of arts for health activities with older people who reside in care homes. The outcome measures focused upon were health, wellbeing and quality of life.

2.4.2 Registration of the Review

Prior to conducting the review, a protocol was developed and registered with the University's Faculty Research Ethics Committee (FREC) and the International Prospective Register of Systematic Reviews (PROSPERO, Curtis et al 2015, CRD42015026264). PRISMA guidelines were adhered to when conducting and reporting this systematic review (Liberati et al. 2009; Moher et al.2009).

2.4.3 Research Question

Is there evidence from published empirical studies that arts for health activities are beneficial to the health, wellbeing and quality of life of older people who reside in care homes?

2.4.4 Objectives

1. To identify published empirical studies that focuses upon arts for health and the benefits to older people in respect of health, wellbeing and quality of life.
2. To create a typology of arts for health activities and their perceived benefits.
3. If possible, gather economic data in relation to arts interventions to explore cost/benefit.

2.4.5 Inclusion Criteria

Both quantitative and qualitative published empirical studies were considered for inclusion. Published studies had to be in English, or translated to English and had to meet the following criteria to be included:

Participants: the study population had to have a mean age of 60 or older. Participants had to reside within a care home facility. The term 'care home' included nursing homes, dual registered care homes, aged care facilities and assisted living facilities. Not-for-profit, private

and public establishments were all eligible for inclusion. The intervention did not have to be carried out within the care home environment.

Intervention: 'Arts for Health' was defined as any activity which may be classed as performing arts, creative arts, visual arts and/or recreational activities, as defined by Arts Council England (2015). Such activities include: performing (music, dance, drama, reading, poetry telling, life story work), creative arts (creative writing, painting, collage, pottery, sewing, knitting and gardening) or recreational activities (visiting museums, galleries, exhibitions, the theatre). Both arts activities that required active involvement by participants and those which included passive involvement were eligible for inclusion. Arts activities could be experienced either as part of a group or individually.

Comparisons: quantitative studies had to include a comparison with either usual care or a different intervention. Studies utilising a within-subjects design were included where there were clear pre and post measures within the results section.

Outcomes: any measure which was related to quality of life, wellbeing or health was eligible for inclusion (including mood, physical health measures, improved socialisation etc).

Study Design: all empirical study designs were eligible for inclusion and both quantitative and qualitative studies were included.

2.4.6 Exclusion Criteria

Studies were excluded where there was a failure to meet one or more of the specified inclusion criteria. Those which involved older people who lived within their own homes were also excluded, unless there were also a proportion of participants who lived within a care home included and their data could be extracted. Studies were also excluded if they focused on traditional reminiscence therapy; only reminiscence studies which involved life

story work and creative activities or outputs were eligible for inclusion. In addition, studies were excluded if they were not empirical (such as reviews, magazine articles with no original data or expert opinions).

2.4.7 Search Strategy

No date restrictions were applied to the search, key databases were searched from inception to December 2016, with continual updates obtained monthly (up until August 2018). Databases included PsychInfo, AHMED, MEDLINE, EMBASE and CINAHL. The Cochrane Library, Joanna Briggs Foundation, and Web of Science Arts and Humanities and Social Sciences Citation Index were also searched. Additional hand-searching of reference lists from systematic reviews and relevant papers was also conducted. It was specified that the returned articles were available in English (where the database search allowed for this specification). Search terms included arts activities: 'dance therapy', 'music therapy', 'art therapy', 'singing', 'reading', 'poetry', 'reminiscence therapy', 'memory box', 'sew', 'knit', 'creative art(s)', 'paint(ing), 'performing art', 'arts and health', 'arts and humanities', 'life story', 'life narrative', 'life review', 'theatre', 'visual art', and 'arts and wellbeing' combined with terms related to care homes: 'nursing home', 'residential home' and 'care home' (see appendix 1 for an example search string). Search terms were piloted prior to use.

2.4.8 Study Selection

The inclusion/exclusion screening form (Figure 2.2) was developed specifically for this study and based on previous examples available from other systematic reviews (Roe 2007, p 11). To ensure relevant studies are identified (Pope, Mays and Popay, 2007; Roe et al. 2007) the form assesses whether the published articles met the PICOS criteria (i.e. participants aged 60 and over, participants who reside in care homes, arts activities and an outcome measure related to quality of life). In order to provide transparency, the reasoning for excluded studies

was also recorded on the form (Pope, Mays and Popay, 2007; Roe, 2007) and the names of each of the reviewers assessing eligibility (Higgins and Green, 2011). The form was then circulated and agreed with members of the team and piloted on an initial set of identified studies. No amendments were required and it has been used to assess all of the potentially relevant studies identified from the searches.

Where there were discrepancies as to whether a study was eligible for inclusion or not, a discussion took place around the reasoning for these disagreements and a consensus was sought. If there was still no agreement, or if there was uncertainty around the eligibility of a study, then this was discussed with another member of the supervisory team in order to resolve this, in accordance with recommended systematic review methods and best practice (Egger, Davey Smith and Altman, 2005; Egger, Davey Smith and O'Rourke, 2008; Webb and Roe, 2007; Liberati et al. 2009; Moher et al. 2009; Higgins and Green, 2011). All studies agreed for inclusion were then eligible for data extraction and quality appraisal.

2.4.9 Data Extraction

All relevant data should be extracted for inclusion within the review, and the same data needs to be extracted and recorded for each of the eligible studies (Pope, Mays and Popay, 2007). Use of a standardised data extraction form lowers the risk of bias and increases the reliability of results (Centre for Reviews and Dissemination, 2009; Moher et al. 2009; Liberati et al, 2009.)

INCLUSION/EXCLUSION FORM

Study ID _____

Title of potential included study

_____
_____
_____

First Author _____
Journal _____
Year _____
Vol _____
Pages_____

First reviewer_____
Second reviewer _____
Date of agreement_____

Study Design _____

Does the study focus on arts for health?

Yes

No

Unsure

If yes, which type?

_____

Does the study include an outcome measure related to Quality of Life/Wellbeing/Health?

Yes

No

Unsure

If yes, what was measured and how?

_____
_____

Does the study involve older adults (60+, or a mean age of 60)?

Yes

No

Unsure

Were the study population care home residents?

Yes

No

Unsure

Decision by first reviewer

Include

Exclude

If excluding, rational for exclusion

_____
_____
_____

Decision by second reviewer

Include

Exclude

Joint Decision

Include

Exclude

Figure 2.2 Inclusion/Exclusion form used to assess the eligibility of studies for inclusion.

An electronic data extraction form was developed and circulated to co-reviewers/ members of the supervisory team for comment and agreement as per recommended practice (Pope et al 2007; Roe 2007; Higgins and Green, 2011). The data extraction form was then used on five initial studies in order to assess its accuracy and structure for documenting relevant data. Following this a final version of the data extraction form was created. An example of the form is shown below (Figure 2.3).

Data extraction was undertaken by the lead reviewer from all included studies using the specifically designed data extraction form (Figure 2.3). Full papers of included published articles were obtained from the Edge Hill University online library, where an article was not available through this medium it was requested through the university's inter-library loan system. The full paper was then read and items were extracted and recorded on the form. The data extraction forms were completed electronically to allow for free text to be typed into each of the boxes.

As data extraction errors are common when carrying out a systematic review (Eden et al. 2011) it is recommended that more than one member of the research team carry out data extraction (Roe, 2007; Centre for Reviews and Dissemination, 2009). Data extraction forms along with the electronic copies of the included studies were distributed equally between the three members of the supervisory team for independent cross-checking. In order to conform to transparency, it is best practice to record all corrections to the data extraction form when checked by a second reviewer (Higgins and Green, 2011) and so these corrections and additional missed items were marked on the original data extraction forms in red.

Study ID	
Author	
Date	
Country of origin	
Study design and dates of study	
Setting	
Aims	
Participant characteristics	
Inclusion criteria	
Exclusion criteria	
Methods (arts activity)	
Outcome measures	
Results	
Intention to treat analysis	
Loss to follow up	
Power calculation/Sample size justification	
Economic data	
Summary of findings	
Notes	
Allocation concealment RCT/CCT	
Quality appraisal score	
Funding source	
Reference	

Figure 2.3 data extraction form used for included studies

One of the papers was missing an average age for the study population and for this case the author was contacted to provide clarification of the age range of the participants and this was recorded on the initial data extraction form. This again adheres to best practice when carrying out a systematic review to ensure that all data are recorded as well as the sources of such information (Liberati et al. 2009; Moher et al. 2009).

2.4.10 Data Items

All data relevant to the research question and PICOS should be identified and extracted for each of the studies (Petticrew and Roberts, 2006; Webb and Roe, 2007) and the data extraction form was created to facilitate this process.

The study ID assigned on the data extraction form was the same as the one assigned for the inclusion/exclusion form to allow for forms to be matched along with the lead author of the study. Further to this the full reference of the paper was recorded. The country of origin was reported along with the design of the study and dates upon which the study took place. The setting and aims of the study were also reported along with the aims of the study.

All reported demographic information relating to the participants were extracted and recorded within the participant characteristics section, including percentages of participants belonging to each (i.e. percentage of males and females). If not reported, then percentages were calculated so that each of the forms contained the same information for each of the studies. Where included, the inclusion and exclusion criteria for participating within the research study were also recorded as reported within the paper, this allowed for additional information of the participants taking part in the study.

The full methods for carrying out the arts activity were extracted and recorded on the extraction form. All information was recorded including the frequency of sessions, what days the sessions were carried out on, whom the sessions were delivered by and the way in which sessions were delivered. All reported outcome measures were also recorded, such information included scales and tools used, observational checklists and qualitative interview schedules. Additionally, all results were extracted from the results sections of the paper and reported. Ensuring that all data from results is obtained and reported can allow

for assessment of the data and meta-analysis to be carried out if appropriate (Khan et al. 2011; Higgins and Green et al. 2011). Any economic information and funding sources was also extracted from the papers. Also recorded was whether an intention to treat analysis was performed, whether there were any participants lost to follow up and details of any reported sample size calculations or justifications and whether allocation concealment took place; all of these items can also help inform the quality appraisal of the study which is discussed further in the next session.

2.4.11 Quality Appraisal- assessing the risk of bias

Inclusion of studies that are poor in quality can result in recommendations and conclusions from a systematic review that are not based upon the best possible evidence (Higgins and Green 2011). Appraising the quality of included studies identifies the methodological flaws within included studies which may impact upon the strength of findings (Petticrew and Roberts, 2006; Khan et al. 2011).

All studies were graded and scored according to their level of evidence using an established level of evidence scoring system (as specified in NICE guidelines, 2012). This scores Level 1a (grade A) evidence as that obtained from systematic reviews, these were not eligible for inclusion within this systematic review. Level 1b (representing high quality) evidence as results obtained from a randomised controlled trial, Levels 2a and 2b (representing moderate quality) as evidence from a well-conducted study without randomisation, level 3 (again, moderate quality) as evidence from a well-designed and carried out descriptive, comparative or correlational study and Level 4 (grade C) which relates to evidence obtained from experts (not eligible for inclusion within this review).

For quantitative intervention studies, the Cochrane GRADE tool was used (Higgins and Green, 2011, shown in Box 2.1). The tool examines risk of bias in relation to several factors which may affect the credibility of results: whether random sequence generation took place and if allocation was adequately concealed, if participants and/or researchers were blinded, completeness of outcome data was complete, any evidence of selective reporting and whether there were any other biases to be considered that may hinder the study in any way. For this review, as blinding of participants and those carrying out the intervention would not be appropriate, blinding was deemed adequate when it was carried out for the persons scoring and/or administering the outcome measures (Schunemann et al. 2013). A score of low risk (+), high risk (-) or unclear risk (?) was determined for each of the factors with rationale explaining each decision, in keeping with the standards for appraising quality in a systematic review (Higgins and Green, 2011; GSchunemann et al. 2013). Meaning, studies that score low risk are of higher quality or there is certainty in accepting the results compared to those that are deemed high risk (lesser quality, higher uncertainty of the results) or uncertain, unable to judge the overall quality or certainty of the results/ evidence.

Criteria:

- Evidence adequate randomisation took place
- Evidence of allocation concealment
- Evidence of blinding of outcome assessment- given the nature of the arts activities it would not be possible to blind those taking part and therefore this was judged to have occurred if the person scoring the outcome assessment was blind to group allocation.
- Complete/Incomplete outcome data
- Selective reporting
- Any other apparent risk of bias

Box 2.1. Criteria based on GRADE (Higgins and Green, 2011) to assess the risk of bias in quantitative studies.

Qualitative empirical studies were assessed using a checklist previously used by Cooke, Mills and Lavender (2010) and developed by Walsh and Downe (2006). An overall grading system created by Downe et al (2009, based on work by Lincoln and Guba, 1985) was used to score each paper based on the above evidence. The criteria for assessing bias and grades assigned are shown in Box 2.2.

Criteria:

- Whether the scope, purpose and aims were presented and clear
- If the study design was appropriate and well explained
- Sampling size adequacy and characteristics
- Whether there was data and evidence to support the authors interpretation of the results
- Evidence of reflexivity from the author
- Whether ethical considerations had been adhered to
- Evidence of relevance and transferability.

Grading:

A- No, or few flaws. The study credibility, transferability, dependability and confirmability is high.
B- Some flaws, unlikely to affect the credibility, transferability, dependability and/or confirmability of the study.
C- Some flaws that may affect the credibility, transferability, dependability and/or confirmability of the study.
D- Significant flaws that are very likely to affect the credibility, transferability, dependability and/or confirmability of the study.

Box 2.2 Quality appraisal criteria for judgement and scoring of included qualitative studies. See Downe et al 2009, Cooke, Mills & Lavender 2010.

Quality appraisal was carried out by the lead reviewer, concurrently with the data extraction. Carrying out the data extraction and quality appraisal process in parallel allows the researcher to focus upon the data and become familiar with the study and its methodology without having to revisit the study at a later date (Centre for Reviews and Dissemination, 2009). In order to maintain reliability and transferability (Webb and Roe, 2007; Schunemann et al. 2013), quality appraisal was cross checked with a second member of the supervisory team (forms were divided equally between the three members of the

team). In accordance with PRISMA guidelines (Liberati et al. 2009; Moher et al. 2009), all quality appraisal scores were recorded on both the data extraction sheet and the corresponding quality appraisal sheet itself and are reported within the results section.

2.4.12 Synthesis of Results

Given the number and heterogeneity of included studies, their differing study designs, methods and outcome measures, data are presented in a narrative descriptive synthesis. This provides an overview and high-level comparison of the different studies, their characteristics, key findings in relation to health, wellbeing and quality of life and their quality appraisal in the first chapter of systematic review results (Chapter 3) and comparison of studies grouped by arts-type in the second chapter of systematic review results (Chapter 4). The quality of included studies is also displayed in Tables and this allowed for synthesis of results from both high- and low-quality evidence to be conducted and compared.

2.5 Summary

This chapter has presented the methodology for carrying out a systematic review, in accordance with guidelines specified by the PRISMA statement (Liberati et al. 2009; Moher et al. 2009) and guidance from the Cochrane Collaboration (Higgins and Green, 2011). An overview of the importance in carrying out a systematic review to inform recommendations for practice is provided in the first half of this chapter. The second half discusses the methods for this systematic review, assessing the impact of arts for health activities on those who reside in care homes. The following chapter (Chapter 3) presents an overall descriptive narrative synthesis of all included studies along with appraisals of quality. This chapter has also been published as a paper (Curtis et al. 2018) (See Appendix 2). Chapter 4 then provides a descriptive narrative synthesis by arts type of included studies, again with appraisals of methodological quality of included studies.

Chapter 3- Study 1: Systematic Review of the arts for health activities and their impact upon the health, wellbeing and quality of life of older people in care homes. Results of the overall high level narrative descriptive synthesis.

3.1 Introduction

Following on from Chapter 2 which outlines the methods for the systematic review, this chapter presents and interprets the results of the systematic review of included studies on arts for health interventions for older people in care homes. The chapter (Chapter 3) begins with an overview and outcome of the search for empirical research and then describes studies selected for inclusion within the review. Given the number and heterogeneity of included studies, their differing study designs, methods and outcome measures, data are presented in a narrative descriptive summary. This provides an overview and high-level comparison of the different studies, their characteristics, key findings in relation to health, wellbeing and quality of life and their quality appraisal. Therefore, an overall synthesis of all included studies is presented along with characteristics of included studies, research designs, aims, locations and settings, years of study, the research aims, study populations, findings and quality appraisal (see Curtis et al. 2018, published paper in Appendix 2). Statistical results of outcomes (where reported for all included studies) are in the included studies table (Appendix 3). This is followed by Chapter 4 which consists of a descriptive narrative synthesis of included studies grouped by arts type, with statistical results (where reported by studies) appear in both the text and each included studies table according to type of arts activities. Where included studies are cited throughout this thesis, they adhere to the Cochrane convention of citing first author and date only in the text and tables.

3.2 Search Outcome

A total of 1090 papers were returned from the electronic database searches and an additional paper was sourced from hand searching, providing a total of 1091 potentially relevant papers. Following removal of duplicates this left a total of 790 papers that were screened for eligibility. A further 624 studies were excluded following title and abstract screening which left 166 papers which underwent full-text screening with the use of an inclusion/exclusion form (see Figure 2.2 in Chapter 2). Of these potentially eligible papers, 74 met the inclusion criteria for the review (see Appendix 4 for reasons for exclusion at full-text screening). A PRISMA flow chart depicting the search, selection and exclusion of potentially eligible studies is shown in Figure 3.1 below.

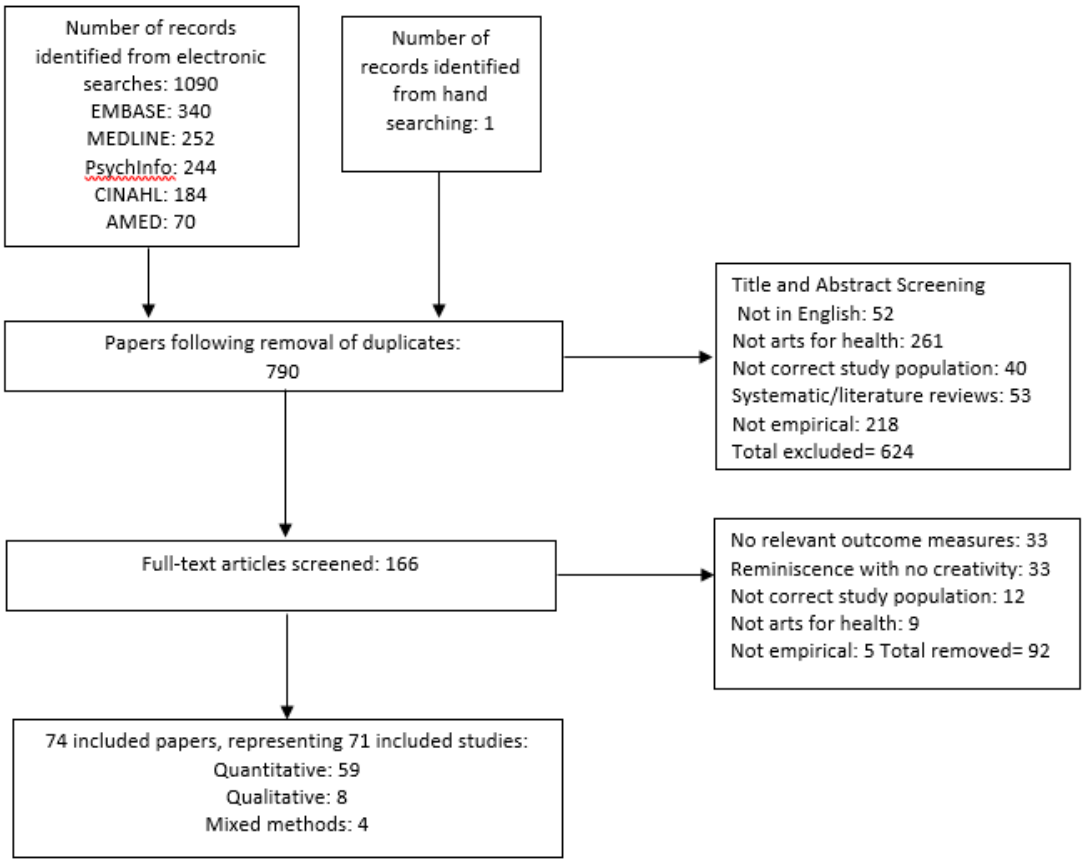


Figure 3.1 PRISMA flow chart of papers retrieved from electronic databases, inclusion and exclusion at each stage and included studies.

The 74 papers represented a total of 71 studies (see included studies Table, Appendix 3). Most (n=59, 83%) were quantitative studies, eight (11%) qualitative and four mixed methods. Of the mixed methods studies, two had separate papers for qualitative (Hammar, 2010; George, 2014) and quantitative findings (Hammar, 2011; Houser, 2014) from the same study. The findings from an additional study were reported across two papers, one reporting outcomes related to mood and depression (Cooke 2010a) and a second reporting outcomes related to agitation and anxiety (Cooke 2010b).

As continual search updates were run until August 2018 potentially eligible studies were still identified until this date. Publication (in January 2018) meant Tables and figures for the systematic review had already been created and therefore an additional study sourced after this date was eligible for inclusion but has not been included in this descriptive narrative synthesis (but will be included in future publications). The study assessed use of visual arts in those living with dementia (Windle et al. 2018) and utilised a mixed methods design. Included were 125 older adults who resided within the UK, with participants recruited from three sites, a hospital, a residential home and those living within their own homes within the community. The wellbeing of those taking part was quantitatively assessed with reference to domains of interest, attention, pleasure, self-esteem, negative affect and sadness (through a developed questionnaire). Interest was shown to statistically improve from baseline to following the programme (28.3 vs. 38.3, $F(4, 173)= 4.3$, $p<0.001$) as did self-esteem (25.5 vs 27.3, $F(4, 1, =2.1$, $p=0.009$). Both attention and pleasure statistically improved following 2 weeks of the activity (55.6 vs. 63.3, $F(4, 158.2, =2.2$, $p=0.03$ and 22.5 vs. 22, $F(4, 162.5, =2.2$, $p=0.02$ respectively) however neither were sustained to the end of the programme. Sadness and negative affect also reduced after 2 weeks (1.7 vs. 0.8, $F(4, 1, =5.1$, $p=0.02$ and 2.3 vs 1, $F(4,$

1, =8.4, *p*=0.004) but again this was not sustained to post-test. Quality of life was also quantitatively assessed (via use of the dementia quality of life scale (DEMQOL) and no significant improvements were reported for this measure. Qualitative changes with respect of quality of life were reported, along with observed behavioural changes.

3.3 Characteristics of included studies

A total of 71 studies were eligible for inclusion (included studies Table, Appendix 3). Published dates of studies ranged from 1983-2015, with increases in publications shown for each decade, a total of 4 (6%) from the 1980s, 14 (20%) from the 1990s, 29 (41%) from 00s – the first decade of the millennium and 28 (39%) within the five-year period 2010 to 2015. Only 15 (21%) of the papers reported actual dates during which the study took place (Ragneskog, 1996; Doric-Henry, 1997; Palo-Bengtsson, 1998; Myskja, 2008; Guetin, 2009; Cooke, 2010a; Cooke, 2010b; Philips, 2010; Chang, 2013; Ridder, 2013; George, 2014; Houser, 2014; Lin, 2014; Liu, 2014; Roe, 2015) and these ranged from 'spring' 1991 to December 2012.

Sixty-three (89%) were intervention studies and eight (11%) descriptive. Of the intervention studies, 59 (83%) were quantitative, four (6%) mixed methods and all descriptive studies (n=8, 11%) were qualitative (see included studies Table in Appendix 3 for included studies and characteristics). The most used study design for obtaining quantitative data (including the mixed methods studies) was uncontrolled pre/post-test, reported by 27 (38%) studies, followed by randomised controlled trials adopted for 23 (32%), 12 (17%) used a quasi-experimental design and there was one retrospective cohort study.

Each included study was categorised into a broad arts type, these will each be presented within the following chapter (Chapter 4) however to provide an overview and context for this overall synthesis, 48 (68%) were classified as the arts type 'music', six (8%) as spoken

and written word, six (8%) as performance, eight (11%) as multisensory activities and three (4%) as applied arts and crafts (see Table 3.1 for included studies and their classified arts type).

Table 3.1 Characteristics of included studies and their identified arts type

Arts Type	Studies classified	
Music	034. Garland (2007)	084. Vink (2013)
	040 Hagen (2003)	085. Lord (1993)
	041. Nair (2011).	090. Bennett (1988)
	043. Casby (1994)	091. Smith (1986)
	046. Ridder (2013)	094. Vanderark (1983)
	047a. Cooke et al. (2010)- Tagged with 047b.	101. Sole (2014)
	047b Cooke et al. (2010)- Tagged with 047a.	106. Hammar (2011)- Tagged with 130.
	048. Sung (2010)	108 Gotell (2007)
	049. Chen (2009)	112 Myskja (2008)
	053. Millard, (1989)	115. Mohammadi (2011)
	055 Liu (2014)	116. Hicks-Moore (2007)
	056. Lin (2010)	120 Suzuki (1998)
	057. Ziv (2007)	121 Carruth (1997)
	058. Holmes (2006)	122. Brontons (1996)
	059. Sung et al. (2006)	123 Clair (1996)
	061 Goddaer (1994)	130 Hammar (2010)- Tagged with 106
	063. Ho (2011)	132 Guetin (2009)
	064. Chang (2010)	139 Burrack (2004)
	068. Svansdottir (2006)	144 Koike (2012)
	069. Ashida (2000)	146 Hicks-Moore (2005)
	070 Ragneskog (1996a)	164 Hong (2011)
	071 Ragneskog (1996b)	168 Clarkson (2007)
	075 Ledger (2007)	169 Cohen-Mansfield (1997)
	082 Remington (2002)	
Spoken and Written Word	010. Subramaniam (2014)	
	017. Billington (2013)	
	031. George (2014)- Tagged with 032	
	032. Houser (2014)- Tagged with 031	
	035. Fritsch (2009)	
	044. Skrajner (2007)	
	173. Philips (2010)	
Performance	098. Houston (1998)	
	074. Palo-Bengtsson (1998)	
	037. Heyn (2003)	
	135. Noice (2006)	
	054. Guzman-Garcia (2013)	
	Vankova (2014)	
Multisensory	065 Simard (2010)	
	076 Chang (2013)	
	118. Kincaid (2003)	
	125 Scott (2014)	
	140 Orsulic-Jeras (2000)	
	162 Klages (2011)	
	202 Cox (2004)	
	200. Roe (2015)	
Applied arts and crafts	097. Doric-Henry (1997)	
	102 de Guzman (2011)	
	153 La Cour (2005)	

3.4 Aims of included studies

Most studies aimed to assess the impact of the activity on more than one outcome measure (for outcome measures of individual studies an included studies Table is located in Appendix

3). The most cited outcome was reduction in behavioural symptoms, including agitation and verbally disruptive behaviours (VDB), reported by nearly half the studies (n= 31, 44%). Second most popular was to determine the impact on psychological wellbeing measures, including depression, anxiety and general wellbeing measures, cited by a total of 24 studies (34%). Eleven (15%) studies reported the aim of assessing cognition, eight (11%) on impacts to quality of life and nine (13%) focused on assessing the influence of the activity on improving care. Lesser reported aims included the feasibility of carrying out the activity (n=6, 8%), engagement of residents (n=5, 7%), physical health (n=4), satisfaction with living environment (n=2), socialisation (n=2), and neuropsychiatric symptoms associated with dementia (n=2).

3.5 Location and setting

A third of the studies took place within the USA (n=24, 33%). The second most cited countries were the UK, Australia and Taiwan, with seven (10%) studies carried out in each. Six (8%) took place in Sweden, five (7%) in Canada and two in Israel. The remaining studies took place in Belgium, Iceland, Norway, France, Holland, Spain, China, Japan, South Korea, Iran, Czech Republic and the Philippines, with one study originating from each country. Only one study included data from more than one country, with participants included from 14 care homes in total, ten from Norway and four from Denmark (Ridder, 2013).

The term used to refer to the care home facility differed across the included studies. Twenty-nine (41%) of the studies used the term 'nursing home', eleven (15%) referred to 'care homes', seven (10%) to 'long term care facilities' and five (7%) reported that residents lived within 'residential homes'. Lesser used terms included 'dementia specialty facility' (n=3), 'aged care facility' (n=3), 'residential facility' (n=2), senior living residence (n=2) 'residential units' (n=1), 'care facility' (n=1), care retirement community (n=1), 'continuing care facility'

(n=1), ‘memory care residence’ (n=1) and ‘supported living facility (n=1). Five (7%) did not specify a term for where the participants resided.

Just one study reported that the activity occurred outside of the care home (Roe, 2015), where residents were taken to a museum and art gallery. Sixty-nine percent (n=49) of the studies reported the activity took place within the care home itself. The remaining 21 (%) did not specify the care home, however from the information given on the method behind the arts activity it would be implied that they were conducted within the homes.

3.6 Populations

The total number of residents taking part in the studies was 2086 (reported for 70 of the 71 included). Sample sizes ranged widely from three up to 117. Just over 70% (n= 51) of the included studies reported the mean age of residents and across these the mean age of residents was 84 years, with a range in mean ages between 69 and 87.5 years. A total of 44 (86%) out of these 51 studies recorded their mean sample as aged 80 years and above. Six (12%) studies reported a mean age greater than seventy (Casby, 1994; Kincaid, 2003; Garland, 2006; Sung, 2006; Ho, 2011; Hong, 2011). There was only one study which reported a mean age of lower than seventy (Mohammadi, 2009) however this was extremely close at 69.4 years.

Gender of those taking part were reported by 62 (87%) studies, the total number of women was 1560 (75% of the 2068 participants reported) and the number of men was therefore 508 (25%). Three studies reported more men than women (Sung, 2006; Mohammadi, 2009; Sung, 2010), two reported equal numbers of men and women (Ragneskog, 1996b; Hilliard, 2004) and two featured women only (Carruth, 1997; de Guzman, 2009).

Almost half the studies (n=35, 48%) specified inclusion criteria that residents required a diagnosis of dementia. One of the studies whilst not specifying dementia as an inclusion criterion did specify the requirement for participants to have 'cognitive difficulties' (Klages 2011). In addition to this Simard (2010) reported most participants to have 'severe' Cognitive Performance Scale (CPS) scores and Martin (2004) reported that no one was excluded from the study because of their cognitive abilities.

Five studies (Bennett 1988; Houston, 1998; Suzuki, 1998; Burrack, 2004; Chen, 2009) featured inclusion and exclusion criteria that may have excluded those with dementia from being eligible to take part, particularly during the latter stages of the disease. Houston (1998) and Hagen (2004) both excluded participants based on 'severe cognitive impairment'. Similarly, Bennett (1988) specified 'speech or hearing impairments' and 'mental dysfunction or confusion' as an exclusion criteria and Chang (2013) included only those with 'good cognitive functioning'. Suzuki (1998) outlined the need for 'adequate reality orientation' for participants to be able to complete memory and psychological testing. Three studies used mini mental state examination (MMSE) scores as part of inclusion criteria, Burrack (2004), Chen (2009) and Vankova (2014) specified cut-off score of 21, 24 and 15 respectively. Scott (2014) also specified inclusion was only eligible for participants with adequate cognition, with the MMSE used to determine this, however they did not report a cut-off score.

3.7 Research methods and outcome measures

Assessing the impact of the activity on changes in behaviour was the most commonly reported aim (n=31, 44%). Sixteen studies (23%, fifteen of which featured music, see included studies Table, Appendix 3) assessed this via use of the Cohen-Mansfield Agitation Inventory (CMAI). Five used observational methods to determine behavioural changes, of which only one used a structured observation outcome measure, the Behavioural Pathology in

Alzheimer's Disease rating scale (BEHAVE-AD) (Svansdottir, 2006), the remaining four studies devised their own observation schedule. Kincaid (2003) assessed behavioural changes through recording incidences of door-testing behaviours with use of a specifically created observation measure. How engaged residents were with the activity were also assessed, using the Menorah Park Engagement Scale (MPES) (Heyn, 2003; Skrajner, 2007) and the Myers-Research Institute Engagement Scale (MRI-ES) (Orsulic-Jeras, 2000). Both staff and resident engagement were also rated via specifically-devised questionnaire in Scott (2014) and symptoms associated with dementia explored via the Neuropsychiatric Inventory Questionnaire (Philips, 2010; Billington, 2013).

The second most explored outcome was the impact of the activity on psychological wellbeing, there was a wide variation in outcome measures used to evaluate this outcome, with most using validated outcome measures. The most commonly assessed area of psychological wellbeing was changes in depression levels. Five studies evaluated changes in depression via the Geriatric Depression Scale (GDS) (Guetin, 2009; Cooke 2010a; Scott, 2014; Subramaniam, 2014; Vankova, 2014) and four used the Cornell Scale for Depression (CSD) (Ashida, 2000; Orsulic-Jeras, 2000; Clarkson, 2009; Philips, 2010). Other measures of depression included the Montgomery Asberg Rating Scale (Myskja, 2006), Gottfrees-Brane-Steen Scale (GBS) (Ragneskog, 1996b), Self-Rating Depression Scale (SDS) (Liu, 2009) and Beck Depression Inventory (Doric-Henry, 1997). Anxiety were measured using the Rating Anxiety in Depression tool (RAID) (Cooke 2010a; Sung, 2010), Hamilton Scale for Anxiety (Guetin, 2009) and the Geriatric Anxiety Inventory (GAI) (Scott, 2014). Five studies collected information recorded on prescriptions of psychotropic medications often used to moderate anxiety and agitation (Brontons, 2009; Clarkson, 2009; Simard, 2010; Ridder, 2013; Houser, 2014). Two studies used

one outcome measure to assess both anxiety and depression using the Depression and Anxiety Stress Scale (DASS-21) (Mohammadi, 2009) and the Hospital Anxiety Depression Scale (HADS) (Houston, 1998) respectively. Measures of general mood and affect were assessed almost exclusively obtained via observational methods using the Observed Emotion Rating Scale (OERS) (Philips, 2010; Hammar, 2011), Positive and Negative Affect Schedule (PANAS) (Suzuki, 2006), Philadelphia Geriatric Affect Rating Scale (Fritsch, 2009) and the Affect Rating Scale (Cox, 2004). The emotional wellbeing subscale of the GENCAT scale was scored in Sole (2014). The Dementia Mood Assessment Scale (DMAS) was completed in Koike (2012), although it is unclear if this were completed by the residents themselves. Un-validated mood measures were reported for just three studies (Lord, 1993; Heyn, 2003; Sole, 2014) and Houser (2014) used patient notes to track potential mood changes.

Measures adopted to assess quality of life included the Quality of Life in Alzheimer's Disease (QoL-AD) (Burrack, 2004), Dementia Quality of Life Scale (DQoL) (Cooke, 2010a), Alzheimer's Disease Rated Quality of Life (ADRQ) (Ridder, 2013) and the Life-Satisfaction Index-A (LISA) (Bennett, 1988), Vanderark (1983) created their own measure and Hilliard (2004) assessed length of life using patient notes. The two studies reporting on satisfaction with living environment both devised their own outcome measures for assessing this outcome (Chang, 2013; Scott, 2014), with the first also using the study to validate the questionnaire (the Satisfaction with Living Environment Nursing Home (SLE-NHS)).

Cognition was explored via scoring of the Mini Mental State Examination (MMSE) (Smith, 1986; Hong, 2011; Koike, 2014), Memory Retrieval Test (Suzuki, 2006), Cognitive Assessment Scale (CAS) (Hagen, 2003), Functional Assessment of Communication Skills (FACS) (Philips, 2010) and Autobiographical Memory Interview (AMI) (Subramaniam, 2014). Simply recording

incidences of face-name-recognition (Carruth, 1997) and 'alert responses' (Clair, 1996) were also used. Noice (2006) featured a cognitive battery of tests which included measures of word recall, listening span and the means end problem solving procedure (MEPS). Measures of physical health were less commonly reported, but measures were balance and joint range (Hagen, 2003), oxygen saturation, temperature, blood pressure and pulse (Koike, 2012) and balance tests (Klages, 2011). Only one study measuring this outcome used a standardised measure, the Multidimensional Observational Scale of Elderly Subjects (MOSES) to assess 'functioning' (Orsulic-Jeras, 2000).

Measures of behavioural changes and agitation were therefore always obtained via observations of residents. Expressions of 'emotion' or 'affect' were mostly obtained through observational methods, aside from Koike (2012) who did not report who completed the outcome measure and those which devised their own measures, where the resident themselves completed the scales. Quality of Life measures were always administered to the residents themselves, as were most depression measures, except for Orsulic-Jeras (2000) and Myskja (2009), both of whom reported care staff had completed perceptions of depression in residents. Similarly, anxiety was rated by the residents in all but one of the studies (Sung, 2006).

3.8 Quantitative Findings

3.8.1 Behaviour

With respect of agitated behaviours assessed via the CMAI, Goddaer (1994) Hicks-Moore (2007), Garland (2007), Ledger (2007), Lin (2010), Ho (2011) and Vink (2013) all reported changes over time such that at the end of the study scores were significantly improved for those taking part in the arts activity (included studies table, Appendix 4). Despite this, four of the five studies that featured a comparison group (Remington, 2002; Hicks-Moore, 2007;

Ledger, 2007; Vink, 2013) failed to show a significant difference between the control and comparison group. The comparison group in Vink (2013) were reported as taking part in occupational therapy type activities and Remington (2002) and Hicks-Moore (2007) both compared the music activity with hand massage. Ledger (2007) did not report information as to what activities were offered within the five homes not taking part in music therapy. In contrast, Sung (2006) and Lin (2010) did show significant differences between the scores obtained following the programme for the intervention and control group, who were reported as partaking in 'usual daily activities', suggestive that music therapy was beneficial in the absence of any other form of activity or structured programme. Further evidencing this, Garland (2007) showed both the arts activity (music) and comparison activity (simulated family presence) to significantly reduce agitation, but not one more so than the other. Despite this, both were more effective than the third control condition which involved no activity. Comparison of classical and preferred music revealed both were effective in reducing CMAI scores, but individualised music reduced throughout the time they were listening, whereas effects for the classical music condition were noted only within the final ten minutes (Gerdner, 2000).

Svansdottir (2006) failed to record any significant difference in BEHAVE-AD scores, these fell for both groups over time and Hagen (2003) also reported both groups displayed reduced unwanted behaviours over time, however these were not compared statistically. Orsulic-Jeras (2000) and Houser (2014) also did not report a significant effect of the activity on behaviours whereas Brontons (2009) did report significantly reduced scores at post-test, however these did not reduce over time and fluctuated throughout the course of the study. Simard (2010)

also recorded significant changes in behaviour, however this was determined via sub analysis and was based upon low baseline levels of social interaction.

Hicks-Moore (2005) also described decreases in prevalence over time, however no statistical analysis of these results is offered and therefore the significance of such changes cannot be determined whereas Cohen-Mansfield (1997) noted there were no significant differences over time, nor between the two groups, however p values were not reported within this study. Disruptiveness and agitation were described separately in Ridder (2013), disruptiveness values significantly increased for the control group and decreased in the music therapy group, for agitation a similar pattern was also described, however not for a significant degree. Cooke (2010b) described increases in verbal agitation for those taking part in the activity, however only for those attending 50% or more of the sessions, suggesting residents became more vocal within the sessions as they got used to them.

Those specifically focused upon verbally disruptive behaviours reported differing results. Cohen-Mansfield (1997) evidenced a decrease in these behaviours over time, although these were not significantly more so than the control group. Very limited findings were reported by Casby (1994), who featured only four participants and described how agitation fell for the one of the residents, but not the other. Again, this was evidenced only by celebration lines as the small sample size would have restricted any statistical analysis. Evidence of verbally disruptive behaviours increasing were reported for Nair (2011) who described a significant increase in records of such behaviours within patient charts immediately following the music (afternoon shift) which did not follow on to the evening shift, therefore indicating this effect was short lived. Increases in verbal agitation were also reported in Clarkson (2007) although it is not known whether these were with respect of positive or negative behaviours, with anecdotal

evidence from nurses lending evidence for positive verbal expressions. Further evidence for this is offered by Ziv (2007) which was the only study to statistically analyse expressions of both positive and negative verbal behaviours who determined significantly more positive behaviours occurred with music than without. Furthermore, Sole (2014) reported 'all behaviours which occurred more than once were positive' but did not analyse these results further and Millard (1989) whilst not explicitly stating changes in VDB as an aim, reported statistically more incidences of verbal participation.

Engagement was significantly improved for two studies focused upon creative storytelling (Skrajner, 2007; Fritsch, 2009). Similar findings were reported by Orsulic-Jeras (2000) who found that levels of constructive engagement increased significantly, whereas passive engagement decreased, indicative of greater focus and interaction during the activity (Montessori programming). Limited evidence was provided for engagement with music activities, descriptive accounts were provided by Heyn (2003) who reported '69.4% of residents were engaged in more than half the activity' and Millard (1989) concluded greater incidences of 'sitting behaviours' were indicative of increased engagement, along with descriptions of greater attendance at the sessions compared with other activities within the home.

None of the studies specifically focused on dementia symptomatology showed significant improvements (Philips, 2010; Billington, 2013; Houser, 2014). The only study to assess door testing (Kincaid, 2003) showed a statistically significant fall in exit-seeking behaviours.

3.8.2 Psychological Wellbeing

Nine studies (13%) analysing depression found significantly reduced scores over time (Ragneskog, 1996b; Doric-Henry, 1997; Houston, 1998; Myskja, 2006; Guetin, 2009;

Mohammadi, 2011; Koike, 2012; Liu, 2014; Vankova, 2014). Ragneskog (1996b) showed limited findings of improvements for one type of music only and Guetin (2009) did not show improvements at the end of the study but did at six-week follow-up. Both Suzuki (2006) and Liu (2010) featured participants who had pre-existing clinical depression and found depression levels in these individuals to be significantly improved by the activity. Two studies which failed to show significant improvements for depression levels were different in their study design (one RCT, one uncontrolled pre/post-test), the outcome measures used and who these were completed by (i.e. resident or proxy) however both used live music performances within their studies (Clarkson, 2007; Cooke 2010a). It may therefore be any positive effects of taking part in the activity did not carry over past the experience of the music. Philips (2010), Houser (2014) and Subramaniam (2014) described no significant changes in depression scores for those taking part in spoken and written word activities.

All but one of the studies (Cooke 2010) which assessed changes in anxiety showed a significant reduction in related scores over time (Houston, 1998; Svansdottir, 2006; Guetin, 2009; Sung, 2010; Mohammadi, 2011). Additionally, the four studies which featured a comparison group all showed significantly lower anxiety scores at follow-up for those who had taken part in the arts activity (Houston, 1998; Svansdottir, 2006; Guetin, 2009; Sung, 2010).

Three studies focusing on general wellbeing and affect demonstrated significantly improved scores after taking part in the activity (Lord, 1993; Suzuki, 2006; Hammar, 2010; Sole, 2014) with Lord (1993) also describing improved scores compared with the control group. Suzuki (2006) did not show improvements in apparent positive affect but did described significant reductions in expressions of negative affect. In direct contrast, Hammar (2010) described significantly increased positive emotional expression, with no changes shown in negative

expressions. Significantly improved emotional wellbeing scores were described for the whole sample in Sole (2014), but sub analysis revealed this was not the case for those with moderate dementia. This study used observational methods and improvements in expression of emotion may have been more difficult to observe in those with more advanced dementia.

Of the studies which reported on p.r.n. (as needed) medication usage, two concluded there was not enough information for analysis (Clarkson, 2007; Brontons, 2009) and one reported no significant changes (Houser, 2014). Two evidenced an impact of the activity, one via increased prescriptions of anti-psychotic medications for the control group (Ridder, 2013) and the second through decreased anti-anxiety prescriptions during the study (Simard, 2010).

3.8.3 Cognition, quality of life and socialisation

Evidence of benefit on general quality of life and wellbeing measures was limited, with improvements report by six (12%) (Bennett, 1988; Hagen, 2003; Hilliard, 2004; Philips 2010; Chang, 2013; Subramaniam, 2014). However, even these findings were tentative, Hilliard (2004) assessed length of life and determined that those taking part in music therapy lived longer, however, they failed to account for the fact these individuals may have been more well than those not taking part and therefore may have had greater life expectancy related to advanced disease and irrespective of the activity. Chang (2013) also reported positive findings for only one measure of satisfaction with the living environment, associated with recalling pleasant memories which may indicate only improvements in memory rather than quality of life. Sole (2014) reported a worsening of quality of life scores, which occurred for both the intervention and control group, although not to a significant degree. Both Hagen (2003) and Subramaniam (2014) reported improvements for those who took part in the activity at post-test were not sustained to follow-up, where both groups revealed equal improvements.

With respect of cognition, two studies using the MMSE found significant improvements over time for those taking part in the music activity (Smith, 1986; Hong, 2011) and the third failed to show significant differences (Koike, 2014). They were also significantly improved compared with the control group for the randomised controlled trial (Hong, 2011). Both Hagen (2003) and Subramaniam (2014) reported improvements in cognitive assessment and memory scores respectively over time, however comparison with OT activities (Hagen, 2003) and those who received autobiographical books as gifts (Subramaniam, 2014) reported the groups were not significantly different at follow-up. Hagen (2003) did report a signficiant difference with a control group who took part in no activities. Similarly, Clair (1996) found alert responses increased for the music condition significantly more than during silence, but not significantly more so than for the reading condition. Unpleasant memories decreased for those taking part in music reminiscence, but the percentage of pleasant memories did not increase (Suzuki, 2006) whilst working memory did not significantly improve as a result of drama in Noice (2006), although word recall and problem solving did. Social communication was not shown to significantly increase (Philips, 2010) and there were very limited findings for face-name-recognition where it was reported 'four participants showed an increase, three did not' with no statistical comparisons, possibly due to the limited sample size.

3.8.4 Improvements to Care and Physical Health

Three quantitative studies reported the aim of assessing quality of care. Two of these explored similar outcomes (Ragneskog 1996a; Ragneskog, 1996b) focusing on nutritional intake during mealtimes. The first reported nurses fed the residents significantly more during the music conditions and residents spent more time with their meals. The second that residents were given larger portions and consumed more food. Hammar (2011) was the only study to focus on restiveness to care. It was found that significantly less pulling away, grabbing

objects and behaviour adduction were observed during the MTC condition as opposed to the normal morning care routines.

With respect of physiological measures, Hagen (2003) did show significant improvements for joint function and balance, although balance measures did decrease again following the programme and significant reductions in temperature and pulse rate were found for Koike (2012), although oxygen saturation and resting heart rate were not significantly changed. For Heyn (2003) resting heart rate was shown to decrease, as would be expected given this study evaluated musical exercise, but weight and blood pressure were not significantly impacted upon.

3.9 Qualitative Findings

Qualitative findings overlapped with respect of identified themes and benefits. All the qualitative studies outlined improvements to wellbeing. Specified by several of these was the ability for the activity to improve sense of self in residents (Martin, 2004; Chen, 2009; DeGuzman, 2009; La Cour, 2010; Billington, 2013). These five studies also all reported on the participation of residents, despite physical barriers which may exist. Creativity and the opportunity to take part in meaningful activity were reported by La Cour (2005), Chen (2009), Billington (2013) and George (2014). Cognitive improvements were reported by three studies including benefits to listening, memory and attention (Billington, 2013) and concentration (Palo-Bengtsson, 1998).

Difficulties with the usual care process were reported by studies which conducted interviews with staff members (Gotell, 2007; Hammar, 2010; George, 2014) and this was contrasted with caring during and following the activity which was reported as being more unified and cooperative. The ability for staff members to use their creativity from the sessions for

problem solving during work was also highlighted by both Billington (2013) and Guzman-Garcia (2013) along with general improvements to the overall atmosphere of the care home (George, 2010). Improved communication between the residents and staff members was also evidenced (Gotell, 2007; Hammar, 2010; Roe, 2015).

Increased socialisation was also reported in terms of interacting with other residents and between the care staff and residents (Martin, 2004; Chen, 2009; Billington, 2013; Guzman-Garcia, 2013; Roe, 2015). Sharing of memories prompted by the activity were highlighted in all five of these studies as aiding and promoting these interactions.

The feasibility of carrying out the activity was described within Cox (2004), Billington (2013) and Roe (2015). Aspects of the programme discussed included physically situating residents at the activity, staff requirements, physical barriers to participation and carrying out the activity itself. Two studies which featured changes in environments Cox (2004, snozelen and garden activity) and Roe (2015, museum and gallery) both described the benefits of the experience of novel surroundings, however both also contained accounts of times when this was distressing for certain residents.

3.10 Quality appraisal of included studies

3.10.1 Levels of evidence

Levels of evidence scores were assigned to each of the studies. Twenty-three (32%) studies provided level Ib evidence (the second highest level and highest eligible for inclusion within this review) based on an RCT, eleven (15%) provided level IIa (moderate evidence) results based on well-designed controlled study without randomisation and most, 29 (41%) received a level IIb (moderate quality) based on well-designed quasi-experimental studies. The remaining eight studies (11%) corresponded to an evidence score of level III (the lowest quality acceptable for inclusion in this review), well designed descriptive studies.

3.10.2 Randomised Controlled Trials

The summary of RCT judgements of quality can be found in Figure 3.2. Twelve of the RCTs provided adequate accounts of randomisation to obtain a low risk score for this measure and nine of these also gave detail as to methods of allocation concealment. Blinding of outcome assessment was reported for a total of four of the studies, with nine reporting no blinding to have taken place. The remaining studies did not provide enough information to make a judgement either way.

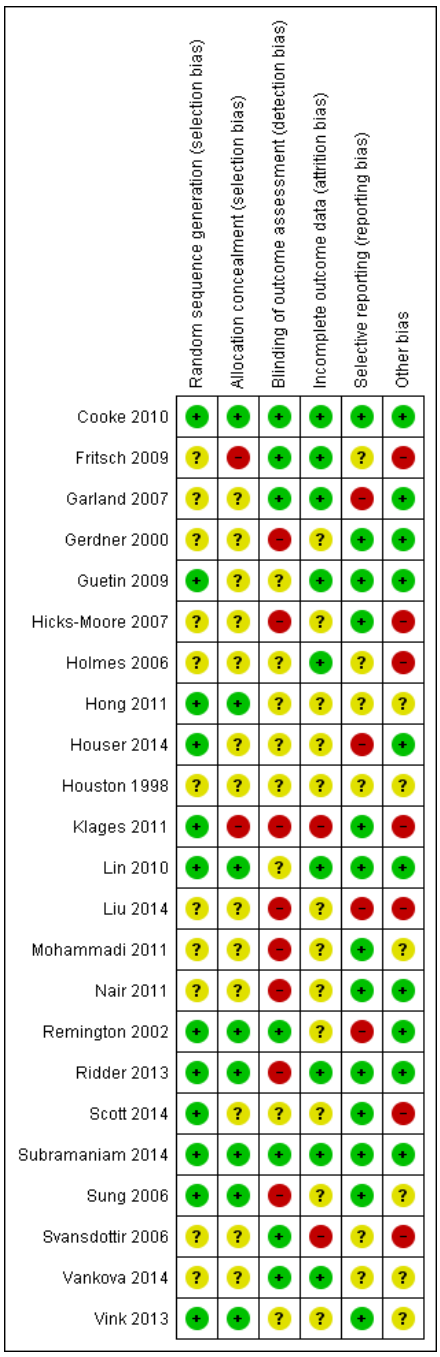


Figure 3.2 Summary of Risk of Bias for RCTs assessed using GRADE criteria where green indicates low risk, yellow unclear and red high.

Nine studies received judgements of low risk for incomplete data, just two receiving high risk judgements, and the remaining studies did not contain enough data to make a judgement. Fourteen appeared to be free of selective reporting, four showed selective reporting and the remaining studies did not provide enough information. Finally, eight studies possessed other biases which may affect the credibility of the studies.

3.10.3 Quasi-randomised controlled trials

Summary of quasi-RCT judgements is presented in Figure 3.3. Two of the quasi-RCTs demonstrated blinding of the individual scoring the outcome assessment, five stated this did not happen (therefore judged as high risk) and four did not contain enough information to make a judgement.

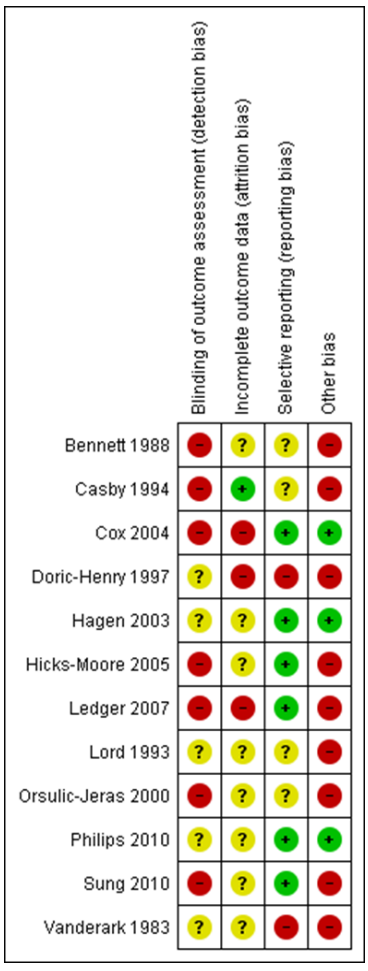


Figure 3.3 Summary of Risk of Bias for quasi-RCTs assessed using GRADE criteria where green indicates low risk, yellow unclear and red high.

Eight were unclear as to whether there was any missing outcome data, two were judged as high risk and just one was deemed to be low risk. In terms of selective reporting, there was only evidence of this occurring within one study, a further six were judged as low risk and it was not clear whether this had taken place in the remaining two studies. Two demonstrated risks of other biases, six were deemed low risk and two did not provide enough information.

3.10.4 Uncontrolled Pre/Post-test

Summary of uncontrolled pre/post-test judgements is shown in Figure 3.4. Evidence of blinding of outcome assessments was shown for just two studies.

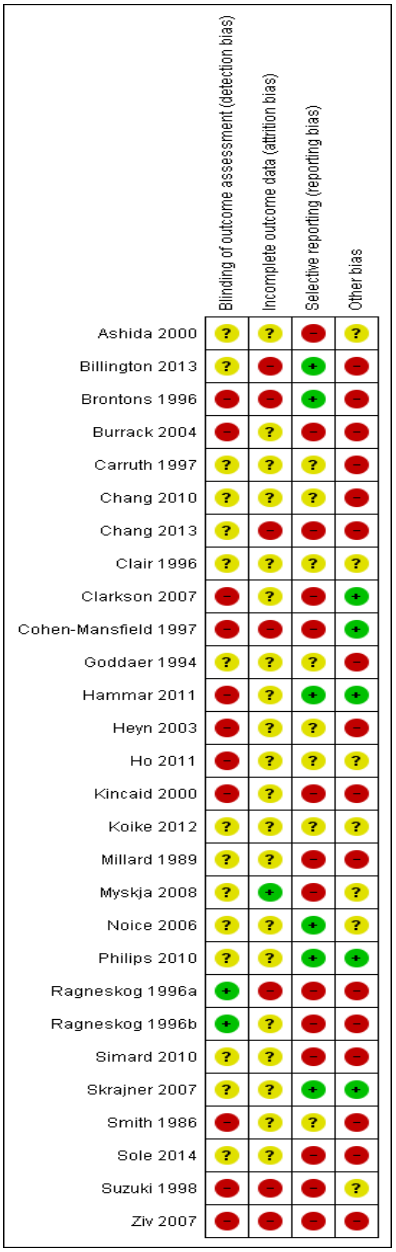


Figure 3.4 Summary of Risk of Bias for uncontrolled pre/post-test studies assessed using GRADE criteria where green indicates low risk, yellow unclear and red high.

Blinding of outcome assessment was reported to have not occured for ten studies, and there was insufficient information to form a judgement about the others.

Incomplete data was evidenced for seven studies, with eighteen unclear and just one judged as having a low risk of bias. Selective reporting was also apparent in fourteen studies, with just five receiving a judgement of low risk for this category. Finally, most of the studies (15 in total) showed evidence of other biases, with only five low risk judgements for this measure.

3.10.4 Qualitative studies

Qualitative studies were appraised using a checklist (Cooke, Mills & Lavender, 2010, developed by Walsh & Downe, 2006, see Chapter 2) where studies were awarded a grade A if they had no or few flaws, grade B if they contained some flaws but these were unlikely to affect confidence in findings, grade C if they had some flaws that were likely to affect confidence in findings and grade D if they had significant flaws which were highly likely to impact on the confidence which could be placed in the study's findings (Downe et al. 2010; Cooke, Mills & Lavender, 2010). Quality of qualitative studies was judged to be good, although none received a grade A. Five studies (La Cour 2005; Gotell, 2007; deGuzman, 2013; Guzman-Garcia, 2013; Roe, 2015) were classified as Grade B for risk of bias and three judged at grade C (Palo-Bengtsson, 1998; Martin, 2004; Chen, 2009). None were judged at a grade D.

3.11 Summary

This chapter presented the results from the systematic review in the form of a high-level descriptive narrative synthesis. In total 71 studies were included within the review providing both quantitative and qualitative findings. Studies adopted a range of designs and methods, along with a number of different aims and outcome measures. Different arts for health activities were evaluated, although the largest proportion of studies were focused on music.

Even within those focused on this arts type there was a large variation in the type of activity offered and methods of delivery.

Studies assessed the impact of arts for health activities on behaviour, psychological wellbeing, cognition, quality of life, socialisation, improvements to care and physical health. There was evidence of benefit for older adults for agitation, with less benefits associated with other behavioural symptoms (verbally disruptive behaviours, irritability and restlessness). Levels of depression and anxiety were also shown to be reduced in those taking part in arts for health, particularly for those with pre-existing clinical need. Quantitative evidence for improved quality of life and general wellbeing was limited, however this benefit was reported qualitatively. Cognitive changes were also evidence, although not widely explored within the studies.

As with the aims, methods and findings there were varying levels of quality across the included studies. RCT's tended to feature less high-risk judgements than the other studies, however only two (Subramaniam, 2014; Cooke 2010) were judged to be low risk of bias for all domains. Qualitative evidence was of a good quality, with most (n=5) being judged at Grade B and the remaining three at Grade C, although none received a judgement of Grade A (with no flaws).

This chapter has provided a high-level descriptive narrative synthesis of all included studies within the systematic review. Chapter 4 which follows is split into five subsections, each providing a descriptive narrative synthesis of the studies featuring that arts type. As with this chapter, the characteristics and findings of included studies will be synthesised, along with judgements of methodological quality.

Chapter 4 Study 1: Systematic review of the arts for health activities and their impact upon the health, wellbeing and quality of life of older people in care homes. Results of the narrative descriptive synthesis by arts type.

4.1 Introduction

This thesis consists of two studies that assessed the impact of arts for health on the health, wellbeing and quality of life of older adults in care homes. The preceding two chapters focused on Study 1, the systematic review of potential benefits to older people residing in care homes who participate in arts for health activities. An overview of systematic review methods is given in chapter 2 and the previous chapter (chapter 3) provides a narrative descriptive synthesis of all the included studies.

In this chapter a narrative synthesis of included studies is reported according to arts type. To determine whether there were any differences between different arts types, studies were categorised based upon the arts activity they focused upon. Five categories were identified: music, literary spoken and written word, multisensory activities, performance and applied arts and crafts. The largest of these categories was music, which featured 48 studies (68% of the total number of studies included within the systematic review). The categories of 'multisensory activities' and 'performance' each had eight studies (each 11% of the total number) and literary spoken and written word featured six studies (8%). Just three studies were categorised as assessing 'applied arts and crafts' (4% of the total number). A narrative synthesis of studies according to arts type is reported within this chapter for each of these categories.

As with the overall high-level descriptive narrative synthesis each subsection outlines the characteristics of the included studies, their aims, location and population. The results from included studies are also presented, along with quality appraisal of study methods.

4.2 Narrative descriptive synthesis of studies featuring music.

Studies were classified within this category when the activity primarily consisted of any form of musical content with either active or passive participation from older adults. Most studies within the systematic review evaluated this arts type (n= 48, 68% of the total number of included studies). Table 4.1 shows the included studies Table for the arts type of music. A descriptive narrative synthesis of included music studies is presented, along with appraisal of methodological quality. As this section focuses upon the music studies only percentages within this section will be relative to the total number of music studies unless otherwise stated.

4.2.1 Characteristics of included studies

Publication dates of studies ranged from 1983 to 2014, with four (8%) published in the 1980s, 10 (21%) during the 1990s, 20 (42%) during the 2000s and 16 (33%) between 2010 and 2014 (see Table 4.1). A total of 45 (94%) were intervention studies and three were descriptive (6%). All but four studies were exclusively quantitative (n=44, 92%), three presented qualitative findings (all descriptive studies, Martin, 2004; Gotell, 2007; Chen, 2009) and one used mixed methods (Hammar, 2010; Hammar, 2011). An additional study (Liu, 2014) reported to contain 'qualitative elements', however no qualitative data were included and so it has been classified as quantitative only (see Table 4.1).

Table 4.1 Characteristics of included studies featuring the arts type music

Reference of Paper, Study ID & Research Design	Location	Population	Activity	Aims of the study	Outcome measures	Findings	Quality
034. Garland (2007) Intervention Dates not specified Quantitative Cross-over RCT with placebo	Australia Participants were residents at a total of 9 nursing homes within Australia.	30 participants in total. Mean age: 79 years, range 66-93. 19 (63%) female and 11 males (37%).	15 minute audiotapes of simulated family presence and preferred music (obtained from family members) were compared with a neutral audiotape. Interventions were applied once a day for three days during weeks 2, 3 and 4 and participants were randomly assigned to a group (A B C) and rotated through treatments. This was compared with usual care.	To compare two interventions; preferred music and stimulated family presence to determine which, if any, was more effective at reducing behavioural disturbances in individuals with dementia who resided within a care home. A comparison group receiving usual care was also included.	Trained researchers observed the participants and recorded whether or not behaviours were present or absent during two minute intervals prior, during and following exposure to the 15 minute tapes, each observation period lasted for a total of 45 minutes. Behaviours were categorised into physically aggressive agitation, physically nonaggressive agitation, verbally aggressive agitation and verbally nonaggressive agitation but varied amongst the participants.	2880 target behaviours were observed in total for this phase. 3.8% were physically aggressive, 64.8% were physically non-aggressive and the remainder were non-aggressive and aggressive verbal behaviours. Physically aggressive behaviours were not included in the analysis. Actual counts of physically agitated behaviours decreased by 30% for the simulated family presence, 25% for music and 15% for placebo. Actual counts of verbally agitated behaviours decreased by 33% for the simulated family presence, 18% for music and 29% for placebo. MANOVA showed a significant time-treatment interaction (F12, 18 =2.62, p=0.032).Separate analysis of each of the types of agitation uncovered significant time-treatment interactions for both physical agitation (F6, 174=2.42, p=0.029) and verbal agitation (F3.71,107.42 =3.37, p=0.014). *Difference in physical mean behaviour counts before and during treatments*: simulated presence vs. music difference in means, 0.31, F 0.77, p 0.388. Simulated presence vs. placebo difference in means 0.84, F 8.29, p 0.007. Simulated presence vs. usual care difference in means 1.12, F 10.23, p 0.003. Music vs. placebo difference in means 0.53, F 3.06, p 0.091. Music vs. usual care difference in means 0.81, F 4.67, p 0.039. Placebo vs. usual care difference in means 0.28, F 2.42, p 0.369. *Difference in mean physical behaviour counts during and after treatments*: simulated presence vs. music difference in means, -0.26, F 0.51, p 0.483. Simulated presence vs. placebo difference in means -0.34, F 1.13, p 0.296. Simulated presence vs. usual care difference in means -0.44, F 1.46, p 0.237. Music vs. placebo difference in means -0.09, F 0.07, p 0.790. Music vs. usual care difference in means -0.19, F 0.27, p 0.605. Placebo vs. usual care difference in means -0.10, F 0.15, p 0.705. *Difference in mean verbal behaviour before and during treatments:* simulated presence vs. music difference in means, 0.43, F 1.44, p 0.239. Simulated presence vs. placebo difference in means 0.23, F 0.46, p 0.504. Simulated presence vs. usual care difference in means 0.96, F 4.78, p 0.037. Music vs. placebo difference in means -0.19, F 0.57, p 0.458. Music vs. usual care difference in means 0.53, F 2.86, p 0.101. Placebo vs. usual care difference in means 0.73, F 5.18, p 0.030. *Difference in mean verbal behaviour during and after treatments*: simulated presence vs. music difference in means, -0.39, F 3.59, p 0.068. Simulated presence vs. placebo difference in means 0.26, F 1.18, p 0.287. Simulated presence vs. usual care difference in means -0.11, F 0.21, p 0.654. Music vs. placebo difference in means 0.66, F 8.29, p 0.007. Music vs. usual care difference in means 0.29, F 2.46, p 0.127. Placebo vs. usual care difference in means -0.37, F 2.52, p 0.124 *Magnitude of change:* of the 30 participants, 11 experienced a fall of 50% or more in physical and/or verbal agitation while simulated presence tapes were played and 15 experienced such a fall while listening to preferred music.	GRADE score: ? ? + + - + Level of evidence II:2, grade B.
040 Hagen (2003) Intervention Dates not specified. Quantitative Quasi-RCT	UK Three 30 bedded wards within long term care facilities.	20 from each ward. Each ward was assigned a condition (Exercise Therapy (ET), Occupational Therapy (OT) or Control (C)) Demographic information not reported.	Musical exercise ET group: sessions lasting 40 minutes, three times a week for a total of 10 weeks. The programme involved movement to music from the 20s-40s and involved mild stretching and aerobics. The OT group: lasted for an hour, three times a week for 10 weeks. Involved crafts, social activities and activities of daily living.	To assess the immediate impact of a ten-week musical exercise therapy intervention on physical, emotional, behavioural and cognitive functioning in older people and to evaluate these changes at a 10-week follow up timeframe.	Cognitive Assessment Scale (CAS) Behaviour Rating Scale (BRS) Overall Dependency Scores (ODS) Life Satisfaction Index (LSI) Physical Assessment Administered at baseline, posttest 1 (following ten weeks of treatment) and at posttest 2 (10 week follow-up).	CAS scores showed a significant interaction between group and time of assessment (F 2, 144= 19.6, p<0.001). Scores for the C group decreased at both post-tests whereas the OT and ET groups showed improved scores at first follow-up with slight declines at 10 week follow up. The BRS also showed a significant interaction between group and time of assessment (F (4, 114)= 4.11, p<0.019. The C group showed increased behavioural disturbances at both follow-ups, whereas the ET and OT groups showed improvements immediately following which were further increased at 10 week follow-up. The ODS showed a significant interaction between group and time of assessment (F (4, 114)= 19.38, p<0.001. LSI scores also showed a significant group and time interaction (F (4, 114)= 4.11, p<0.019. Score remained the same for the C and OT group, but rose immediately following the intervention for the ET group before falling at 10 week follow up.	Levels of evidence score LII-a, grade B.
041. Nair (2011). Intervention. Dates not specified.	Australia A dementia specific aged care facility located in	75 in total: 38 from Unit A and 37 from Unit B.	Liistening to Baroque music. A selection of Baroque music was played between 3 and 7pm within the	This study tested the effects of listening to Baroque music on the behavioural disturbances exhibited	Behaviour chart documenting physical aggressiveness, verbal abuse, agitation, wandering and inappropriate sexual advances.	Music played High level ward (afternoon) 2.39 (SD 2.104). Music played High level ward (night) 0.42 (SD 0.783). Music played Low level ward (afternoon) 1.16 (SD 1.451). Music played Low level ward night 0.98 (SD 1.351). No music High level ward (afternoon) 1.76 (SD 1.927). No music High level ward (night) 0.5 (SD 0.906). No music Low level ward (afternoon) 1.14 (SD 1.398). No music Low level ward (night) 0.54 (SD 1.13).	GRADE score ? ? - + + + Levels of evidence

Reference of Paper, Study ID & Research Design	Location	Population	Activity	Aims of the study	Outcome measures	Findings	Quality
Quantitative Cross-over RCT	Newcastle, Australia. Two units were selected for the study, Unit A which provided high-level care and Unit B which offered low-level care.	Unit A- high level care Mean age 85.8 (± 5.37). 29 (75%) were female and 9 (25%) were male. Unit B- low level care Mean age 81.7 (±8.52). 28 were female (75%) and 9 (25%) were male.	common room areas, loud enough so that it could be heard within the room but not in the participants'' bedrooms. Residents could come and go from the common room area as they pleased.	by older people who lived within an aged care facility for people with dementia.	This was completed at the end of the afternoon shift (3-11pm) and following the night shift (11pm-8am) by nurses who were working that shift.	The lower care unit had 0.3 fewer episodes per week than higher (p=0.001), women had 0.2 more episodes per week than men (p=0.04) and afternoon shifts had one more episode per week than night shifts (p<0.001). Weeks where music was played showed significantly more episodes than observation weeks (0.23 more episodes per weeks, p=0.004). When multivariate analysis was carried out Sex and unit type did not remain significant (p=0.39 and p=0.26 respectively). Shift did remain significant, with one extra episode of behavioural disturbance per week in the afternoon shift (p<0.0001). Music also remained significant with 0.23 extra episodes per week compared with no music (p=0.01).	score LI, grade A.
043. Casby (1994) Intervention Dates not specified Quantitative Quasi-RCT	USA A long term care setting in Washington State	Three participants in total. Participant 1: 87 year's old, female. Participant 2: 77 year's old, female. Participant 3: 69 year's old, male.	Music- listening to music. Three phases: Phase A- baseline, no music and no headsets. Phase B- Intervention listening to relaxing classical music (Pachelbel's *Canon in D)* Phase C- Intervention listening to favourite music chosen by the participants.	To determine whether listening to music had an effect on the disruptive vocalisations of three individuals with dementia who resided within a long term care facility.	Data were collected during the times of day when instances of disruptive vocalisations were normally the greatest. For each phase data were collected during two 10 minute sessions each day over four days for a total of 12 observations days for subjects 2 and 3 (who took part in phases ACA and ABA respectively) and 16 for subject one (who took part in phases ABCA).	Subject 1: Phase A mean 15 (± 19), Phase B mean 0.5 (± 9), Phase C 2.9 (± 5) Phase A mean 0. Subject 2: Phase A mean 45 (± 51), Phase C 15 (± 15), Phase A 12 (± 16). Subject 3- Phase A mean 43 (± 48), Phase B mean 32 (± 52), Phase A 1 (± 4). Celeration lines were presented for each of the subjects, which showed the number of disruptive vocalisations that were observed over a 10 minute period. Celeration lines show the variability of the data, the more varied the data points are (i.e. they do not show a distinct pattern) then the less valid the final results are. In this case, the data did not show a sTable pattern at baseline. For subject's 2 and 3 there was an accelerating trend at baseline, which suggests that verbally disruptive behaviours rose over the four days at baseline level. However, for subject 1 there was a decrease in verbally disruptive behaviours, as shown by a decline in the data points at baseline. For subject 3 there was no significant difference (p > 0.05) in the intervention (phase B- classical music) and all points fell below the celeration line. For subject 2 all data points during the intervention phase (phase C- favourite music) fell below the celeration line and this was significant (p <0.05) For subject 1, data points fell below the celeration line for both the classical music phase and the favourite music phase, and this was significant (p<0.05).	GRADE score ? - - - ? + Levels of evidence score, LII, grade B.
046. Ridder (2013) Intervention Data were collected in three 15-week periods during Autumn 2010, Spring 2011 and Autumn 2011. Quantitative RCT- two-armed cross-over.	Denmark and Norway. 14 nursing homes in total: 10 in Norway and 4 in Denmark.	42 participants. 21 in each group. 29 (69%) were female and 13 were male (31%). Age data (only provided for 26% of participants) mean age of 81, range 66-96 years.	Music therapy. Participants had received referrals from their homes to take part in music therapy. All completed a music therapy protocol carried out with a music therapist employed within their home. Sessions were delivered by a music therapist.	To carry out individual music therapy with individuals who had moderate to severe dementia and lived within a care home and determine its effects on frequency and disruptiveness of agitation. The study also aimed to explore the effects of the intervention on psychotropic medication use and quality of life of residents taking part.	Agitation was assessed using the CMAI-fr (frequency) and CMAI-di (disruptiveness) at baseline, following 7 weeks (crossover) and following 14 weeks. Quality of Life was assessed using the Alzheimer's Disease Rated Quality of Life (ADRQL). These scales were adapted so that the respondent only needed to focus on the previous week. Anxiolytic medication prescriptions were also assessed at baseline and week 14.	During standard care the frequency of agitation slightly increased (0.46) whereas it decreased during music therapy (-2.96), the difference was therefore -3.41 which shows a small effect size (-0.21) which was not significant (p=0.38). Agitation disruptiveness increased during standard care (3.26) and decreased during music therapy (-3.51) and this difference reached significance (p=0.027). Sensitivity analysis using LOCF confirmed the results, the difference in perceived agitation disruptiveness remained significant (p=0.03). Per-protocol analysis (excluding those who received less than 8 music therapy sessions) showed similar effect sizes and significance (p=0.02). Quality of life showed a decrease during standard care (-5.88) and an increase during music therapy (10.42) (p=0.44). Those who received music therapy later may have had greater improvements (by 8 points from 20.81 to 12.59) then those who received music therapy first (who had a smaller reduction from 15.71 to 15.65). Medications were recorded at baseline and week 14. For this period there was no change in 48% of the participants. Increases in psychotropic medication occurred significantly more often during standard care than during music therapy (McNemar's X2=5.14, df=1, p=0.02). It also remained significant in the per-protocol analysis excluding participants who received fewer than eight sessions (p= 0.02).	GRADE score + + + + + - Levels of evidence score, LI, grade A.
047a. Cooke et al. (2010) Intervention October 2008-March 2009 Quantitative	Australia Two mixed gender aged care facilities, site A and site B. Both were in	47 in total. 33 (70%) were female and 14 (30%) were male. 3 aged 65-74, 13 aged 75-84,	Music and reading groups were carried out three times a week (Mon, Weds and Fri) for a period of 8 weeks. The music session comprised 30 minutes of live music and 10 minutes	To investigate the effects of a live music therapy programme on depression and quality of life for individuals with dementia. The study also looked at the	Quality of life as measured using the Dementia Quality of Life Scale (DQOL) and changes in depression measured using the Geriatric Depression Scale (GDS).	The only significant finding was that there was a significant difference in mid-point QOL belonging scores between the music and reading groups (F(1, 45)= 6.672, p<0.05. Participants who had experienced the reading control group first reported higher feelings of belonging (3.61) than those who experienced the music first (3.17). When the reading group crossed over to the music groups their scores decreased (3.61 to 3.46) but those in the music group first who crossed over to the reading group showed an increase (3.17 to 3.57). Two sub-analyses were carried out which showed significant differences: 1. In participants who attended greater than half of the music sessions (in either order, n=24) there were significant improvements in QoL self-esteem scored	GRADE score + + + + + + Levels of evidence

Reference of Paper, Study ID & Research Design	Location	Population	Activity	Aims of the study	Outcome measures	Findings	Quality
Cross-Over RCT	Queensland and provided both low (assisted) and high (nursing) care.	28 aged 85-94 and 3 were over the age of 95.	listening to pre-recorded music. Music activity delivered by musicians and reading activityby a research assistant.	long-term effects of the programme over six months.	Administered at baseline, midway point (3 months) and immediately following the intervention. Sub-analysis was carried out with those who attended 50% or more of the sessions.	over time (F(2, 46) = 4.471, p<0.05) in scores from mid-point (3.36) to post-intervention (3.75). 2. In participants who had scores of less than 5 on the GDS (n=12) there was a significant difference in depression scores over time (F (2, 22)= 8.129, p<0.01. Depression scores decreased (8.25, 6.50, 4.42 respectively). This was more noticeable for those experiencing the music (9.00, 6.20, 4.40) compared with the reading group (7.71, 6.71, 4.43). This was analysed (GLM) and these findings occurred independent of carry-over effects as a non-significant order by treatment interaction was determined (p=0.649).	score LI, grade A.
047b Cooke et al. (2010) Intervention October 2008-March 2009 Quantitative Cross-Over RCT	Australia Two mixed gender aged care facilities, site A and site B. Both were in Queensland and provided both low (assisted) and high (nursing) care.	47 in total. 33 (70%) were female and 14 (30%) were male. 3 aged 65-74, 13 aged 75-84, 28 aged 85-94 and 3 were over the age of 95. Loss to follow up: Intervention/Control: 8 and control/ intervention: 8	Music- live music. Music and reading groups were carried out three times a week (Mon, Weds and Fri) for a period of 8 weeks. The music session comprised 30 minutes of live music and 10 minutes listening to pre-recorded music. The music activity was delivered by two musicians and the reading group by a research assistant.	This study is tagged with 047b and features the same study population. The overall aim was to investigate the effects of a live music therapy programme on agitation and anxiety.	Changes in anxiety, as measured using the Rating Anxiety in Dementia (RAID) tool and anxiety, as measured using the Cohen-Mansfield Agitation Inventory Short Form (CMAI-SF). Sub-analysis was carried out with those who attended 50% or more of the sessions.	Analysis did not find an overall effect of the music programme in reducing agitation and anxiety over a six-month period. There was no evidence of any interaction effects of the first intervention arm (music or reading control group) at mid-point therefore suggesting there were no carry-over effects of either the music or reading group. Just one significant finding was shown for those who attending 50% or more of the music therapy sessions (n=24). Analysis showed a significant increase in frequency of verbal aggression over time regardless of group (F(1, 46)= 3.534, p<0.05). Mean scores increased from baseline (1.26, SD=0.590) to post-intervention (1.64, SD=0.767). A series of multiple regressions showed agitation levels at baseline and post-intervention were predicted by a number of factors:1. Participant MMSE scores (β= -0.409, p<0.01) significantly predicted overall agitation (F(4,39)= 2.952, p<0.05, adjusted R2= 0.154). Specifically, a higher level of agitation was associated with greater cognitive impairment. A shorter length of time living in the facility also appeared to be a predictor, although not at a significant level, of greater overall agitation (β 0.271, p= 0.064). 2. Although the models for baseline physical aggression (adjusted R2= 0.124, p= 0.057) and baseline physical non-aggression (adjusted R2= 0.114, p= 0.069) were just outside of significance, inspection of values suggested that lower MMSE scores predicted higher levels of these two subtypes of agitation (β= -0.383, p= 0.013; β= -0.354, p= 0.021, respectively). 3. Non-significant results were found for baseline verbal aggression (adjusted R2= -0.026, p=0.579) and verbal non-aggression (adjusted R2= 0.055, p=0.187). Post-intervention regression analysis found the following: 1. Although the model for overall agitation was marginally outside significance (F(4, 36)= 2.340, p=0.075, adjusted R2=0.118) the results mirrored those found at baseline. Indeed participant MMSE scores (β= -0.416), p<0.05) predicted overall agitation at post-intervention. A shorter length of time living in the facility also appeared to be a predictor of greater overall agitation, although not at a significant level (β= -0.302, p=0.059). 2. Greater cognitive impairment was measured by participant MMSE scores (β= -0.288, p= 0.081) and being male (β= -0.285, p= -0.078) were predictors, however not significantly so (F (4, 36)= 2.271, p=0.081, adjusted R2= 0.113) of greater physical aggression. 3.Participant MMSE scores (β= -0.511, p<0.01) significantly predicted verbal aggression (F (4, 36)= 2.815), p<0.05, adjusted R2= 0.154) with higher levels of verbal aggression associated with greater cognitive impairment. 4. The model for verbal non-aggression was non-significant (adjusted R2= 0.123, p=0.068), however inspection of β values suggested greater cognitive impairment (β = -0.356, p= 0.032) and a shorter length of time living the in the facility (β= -0.386, p=0.017) predicted higher levels of this agitation subtype. 5. A non-significant model was found for post-intervention physical non-aggression (adjusted R2= -0.035, p=0.621.	GRADE score + + + + + + Levels of evidence score LI, grade A.
048. Sung (2010) Intervention Dates not specified Quantitative Quasi-RCT	Taiwan A long-term care facility with participants being selected from two units.	52 participants in total, 29 in the experimental group and 23 in the control group. Experimental group: 18 males (62%) and 22 females (38%). Mean age was 78.1 years (± 7.2). Control group: 11 males (48%) and 12 females (52%). Mean age was 82.7 years (± 7.4).	Music- preferred music. Those in the intervention group listened to preferred music for 30 minutes a day in mid-afternoon twice a week (total of 12 sessions over 6 weeks). Those in the control group continued with usual care. Music sessions were carried out by trained nurses and nursing assistants.	To explore the effects of a preferred music listening intervention on anxiety symptoms in individuals with moderate to severe dementia who lived within a nursing facility.	Changes in anxiety symptoms as measured using the Rating Anxiety in Dementia Tool (RAID). Administered at baseline and following the intervention period.	The experimental pre-test mean score was 10.93 (SD 5.46) and this decreased post-test to 8.93 (SD 4.86) which was a significant reduction (t=5.64, p <0.001). The mean control group was 9.52 (SD 4.09) and this decreased to 9.35 post-test, however this was not significant (t= 0.68, p=0.51). An ANCOVA was carried out which showed those in the intervention group had significantly lower anxiety scores following 6 weeks of a preferred music intervention in comparison to the control group who received usual care only (F=12.15, p=0.001).	GRADE score ? - - ? + - Levels of evidence score, LII, grade B.

Reference of Paper, Study ID & Research Design	Location	Population	Activity	Aims of the study	Outcome measures	Findings	Quality
049. Chen (2009) Intervention May 2004-May 2005 Qualitative Focus-group interviews	Taiwan A 350 bed nursing home in Changhua City, central Taiwan for older adults who are wheelchair users.	17 participants, 11 females (65%) and 6 males (35%). Mean age was 80.5, range 70-90.	Music Therapy Group music therapy carried out once a week for an hour for a total of three months, then once every 2 weeks for a year. Structured music therapy programme: warm up, dancing, moving, group play, listening to relaxing music, listening to a musical performance by a guest performer and a concluding phase. Sessions delivered by the authors (unclear if they are MTs or not)	To explore the perceptions of wheelchair-bound older adults within care homes when taking part in a music therapy intervention.	Focus Group Interviews carried out 6 months following the intervention. Guided on three questions: Can you please tell me your thoughts and feelings when you attended the group music activity? Which aspect of the group activity influenced you the most? What part of this music activity did you like best and why? These underwent thematic analysis.	Major themes and subthemes Strength derived from the group dynamic- sense of energy, distraction from suffering and confirmation as a person. Enhanced quality of life- variety added to lifestyle, motivation to exercise, learning positive behaviour and greater life satisfaction.	QA Grade C Levels of evidence LIII.
053. Millard, (1989) Intervention Dates not specified Quantitative Quasi- RCT	USA A special unit for individuals with dementia within a nursing home.	10 participants in total were observed (however more attended the sessions). Mean age of the participants was 81.4 years (± 7.3), range 71-98 years. Seven were female (70%) and 3 were male (30%). Loss to follow-up: None reported.	Music Therapy Sessions were held twice a week at 3.00pm for a total of five weeks. The experimenter facilitated these sessions which involved structured singing.	To explore the potential for singing to improve "quality of care" for individuals with Alzheimer's Disease.	Observations were carried out using 'behaviour mapping'. A checklist was used that measures frequencies of certain behaviours and this was modified to be completed by two observers with each subject being observed for 1 minute intervals, four times per session. Observers also collected data for 30 minutes post-intervention.	Within-subjects ANOVA's were calculated (A x B x S) where A served as the baseline condition (discussion) and B served as the treatment condition (singing) and S represented the subjects. Sitting ($F=3.13$, $df\ 4, 36$, $p<0.026$) and walking with others ($F= 3.13$, $df=4, 36$, $p<0.026$) were significantly different for the singing condition. A significantly higher number of vocal and verbal participation was observed for the singing sessions ($F= 4.435$, $df\ 4, 36$, $p<0.005$). Attendance was found to be significantly higher for the singing sessions compared with the discussion sessions ($F= 2.61$, $df\ 4, 76$, $p<0.05$).	GRADE score ? ? - - Levels of evidence score, LII, grade B.
055 Liu (2014) Intervention November 2012- February 2013 Quantitative (with 'qualitative elements). RCT	China A nursing home in the Shijingshan disctric of Beijing.	No demographic information reported, just that 25 were assigned to the control condition and 25 to the treatment condition. Does state no significant difference in age of groups at baseline ($p<0.05$).	Music- Chinese 5 element. This involved both active and passive music therapy content, divided into 4 parts. Sessions lasted for 1-2 hours a week (Nov '12-Feb '13). Not clear who delivered sessions.	To explore the effect of music therapy consisting of five-element Chinese music, on older people with a diagnosis of seasonal affective disorder.	The Self-Rating Depression Scale (SDS) and Hamilton Depression Scale (HAMAD) were used to assess changes in depression levels. States qualitative elements but no methods for this reported.	T-tests were used to evaluate the difference between group means. Baseline mean SDS scores for control group 56.2 (± 19.3) for the treatment group 56.2 (± 19.1), ($p<0.05$). Baseline mean HAMAD scores for control group 14.7 (± 4.5) for the treatment group 15.1 (±5.9) ($p<0.05$). Following treatment mean SDS scores for control group 49.9 (± 18.8) for the treatment group 40.2 (± 18.1) ($p<0.05$). Mean HAMD scores 11.2 (±3.1) for the control group) and 8.8 (±4.9) for the treatment group. N.B. Reports there were qualitative elements, however simply states in the results section '*strength derived from the five-element group music therapy and emotional adjustment. The five-element group music therapy can reduce patients' psychological distress and let them feel inner peace and enhance their life satisfaction.* No actual analysis reported.	GRADE score ? ? - ? - - Levels of evidence score LI, grade A.

Reference of Paper, Study ID & Research Design	Location	Population	Activity	Aims of the study	Outcome measures	Findings	Quality
056. Lin (2010) Intervention August 2008-January 2009 Quantitative RCT	Taiwan Three nursing facilities for older adults with dementia.	Experimental group: n=49. 26 female (53%) and 23 were male (47%). Mean age for the group was 81.5 years. Control group: n=51. 27 were female (53%) and 24 were male (47%). Mean age 82.2.	Music- listening to music. Information around music preferences was collected from participants and a music therapy protocol developed based on previous research. Sessions delivered by the researchers who had attended a music therapy course at university.	To ascertain the effects, if any, of a group music listening intervention on agitated behaviour exhibited by adults with dementia living within a nursing facility.	Chinese Cohen Mansfield Agitation Inventory (C-CMAI). Collected at baseline, following the sixth session and following the 12th session. Then follow-up one month post-intervention.	Mann-Whitney U-tests showed little difference between the average C-CMAI for the two groups for total score and each of the subscales at baseline. Longitudinal effects were analysed by using generalized estimating equations (GEE). There was no significant difference in C-CMAI scores for the control group (38.25, 38.55, and 37.75). *Changes in overall score:* significant decrease in agitated scores for the intervention group at each of the time points: 6th session scores versus pretest- the average score of the experimental group was 0.47 points lower (p<0.001). 12th session versus pretest- the average score of the intervention group was 0.44 points lower (p<0.001). One month follow up- the average score was 0.47 points lower (p<0.001). *Changes in physically non-aggressive behaviours*: There was a statistically significant decrease in physically non-aggressive behaviours at each of the time points for those participating in the intervention: 6th session versus pretest- the average score was 0.31 points lower (p=0.004). 12th session versus pretest- the average score was 0.26 points lower (p=0.015). 1 month follow-up- the average score was 0.34 points lower (p=0.006). *Changes in physically aggressive behaviours*: There was a statistically significant decrease in physically aggressive behaviours in those taking part in the intervention at the different time points: 6th session versus pretest- the average score was 0.23 points lower (p=0.28). 12th session versus pretest- the average score was 0.20 points lower (p=0.025). At 1 month follow-up versus pretest the average score was 0.21 points lower (p=0.018). *Changes in verbally non-aggressive behaviours*: There was a statistically significant difference in verbally non-aggressive behaviours at each of the time points for those who took part in the intervention: 6th session versus pretest- the average score of was 0.22 points lower (p=0.042). 12th session versus pretest- the average score was 0.28 points lower (p=0.010). 1 month follow-up- the average score was 0.26 points lower (p=0.037) *Changes in verbally aggressive behaviours:* There was a statistically significant decrease in verbally aggressive behaviours in the intervention group at 6th session versus pretest with an average of 0.11 points (p=0.021). However the decrease at 12th session versus pretest was 0.09 points and this did not reach significance (p=0.104) and the difference at follow up (0.02 points) did also not reach significance (p=0.764).	GRADE score + + ? + + + Levels of evidence score LI, grade A.
057. Ziv (2007) Intervention Dates not specified Quantitative Uncontrolled pre/post-test	Israel The Sophie and Abram Stuchynski Alzheimer Research and Treatment Centre (2 different residential wards).	28 participants. 25 were female (89%) and 3 male (11%). Mean age of the participants 82.6 (± 4.89).	Music- listening to music A disk with songs totalling 16.48 minutes' duration comprised songs 'familiar' to participants. Took place on three separate Wednesdays between 13.15 and 14.25.	To explore the role of a music listening intervention on the behaviour of individuals with 'medium to advanced' Alzheimer's Disease, with focus on agitation and socialisation levels and affect.	Observation schedule devised using Cohen-Mansfield and Billing (1986- CMAI) conceptual framework observing three types of behaviours: positive, negative and neutral. Two observation sessions took place in each ward.	The mean number of positive behaviours for the group when listening to music was 5.18 (SD 3.62) and without music was 1.36 (SD 2.08), t=6.75, p=0.001. The mean number of negative behaviours when listening to music was 2.43 (SD 3.2) and without music was 5.96 (SD 4.13), t=-5.05, p=0.001. The mean number of neutral behaviours when listening to music was 4.39 (SD 3.61) and without music was 4.68 (SD 3.61), t=-0.37, (ns). Secondary analysis removed all behaviours that would only occur with music (i.e. swaying with music). 10 participants showed positive social behaviour only in the presence of music, 6 showed more positive behaviour in the presence of music than in absence and 9 participants did not show any difference in the amount of positive social behaviour. No participants showed more positive social behaviour without music than with it. Most negative behaviours were classified as repetitive. Aggressive and negative social behaviours were less frequent in general. Repetitive behaviours stopped during the music condition for 9 participants and were reduced with background music for 8 participants. 4 participants showed no difference and 2 showed repetitive behaviours only in the music condition. Aggressive behaviours appeared only without music and stopped during music in 4 participants and in a further 4 participants' music reduced aggressive behaviour. None of the participants showed aggressive behaviours when there was music playing.	Levels of evidence score LII-b, grade B.
058. Holmes (2006) Intervention Dates not specified Quantitative Cross-over RCT	UK Residential homes within the south of England	32 in total. 28 were female (88%) and 4 (12%) were male. Mean age 84.9 (± 4.7). Range not reported.	Music- both live interactive and passive listening. Participants took part in three sessions, each lasting for 30 minutes. Sessions were delivered by musicians.	To explore and compare the effects of a live music interactive session, passively listening to music and silence on behavioural symptoms of individuals with 'moderate to severe' dementia.	Dementia Care Mapping. Looks at 26 activities with a six point Likert-Scale. Category E of the scale correspond with engagement with a creative activity and was used for this study. Quality of engagement was rated every three minutes for each session from videotapes which were muted and had the live music activity obscured.	Live music showed significantly higher percentage of subjects showing positive engagement in comparison to silence X2 p<0.0001 for all participants and for those with moderate dementia only X2p<0.01 and severe dementia only X2p<0.01. There were no significant results when comparing the percentage of subjects engaged with pre-recorded music versus silence (all X2 >0.1). For those with severe dementia a greater level of participants showing engagement was found for the live music condition in comparison with the pre-recorded music (X2<0.01). In those with moderate dementia a greater percentage of engagement was found for live music compared with pre-recorded music (71% compared with 36%), but this was nonsignificant, X2 p=0.12. Wilcoxon signed rank tests for category E scores showed similar findings: *All residents:* Live music median +1 (range 0-3) vs. silence median 0 (range 0-1) p<0.001. Live music median vs. pre-recorded music median 0 (0-3) p<0.001. *Moderate dementia group:* live music median +1 (range 0-3) vs. silence median 0 (range 0-3), p=0.04. Live music vs. pre-recorded music median 0 (0-3), p=0.09. *Severe dementia group*= live music median +1 (0-3) vs.silence median 0 (range 0-1), p=0.04) and live music vs. pre-recorded music median 0 (0-1), p=0.04.	GRADE score ? + + ? – Levels of evidence score LI, grade A.
059. Sung et al. (2006) Intervention Dates not specified Quantitative	Taiwan A 160-bedded residential care home.	36 in total. Experimental group: n=18. 11 males (61%) and 7 females (39%). Mean	Music- music intervention with movement. Sessions were delivered twice a week for a total of four weeks and lasted for 30 minutes. Music familiar to the participants was	To examine the effects of a group music with movement intervention on the frequency of agitated behaviours displayed by individuals	Cohen-Mansfield Agitation Inventory (CMAI). This was modified to make observations for 60 minutes, counting behaviours that occurred using 10 minute periods.	Baseline CMAI scores showed no statistical differences in agitated behaviours between the two groups (t=0.54, p=0.59). The mean number of agitated behaviours in the music intervention group significantly decreased by 1.17 at week 2 (m= 5.11 baseline, m=3.94 at week 2) and by a further 0.5 at week 4 (m=3.44). The mean number of occurrences of agitated behaviours in the music intervention group at week 4 decreased significantly in comparison to the control group (t=3.85, p=0.001). At week 2 the mean occurrence of agitated behaviours also decreased significantly in	GRADE score + + - ? + - Levels of evidence score LI, grade A.

Reference of Paper, Study ID & Research Design	Location	Population	Activity	Aims of the study	Outcome measures	Findings	Quality
RCT		age was 78.4 years (± 7.85). Control group: n=18. 15 males (83%) and 3 females (17%). Mean age was 78.4 years (± 7.9).	okayed on a CD for 30 minutes and they were instructed on how to move. Sessions were delivered by were delivered in the afternoon by a nursing researcher and two research assistants who were trained to carry out the intervention	with moderate to severe dementia.	The CMAI was used at baseline, week 2 and week 4.	comparison with the control group (t=3.65, p=0.001), however, the control group did experience a slight decrease in mean occurrence of 0.16. The changes in occurrence of agitated behaviours were significantly different between the groups (F (2,33)=15.03, p<0.001. The music intervention group saw a mean decrease in occurrences of agitation of 1.67 and the control group saw a mean decrease of 0.22.	
061 Goddaer (1994) Intervention Dates not specified Quantitative Uncontrolled pre/post-test	Belgium Further information not specified.	29 completed the study. 23 were female (79.3%) and six male (20.7%). Mean age of the participants was 81.3 years (± 6.9).	Music- during dinner. Relaxing music" was chosen for the intervention. This was defined as music with a slow tempo; slow, irregular and unpredicTable rhythm; no sound impulses; linear melody no evocation; a certain degree of homogenous monotony from the beginning to the end; no variation in intensity; and situated in the register of bass frequencies.	To evaluate the impact of music listening during dining on agitated behaviours in older adults who had severe cognitive impairment.	The CMAI, translated in to Dutch. The scale was reduced to a 0-1 scale, from its original format, to record presence (1) and absence (0) of behaviours. Scores were obtained at baseline (week 1), during music being played at mealtimes (week 2), during a week of no music (week 3) and when music was reintroduced (week 4).	A significant change was observed over the four week period on cumulative incidence of agitated behaviours (F(3, 78)= 8.52, p<0.001. Total agitated behaviours decreased by an average of 54% from week 1 to week 2. In week 3, the music was removed and the incidence of total agitated behaviours increased by 38.4%, giving a net loss of improvement of 15.6%. Reintroducing music at week 4 decreased scores by 43% from week 3 levels. The total reduction in agitation over the four weeks was 63.4%. Significant changes were observed on the dimension of physically nonaggressive behaviours (F(3, 78)= 5.16, p<.003). When music was introduced at mealtimes there was a 40.7% reduction in the level of this agitation measure. This level rose again by 26.3% (net loss of 14.4%) in week 3. During the final week there was a drop of 41.7%, indicating a total decrease of 56.3% for physically nonaggressive actions by those taking part. There were significant changes in verbally agitated behaviours (F(3, 78)= 4.01, p<.01. The decrease of these behaviours from week 1 to week 2 was 74.5%. Verbally agitated behaviours then rose again by 33.3% when the music was stopped in week 3 (net loss of 41.2%). By week four levels were the same as for week 2. Therefore from week 1 to week 4 there was a 74.5% reduction in these behaviours. There were no significant changes over time for physically aggressive behaviour (F (3, 78) = 2.60, ns) and hiding/hoarding (F (3, 78) = .57, ns). No percentage reductions were calculated for these domains.	GRADE score ? ? ? – Levels of evidence score LII-2, grade B.
063. Ho (2011) Intervention May-December 2010 Quantitative Uncontrolled pre/post-test.	Taiwan A hospital based nursing home in Northeast Taiwan which had a 60-bed capacity.	22 in total. 12 were female (54.5%) and 10 male (45.5%). Mean age of the participants was 77.3 years (± 7.64), range 62-91 years.	Music Six different piano pieces were played on a CD player during mealtimes for a total of 4 weeks. The intervention took place twice a day between 11.30am and 12.30pm and 5.00pm and 6.00pm.	To evaluate a group music listening intervention consisting of music composed by the researcher and determine its effectiveness on agitation frequency in older adults with dementia.	The CMAI which was assessed at baseline, throughout the 4 week duration period and for a 2-week follow-up period. Staff nurses were trained by the researchers to use the CMAI prior to the intervention and kept a 24-hour record of agitated behaviour for the study duration.	Friedman tests revealed treatment effects across the different time points for the CMAI scale and subscale. The CMAI significantly changed from T1 (total mean 60.64) to T5 (total mean 42.99), showing a cumulative dose effect. At T5, CMAI scores had declined by 29.1% from baseline. The subscales had also decreased by 25.09%-35.91% by T5. At T6 the CMAI scores had increased slightly (total mean 46.14), but were still significantly lower than baseline data (all p<0.008). 15 participants (68.2%) said they liked the music played at mealtimes, 6 (27.3%) said they had no opinion and one participant said they did not like the music.	GRADE score ? ? – Levels of evidence score LII-2, grade B.
064. Chang (2010) Intervention Dates not specified Quantitative Uncontrolled pre/post-test	Taiwan Further details not reported.	41 in total. 15 were male (36.6%) and 26 were female (63.4%). Mean age 81.7 (± 6.4). Age range: 69-94 years.	Music. Nature music was played during the time immediately following lunch, where agitated behaviours were identified as being most prevalent.	To implement and evaluate a lunch time music programme for individuals with dementia living within a care home. The effects of the programme on agitation levels in those taking part were also explored.	The Chinese CMAI (C-CMAI). This was administered by six students (either nurses or social workers) at baseline and during Week 2, Week 4, Week 6 and Week 8 (music weeks) and Week 3, Week 5 and Week 7 (non-music weeks).	Line plots and GEE analysis are provided. Following adjustment for time trend and covariates the weeks CMAI scores during music player were significantly lower (*B*= 1.06, *p*=0.04) and physical aggression (*B*=0.39, *p*=0.04) and verbal aggression (*B*=0.49, *p*=0.02). This also carried over to the following week for physical aggression (*B*=-0.55, *p*=0.01) and verbal aggression (*B*-0.49, *p*0.03).	GRADE score ? ? ? – Levels of evidence score LII-b, grade B.
068. Svansdottir (2006) Intervention Dates not specified Quantitative RCT	Iceland Two nursing homes and two psycho-geriatric wards.	47 participants in total. Age range 71-87 years. No further demographic information provided.	Music therapy sessions. Carried out in groups of 3-4 participants by a registered music therapist, 3 times a week for a total of 6 weeks (sessions lasted 30 mins). Involved the music therapist playing instruments and singing	To assess the effects of a group music therapy programme on the behavioural and psychological symptoms of those with 'moderate to severe' Alzheimer's Disease.	Behavioural Pathology in Alzheimer's Disease Rating Scale (BEHAVE-AD). Participants were rated at baseline, following the intervention (6 weeks) and follow-up (four weeks after completion of the programme).	Most reported BEHAVE-AD symptoms presented at baseline were activity disturbances and paranoid and delusional ideation. Wilcoxon signed ranks showed a significant decrease for activity disturbance scores in the therapy group (mean 1.6 pre vs.0.7 immediately following, p<0.02) after six weeks but not for the control group (mean 1.4 pre vs. 1.0 immediately following, p>0.5). This effect decreased over the next four weeks and was not significantly lower than baseline. Total BEHAVE-AD scores were lower at six week follow up in the therapy group, however this was not significant (mean 5.5 pre vs. 4.4 at six weeks, p=0.3) and a small decrease in the control group which was also not significant (mean 5.4 pre vs. 4.7 at six weeks, p=<0.5). No other subscales showed a significant decrease in symptoms for either group.	GRADE score ? ? + - ? - Levels of evidence score, L-I grade A.

Reference of Paper, Study ID & Research Design	Location	Population	Activity	Aims of the study	Outcome measures	Findings	Quality
			songs, with residents encouraged to join in.			Grouping together activity disturbances, aggressiveness and anxiety scores showed a significant decrease in symptoms for the therapy group ($p<0.01$) but not the control group ($p=0.5$). At 4 week follow-up all of the scales showed no lasting benefits.	
069. Ashida (2000) Intervention Dates not specified Quantitative Uncontrolled pre/post-test.	USA Two care homes located in Florida.	20 in total, mean age of the group was 86.2 years with a range of 73-94 years. 17 females (85%) and 3 males (15%).	Music- drumming, singing and reminiscence with discussions. Participants were divided in four small groups based upon the homes they lived in (the average number of participants in a session was 4.8). Participants took part in the music therapy every day for a week. The average length of the sessions was 42.95 minutes, and they ranged from 38 to 45 minutes in duration.	To explore whether there was any effect on depressive symptoms when a group of older people with dementia took part in music therapy sessions which involved playing instruments (drums), singing and using reminiscence to shape the content of the sessions.	Cornell Scale for Depression completed by activity staff who interviewed the staff responsible for the residents. Data were collected at the end of the first week (pre-test), the end of the second week (no treatment) and the end of the third week (post-test).	*Cornell Scale for Depression in Dementia*: Data were only included for participants who had attended at least four of the music therapy sessions, the total number included was 20. Ten of the participants had attended four sessions and the others had attended five. One-way ANOVA, F test was used to test significance of differences. Significant difference between groups was assumed given the obtained F value (3.77) was larger than the F critical value (3.23) for df 2, 57, $p<.05$. A Newman-Keuls Multiple Comparison Procedure was then used to determine which weeks produced significant differences in scores. Values for this test were df=57, $\alpha=.05$ and critical values 3.032 (two-step) and 3.646 (three-step). Obtained values were .3 between O1 (pretest) and O2 (no treatment phase); 3.4 between O1 and O3 (follow up); and 3.7 between O2 and O3. Significant differences were found between baseline and at the end of the intervention week (O1 and O3) and between the control week (with no music, O2) and treatment week (O3). No significant differences were observed between baseline and the control week (O1 and O2). Data collection sheets were used to ascertain the levels and characteristics of participants in each group and to see if any changes occurred in these as the weeks progressed. Each of the participants were observed for 30 second periods at intervals of 2.5 minutes. Observed behaviours included positive affect, on-task, active participation and passive participation	GRADE score ? ? - - Levels of evidence score LII-b, grade B.
070 Ragneskog (1996a) Dates not specified. Quantitative Observation	Sweden A nursing home located near Gothenburg, Sweden, with a special care Dementia ward.	10 participants from a potential 25 were selected (those with the highest Multi-Dimensional Dementia Assessment Scale Scores- therefore the highest level of behavioural disturbances). 5 were analysed Mean age 80 (SD 2.4) 4 female (80%) and one male (20%).	Music- during dining. The music intervention lasted for around 30-45 minutes and was played in a group setting during lunch time (1pm). Three types of music were played (soothing, music from the 20s and 30s and rock and pop) and the sequence was decided randomly. During period 1 (baseline) no music was played.	To assess the effects of different genres of dinner music on symptoms of irritability, restlessness and uninhibited behaviours in nursing home residents with dementia.	Recorded observations were coded such that the researchers analysing them were blind as to treatment. For each segment the following behaviours were noted as present or absent: eating the food, sitting at the Table without touching the food, pushing away the plate, taking or trying to take a fellow patients food, seeking attention, rising from the Table, eating food, smearing food, picking, being spoon fed successfully/unsuccessfully. Time spent with dinner was recorded also.	Four of the five participants spent more time with their food during the music periods than the control period. Longest time spent with dinner was during soothing music, second longest during the time when music from the 20s and 30s were played, third longest during pop music being played. Total mean time spent with a meal increased by 22% during music being played compared with the control period. Dinner time decreases from playing of the soothing music towards the control period ($p<.05$). Nursing staff fed the participants significantly more often during the soothing music periods ($p<0.05$). Individual case studies are presented for each of the five participants.	GRADE score + - - - Levels of evidence score LII-b, grade B.
071 Ragneskog (1996b) Intervention Spring 1991 Quantitative Uncontrolled pre/post test	Sweden A nursing home located in Bohslan	20 completed, from an initial 25 recruited. 10 (50%) were male and 10 (50%) female. Mean age of participants was 80 years. Range 69-94 years.	Music- during dining. No music played during period 1 (5 days), soothing music during period 2 (8 working days), no activity for a week, period 3 where tunes for the 20s and 30s were played for 10 work days, no activity for a week, period 4 (pop & rock music from the 80s) for 8 days followed by a control period of 9 days.	To evaluate the effects of music being played at dinner time on symptoms commonly seen in dementia, such as depressed mood, irritability and restlessness and determine whether a certain type of genre were more beneficial than others. Also to evaluate the effects of the music intervention on food intake.	Mood changes assessed via the GBS (scored by a blinded psychologist) Amount of food consumed, measured by weighing the plate of food before dinner and then following. Weight of participants was recorded at the end of each mealtime session. Questionnaires on music preferences were administered to staff members pre-study.	Friedman rank sum tests showed significant improvements for dimensions of irritability, fear-panic and depressed mood compared with the control period ($p<0.05$), with results most pronounced for when soothing music was played. All the other dimensions were reported as not significant. A two-way ANOVA showed food intake performance was significantly worse when turns from the 20s and 30s were played ($p<0.05$) and emotional blunting was significantly poorer when soothing music was 20s and 30s were played ($p<0.05$). No significant differences were observed for impaired motor performance, intellectual impairment and emotional impairment. Fischers test showed staff members served larger portions to the participants when music was playing, compared with the control period ($p<0.001$) for both main course and dessert. Participants consumed more food during the music period than the control period, however this was mainly found for dessert ($p<0.001$). There was a significant correlation between served and consumed food ($r=0.85$, $p<0.001$). Heavier participants were administered larger helpings ($r=0.63$, $p<0.01$) and consumed more food ($r=0.58$, $p<0.01$).	GRADE score + ? ? ? Levels of evidence score LII-b, grade B.

Reference of Paper, Study ID & Research Design	Location	Population	Activity	Aims of the study	Outcome measures	Findings	Quality
					Measured at baseline, during soothing music (period 2), during 20s and 30s music (period 3), during pop and rock music (period 4).		
075 Ledger (2007) Intervention Dates not specified Quantitative Quasi-RCT	Australia 12 Nursing Homes in Queensland and 1 in Victoria. Homes were made up on community, church and privately owned facilities and offered a mixture of high and low level nursing care.	45 in total. Experimental group: n=26. 22 were female (84.6%) and 4 male (15.4%). 3 were aged 71-75, 4 aged 76-80, 6 aged 81-85, 9 aged 86-90, 3 aged 91-95 and one aged between 96-100. Control group: n=19. 18 were female (94.7%) and one was male (5.3%). Three were between the ages of 76 and 80, 7 aged 81-85, six aged 86-90, 2 between 91-95 and 1 between 96-100	Music therapy- group sessions. Sessions were carried out weekly, lasting between 30 and 45 minutes and for a minimum of 42 weeks of the year. Each site varied in the time sessions were carried out and group sizes varied from 2-10. Four of the music therapy groups were conducted by the first author (a qualified music therapist) and one group was conducted by a research assistant (also a qualified music therapist).	To assess whether participants who took part in a music therapy programme would have a reduction in the frequency and severity of agitated behaviours compared with a control group not taking part in the intervention.	The CMAI was used to assess the agitation levels of the participants. Levels were recorded at baseline, at three months, six months, nine months and a year. The CMAI was completed with staff members who had spent time with the participants to provide information as to agitation levels during the previous two weeks.	Baseline CMAI scores were significantly higher for the control group (t=2.17, p<0.05), with both groups showing high standard deviations (control group mean= 39.05, ± 22.15, experimental group mean= 25.95, ± 15.42). CMAI means for both groups changed throughout the data collection points. The control group showed an increase in CMAI scores at time point 3 and then a decline towards the end of the study (time points 4 and 5). The experimental group showed increased CMAI scores at the first assessment and then a small decrease at time point 3, before increasing again at time point 4 before returning to the same level as time points 2 and 3. Repeated measures multivariate analysis showed significant effects within-participants over time (F=2.61, p<0.05) but not within-participants over time by group (experimental or control F=1.61, p=0.432), showing there was no significant difference between the groups. There was no significant difference between the groups for any of the subscales verbal non-aggressive F=0.33, p=0.57; verbal aggressive F=0.59; p=0.45; physical non-aggressive F=0.62, p=0.44; physical aggressive F=0.78, p=0.38. Univariate analysis did show a significant effect for time and group for the verbal aggressive scale (F=2.70, p<0.05) which may show that the experimental group showed a decrease in verbally aggressive behaviours over time but the effect was too small to be detected by multivariate analysis, which is also supported by the fluctuation in levels observed across time points. Anecdotally, the log kept by the music therapist reported that those taking part in the intervention showed less wandering, fidgeting, grabbing, insults, complaints and anxious statements after sessions.	GRADE score - - + - Levels of evidence score LII-a, grade B.
082 Remington (2002) Intervention Dates not specified Quantitative RCT	USA Four different long term care facilities.	68 in total. 59 female (87%) and 9 (13%) male. Mean age 82.4 years. Range 62-99 years.	Music- calming music, hand massage, both interventions together. CM condition: music was played during this condition for ten minutes using a CD player. HM condition: participants received ten minutes of hand massage based on a protocol from previous research. CM and HM: those in this condition received both interventions at the same time. Those in the control condition received no intervention. Hand massage was carried out by the author.	To assess the effectiveness of two interventions; exposure to calming music and hand massage in reducing agitated behaviours of individuals with Alzheimer's Disease. The effectiveness of combining both interventions was also assessed.	The CMAI recorded frequency of agitated behaviours on four occasions: immediately before the intervention, during the intervention, immediately following the intervention and one hour following the intervention. The CMAI was administered by a research assistant who was trained in its use.	A one-way ANOVA showed no difference in agitation level at baseline (F test for group differences F=1.1, p=.36), showing that the participants had been adequately randomly allocated (Levene test for homogeneity F=.63, p=.60). A one-way ANOVA compared the frequency of agitated behaviours across all four time points. As Mauchley's sphericity test was significant then Fcons represents the F tests. There was a significant difference over time in agitation levels for the four groups (Fcons=6.47, p<.01). Tukey's HSD test found the control group was had significantly higher agitation scores than any of the experimental groups during the intervention period, no difference was found between the three intervention groups at any of the intervention time-points (2, 3 or 4). There was no significant difference over time for physically aggressive agitated behaviours (Fcons=1.93, p=.09) for the groups. Physically nonaggressive behaviours showed a significant difference (Fcons=3.78, p<.01). At time point 2, 3 and 4 the control group exhibited significantly greater scores for this subscale (Tukey's HSD post-hoc tests with a significance level of .05). Verbally agitated behaviours were also not effected by time (Fcons=1.92, p=.10). Tukey's post-hoc analysis (with a significance level of 0.5) showed there was a significant difference between the control group and the hand massage group at time points 2 and 3. A significant difference in scores was also found for all of the treatment groups and the control group.	GRADE score + + + ? - ? Levels of evidence score LI, grade A.
084. Vink (2013) Intervention	Holland 6 nursing homes, 4	Intervention group: n=43. Mean age 82.4	Music- group music therapy.	To determine the immediate short-term effects of a music	Cohen-Mansfield Agitation Inventory (CMAI).	Results were calculated excluding those where agitated behaviours were not present prior to the intervention (n=) and then again with these participants included.	GRADE score ? + - + + -

Reference of Paper, Study ID & Research Design	Location	Population	Activity	Aims of the study	Outcome measures	Findings	Quality
Dates not specified. Quantitative RCT	described as 'typical Dutch nursing homes' and two others based on alternative medicine (anthroposophic care.	(SD 7.6). 29 (67%) female and 14 (33%) male. Control group: n=34. Mean age 81.8 (± 5.7). 25 (74%) were female and 9 were male (26%).	All of the homes had a music therapist employed. Sessions took part over a four month period in a small group setting (no more than 5), twice a week for 40 minutes. Frequency, content and duration of sessions was based on a previous study. OTs provided content for the control condition (board games, puzzles etc.).	therapy intervention with regards to reduction of agitated behaviours in individuals with Dementia.	Recorded by nurse carers who were trained in its use. Data were obtained for each session 1 hour prior to the start, 2 hour following the session, 2 hours following and 4 hours following.	Following adjustments for agitated behaviours before the intervention and session numbers the reduction in mean CMAI scores (in residents who demonstrated agitated behaviours prior to taking part in the intervention) was greater for the music therapy group rather than the general activities group, however this effect was not significant (F=2.885; p=0.090) This effect was lessened when GDS was adjusted for (F=1.500, p=0.222). The median CMAI score in those taking part in music therapy was 2 (IQR: 1-3, range 1-9) 1 hour before the session commenced. This was 1 (IQR 0-2, range 0-7) at 4 hours following the session. For those in the control group the median score was 2 (IQR 1-3, range 1-13) 1 hour before the start of the sessions and was 1 (IQR 0-2, range 0-7) 4 hours post-intervention. No interaction was shown between the type of intervention and the session number for CMAI scores 4 hours following the intervention (F=0.275, p=0.603). There was also no interaction effect between the type of intervention and CMAI score before the session for the CMAI scores 4 hours after it took place (F=0.225, p=0.635).	Levels of evidence score LI, grade A.
085. Lord (1993) Intervention Dates not specified Quantitative Quasi-RCT	USA A privately funded care home for older adults.	60 recruited. 42 females (70%) and 18 males (30%). Age range 72-103. Participants were non-systematically separated into three groups of 20 participants each with 14 females and 6 males. One group took part in the intervention, one carried out puzzle activities and one participated in usual care.	Music- group music therapy Carried out in a group of 20. Each session lasted 30 minutes and took place for a total of six sessions. Music group: big band music from 20s and 30s played and participants given 'children's' instruments to play. Sessions were delivered by the authors. Nursing assistants (two) and an activities coordinator accompanied the sessions.	To assess the impact of music therapy on the mood, memory and social interactions of persons with Alzheimer's Disease.	Questionnaire developed for the study specifically (and un-validated). The beginning of the questionnaire asked demographic information, the second section contained questions to aid memory and the third section focused on social interactions. 30 second observations were carried out for each participant for the first two weeks of the study.	ANOVA showed a significant effect between groups was observed for recall, social interactions and general mood (F= 4.11, p<0.01). T-ratios for scores on recall, social interaction and mood questionnaire, showed a significant difference for the music groups scores when compared with both the placebo (M 2.0, t 1.93, p<0.05) and control group (M 2.3, t 2.7, p<0.05). There was no significant difference between scores for the control and placebo group. Correlation calculations showed there to be a significant correlation between pre and post questionnaire scores for the music group (*r*= 0.468, p<0.01), but no effect for either the control (*r*= 0.712, p=reported ns) or placebo group (*r*= 0.728, p= reported ns). Within the music group significant mean gain scores were observed for pre and post observations for patient mood (M=1.51, t= 6.71, p<0.01) and social interaction (m=1.6, t= 6.71, p<0.01).	GRADE score ? ? ? – Levels of evidence score LII-b, grade B.
090. Bennett (1988) Intervention Dates not specified Quantitative Quasi-RCT	Australia Participants resided in either a hostel or care home environment	26 in total. Mean age of the treatment group was 82.2 (± 7.8) and for the control group was 81.2 (± 7.8). Gender differences not reported.	Music- music based life review. Carried out within a group setting once a week for six weeks. 45 minutes in duration. The sessions involved listening to a piece of music and then having a discussion about their feelings in relation to the song. The control group took part in a verbal life review process. Sessions were delivered by the researcher.	To explore the effectiveness of a music-based life review carried out with individuals who resided either within a care home or hostel. The effectiveness of the music-based life review was also compared with that of a verbal life review therapy (the most common form of life review therapy).	Life Satisfaction Index A (LISA) and Ego Integrity Subscale of the Ego Adjustment Scale. Administered pre-treatment and post-treatment. Participants were also given a questionnaire asking about their enjoyment of the intervention.	*LISA:* 2x2 repeated measures ANOVA with two within-subjects factors; treatment (music based life review) and time (pre and post treatment), showed a significant F statistic for the interaction between treatment and time for the LISA (f (1, 12)= 26.674, p<0.001). The pre-treatment scores for the life review group were higher, but this trend was reversed at post-treatment. *Ego Integrity:* Participants in the music group showed significantly higher scores compared with the verbal life review group (F (1, 12)= 9.07, p<0.05), however they also showed higher pre-intervention scores. There was no effect of treatment type and time for ego integrity scores. *Enjoyment Questionnaire:* those in the music-based condition showed higher enjoyment scores (m= 3.23) than those in the verbal review group (m= 2.31), t (112)= 3.21, p<0.01. Those in the music condition also reported they found the treatment significantly more helpful (m= 2.07) than the verbal life review group (m= 2.00), t (12)= 3/27, p<0.01. A highly significant moderate correlation was shown between life satisfaction scores and enjoyment and perceived helpfulness of the intervention.	GRADE score ? - - ? ? – Levels of evidence score LII-a, grade B.
091. Smith (1986) Intervention Dates not specified	USA Two homes located in Philadelphia	12 participants, four of whom were located from one of the homes and	Music- musically cued reminiscence. Three groups met for a total of six sessions, each lasting for half an hour.	To determine the effects on cognitive functioning during three separate interventions: musically cued	Mini Mental State Questionnaire (MMSQ- now the MMSE).	Total MMSQ scores were increased following musically cued reminiscence sessions (mean 10.9 pre VS. 12.1 post), however this was not significant (t (11)= 1.07, p>0.05). No significant differences for orientation or attention subscales (t (11)= .37, t (11)= .80, p<0.05), but a significant improvement in the language subscale (mean 5.5 pre vs. 6.1 post, t (11)= 1.83, p>0.05).	GRADE score - ? ? – Levels of evidence

Reference of Paper, Study ID & Research Design	Location	Population	Activity	Aims of the study	Outcome measures	Findings	Quality
Quantitative Uncontrolled 3-way repeated measures.		eight from the other. No further demographic information was provided. Loss to follow-up: 1 participant became ill for a session for the musically cued reminiscence.	They received musically cued reminiscence, verbally cued reminiscence and a musical activity on its own. Each group took part in the activities in a different order. Sessions were delivered by the researcher.	reminiscence, verbally cued reminiscence and musical activity (without the reminiscence aspect). To compare the interventions to assess which had the greatest effect on cognitive functioning and then determine which aspect of cognitive functioning was most impacted upon (orientation, attention, language or total score).	Administered at baseline and then to 2 members of the group (from a potential four)	Verbally cued reminiscence scores were also improved (10.9 pre compared with 12.0 post), however not significantly so (t (10)= 1.42, p >0.05). There were no significant increases for either the orientation or attention subscales (t (10= .44, t (10)= .56, p >0.05). A significant improvement was shown for the language subscale (t (10)= 1.84, p<0.05). Musical activity alone showed a significant increase in total scores (t (11)= 1.83, p<0.05), but no significant increase for any of the subscales (orientation t (11)=0.92; attention t (11)=0.18 and language t (11)=1.68, p>0.05.) An ANOVA determined that no one intervention improved MMSQ scores significantly more so than the other (total F (2, 32)=0.003; orientation F (2, 32)=0.028, attention F (2, 32)-0.022 and language F (2, 32)=0.033, p >0.05). Pearsons correlation showed that all the scales were significantly different from zero (p<0.05), but they were not significantly different from one another (p>0.05).	score LII-b, grade B.
094. Vanderark (1983) Intervention Dates not specified Quantitative Quasi-RCT	USA Location not specified	43 in total, 20 from the experimental group and 23 from the control group. Mean age of 78 for the intervention group and 82 for the control group.	Music- group music programme Two music sessions per week for a total of five weeks. Treatment was planned by the experimenter and modified accordingly. Sessions were delivered by the researchers.	To examine whether there would be an effect on 'concerns and attitudes' (measured by self-concept, life satisfaction, socialisation, music attitudes and/or music self-concept) in those taking part in a music participation programme.	The authors devised the measure used within this study. The measures included Items that would indicate a change in attitude about life satisfaction, socialization, self-concept, music or music self-concept.	T-tests on pre-test mean scores showed the groups to be similar at baseline, however significant differences were found for measures of self-concept (t=1.75, p<0.05) and music self-concept (t= 2.31 p<0.05). The music intervention group showed significant gains for measures of life satisfaction (t= 3.74, p<0.05), music attitude (t= 2.64, p<0.05) and self-concept in music (t= 2.54, p<0.05). It was found that the experimental group was significantly improved over the control group when all criterion variables were considered simultaneously (p = .031).	Levels of evidence score LII-b, grade B.
101. Sole (2014) Intervention Dates not specified. Quantitative Uncontrolled pre/post test	Spain Setting not specified.	16 in total 15 female (94%) and 1 male (6%). Mean age of participants 87.5 (SD 5.6) and range 76-91. Group 1- GDS score 3-4, n=9 Group 2- GDS score 5, n=5 Group 3- GDS score 6-7, n=2	Music- group music therapy. Activities included: singing, listening to music, playing instruments, composition and movement. Took place once a week for 12 weeks, lasting 45-60 minutes. All sessions were carried out by a qualified music therapist.	To analyse whether there were changes in affect and participation levels in individuals with dementia (mild, moderate and severe) who were taking part in group music therapy sessions.	GENCAT-Quality of Life, administered at baseline and then follow-up three months later (end of programme). Observation schedule focusing on verbalisations, physical contact, visual contact, active participation and emotions.	Wilcoxon signed ranks was used to analyse differences in quality of life scores pre and post-test. QoL scores pre-test median 28 (IQ 24) and post-test median 21 (IQ 19), (Z=-.82, p=0.410). Several GENCAT measures were reportedly 'not addressed' by the intervention so only dimensions of emotional wellbeing, interpersonal relations and personal development were analysed. *Emotional wellbeing:* entire group pre-test median 12 (IQ 4) post-test median 23 (IQ 4), Z= -2.176, p=0.03. Median for 'mild' Group 1 at pre-test 22 (IQ 4) and post-test 23(4): z=-2.047, p=0.041. No other significant effects for different groups. *Interpersonal relations:* whole group pre-test median= 27 (IQ 6) post-test 25.5 (IQ 6), Z= -2.074, p=0.03). Behaviours observed for a total of four or more times per session were all positive, aside from agitation which was present for two of the participants with GDS scores of 5 and 7. Positive behaviours occurred when the participants were interacting with the music therapist (V1, V5 and L1) and active participation in activities when playing musical instruments (P1, P3, P4). Negative verbalisations (V2, V6) had low values whereas expression of negative emotions did not appear (aside from agitation). Positive emotions with smiles (E1) were low level throughout. Initiation of spontaneous physical contact did not occur. A total of 10 behaviours were recorded for Group 1. Behaviours appearing on four or more occasions: positive verbal responses to the music therapist, playing musical instruments and improvisation. There were no negative verbalisations. There was a slight trend in the appearance of positive emotions, with smiles in particular. There were no signs of negative emotions aside from one session where crying was observed. No positive spontaneous physical contact occurred with the therapist or other members of the group. Group 2: 11 behaviours were observed. Behaviours recorded four or more times: interaction with the music therapist, in particular verbalisations and looks and participation in playing instruments. There was also participation in singing activities but to a lesser extent than those in the mild group. No negative verbalisations were seen. One case of agitation was observed. Signs of restfulness were high, however there was a decrease in these behaviours. No sign of spontaneous physical contact with the therapist or others. 10 behaviours recorded for Group 3 These were observed on a lower level to those in the other. Nonverbal responses to demands of the music therapist were observed. Non-verbal communication with the therapist reduced from the start of the program to the end where there was almost no eye contact between participants. Instrument playing either spontaneously or directed remained high throughout. Once case of agitation was observed, but no other emotional expressions.	GRADE score ? ? - - Levels of evidence score LII-b, grade B.

Reference of Paper, Study ID & Research Design	Location	Population	Activity	Aims of the study	Outcome measures	Findings	Quality
106. Hammar (2011) Intervention Dates not specified. Quantitative Uncontrolled pre/post-test	Sweden Two care homes for people with dementia.	Ten participants took part in each stage of the study. Mean age was 81.3 years with a range of 66-92 years. Six were female (60%) and four males (40%).	Music- music therapeutic caregiving Two Caregivers attended an MTC course) and the remainder were taught how to undertake MTC by the study author. Morning care sessions (washing, dressing etc.) were then carried out along with caregiver singing of 'familiar' songs.	To incorporate music therapeutic caregiving (MTC) into the morning routines of people with dementia and assess the impact on expression of emotions and restiveness to care.	The Observed Emotion Rating Scale (OERS) and the Restiveness to Care Scale (RTCS). A total of 8- sessions were recorded, eight for each participant (four at baseline and four during the intervention).	<u>Restiveness to Care Scale</u> At baseline, there was a longer duration of pulling away in comparison to when the MTC was being carried out (148.8 seconds compared with 49.3 seconds) and this result was significant (p=0.031). Grabbing objects: The mean length of time for grabbing objects was significantly greater in the baseline phase (81.3), compared with the intervention phase (32.7), p=0.020. Behaviour adduction: Mean length of time was significantly lower for those in the MTC phase (30.8) compared with the baseline (78.5). Screaming: The only behaviour which lasted for longer than 2 minutes, observed in both phases. 55% (n=22) of observations within the baseline phase involved a participant screaming, however only 32.5% (n=13) of observations in the MTC did. Length of time was 205.8 for the no singing condition and decreased to 104.5 for the singing condition, however this result was not statistically significant. <u>Observed Emotion Rating Scale</u> Mean length of time that pleasure was observed for was 281.8 for the no singing condition and 1387.5 seconds for the singing condition and this was a statistically significant difference (p=0.016). 20% of the participants expressed pleasure for more than one minute at baseline, compared with 50% during the intervention. General alertness occurred for 2010.2 seconds on average, compared with 2703.3 seconds during the intervention (p=0.042).. For negative expressions of emotion showed that anger lasting longer than 16 seconds decreased from 25 to 40 observations within the baseline condition to 14 out of 40 during the intervention. None of the participants showed anger for longer than five minutes during the MTC. Anxiety/fear lasting longer than 16 seconds occurred for half of the observations at baseline, versus only 37.5% during the intervention. Sadness was the least common expression for both conditions, 75% expressed no sadness during the usual care condition and 85% expressed no sadness during the MTC condition.	GRADE score - ? + + Levels of evidence score LII-b, grade B.
108 Gotell (2007) Descriptive Dates not specified. Qualitative Content analysis	Sweden A special care unit within an Urban area of Sweden.	9 residents, with a mean age of 84 years and a range of 80-90 years. Seven were female (78%) and two (22%) were male. Five female caregivers also participated within the study.	Music- carer singing Three conditions were observed: the usual morning care routine, the morning care routine with background music and the morning care routine with care-giver singing.	To explore whether there were any effects for people with dementia who were exposed to carer singing during their morning care routines.	Observations- each of the participants were observed for each of the conditions, giving 27 observation periods in total. Video-recordings were transcribed into text and then analysed using content analysis.	A major theme was developed for each of the three conditions which represented the emotions and moods that occurred during the interactions between the carer and the resident. *The usual morning care session*: disjoint vitality *Background music:* mutual vitality infused with playfulness *Caregiver singing:* mutual vitality infused with sincerity.	QA score Grade B Levels of evidence score, LIII, grade C.
112 Myskja (2008) Intervention October 2003-January 2004 Quantitative Uncontrolled Pre/post- test	Norway A nursing home in Oslo (Valerengen bo-og service-senter) with three wards.	72 participants regularly (85.7% of all at the home). 51 (71%) were female and 21 (29%) were male. Mean age was 87.5 and range was 63.2-95.8 for females and 80.9 (57.5-93.2) for males.	Music- group music therapy. 45 minutes in duration carried out twice a week. Delivered by a music therapy aide who played the piano and led singing of familiar and preferred songs.	The aim of this study was to restart a music therapy programme that had been previously discontinued due to staff changes and assess changes in depressive symptoms of residents following the programme starting again.	Montgomery Asberg Rating Scale (MADRS) administered pre-test when the usual music therapist had been absent for 11 weeks and 2 months after the MT had been reinstated. Scored by the nurse leading the programme and the named nurse of the participant. Level of participation: 3-always/nearly always; 2-usually; 3- sometimes; 1-never.	Baseline depression scores varied for each ward; Ward 2 (mixed somatic and dementia ward) = 27.5 (± 3.8); Ward 1 (somatic)= 18.6 (± 5.8) and Ward 3 (dementia)= 21.7 (±5.8). Frequency analysis showed significant reduction in depression scores (all only reported as p<0.05) for the music condition for all results combined (mean 20.4 vs 12.2 following) and for each individual ward (ward 1 18.6 vs 11.1; ward 2 27.5 vs. 15.5 and ward 3 21.7 vs. 12.7). The relationship between participation levels and MADRS values, linear model showed participation levels predicted a change in depression scores (t -2.18, p<0.05). N.B participation measures are not reported in the outcome measure section of the paper but are reported in the results section. The authors also refer to semi-structured interviews within the discussion section, but again theseare not presented in the results section.	GRADE score ? + - - Levels of evidence score LII-b, grade B.
115. Mohammadi (2011) Intervention	Iran 'Long term care facility'.	19 in total. 11 males (58%) and 8 females (42%).	Music- Iranian music therapy. A total of ten sessions, once a week lasting for 1 and a half hours.	To assess the effects of an Iranian group music therapy programme on levels of stress, anxiety and depression in older	Depression Anxiety Stress Scale (DASS-21) Measures were determined at baseline and then	No significant differences between groups at baseline for depression Z-0.91 (p=0.36), anxiety Z- 0.96 (p=0.96) or stress Z- 0.48(p=0.48). Post-test significant differences between groups for depression Z=-3.60, p <0.001; anxiety Z=-2.91, p=0.004 and stress Z=-3.20, p=0.001.	GRADE score ? ? ? ? - ? Levels of evidence

Reference of Paper, Study ID & Research Design	Location	Population	Activity	Aims of the study	Outcome measures	Findings	Quality
Quantitative RCT		Intervention group 11, six males 55% and five females (45%) mean age 69.9. Control group 5 males (63%) and 3 females (37%). Mean age 68.9	Those taking part were encouraged to be involved with the intervention, as well as taking part in group discussions prompted by the music. They were also encouraged to sing the words to popular songs and play the instruments with the music. It is implied but not explicitly stated that the researchers carried out the programme.	adults living within a care home.	immediately following the intervention.		score LI, grade A.
116. Hicks-Moore (2007) Intervention Dates not specified. Quantitative RCT	Canada Three large care homes located within New Brunswick, Canada. Each had special care units, two of which had 24 beds and one which had 34 beds.	40 in total. 32 were female (78%) and 9 male (22%). Mean age of participants was 84.5 years (SD 6), range 67-92 years. Consent was obtained for 56 residents. Of these, 41 completed the entire study.	Music and hand massage (individually then both at once). Favourite music-preferences were collected from family members and a 10 minute disk of favourite music created. This was then played on a CD player to the residents. Hand massage- five minutes for each hand based on a protocol developed by a previous study. The hand massage was carried out by the author.	To explore the effects of hand massage or favourite music (carried out individually) on agitated behaviours in individuals with dementia. Then, to assess the effects of both interventions carried out at the same time on agitation levels in those with dementia. Finally, to compare whether both treatments carried out simultaneously were more or less effective in reducing agitation levels than either of the interventions carried out on their own.	A modified version of the CMAI which records the frequency of agitation as it occurs rather than retrospectively. The CMAI was measured at baseline (10 minutes prior to treatment), immediately following the intervention and then one hour post-treatment.	A 3x3 repeated measures ANOVA was used to determine whether any change in aggressive, physically non-aggressive and verbally agitated behaviours. There was no significant difference in aggression scores over time (Fcorr =1.91, p=.17) or across the three treatment types (Fcorr=2.97, p=0.8). There was also no significant interaction between treatment type and time. Therefore neither of the treatments was shown to significantly reduce physically aggressive behaviours. There was a significant reduction in verbally agitated behaviours over time (Fcorr =24.74, p=0.001). Scores decreased following treatment and then levelled off between post-treatment and follow up. Bonferroni-corrected pairwise comparisons also indicated significant differences between pre-treatment and the follow-up (p<0.001). There were no significant differences or interactions between the treatment types. There was a significant effect over time for non-aggressive behaviours (F= 123.38, p=.001) but not over the three treatment conditions (F=.47, p=.63). Pairwise comparisons confirmed this significant difference between pre-treatment and post-treatment/follow up (p <0.001). Scores for both this measure and that of non-aggressive behaviours decreased quickly following treatment and then remained sTable between post-treatment and follow up.	GRADE score ? ? - ? + - Levels of evidence score LI, grade A.
120 Suzuki (1998) Intervention Dates not specified Quantitative Uncontrolled Pre/Post test	USA Residential care facility in the North-West.	8 took part in the study from an initial 11 recruited. Five were female (63%) and three male (37%). Mean age 83.5 (SD not reported). Range 75-93.	Music- group singing and lyric composition. Sessions delivered 3 time a week for total of 9 times over 3 weeks. Sessions made up of three parts- sing-along; music making and relaxation. Sessions were delivered by the researcher.	The overall aim of the study was to assess the role of music therapy on mood and congruent memory recall in older adults who had depression.	Positive and Negative Affect Schedule (PANAS). Memory Retrieval Test. Pre-test took place before the intervention and post-test no more than 90 minutes following.	There was no significant difference in scores for the positive affect aspect of the scale, however the change in negative affect were significant (p=0.26). This show there was a significant decrease in negative mood following the music therapy. There was also no significant difference in the percentage of pleasant memories pre and post-test. There was a significant change in the percentage on unpleasant memories (p=0.034), this shows that following the music therapy there was a significant decrease in negatively recalled memories (Table 2). Spearman rank correlation coefficient was determined to test mood congruent effect. There was no significant correlation between Positive affect and pleasant memories (r=.71, p>0.5), Positive Affect and unpleasant memories (r=-.66, p>0.5), Negative Affect and pleasant memories (r=-.60, p>0.5) and Negative Affect and unpleasant memories (r=.31, p>.05) The discussion section reported analysis by gender but this is not included within the results section and values are not given. N.B used GDS to select participants but not as an outcome measure.	GRADE score - - + ? Levels of evidence score LII-b, grade B.
121 Carruth (1997) Intention Dates not specified. Quantitative	USA Florida nursing home- an Alzheimer's specific unit and skilled nursing unit.	Seven participants in total, all were female. Mean age was 87 (range 79-90).	Music therapy 14 sessions in total, sessions 7-10 involved music. Participants were given a photo of the staff member who then gave their name and sang a song, participant was encouraged to join in. if	To determine whether there was any improvement in face-name recognition when older people with dementia took part in a music therapy intervention.	Target-stimulus recall. This tested whether the participant could successfully recall the name of the care giver. If the participant failed to correctly recall the name on the photograph then further timed intervals were carried	Only four participants showed greater mean response times for the music condition (S1 88% music vs. 83% none; S2 96% music vs. 86% none; S3 100% music vs. 67% none and S7 88% music vs. 75% none). No statistical comparisons for any means.	GRADE score ? ? ? – Levels of evidence score LII-b, grade B.

Reference of Paper, Study ID & Research Design	Location	Population	Activity	Aims of the study	Outcome measures	Findings	Quality
Uncontrolled pre-test/post-test.			they did not the carer sang the song again.		out (10 seconds, 20 seconds, 40 seconds, 1 minute, 1.5 minutes, 2 minutes, 3 minutes, 4.5 minutes, 7 minutes, 10 minutes, 1 day). Participants were given three chances for each criterion, if they were still unable to recall the name then the data collection was stopped. An observation form was also completed by a graduate student.		
122. Brontons (1996) Intervention Dates not specified Quantitative Uncontrolled pre-post test	USA Four North-Western care facilities.	47 initially recruited, 20 completed ad data only given for these. 17 (85%) female and 3 (15%) male. Mean age of participant was 82 (± 6.6). Range was 70-96 years.	Music therapy- group. Took place in groups of 3-4 participants. 5 sessions in total lasting for 30 minutes. Sessions started with a hello song, then dance or movement based activity, then composition, followed by percussion and ending with singing songs. Sessions were carried out by two certified music therapists who had more than five years' worth of experience in working with older adults.	To carry out a live music therapy intervention and assess its impact on agitated behaviours in older people with Alzheimer's Disease.	Agitation behaviour scale section of the Disruptive Behavior Rating Scale (DBRS) administered at baseline, during the intervention and follow-up. Medical charts to obtain P.R.N medication use. Musical backgrounds of participants were collected via questionnaire.	9 of the participants= musical background, 5= no musical background, no information for the other 6. 40% of the sessions were analysed by two independent observers and interrater reliability was established to be .99. There was no effects with regard to music background (F1=1.79, p=.20) but there was a significant main effect of agitation behaviour (F3=16.33, p=.0001). Fisher's PLSD tests showed participants to be significantly more agitated before music therapy (m=11.46) than either of the two time points during the intervention (m=7.68 and m=7.52) and following the music therapy sessions (m=8.37). There was no two-way interaction between music background and agitation behaviour (F3=1.34, p=.28). 40% of the total observations were independently observed. A t-test for independent samples was calculated and showed there was no significant differences in agitation scores between music therapists and caregivers [t(78)=.67, p=.50]. The scores means and standard deviations for music therapists and caregivers were m=7.90, SD=2.01 and m=7.62, SD=1.63 respectively. Not enough information on P.R.N medications for an analysis.	GRADE score - - + - Levels of evidence score LII-b, grade B.
123 Clair (1996) Intervention Dates not specified Quantitative Uncontrolled pre/post-test	USA A 'Midwest residential facility'.	26 in total. Age range was 62-83 years. 22 were female (85%) and 4 were male (15%).	Music- live singing. Sessions involved the participants being exposed to all of the conditions, each lasting for two minutes each, one after the other: 4 x 2 minutes of singing, 4 x 2 minutes of reading, 4 x 2 minutes of silence, preceded by 2 minutes of baseline. These sessions were carried out by the researcher.	To assess alert-responses in individuals with dementia who were non-ambulatory and non-verbal during a live singing intervention and compare this with silence and listening to a reading.	Graduate students recorded presence or absence of behaviours at 15 minute intervals. The following items were identified: rotating the head to localise sound, rotating the head to localise the experimenter, opening eyes, turning eyes in the direction of the experimenter, moving arms, hands, legs and/or feet. Making vocal sounds and changing facial expression.	Participants were separated into three groups to conduct within-group analysis, Group 1 (n=3) were described as being 'agitated'. Group 2 (n= 16, including all males) were subdued/closed and Group 3 (n=7) were classified as subdued/open. T-tests were calculated in order to establish whether there was any difference between each of the three conditions (all participants in all groups). There were significant differences between singing and silence and silence and reading (p values not reported). No statistically significant effect was shown for the singing and reading comparison. One-way ANOVA showed there to be no statistically significant difference between the three groups at baseline. There was also no statistically significant differences following the first session. *Second Session:* For the second session, there was a statistically significant difference between the groups for the singing and reading conditions, but not for the silence condition. The singing mean scores were significantly different between the groups (df=2 & 23, F=3.743, p=.039). Reading mean scores were also significantly different for the second session (df=2 & 23, F=3.45, p=.048). According to the multiple range test, Group 1 (agitated, M=8.67) and Group 3 (subdued/open, M=29.43) were significantly different (at .05 confidence level). Group 2 (subdued /closed, M=17.31) and Group 3 (subdued/open, M=29.43) were also significantly different (at .05 confidence level). Therefore those who were open were more likely to respond to the reading condition than those who were agitated or showed closed behaviours. *Third session*: there was a significant difference between the groups for singing and silence, but not for reading. Singing differences (df=2 & 23, F=5.06, p=.015) among the groups occurred between Group 2 (subdued/closed, M=11.56) and Group 3 (subdued/open, M=29.14), but no significant difference was observed for the agitated Group 1 (M=12.67) and any of the other groups. An ANOVA also showed significant differences among the groups (df= 2 & 23, F=5.78, p=.009) for responses during the silence with follow-up tests confirming differences between Groups 1 (agitated, M=5.33) and 3 (open, M=24.00) as well as between Groups 2 (subdued/closed, M=8.13) and Group 3 (subdued/open, M=24.00), at the .05 confidence level. Those in Group 3 who had their eyes open were not very active responders during the silence condition. *Fourth Session:* At session 4 there were no significant differences among the behavioural groups for any condition and the mean scores showed that Group 1 (agitated) and Group 2 (subdued/closed) had increased in their responses under all the conditions.	GRADE score ? ? ? ? Levels of evidence score LII-b, grade B.

Reference of Paper, Study ID & Research Design	Location	Population	Activity	Aims of the study	Outcome measures	Findings	Quality
						Total: statistically significant differences were found for singing (df= 2 & 23, F=4.4682, p=.022) and reading (df=2 & 23, F=3.44376, p=.049) but not for silence. Multiple range follow up tests for singing showed Groups 1 (agitated, M=64.33) and Group 2 (subdued/closed, M=68.75) were not significantly different, they were both significantly different from Group 3 (subdued/open, M=116.43). This was shown for the reading condition where it was found that Group 1 subjects who showed agitation (M= 70.33) and Group 2 subjects (subdued/closed, M=65.81) showed significantly fewer responses than those in Group 3 (subdued/open, M=106.86).	
130 Hammar (2010)- this study is tagged with 106. Intervention Dates not specified. Qualitative Group interviews	Sweden Two care homes in urban Sweden.	Older people- 10 in total, demographic information not provided. Carers at the home- six in total. Age range 31-54.	Music- therapeutic caregiving (MTC). Two Caregivers attended an MTC course) and the remainder were taught how to undertake MTC by the study author. Morning care sessions (washing, dressing etc.) were then carried out along with caregiver singing of 'familiar' songs.	This study was part of a larger research project aimed at understanding the influences of music therapeutic caregiving on people with dementia and their caregivers.	Group interviews were carried out with the caregivers, interviews lasted around an hour each (four in total). Observations- each person with dementia was observed for a total of four times (once a week) during usual care and four times (once a week) during MTC taking place- these are described not analysed in this paper.	Main theme arising from the interviews related to the normal morning care situation- 'Struggling for Care in Communion', with four subthemes: hampered communication, physical and mental struggle with aggression, struggling with ethical demands and the reward- consolation and love. The second main theme which arose from the interviews related to the MTC morning care situation was 'consolidating care in communion' with two subthemes: awakening cooperation and feelings of wellbeing.	
132 Guetin (2009) Intervention September 2007-April 2008. Quantitative RCT	France Les Violettes Nursing Home.	30 in total. 22 females (73%) and 8 males (27%). 15 assigned to the music therapy group. 13 females (87%) and 2 males (13%) with a mean age of 85.2 (± 6) range 75-93. 15 assigned to the control group 9 females (60%) and 6 males (40%).	Music therapy Sessions carried out weekly, for 16 weeks in total. A control group were offered sessions such as reading at the same time. A "U sequence" was created using preferred music which was then listened to via headphones in the participants' bedrooms or preferred areas.	To assess the impact of a short and medium term music therapy programme on anxiety in those with mild to moderate Alzheimer's disease who lived within a nursing home. The study also aimed to assess depression scores at 2-months post intervention to assess long-term impact.	Hamilton Scale for Anxiety Geriatric Depression Scale (GDS) Took place at baseline (D0), week 4 (W4), week 8 (W8), week 16 (W16) and week 24 (W24- follow up). Carried out by a neurologist and psychologist.	An ANOVA with repeated measures showed a significant difference (p<0.001), showing that the two groups progressed differently during the follow up period. At baseline there were comparable levels of anxiety between the two groups (22, SD 5.3 for the music therapy group and 21.1, SD 5.6 for the control group). However, at week 16 the anxiety scores for the music therapy had decreased to 8.4 (SD, 3.7) whereas the control group had only decreased to 20.8 (SD, 6.2) (p>0.001). *Anxiety:* 2 months following cessation of the sessions there was a significant difference between the two groups with respect to their anxiety scores (p<0.0001). Week 24 mean music therapy 10.6 (± 6.3) and control mean 20.5 (± 5.4) (p<0.001). Week 16-24 music therapy mean change in points 2.1 (± 3.7) and control mean change points -.08 (± 2.8) and baseline-week 24 change in mean score music therapy group -11.5 (7.2) and control group -1.5 (± 6.8). *Depression:* Baseline music therapy mean score 16.7 (± 6.2) control mean 11.8 (SD 7.4) repeated ANOVA was significant (p=0.001). Overall changes were not significant over time, however each group progressed in a different way during follow-up (significant time/group interaction p= 0.0095). W16 scores for music therapy group 8.9 (± 3.3) and control mean 11.2 (± 6.1). ANOVA significant (p=0.002). After 16 weeks the improvement was approximately 7.7 (± 4.6) points (47%) in the music therapy group, with a mean depression score of 16.7 (± 6.2) versus an improvement of around 0.2 (± 4.4) points (1.7%) on the control group with a mean depression score of 11.8 (± 7.4). Scores at W24 were compared between the two groups. ANOVA with repeated measures showed a significant difference (p=0.006). Scores progressed differently during follow-up up to the six-month point. Depression score at W24 was 12.5 (± 6.4) for music group and 12.1 (± 7.6) in the control group (p=0.03).	GRADE score + ? + + + + Levels of evidence score LI, grade A.
139 Burrack (2004) Intervention Dates not specified Quantitative Uncontrolled pre-test/post-test	USA 'Large urban nursing home'	16 agreed to take part, 13 completed. Of these 9 female (69%) and 4 males (31%). Mean age 82 (SD 8.92) range 64-93 years.	Music- therapy, individualised individually listened to 'individualised' music within their rooms through a personal CD player for a one-off session which lasted for 30 minutes	To explore the effects of individualised music therapy on immediate satisfaction and overall quality of life.	Global quality of life assessed via the Quality of Life Alzheimer's Disease (QoL-AD). Music background questionnaire and music intervention questionnaire. Administered at pre-test and then immediately following the arts activity.	*Access and Frequency of Listening to Music*: participants reported that they listened to music significantly less often following entering the nursing home (m=3.58, SD=1.16) than they had before (m=4.50, SD-0.52), t(11) = 2.73, p<.05. Twelve participants reported that they used to attend live music performances. Eight participants used to play a musical instrument and five stated that they used to sing. In the nursing home the most frequently reported mode of listening to music was by TV (8 participants). Seven participants said they listened to music by radio and three through a tape recorder. Seven reported they had the opportunity within the nursing home to listen to music they enjoyed, one reported they sometimes had the opportunity and five reported that they did not have the opportunity to listen to music they liked. When asked if they attended music activities, four responded that they did, seven said no and one said that they liked listening to music in their own room. There were no significant differences between the pre and post-test on any of the global quality of life items or combined scales. Descriptive data based on participants' enjoyment was also reported.	GRADE score - ? - - Levels of evidence score LII-b, grade B.
144 Koike (2012) Intervention Dates not specified	Japan Details not specified	15, 10 females (67%) and 5 males (33%).	Vibroacoustic therapy Delivered in the afternoon between 15.30 and 16.00 using classical music.	To determine whether vibroacoustic therapy would improve the levels of psychological symptoms in individuals	Dementia Mood Assessment Scale (DMAS) for affect changes. Dementia Behavioural Disturbance measure.	Wilcoxon signed rank tests. *DMAS:* significant changes from pre to post intervention for total scores obtained, mean pre 49.66 (± 16.17) and mean post 43.8(± 17.93), p< 0.05. Significant changes for depression and sadness mean pre 36.66 (± 11.13) and post mean 32.93 (± 14.05). No significant MMSE, DBD or dementia severity changes (all p>0.05). *Physiological changes:* significant change in temperature mean, pre 36.46 vs. post 36.31, p<0.001 and pulse rate, mean pre 72.28 (± 9.88) and post 70.66 (± 9.44), p<0.001. Sp02 and blood pressure not significant.	GRADE score ? ? + ? Levels of evidence score LII-

Reference of Paper, Study ID & Research Design	Location	Population	Activity	Aims of the study	Outcome measures	Findings	Quality
Uncontrolled pre/post-tes				with depressive symptoms.	MMSE scale for cognition changes Physiological changes: temperature, Sp02, blood pressure and pulse rate. Sleeping hours using a wristwatch type actigraphy.	*Sleep time:* total sleep in the second week was significantly decrease, compared with the first (626.54 vs. 559.61 mins, p<0.05). Wake time was significantly increased in the second week compared with the first (813.45 mins vs. 838.06 mins, p<0.05). Night time sleep and night time sleep efficiency were not significantly different.	a, grade B.
146 Hicks-Moore (2005) Intervention Dates not specified Quasi-RCT crossover	Canada A Specialised Care Unit (SCU) within a 120 bed nursing home in urban South Eastern Canada. The unit has 34 beds for residents with severe cognitive impairment.	30 in total. Mean age 82.4, range 70-101. 21 female (70%) and 9 male (30%). .	Music- relaxing music during meal times. Baseline Week 1 (days 1-7) the music was played during week 2 (8-14), a control period during week 3 (15-21) and week 4 (22-28) music was played again.	To explore the impact of a relaxing music intervention on agitation levels in older people who had dementia and resided within a nursing home.	The CMAI- recorded the presence or absence of behaviours during the meal period. Baseline data were obtained during baseline (week 1) and at throughout each of the weeks, with a mean per week (i.e. week 2 score, week 3 score etc).	Total agitated behaviours had a potential range of 0-203 (29 possible behaviours x 7 days). The actual range of behaviours observed for week 1 (no music) was 0-69 and there was a mean incidence of 9.85. For week 2 (music) the range of behaviours was 0-32 and the mean incidence was 4.57. During week 3 (no music again) the mean incidence of total agitated behaviours had increased to 7.29 with a range of 0-51 and in week 4 (music) had decreased again to a mean of 3.34 with a range of 0-24. No statistical analysis carried out using these means.	GRADE score - ? + - Levels of evidence score LII-a, grade B.
164 Hong (2011) Intervention Dates not specified. Quantitative RCT	South Korea Not specified, just referred to as 'nursing home'.	30 in total. Mean age of the group was 78.3 years (SD 6.3). 28 (93.3%) of the participants were female and 2 (6.7%) were male.	Creative song writing (3 stages) Preparing songwriting, songwriting and then reinforcing the songwriting. Sessions were carried out once a week for a 16 week period and lasted for an hour. Those in the control group were given free time during the sessions. Not reported who carried out the arts activity.	To explore the effects of a song-writing intervention on cognitive functioning of older people with dementia living within a nursing home.	Korean Mini Mental State Examination (MMSE-K) scores. Baseline scores were obtained two weeks prior to the programme and the final assessment was carried out one week following the programme. Each participant was rated by a researcher and a social worker that was employed by the care home, with average scores being calculated.	For the intervention group scores rose for all of the five subscales, whereas in the control group they either decreased slightly or remained unchanged (memory and comprehension and judgement). Changes over time: Overall mean MMSE-K scores rose for the intervention group by a percentage of 22.16%, compared with a decline of 0.06% recorded for the control group. MMSE-K percentage changes for the intervention group were 22.16% for orientation, 15.78% for memory, 206.69% for attention and calculation, 40.35% for language functions and 37.93% for comprehension and judgement. While large relative changes were observed for the categories of 'memory' and 'comprehension and judgement' their magnitude meant that their overall MMSE-K score contribution was low. MMSE-K scores were significantly increased from 14.6 to 18.40 for the intervention group following the programme (p<0.001). The control group showed a mean total MMSE-K score of 15.00 prior to the intervention period and this decreased significantly to 14.13 at the end (p<0.001). Significant changes were observed for the 'orientation' scale (from 5.73 to 7.00, p<0.001), 'memory' scale (from 5.07 to 5.87, p<0.02) and 'language' scale (from 2.80 to 3.93, p<0.01). Significance was set at the p<0.01 level meaning improvements in 'comprehension and judgement' (p=0.02) and 'attention and calculation (p=0.04) were not significant. None of the subscales for the control group reported significant improvements at follow-up. Between groups: Mean-MMSE scores were not different at baseline. Following the intervention period, there were significant differences between the groups for total MMSE-K scores (p=0.001), 'memory' (p<0.001) and 'language function' (p=0.003). Orientation scores were close to significance (p=0.013). 'Attention and calculation and 'comprehension' subscales did not show significant improvements (p=0.499 and p=0.164 respectively).	GRADE score + + ? ? – Levels of evidence score LI, grade A.
168 Clarkson (2007) Intervention Dates not specified. Quantitative Uncontrolled Pre/Post-test	Canada 'Urban nursing home'	17 in total, 8 in the moderate dementia subgroup and 9 in the severe dementia subgroup. Moderate dementia group- mean age 86 (SD 8.9), 7 (88%) female and 1 (22%) male.	Music- live performance The live music concerts were held 5 days a week for a total of three weeks, taking place in the hour after lunch. Music for the intervention was selected based on 'appropriateness for the older adults' generation and all participants were encouraged to join in with the singing.	The overall aim was to determine the effectiveness of a live music intervention carried out by volunteers, on agitation and depression levels in older people with dementia living within a nursing home. Secondary to this to determine whether there were any differences in presence	Cohen-Mansfield Agitation Inventory (CMAI) for agitation, scored by two nurses. P.R.N sedative use to assess any changes in need of anti-anxiety medications, obtained from information contained in patient charts. Cornell Scale for Depression, scored by the nurse who delivered the CMAI.	*Agitation:* total number of concerts attended ranged from 8-15, low levels of agitation were observed. For the first 2, only 2 and 3 participants showed agitation (from an average attendance of 16.6). only 2 participants showed agitation during the third weeks' session. Most common behaviours were nonaggressive vocalisations and mannerisms. *Physical Aggression:* significant effect of phase (F[2, 30]=5.77, p<.001) and a marginally significant effect of phase (F[2, 30]= 3.20, p<.055). One-way ANOVAs comparing the three study phases were significant for the group with moderate dementia only (F[2, 14]=5.02, p<0.23). T-tests showed a near-significant decrease in physical aggression in the music phase compared with baseline (baseline vs. music mean [SD]=14.0 [4.3] vs. 11.7 [1.5], t[7]=2.23, p<.061) for the group with moderate dementia and similar results were shown for the post-intervention phase (post-intervention mean [SD]=11.4 [1.1[, t[7]=2.27, p<0.58. *Verbal aggression*: The two-by three ANOVA revealed a significant interaction effect of phase-by-group (F[2, 30]=3.56, p<0.4). Despite this, one-way ANOVAs did not show any significant changes across the study period for either of the groups.	GRADE score - ? - + Levels of evidence score LII-b.

Reference of Paper, Study ID & Research Design	Location	Population	Activity	Aims of the study	Outcome measures	Findings	Quality
		Severe dementia group- mean age 86 (± 7.6), all female (100%).	A volunteer group carried out the music concerts.	of agitation depending on the stage of dementia.	Actigraphy (rest/movement cycles) measures physical agitation in a random sample (n=4) participants using an actiwatch. These were recorded weekly for the study period duration.	*Verbal non-aggression*: two-by-three ANOVA showed a significant effect of study phase (F[2,30]=4.28, p<.023) and a non-significant interaction of phase-by-group (F[2, 30]=1.06, p>.05). Follow up t-tests (collapsing the severity groups, n=17) examining the phase effect showed a significant rise in verbal nonaggression during the music phase (baseline vs. music mean [SD]=10.7 [5.4] vs.13.0 [5.5], t[17]=2.74, p<0.015) which lasted during the post intervention phase (baseline vs. post intervention mean [SD]= 10.7 [5.4] vs. 12.4 [6.7], t[17]=2.93, p<.01.	
169 Cohen-Mansfield (1997) Intervention Dates not specified. Quantitative Uncontrolled pre/post-test.	Israel 'Nursing home'.	32 in total. 6 (18.8% male) and 26 (81.3%) female. Mean age of the participants was 86.8 (± 1.16), range not reported. .	Music (with three other interventions). Medical examination- to ascertain physical causes of behaviour. Exposure the music- half hour audio tapes based on preferred music. Use of family-generated videotape. One-to-one social interaction.	To assess and compare the effectiveness of four different interventions on incidences of verbally-disruptive behaviours in older people with dementia.	Tape recordings: carried out for an hour a day, 15 minutes prior to the intervention, 30 minutes during and 15 minutes following. Coded by trained research assistants. Standardised observations: using the Screaming Behavioural Mapping Instrument (SBMI). Informant ratings: staff members completed the CMAI for participants by their care givers following the intervention.	*Medical Examination:* for 23 participants (78%) there was no physiological reasons identified for their VDB following assessment by a physician. 7 were found to be in some form of pain (from mild (n= 2) to discomfort (n= 4) to distressing (n= 1). Physicians also noted that despite the pain in these participants this was not the source of their disruptive behaviours. Three sources of data were used to assess the effectiveness of the interventions: recorded tapes (i.e. duration), standardised observations (i.e. frequency) and nurses' assessments. Repeated measures MANOVAs were performed to assess the effectiveness of the interventions. *Audio-tape data:* the effect of time was significant (F (1, 23)= 137.1, p <0.01) as well as the interaction between time and type of intervention (F (3, 124)= 7.7, p <0.01). The effect of intervention was significant (F (3)= 45.5, p <0.001) and all the interventions were significantly better than no treatment. Behaviours decreased by 56% during the social interaction, 46% during the videotape, 31% during the music and 16% during the no-intervention condition. *Observational data*. Shouting, complaining and/or inappropriate verbal and other VDB all showed significant differences between nothing and social interaction (p<0.01), nothing and videotapes (p<0.01) and nothing and music (p<0.01) (shouting: F (3, 318)= 11.8, p<0.05, complaining: F (3, 345)= 11.5, p<0.05 and social interaction F (3, 183= 6.6, p<0.0). the repeating words measure showed significant decreases between nothing and social interaction (p<0.01), nothing and social interaction and videotape (p<0.01) and nothing and music (p<0.01) (F (3, 294)= 18.8, p<0.01. requests for attention showed a significant difference between social interaction, videotape and music (p<0.01) (F (3, 90)= 29.3, p<0.05. Hallucinations showed a significant difference between nothing and videotape (p<0.01) and nothing and music (p<0.01) (F (3, 39)= 3.6, p<0.01). *Nurses assessments:* duration of VDB decreased 'considerably' during the first three minutes of each intervention (apart from the no treatment) then remained sTable for the half hour intervention before declining post intervention. Appropriate verbal behaviours increased significantly during the one-to-one social interaction and videotapes (F (6, 248)= 25.57, p<0.01).	GRADE score - - - + Levels of evidence score LII-b, grade B.
180 Martin (2004) Descriptive Qualitative Observation-service evaluation	Canada A not-for-profit long term care facility which consisted of four 15 bed houses and a adult day programme with 54 clients.	63 in the nursing home, with most taking part at least once. Mean age 84, range 60-102. 46 (73%) were female and 17 (27%) were male. 54 took part from the day programme. Mean age 78, range 54-92.	Music- drum circles The sessions were carried out on average twice a week and the observation time was a period of 5 months. Four project members attended a half-day workshop with the custom drum manufacturer in order to learn how to implement the drum circle. Participants were free to join or leave the drum circle at any time. Any of the instruments chosen by the participants would be accepTable and residents were able to switch from one instrument to another.	To assess the feasibility of a drum circle intervention for older people who lived within a care home and had functional cognitive abilities. Questions focused on the observed experiences of the participants and facilitators.	Ethnographic field notes; both participant and non-participant observations, real time interviews focus groups and interviews with facilitators. 2 facilitators recorded notes in 3 of the 20 sessions.	Group comparison showed nursing home participants were more likely to be female, older and severely regressed (if they had dementia)- statistical values not given. Length of sessions ranged from 21-52 minutes (mean 37) for the nursing home group and 25-46 minutes (32 minutes mean) for the day programme. Day programme group numbers ranged from 12-17 (mean= 15) and nursing home groups from 8-17 participants (mean= 10). Emergent thematic behaviours: empowerment, bringing out and bringing together. These behaviours magnified to create energy and community resonance.	Risk of bias score: Grade C, some flaws that may affect the credibility, transferability, dependability and/or confirmability of the study.
185 Hilliard (2004) Descriptive Dates not specified.	USA 15 different nursing homes, 7 in Tallahassee and 8 in	Total of 80; 40 females (50%) and 40 males (50%). Intervention group mean	Music- therapy Cognitive-Behavioural music therapy carried out within nursing homes by a certified music therapist with all-live music and	To explore the use of music therapy for people receiving hospice care in a nursing home environment with	Patient notes were utilised from individuals who had died. Information on length of life and amount of time spent with the therapist were extracted.	Mean number of days within the hospice (length of life) was significantly greater for the music therapy group (141.28 (± 129.5), than the group not receiving music therapy (24.28, ± 44.44), t-5.867, df=78, p<0.001. Music therapists participated in more visits with participants than social workers (mean 10.88 vs. 6.53 respectively) t=5.447, df=39, p<0.001. Music therapists spent more minutes with the residents than the social workers (mean 399.98 vs 163.88 respectively), t=5.889, df=39, p<0.001.	GRADE score ? ? ? ? - + Levels of evidence

Reference of Paper, Study ID & Research Design	Location	Population	Activity	Aims of the study	Outcome measures	Findings	Quality
Quantitative Retrospective Quasi-RCT.	rural areas of nearby counties.	age male 78 (range 49-95) and mean age female 77 (range 48-98). Control group 50% male (n=20) mean age 70 (range 22-89) and female mean age 82 (range 48-98).	sessions planned by the therapist.	emphasis on length of life.		Care plans showed cognitive issues were most addressed by nurse (n=47), then music therapists (n=3) then social workers (n=1). Physiological needs were also most often addressed byy the nurse (n=273), then music therapist (n=8) then social workers (n=1). Emotional needs were addressed by both the social workers (n=38) and music therapists (n=39). Social needs were addressed by the social worker (n=33) then the music therapist (n=1) but not the nurses (n=0). Music therapists and social workers addressed spiritual need (n=7, n=2 respectively) but not nurses.	score LI, grade A.
201. Gerdner (2000) Intervention Dates not specified. Quantitative Cross-over RCT	USA Six participating long term care facilities located within Iowa.	39 in total. 30 females (67%) and 9 males (33%). Mean age 82.6 years.	Listening to classical relaxation music and individualised music.	To compare immediate and 30 minute residual effects of those exposed to 'classical relaxation music' with those exposed to individualised music with respect of agitated behaviours.	A modified version of the Cohen Mansfield Agitation Inventory (CMAI). This instrument had been modified using 10 minute increments to allow for the 60-minute observation period in order to provide definitive assessment of immediate and residual effects	14 of the 16 demographic variables were not statistically different between the two groups. Significant differences were shown for the variables of urinary incontinence and assistance needed with ambulation but these variables were not associated variables of gait or toileting and therefore the groups were regarded demographically similar. There were occasions of data missing from the MCMAI and when this was the case least square means were calculated using individual values to predict missing values. ANOVA revealed a significant two-way interaction between phase (df (2, 74), F=32.93, p=0.001) and minute (df 5, 2763, F=53.28, p=0.001) and the main effects of phase and minute was also significant (df(10, 2763), F= 4.77, p=0.001). No significant effect was found for phase and week (F= 4.77, p=.07) or phase, week and minute (F= 0.64, p=0.99). Bonferroni Post-Hoc analysis showed that individualised music was associated with a significantly greater decrease in frequency of agitated behaviours, compared with classical music, at each of the three 10-minute intervals that music was played (0-30 minutes) and the 30-minute post-intervention phase. Frequency of agitated behaviours was also significantly less during each of the three 10 minute increments during which individualised music was playing. Frequency of agitated behaviours was significantly less during the 30 minutes immediately following presentation of individualised music compared with the corresponding three 10-minute intervals of baseline assessment. Finally, no significant difference occurred in frequency of agitation between baseline and first 20 minutes of classical music. A significant decrease in agitated behaviours did occur during the final 10 minutes of classical music compared to baseline. This continued only during the first 10 minutes following presentation of classical music.	

Most studies used an uncontrolled pre/post-test design (n=20, 42%). Sixteen (33%) studies were randomised controlled trials (RCTs), six (38% of included RCTs, 13% of total music studies) of which used a cross-over design. Eight (17%) of the studies adopted a quasi-RCT design and one study was a retrospective cohort study (Hilliard, 2004). The remaining three descriptive studies sourced data from observations only (Gotell, 2007), observations combined with interviews and focus groups with staff members (Martin, 2004) and focus groups with residents (Chen, 2009).

4.2.2 Aims of included studies

Many of the studies (n=17, 35%) identified more than one aim (range 1-4) (aims of all included studies are reported in Table 4.1). Over half the studies (n=26, 54%) assessed the impact of the activity on behavioural symptoms. Definitions of behavioural symptoms varied amongst the studies but primarily focused upon agitation, verbally disruptive behaviours (VDB), restlessness and irritability (Table 4.1). A third (n=16, 33%) aimed to determine the impact of the activity on incidences of agitated behaviours (see Table 4.1). Four specified verbally disruptive behaviours (VDB) associated with dementia (Casby, 1994; Cohen-Mansfield, 1997; Garland, 2007; Nair, 2011). Three reported the general term of 'behavioural symptoms' (Hagen, 2003; Holmes, 2006; Svansdottir, 2006). Two studies carried out by the same author (but not using the same data) stated the aim of assessing improvements in 'irritability and restlessness' (Rangeskog, 1996; Ragneskog, 1996b), with the first study also exploring 'uninhibited behaviours'. Only one study focused solely on 'positive' behaviours (Holmes, 2006), although positive behaviours were also assessed by both Ziv (2007) and Sole (2014).

The second most reported aim was the impact of the activity on measures of psychological wellbeing, outlined by 17 (35%) studies. Six (13% of the music studies, 35% of those focused on psychological wellbeing) of these specified depression (Ashida, 2000; Myskja, 2008;

Clarkson, 2007; Cooke, 2010a; Liu, 2014; Koike, 2014), two focused on anxiety (Cooke 2010; Sung, 2010) and two assessed both depression and anxiety levels (Guetin, 2009; Mohammadi, 2011). Remaining studies all referred to non-specific assessments of 'mood' (Lord, 1993), 'psychological symptoms' (Svansdottir, 2006; Koike, 2012), 'affect' (Suzuki, 1998; Sole, 2014) and 'emotional expression' (Hagen, 2003; Hammar, 2011).

Eight studies (17%) assessed the impact on cognition, of these, two focused on memory recall (Lord, 1993; Suzuki, 1998), one on face-name recall (Carruth, 1997) and one on alert-responses (Clair, 1996). The remaining three studies stated aims which simply referred to 'cognition' (Smith, 1986; Hagen, 2003; Hong, 2011; Koike, 2014).

Six studies focused on improvements to 'quality of life', with all but one stating general quality of life measures (Vanderark, 1983; Bennett, 1988; Burrack, 2004; Cooke 2010a; Ridder, 2013) and the last study unusually focused on length of life (Hillard, 2004). Five studies assessed improvements to care, two of which focused on improved food intake (Ragneskog, 1996a; Ragneskog, 1996b) and one on resisting care (Hammar, 2011). The final study did not specify how this would be determined and referred only to 'quality of care' (Millard, 1989). Finally, improvements to socialisation were identified as aims in two of the studies (Lord, 1993; Ziv, 2007) and two studies focused on physiological improvements associated with the activity (Hagen, 2003; Koike, 2012). Bennett (1988) also assessed the enjoyment of participants through an un-validated questionnaire.

For descriptive studies, one sought to explore the perceptions of residents taking part (Chen, 2009) and one to determine the feasibility of the music activity (Martin, 2004). The final qualitative study aimed to assessed changes to morning care routines because of the musical

activity (Gotell, 2007), this was also the aim for the qualitative part of the mixed methods study (Hammar, 2010).

4.2.3 Location and setting

The largest proportion of studies originated from the USA (n=14, 29%) (see Table 4.1). Five studies (10%) were located within Australia and four (8%) in Canada. A further 12 studies (25%) were carried out within Europe: of these four (8%) were from Sweden, two from the UK and the remaining studies took place within Belgium, Iceland, Norway, France, Holland and Spain. Nine (19%) were in East-Asia: six (13%) of these were from Taiwan and the remaining three from China, Japan and South Korea. Two originated in Israel and one in Iran. Only one study included data from more than one country, with participants included from 14 care homes, ten from Norway and four from Denmark (Ridder, 2013).

Operationalisation of 'care homes' different across studies, 21 referred to 'nursing homes', seven to 'long term care facilities' and six used the term 'care home'. Other terms used were 'residential homes' (n=4), 'dementia speciality facility' (n=3), 'care facility' (n=1) and 'aged care facility' (n=1). Five studies did not specify care home type or provide further details as to where participants resided. Seventeen studies (35%) reported data was provided by residents from more than one care home (range 2-14). Holmes (2006) reported data were obtained from 'residential homes' but did not specify the total number included.

Eleven studies (23%) did not specify the location within the care home where the music activity was delivered. The activity took place within the residents' bedrooms for five studies (Clair, 1996; Cohen-Mansfield, 1997; Garland, 2000; Burrack, 2004; Guetin, 2009) with Guetin (2009) also specifying that residents could choose to engage with the activity in another area of the home if they preferred. Remington (2002) and Hicks-Moore (2007) stated the activity

was carried out either within the residents' rooms or the lounge area and Carruth (1997) reported use of the residents' bedrooms for one unit of the residential home and the living area for another. An additional two studies reported the activity was carried out during morning care routines, which took place within the residents' rooms (Gotell, 2007; Hammar 2010; Hammar, 2011). Delivery of the intervention took place within the dining room area for five studies (Millard, 1989; Goddaer, 1994; Ragneskog, 1997a; Ragneskog, 1997b; Hicks-Moore, 2005). Two of these (Ragneskog, 1997a; Rangeskog, 1997b) were specifically assessing the impact of the activity during meal times. Just two specified use of a specific 'activity room' (Martin, 2004; Sung, 2006). Other locations included 'communal areas' (Holmes, 2006; Nair, 2011), 'lobby area' (Ziv, 2007), a 'common room' (Lord, 1993) and where residents 'spent most of their time' (Gerdner, 2001). Chang (2009) reported music was played via an internal broadcast system, so would be heard across the home and Vink (2013) simply stated that the activity took place 'away from the ward'.

4.2.4 Populations

Demographic information relating to participants of individual included studies is shown in Table 4.1. Data were provided by a total number of 1,731 older adults (inclusive of control groups) across the 48 studies, with a range of actual sample sizes of 3-117 (mean sample size 36). The study with the largest sample size (Martin, 2004) included only 63 participants who resided within the care home, additional data were provided by 54 day clients. Data were analysed for the total sample and could not be separated. It is also unclear whether all 117 residents were observed for all the sessions or whether different residents were observed for each session. Two studies also collected qualitative data from caregivers through observations (Gotell, 2007) and semi-structured interviews (Hammar, 2010).

Total sample sizes differed with study design, the 16 RCT's showed largest sample sizes with an average group size of 48 (range 19-100). Average sample sizes for the quasi-RCT's was 40, however there was a large range of 3-60 participants. With the removal of Casby (1994), which had a very small and unjustified sample size (n=3), the average group size for the remaining quasi-RCTs was 45 with a smaller range of 26-60. The smallest mean sample size was shown for the uncontrolled pre/post-test studies (mean= 22, range 8-63), which would be expected given the lack of a control group.

Numbers of those who participated in the music activity totalled 1,355 participants. Sixteen RCTs and quasi-RCTs featured control groups which did not take part in the activity, equalling 376 participants, with a range of 8-40. One of the quasi-RCT studies involved a cross-over design (Hicks-Moore, 2005) and the second featured only three participants who all took part in at least some form of musical activity (Casby, 1994). In addition, six of the RCTs were cross-over studies, therefore both groups experienced the music activity at some point during the study.

For the studies (n= 34, 71%) that reported mean age the total mean age was 81.9 years (range 69-87.5 years). Most studies that reported age featured a mean of 80 years or older (28 out of 34, 82% of studies that reported age), with only one study where a mean age of lower than 70 was reported (Mohammadi, 2011). Only four studies failed to report ages of participants (Bennett, 1986; Hagen, 2003; Koike, 2012; Liu, 2014). An additional three did not report the mean age, but did report age range, Svansdottir (2006) ranged from 71-87 years, Lord (1993) between 72 and 103 years and Clair (1996) between 62 and 83 years. Two studies reported the brackets within which the age of the included participants fell (without recording a mean or total range) with one reporting 60% of the sample to be between 85 and 94 (Cooke 2010a;

Cooke 2010b) and the second reporting 62% (n=28) of the sample to be over the age of 80 (Ledger, 2007).

In total 42 (88%) studies reported proportions of men and women, the majority, 973 were women (70%) and 409 men (30%) (total of 1382 residents). Only five studies did not report more women than men, two featured equal numbers (Ragneskog, 1996b; Hillard, 2004), two featured more males in the experimental group only (Mohammadi, 2009; Sung, 2006) and Sung (2006) had more males in both the control and experimental group (proportions of men and women for all studies shown in Table 4.1). Two studies that did not report participant's ages also failed to report gender (Hagen, 2003; Liu, 2014).

Twenty-five (52%) of the music studies specified inclusion criteria for participants to have a dementia diagnosis, with two specifying an Alzheimer's Disease diagnosis (Svansdottir, 2006; Ledger, 2007) (see Table 4.1 for inclusion criteria of included studies). Studies by both Nair (2011) and Ziv (2007) did not specify inclusion criteria related to a dementia diagnosis or provide demographic information related to dementia, however both took place in dementia-specific facilities where it would be assumed a dementia diagnosis had been given to most if not all participants. An additional four studies did not specify a diagnosis of dementia, however they provided participant demographics which showed the total sample had dementia (Hammar, 2010; Hammar, 2011 Ho, 2011; Hong, 2011). In addition, Sole (2014) reported 56% (n=9) of participants had a dementia diagnosis and a further 31% (n=5) had 'probable dementia'. Myskja (2008) and Carruth (1997) also reported high proportions of individuals within their populations as having dementia, 74% (n=53) and 64% (n=4) respectively and Cohen-Mansfield (1997) reported 30% (n=10) of the sample to have multi-infarct dementia with over half (53%, n=17) having 'probable Alzheimer's'. The only study

reporting on demographics to record a minority with dementia was Hilliard (2004) where only 26% (n=10) of participants were reported to have this diagnosis.

4.2.5 Characteristics of the arts activities

Delivery of the musical activity varied between studies (see Table 4.1), just over a quarter (n=14, 29%) evaluated auditory exposure only where residents listened to pre-recorded music via CD or cassette players, through headphones or speakers. Ten percent (n=5) of included studies actively encouraged residents to participate using musical instruments during all or part of the sessions, adding a tactile element and encouraging creativity through rhythm and expression (Lord, 1993; Suzuki, 1998; Ashida, 2000; Martin, 2004; Mohammadi, 2011). Further to this, four described the encouragement of movement with the music (Vanderark, 1983; Brontons, 1996; Suzuki, 1998; Sung, 2006) and an additional study consisted of a structured 'musical exercise programme' (Hagen, 2003) (Table 4.1.).

Eleven studies (23%) carried out the music activity to coincide with caring situations that may increase distress and/or behavioural disturbances for residents (such as personal care), therefore placing strain on care staff. Four of these used one-on-one caregiver (or researcher) singing (Clair, 1996; Carruth, 1997; Gotell, 2007; Hammar 2010 and Hammar, 2011). Six featured music played during mealtimes to ascertain whether this background music could improve the disruption caused during this time of day (Goddaer, 1994; Ragneskog, 1996a; Ragneskog, 1996b; Hicks-Moore, 2005; Chang, 2010; Ho, 2011).

Only nine studies (19%) identified 'music therapy' was carried out and outlined a protocol for sessions coinciding with this method (Millard, 1989; Ledger, 2007; Myskja, 2008; Guetin, 2009; Chen, 2009; Lin 2010; Ridder, 2013; Vink, 2013; Sole, 2014). All but two of these (Guetin, 2009; Ridder, 2013) were carried out within a group setting. Liu (2014) also reported the

music activity to be 'music therapy' however no detailed protocol was provided within the study to evidence this. Similarly, Hilliard (2004) stated the activity of interest within their study was 'cognitive behavioural music therapy sessions', however there is no information as to the actual structure of any of these sessions. Furthermore, as retrospective information was sourced from different residents who lived within different care homes providing palliative care, consistency of the activity cannot be determined either.

Most studies carried out music activities within a group setting (n=35, 73%) (see Table 4.1). Group size was explicitly recorded by ten of these (21%), with only two studies reporting groups with more than ten participants (Lord, 1993 (group size of 20); Martin, 2004 (group size ranged from eight to 17) (see Table 4.1 for group sizes of each of the studies). Nine (19%) studies delivered the music interventions individually with participants. This was likely due to the aims of the studies, four featured staff singing during the care giving process (Clair, 1996; Carruth, 1997; Gotell, 2007 Hammar, 2010; Hammar, 2011) and five featured individually playing music to residents (Casby, 1994; Cohen-Mansfield, 1997; Gerdner, 2000; Garland, 2006; Hicks-Moore, 2007). Two did not report whether sessions took place in a group (Remington, 2002; Liu, 2014) and Hilliard (2004) contained little information about the music therapy sessions themselves, although it is probable that these also occurred within a group setting as they were live music performances. Hong (2011) indicated participants could choose whether the performance and validation of their songs took place within a group or individual setting, however it is probable the rest of the intervention took place in a group.

Duration of the programme ranged from one-off songs (Casby, 1994) to evaluation of an ongoing music therapy programme carried out weekly for at least 42 weeks of the year (Ledger, 2007) (see Table 4.1). Most popular frequency of sessions was weekly, reported in

11 studies (23%) (Bennett, 1988; Brontons, 1996; Ledger, 2007; Ziv, 2007; Chen, 2009; Guetin, 2009; Mohammadi, 2009; Hong, 2011; Hammar, 2011; Liu, 2014; Sole, 2014), with the second most popular frequency being twice a week, reported by eight studies (17%) (Vanderark, 1983; Millard, 1989; Martin, 2004; Sung, 2006; Myskja, 2009; Lin, 2010; Sung, 2010; Vink, 2013). Frequency of greater than this was reported for nine studies (19%) , five (10%) reported three sessions per week (Hagen, 2003; Svansdottir, 2006; Garland, 2007; Suzuki, 2009; Cooke, 2010), one five days per week (Clarkson, 2009) and three were carried out daily (Ashida, 2000; Garland, 2007; Chang, 2010; Koike, 2012). Length of programme (where music was played) was shortest for those delivering programmes every day (or every weekday), between one and four weeks' duration (Ashida, 2000; Chang, 2009; Clarkson, 2009; Koike, 2012). Conversely, of the eleven studies that took place weekly, only four were less than 10 weeks in duration (Bennett, 1998; Ziv, 2007; Brontons, 2009; Hammar, 2010), with the remaining studies' music sessions being delivered for between 10 weeks and one year (Table 4.1).

Just under a quarter of studies (n=11, 23%) reported use of music preferred by the participants, study designs were mixed for those featuring preferred music, with five RCTs, three uncontrolled pre/post-test studies, two quasi-RCTs and one descriptive study (see Table 4.1). However, only five gained input directly from the residents on their musical preferences (Gerdner, 2000; Burrack, 2004; Sung, 2006; Myskja, 2008; Guetin, 2009), the remaining studies sought information from family members (Casby, 1994; Cohen-Mansfield, 1997; Garland, 2006; Gotell, 2007; Hicks-Moore, 2007; Sung, 2010). There was variation in the level of information collected by studies identifying the use of preferred music. Cohen-Mansfield (1997), Burrack (2004) and Guetin (2009) collected information on preferred genre only. Only one study provided use of any type of data collection tool for ascertaining the musical

preferences of the residents, reporting the use of the Modified Hartsock Music Preference Questionnaire (Gerdner, 2000).

Despite reporting preferences were obtained from residents, two interventions were carried out within a group setting (Sung, 2006; Myskja, 2008). The first stated preference was gained from the 'majority' of participants, however no numbers are given as to how many residents this refers to. In addition, it is not explicitly stated whether all residents were asked to provide information on their preferred music or if certain residents were chosen, and therefore it is not known how this information was obtained (Sung, 2006). The second study (Myskja, 2008) contained clearer information as to how individual music preferences shaped the group sessions. Information was obtained via individual questionnaires and the residents' music collections to ascertain preferences of the group and then modified the sessions based upon the preferences and responses of the whole group (Myskja, 2008). The remaining nine preferred music studies were all carried out individually.

An additional six (13%) studies (two RCTs, two descriptive, one uncontrolled pre/post-test study and one using a quasi-RCT design, see Table 4.1) reported music for sessions was decided by both the participants and the person delivering the sessions. One study (Vanderark, 1983) described familiar songs jointly chosen by both the facilitator and the participants and a similar process was described by Svansdottir (2006) where a list was compiled by the music therapist carrying out the sessions and songs were chosen from that list by participants. Lin (2010) collected information on music preferences of participants and then chose the music, however it is not explained how participants' musical preferences were obtained, or how much input they had in shaping the session. Content was chosen by the music therapist in two studies (Chen, 2009; Sole, 2014) with feedback given at the end as to

content for the following session. Again, it is not clear how much say participants had, whether they were all consulted or simply asked to volunteer their thoughts. Martin (2004) described drumming sessions that were unstructured with rhythms being created by the therapist and participants as the sessions progressed.

Nine (19%) studies did not report who chose the music for the sessions (Lord, 1993; Carruth, 1997; Houston, 1998; Suzuki, 1998; Hilliard, 2004; Mohammadi, 2011; Liu, 2014). The remaining studies all featured music which had been chosen by either the music therapist, or researcher (which was often the same person).

4.2.6 Research methods and outcome measures

Descriptions of the research methods and outcome measures used within each of the studies is found in Table 4.1. The most commonly used validated measure was the Cohen-Mansfield Agitation Inventory (CMAI), used in 15 of the quantitative studies (29%), over half of which were RCTs (n=8), five were uncontrolled pre/post-test designs and two used a quasi-RCT design (Table 4.1). The scale was used in its traditional form for five of these studies (Cohen-Mansfield, 1997; Ledger, 2007; Chang, 2010; Lin, 2010; Ho, 2011) recording retrospective occurrences of agitated behaviours in residents and an additional study (Ridder, 2013) used the CMAI-fr (frequency) and CMAI-di (disruptiveness). Cooke (2010b) made use of the CMAI short form (CMAI-sf) with 14 items, rather than the 29 on the full version. Five studies used a modification of the CMAI for use as an observational tool, recording the presence or absence of behaviours in real time instead of retrospectively (Goddaer, 1994; Remington, 2002; Hicks-Moore, 2005; Hicks-Moore, 2007; Vink, 2013). Both the traditional and modified (presence/absence observational) CMAI was used in Clarkson (2007). A similar modified version that allowed behaviours to be assessed in ten-minute increments during the activity was used by Gerdner (2000) and Sung (2006).

Scoring of the CMAI also differed amongst the studies that used it, staff members within the care homes completed the scale in five of the studies (Cohen-Mansfield, 1997; Ledger, 2007; Chang, 2010; Cooke, 2010b; Ridder, 2013; Vink, 2013) and in Clarkson (2007) staff members completed the traditional version only. Research assistants scored the measure in three studies (Gerdner, 2000; Remington, 2002; Hicks-Moore, 2007) and a member of the research team for the presence CMAI measure in Clarkson (2007). Two of the studies reported the CMAI to be scored by 'trained raters' (Goddaer, 1994; Hicks-Moore, 2005) and in one the measure was scored by nursing and/or social work students (Chang, 2010). The remaining studies were not explicitly clear who completed the CMAI (Sung, 2006; Lin, 2010; Ho, 2011). The CMAI was translated in three studies, a validated Chinese version (C-CMAI) was used for two studies (Chang, 2010; Lin, 2010), with another (Goddaer, 1998) translating the English version of the CMAI into Dutch, with the Dutch version then translated to English to ascertain validity of this version.

Four studies used observation schedules to assess the impact on behaviour (Svansdottir, 2006; Ziv, 2007; Garland, 2007; Nair, 2011). Only one of these used a standardised observation tool, the BEHAVE-AD (Svansdottir, 2006). The remaining three devised their own measures based on the items found on the CMAI (Ziv, 2007; Garland, 2007; Nair, 2011).

The impact on psychological health was also assessed differently across studies (Table 4.1). The Geriatric Depression Scale (GDS) was used by Guetin (2009) and Cooke (2010a) and was completed by the residents before, during and following the music activity. The Cornell Scale for Depression (CDS) was used by Clarkson (2009) and Ashida (2000) and was completed by staff members from the home in both studies. Depression levels were also assessed via the Montgomery Asberg Rating Scale (MARS) (Myskja, 2009) and the Self-Rating Depression Scale

(SDS) (Liu, 2014) at baseline and follow-up. The MARS was completed by care staff (Myskja, 2009) and it is unclear who completed the SDS, although given it is a self-report measure it would seem likely residents completed the measure.

Two studies focused on anxiety used the Rating Anxiety in Dementia (RAID) measure (Cooke 2010a; Sung, 2010), completed by the residents themselves (Cooke 2010a) with respect of their own feelings of anxiety and staff members based on perceptions of anxiety in residents (Sung, 2010). Both studies scored the measure at baseline and following the activity, with Cooke (2010a) also completing mid-way through. Guetin (2009) used the Hamilton Scale for Anxiety, administered at the same times as the GDS (discussed above). Svansdottir (2006) assessed anxiety levels, along with other behavioural symptoms (outlined above) via the BEHAVE-AD. In addition, Clarkson (2009) and Ridder (2013) collected information from patient charts on levels of 'as and when' (p.r.n) anti-anxiety medication use during the study period. Mohammadi (2011) used the Depression Anxiety Stress Scale (DASS-21) to measure levels of depression, anxiety and stress at baseline and immediately following the study period. This self-report measure was completed by the residents but it is not recorded who administered them.

Observed emotion was assessed by Hammar (2011) and Sole (2014) from videotaped sessions. A standardised observation schedule was used in Hammar (2011) (the Observed Emotion Rating Scale- OERS) and rated by members of the research team. Sole (2014) devised a five item schedule and used a professional music therapist to score the items. The latter study also used items from Category E on the GENCAT which reports on emotional wellbeing, this was scored by research assistants.

The Mini Mental State Examination (MMSE) was used in both Smith (1986), Hong (2011) and Koike (2014), pre and post-test. In Hong (2011) the MMSE-K (Korean version) was said to be scored by rated by a researcher and social worker, it is unclear how this took place and what input residents had. Two studies focused on memory with one through a devised questionnaire (Lord, 1993) and the second through the Memory Retrieval Test (Suzuki, 2006), both were scored by researchers. Hagen (2003) used the cognitive assessment scale (CAS) scored at baseline, post-test and then at 10-week follow-up. Two studies focused on responses of residents, the first assessing face-name recall (Carruth, 1997) and the second alert-responses in residents who were non-ambulatory and non-verbal (Clair, 1996).

Quality of life measures also differed among the studies. Four used standardised measures, The Life Satisfaction Index A (LISA) and ego integrity scale (Bennett, 1998), Quality of Life in Alzheimer's Disease Scale (QoL-AD) (Burrack, 2004), Dementia Quality of Life Scale (DQoL) (Cooke, 2010a) and Alzheimer's Disease Rated Quality of Life measure (ADRQOL, adapted to refer to the previous week only) (Ridder, 2013). Vanderark (1983) developed their own measure which contained QoL and self-concept measures. Three of these studies administered outcomes at pre- and post-test (Vanderark, 1983; Bennett, 1988; Burrack, 2004), one at baseline, midway and immediately following (Cooke 2010a) and another at baseline, crossover and following (Ridder, 2013). Only two studies reported who administered the outcome measures and in both instances this was research assistants (Cooke, 2010a; Ridder, 2013). Hilliard (2004) unusually focused upon length of life following taking part in a music therapy programme using retrospective data obtained from residents' notes.

The two studies collecting data related to physiological improvements featured measurements of balance and joint range (Hagen, 2003) and temperature, oxygen saturation,

blood pressure and pulse rate (Koike, 2012). Sleep cycles were collected via actigraphy in Koike (2012) and this method was also used to assess levels of agitation in Clarkson (2007).

Two of the descriptive qualitative studies conducted group interviews. Chen (2009) carried out focus-group interviews directly with participants, with 6-8 residents in each group. Three groups were formed and interviews were carried out until data saturation was reached. Hammar (2011) conducted their unstructured interviews with care staff, who were asked open-ended questions prior to carrying out music therapeutic caregiving and following. A third study also carried out focus-group interviews (Martin, 2004) as part of an ethnographic study which also collected field notes and observations carried out by 2-3 observers for a total of 20 sessions. A total of three observations were videotaped (one for each condition) in Gotell (2009) for each of the nine residents (27 in total).

In total, 19 studies (40%) featured a comparison condition, either comparing different musical genres or through offering a different activity (Table 4.1). Six studies (Casby, 1994; Ragneskog, 1996a; Ragneskog, 1996b; Gerdner, 2000; Holmes, 2006; Gotell, 2007) provided a comparison of different types of music. Two evaluated the effects of 'preferred' and 'relaxing music' (Casby, 1994; Gerdner, 2000), with one featuring a control group (Casby, 1994) and the second also comparing the two genres to participants' baseline scores (Gerdner, 2000). Despite featuring control conditions and two comparisons, Casby (1994) only featured three participants and each condition featured only one song, a chosen preferred song (only listened to by two of the three residents), a calming song and silence. Two uncontrolled pre/post-test studies compared different genres of music (soothing, 'music from the 1920s and 1930s' and rock and pop music) (Ragneskog, 1996a; Ragneskog 1996b). The final two studies compared live music, listening to music and silence in an RCT cross-over (Holmes,

2006) and the effects of background music and music therapeutic caregiving compared with usual care in an uncontrolled pre/post-test study (Gotell, 2007).

A further four studies featured control groups that took part in occupational therapy activity programmes (described as board games, puzzles etc.) (Hagen, 2003; Vink, 2013) and other similar activities (not described as occupational therapy) such as puzzles (Lord, 1993) and reading (Guetin, 2009). Two featured interactive group activities in the form of discussion (Millard, 1989) and reading groups (Cooke 2010a; Cooke 2010b). Another study compared the effects of care-givers sitting in silence with non-ambulatory and non-verbal residents, reading to them and singing to them on levels of alert responses (Clair, 1996). Four studies compared the effectiveness of the music activity with that of a comparison activity, two featured simulated family presence tapes (Cohen-Mansfield, 1997; Garland, 2007). In addition, Cohen-Mansfield (1997) also featured a one-to-one social interaction condition. Hand massage was compared with the music activity in both Remington, (2002) and Hicks-Moore, (2007). Finally, two earlier studies (Smith, 1986; Bennett, 1988) carried out music-based life review and compared this process with traditional verbal-based life review.

4.2.7 Quantitative findings

4.2.7.1. Behaviour

In total, 26 (58%) studies assessed the effects of the intervention on behaviour. Significant changes over time were reported by nine studies that used the CMAI (Goddaer, 1994; Gerdner, 2000; Remington, 2002; Sung, 2006; Hicks-Moore, 2007; Ledger, 2007; Chang, 2010; Lin, 2010; Ho, 2011) (statistical results for studies using the CMAI are summarised in Table 4.2).

Table 4.2 Statistically significant results obtained from studies using the CMAI

Study ID	**Statistical Result**
Goddaer (1994)	Mean 7.2 (SD 7.49) vs. 2.75 (SD 4.24), *F*(3, 78)= 8.52, p<0.001
Gerdner (2000)	Mean values reported as graphs, minute= *F*=32.9(2, 74), p=0.001, phase= *F* (5, 2763)= *p*=0.001
Remington (2002)	Mean 18.41 (SD 11.19) vs. 4.65 (SD 7.87), *F*cons=6.47, p<.01
Sung (2006)	Mean 5.11 baseline, 3.94 at week 2 and 3.44 at week 4, *F*(2,33)=15.03, p<0.001.
Hicks-Moore (2007)	Mean 0.75 (SD 0.72) baseline, 0.08 (SD 0.24), *F*corr=24.74, *p*=0.001
Ledger (2007)	Mean total CMAI scores only shown in a graph, reported *F*=2.61, p<0.05)
Chang (2010)	Mean total CMAI scores only shown in GEE plots, reported (B= 1.06, p=0.04)
Lin (2010)	Mean 43.12 (SD 16.32) at baseline vs. 36.37 (SD 10.64) at week 12. GEE plots, reported *p*=<0.001
Ho (2011)	Mean baseline 60.64 vs. 42.99 at week 4 X^2=97.557, *p*<0.008

Although they report the CMAI was used, little is reported on the findings from this measure within Cohen-Mansfield (1997), just that there were no statistically significant differences in scores at baseline or following the activity. Similarly, Hicks-Moore (2005) contained mainly descriptive data with no statistical comparison of CMAI scores meaning while scores were shown to improve the values of such results cannot be determined. CMAI scores were not significantly improved by the music activity for three studies (Clarkson, 2009; Cooke 2010a; Cooke 2010b; Ridder, 2013) (Table 4.1).

Of those assessing behaviour but not using the CMAI statistically significant changes were shown for reduced agitation (Garland, 2007 (F12, 18 =2.62, p=0.032)) increased positive behaviour (Ziv, 2007 (5.18 vs 3.62, *t*=6.75, *p*=0.001)) and increased engagement with live music (Holmes, 2006, (median +1 vs 0 Wilcoxon (value NR)= p<0.0001, no significant different for pre-recorded music) across time.

Changes over time were suggested by Casby (1994) and Hagen (2003), however both presented their analysis graphically, with no statistical summary of results and therefore magnitude of change cannot be determined. The first study also consisted of only three participants. Three studies did not show statistically favourable results for the music condition over time (Svansdottir, 2006; Nair, 2011). Despite stating the aim of assessing the impact of

the activity on behaviour, both Ragneskog (1996a) and Rangeskog (1996b) provided little evidence of assessing this. Similarly, Hicks-Moore (2005) identified use of the CMAI but did not carry out statistical analysis. A final study by Sole (2014) simply stated that all behaviours observed more than four times by residents were positive, aside from agitation, shown for two residents. No further statistical analysis occurred with these results which could have provided insight into possible increases in these positive behaviours that could have been a benefit of the activity (see Table 4.1).

Findings were tentative when comparing music with another form of arts or social activity, five studies failed to show a significant difference with relation to behavioural changes between two groups post-study (Remington, 2002; Hagen, 2003; Garland, 2007; Hicks-Moore, 2007; Vink, 2013) (Table 4.1). However, music was particularly beneficial when compared to a group who did not receive any form of activity. Sung (2006) showed significantly fewer behavioural disturbances for the music group compared with the control group post-study (m=3.44 (SD 2.45) vs. 4.5 (S1.65) t-3.85, p=0.001). Similarly, Garland (2007) showed both music and the comparison activity (simulated family presence) significantly reduced verbal and physical behavioural symptoms across time with neither activity shown to be more effective than the other (number of statistical tests, shown in Table 4.1). However, the music groups showed significantly fewer incidences of agitated behaviours when compared with a group where only usual care was offered (mean difference across time= 0.31, *F*= 0.77, p=0.04). Similar findings were shown in Remington (2002) using the modified CMAI, while there was no significant difference in agitation scores between music and other forms of activity, those who did not take part in any activity showed significantly higher agitation scores (mean 4.65 vs. 20.47, Fcons=6.47, p<.01) at the end of the study period. Ridder (2013) used the CMAI disruptiveness and frequency scales in a cross-over RCT, reporting agitation

disruptiveness increased during the standard care condition and decreased during music therapy whereas agitation frequency slightly increased for standard care and decreased for music therapy, however this result did not reach significance. Using the BRS, Hagen (2003) also showed both music and occupational therapy decreased behavioural symptoms more so than those who did not take part in an activity (*F* 2,144= 19.6, p<0.001) (Table 4.1). Svansdottir (2006) did not show the music condition to be more beneficial than usual care.

There were differences in the types of behaviour impacted upon by the music activity. Physically non-aggressive behaviours were significantly reduced post-test in four studies (Goddaer, 1994 (4.10 vs. 1.79, *F*(3, 78)= 5.16, *p*<.003); Remington, 2002 (18.4 vs. 4.65 *F*cons=3.78, *p*=0.01); Hicks-Moore, 2007 (1.99 vs. 0.44, *F*corr=123.38, *p*=.001)); Garland, 2007, 27% decrease, (*F*6, 174=2.42, *p*=0.029)), but verbally disruptive behaviours were only significantly reduced for two (Goddaer, 1994, 1.41 vs 0.36, decrease *F*(3, 78)= 4.01, *p*<0.01; Hicks-Moore, 2007, 1.92 vs 0.37, *F*corr= 24.74, *p*=0.001). Furthermore, Nair (2011) (mean change 1.0 (CI -0.09 to 0.73), *p*<0.001) and Clarkson (2009) (10.7 (SD 5.4) vs.13.0 (SD5.5), t[17]=2.74, p<0.015) demonstrated significant increases in verbal agitation levels. Cooke (2010b) determined that increased verbalisations were only shown in those frequently attending sessions (mean 1.39 (CI 1.20-1.58 vs. 1.50 (CI 1.28-1.72) *F*(1, 46)= 3.534, p<0.05), suggesting participants became familiar with the activity over time. Anecdotally, staff members in Clarkson (2009) attributed this to increased positive behaviours but this was not statistically examined. The only study to determine differences in positive and negative behaviours (Ziv, 2006) showed increases in positive behaviours during the music activity (mean =5.18 vs. 3.62, *t*=6.75, *p*=0.001) and fewer negative behaviours (mean= 5.96 vs 2.43, *t*=5.05, *p*=0.001). This is reinforced by findings from Holmes (2006) who found increased engagement with live music versus silence (median +1 vs. 0, *X*2(value not reported) *p*<0.001).

Sole (2014) also described increased positive behaviours, however no statistical analysis was provided for such incidences. Further limited evidence was provided by Ragneskog (1996b) who found significant improvements for the domains of irritability, fear-panic and depressed mood (simply reported as p<0.05), but not for any other domains on the GBS.

4.2.7.2 Psychological wellbeing

Eight studies analysed depression levels (Table 4.1), two of which evaluated depression and anxiety through one outcome measure (Guetin, 2009; Mohammadi, 2011). Statistically significant reductions in depression scores were shown for Ashida (2000) (F=-3.77, *df* 2.57, *p*<0.001), Myskja (2008) (mean 20.4 vs 12.2 following *p*<0.05), Koike (2014) (mean 36.66 (SD 11.13) vs. 32.93 (SD14.05) p<0.05 and Liu (2014) (56.2 (19.1) vs 49.9 (18.1) p<0.05) with the latter two studies only including participants with pre-existing depression. Clarkson (2007), Guetin (2009) and Cooke (2010a) reported no statistically significant improvements in depression levels for their total participants across time and Mohammadi (2011) did not analyse changes pre- and post-test. However, Cooke (2010a) carried out post-hoc analysis on those with scores indicative of mild depression (greater than five on the GDS) and showed significant improvements (F(2, 22)= 8.129, p<0.01). Further to this, Clarkson (2007) commented on low baseline scores which may have affected the likelihood of yielding a statistically significant change in depression levels

Three studies with control groups showed statistically significant differences between the two groups at post-test (Mohammadi (2011), *Z*=3.60, p<0.001, Guetin (2009), 8.9 vs. 11.2 ANOVA *p*=0.006) and Liu (2014), mean= 49.9 vs. 40.2, (test NR) *p*<0.05)). Changes were also significant at six-week follow-up in the latter study (Guetin, 2009) (*p*=0.006). However, Mohammadi (2011) failed to analyse changes over time and Guetin (2009) showed no significant changes over time so it is possible that rather than improving depression levels the activity prevented

decline. Cooke (2010a) showed no significant changes between groups post-test, this was the only study of the that described an activity (reading) for the control group.

No significant changes in anxiety scores over time were observed by Cooke (2010b) or Guetin (2009). Sung (2010) reported significantly improved anxiety levels following the music programme (mean 10.93 (SD 5.46) vs. 8.93 (SD 4.6) t= 5.64, p<0.001) and Mohammadi (2011) did not assess these changes. The intervention group showed significantly lower anxiety scores following the study in both Sung (2010) (mean 8.93 (SD 4.6) vs. 9.35 (SD 4.09) F=2.14, p=0.001) and Mohammadi (2011) (3.27 (SD 2.97) vs 8.13 (SD 2.16) Z=2.91, p=0.004) and were significantly improved at six-week follow up for Guetin (2011) (10.6 (SD6.3) vs. 20.5 (SD 5.4) p<0.001), although not immediately post-test (Table 4.1).

Of the studies assessing general 'wellbeing' Lord (1994) showed improved mood across time (M=1.51, t= 6.71, p<0.01), Sole (2014) showed improved emotional wellbeing (median 12 (IQ 4) vs. 23 (IQ 4) Z=-2.18, p=0.03) following the activity. Hammar (2011) found increased levels of observed emotion for those taking part in the activity in comparison to those who did not (mean 2703.3 vs 2010.2 (CI -7.70-29.35) (test statistics not reported, p=0.042). The final study by Suzuki (1998) showed that there was no significant increase in positive affect, but negative affect did significantly decrease (no mean values reported, test statistic not reported, p=0.26).

4.2.7.3 Cognition, quality of life and socialisation

Two studies both carried out over 20 years ago assessed improvements in memory. Lord (1994) analysed the un-validated questionnaire and no analysis was provided for memory alone. Suzuki (1998) described a decrease in the number of 'negative' memories recalled (test statistic not reported, p=0.034), but no significant increase in the number of 'positive' memories. Clair (1996) showed increased alert responses across time for the singing condition

(total means not reported, *F*(2, 23)=4.4682, p=.022), however there was no significance in post-test scores for the singing or reading condition when compared. Carruth (1997) also reported increased incidences of face-name-recall, however these are not statistically compared and raw scores are not presented meaning these could not be calculated post-hoc (Table 4.1).

For the studies using the MMSE, Smith (1986) showed musical activity alone improved total MMSE scores (mean 10.9 (6.4) vs. 12.1 (5.8), *t*(11)= 1.83, p<0.05), but individual subscales showed no significant changes and musically cued reminiscence did not improve MMSE scores. Hong (2011) showed an overall improvement in MMSE scores over time (14.6 vs. 18.4 test statistic not reported, *p*<0.001) in those participating in the music activity and a significant decrease in the MMSE scores of those in the control group (mean 15 vs. 14.13, test statistic not reported, p<0.001). Cognition scores were described as increasing for the intervention group and decreasing for the control group for Hagen (2003) following the activity, although these had worsened slightly at 10-week follow-up (means shown graphically, *F*2, 144= 19.6, p<0.001). Scores also improved for the group taking part in occupational therapy with neither activity shown as more beneficial than the other (Table 4.1).

Bennett (1990) showed improved life satisfaction scores following the activity for those in the music group, compared with those in the verbal life review group (mean 11.42 vs 15.38, F (1, 12)= 26.674, p<0.001). The un-validated enjoyment questionnaire compared between the two groups showed significantly higher scores for the music group (15.38 (SD 2.9) vs. 10.77 (SD 2.08), *t* (112) = 3.21, p<0.01) post-test. Vanderark (1983) also showed significant mean gains for the intervention group for life satisfaction (41.88 (SD 8.42) vs 47.5 (SD 4.26), t= 3.74,

p<0.05), along with significant improvements in self-concept (8.39 (SD 1.96) vs. 8.69 (SD 1.99) t= 2.54, p<0.05). Three of the four studies which used validated quality of life measures showed no significant improvements over time (Burrack, 2004; Cooke, 2010a; Ridder, 2013). Sub-analysis of those attending 50% or more sessions in Cooke (2010a) did show improved scores over time (m=3.36 vs. 3.375 *F*(2, 46) = 4.471, p<0.05), suggesting a cumulative effect of the activity on quality of life. One of the two studies focused on social interaction showed a significant mean gain in social behaviours from pre- to post-test (m=1.6, t= 6.71, p<0.01) (Lord, 1993). The second provides only descriptive accounts as to how many residents were observed taking part in social behaviours (Ziv, 2007).

4.2.7.3 Improvements to care and physical health

Both Ragneskog 1996a and Ragneskog 1996b assessed quality of care through nutritional intake during meal times. Limited evidence was provided by the first study (Rangeskog, 1996a) which showed soothing music increased the time staff members spent with residents at meal times (just reported as *p*<0.05). The second (Ragneskog, 1996b) showed music of some genres to significantly reduce food intake (music from the 1920s and 1930s, reported only as *p*<0.05). More food was served to residents during the music conditions and these were also consumed more (r=0.85, p<0.001). Hammar (2011) showed significantly less time spent on behaviours indicative of pulling away (*p*=0.013) and grabbing objects (*p*=0.02) during personal care. Millard (1989) provided very limited evidence of assessing quality of care, but did show statistically significant improvements for sitting (F=3.13, df 4, 36, p<0.026) and walking with others (F= 3.13, df=4, 36, p<0.026) during the music activity, tentatively indicative of cooperation with staff.

Two studies assessed the impact on physical health. Koike (2012) reported significant changes in temperature (mean 36.46 vs. post 36.31, $p<0.001$) and pulse rate, (mean 72.28 (SD 9.88) and post 70.66 (SD 9.44), $p<0.001$).

4.2.8 Qualitative findings

There were similarities between findings identified from qualitative evidence (Table 4.1). Overlap was evident between two studies that involved group music activities (Martin, 2004; Chen, 2009). Both these studies identified the inclusivity of the music activity, Martin (2004) highlighted how the unstructured nature of the activity allowed those with cognitive impairment to participate and Chen (2009) described how the music offered a shift of focus from pain experienced by residents. Both also reported an improved sense of self occurring due to increased control over their activity and lives.

Similarities were also found in the results from studies focused upon music therapeutic caregiving (MCT) (Gotell, 2007; Hammar, 2010). Both described how the normal process of morning care was difficult for caregivers in terms of ethics and restraint (Hammar, 2010) and frustration with resistiveness to care (Gotell, 2007). Also discussed in both studies was improved wellbeing to both staff members and residents.

All four studies reported improved energy as an outcome, two between the group taking part (Martin, 2004; Chen, 2009) and two between residents and caregivers (Gotell, 2007; Hammar, 2010). Changes in emotional wellbeing were reported across all studies. The ability to create a shared experience was also universally found within the studies. Martin (2004) particularly commented upon how the nature of the activity promoted other staff members and relatives visiting the home to join in. Also found was the shared experience through the group, particularly with reminiscence promoted by the songs. Chen (2009) also commented on the

changing group dynamic to one full of energy and responsiveness. The two studies focused upon MTC both described how the interactions which took place during morning care became reciprocal and in turn improved understanding and cooperation (Gotell, 2007; Hammar, 2010). Findings comparing background music and MTC revealed that while background music enhanced the mood of residents and elicited humour and playfulness, the MTC encouraged meaningful interactions, which although less light-hearted were calmer than both background music and usual care (Gotell, 2007).

4.2.9 Quality appraisal of included studies

4.2.9.1 Levels of evidence

As discussed in Chapter 2 (Systematic review methodology and methods) all studies were assessed for quality. Based on levels of evidence criteria, 16 (33%) studies offered level 1b evidence, provided by randomised controlled trials (RCTs). The highest score of 1a would not be demonstrated within this study as it reflects findings obtained via systematic reviews which were not eligible for inclusion and therefore 1b refers to the second-highest quality evidence. Eight studies (17%) were level IIa, representing moderate quality evidence obtained from studies without randomisation. Twenty-one provided evidence from studies using other experimental methods (pre/post-test), of a slightly lower quality than level IIa and therefore assigned level IIb. Finally, the three descriptive studies provided evidence at the lowest level of evidence level III.

Quantitative studies were assessed via GRADE criteria which assigns a score of low, medium or high risk of bias for domains reflective of methodological quality (randomisation, allocation concealment, blinding of outcome assessment, incomplete data, selective reporting and other biases). Randomised controlled trials were assessed against all these criteria, with quasi-RCT and pre/post-test studies assessed against aside from randomisation and allocation

concealment. An outline of the process of scoring using GRADE criteria can be found in Chapter 2 (systematic review methodology and methods). Qualitative studies were assessed using a checklist (Cooke, Mills and Lavender, 2010, developed by Walsh & Downe, 2006) determining clarity of aims, appropriateness of study design, sample size adequacy, appropriate interpretation, consideration of ethics and transferability. Scores were assigned (created by Downe et al, 2009, based on work by Lincoln & Guba, 1985) with A indicating no/few flaws, B some flaws (unlikely to affect confidence in findings), C some flaws (which may affect confidence) and D (significant flaws likely to affect confidence in findings.

4.2.9.2 Randomised controlled trials

Only one study (Cooke 2010a) obtained a score of low risk for each of the domains (Figure 4.1). Seven studies reported the individual assessing the activity was not blinded to group allocation. Of these seven, five used an observational study design, meaning blinding would be difficult (Gerdner, 2002; Sung, 2006; Hicks-Moore, 2007; Nair, 2011; Ridder, 2013). However, there were two studies receiving a low risk of bias for this domain that used observational methods and blinded raters. Garland (2007) obscured the activity from view and muted them and Svansdottir (2006) obtained behavioural data from nurses blinded to the activity. Both Cooke (2010a, 2010b) and Garland, 2007 also blinded those scoring outcome measures as to group allocation. The remaining studies received unclear judgements as not enough information were recorded (Figure 4.1).

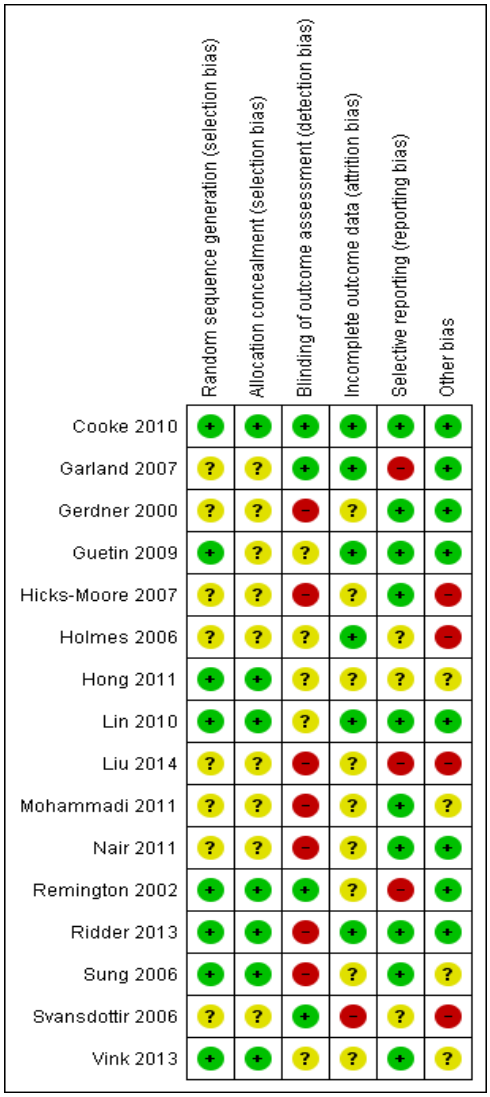


Figure 4.1 Judgements of methodological quality using GRADE criteria. Green corresponds to low risk, yellow to unclear and red to high.

Both Garland (2007) and Remington (2002) contained evidence of selective reporting, with the first study stating 'physically aggressive behaviours' were collected but did not occur enough to warrant inclusion and the second study presenting no mean values for outcome measures. The third study (Liu, 2014) failed to report on qualitative data outlined in the aims and methods section of the study. Only three studies did not contain enough information to receive a judgement of low risk for selective reporting (Holmes, 2006; Svansdottir, 2006; Hong, 2011).

Finally, other biases were recorded for four RCTs, the first featured a disproportionate control and intervention group (with only 9 in the control versus 32 in the intervention) (Hicks-Moore, 2007). Holmes (2006) reported use of only one observer 'to avoid problems associated with inter-rater reliability' and whilst Dementia Care Mapping is a valid measurement, there is no supporting evidence for the use of one category (in this case Category E) on its own. The third study (Svansdottir, 2006) reported no selection criteria for participants and failed to report whether inter-rater reliability was established for the two nurses scoring the outcome measures. The final study (Liu, 2014) reported no baseline characteristics of the included residents, nor how interviews were carried out. Worthy of noting, despite receiving unclear judgements, is two studies did not report how they modified the CMAI (Sung, 2006; Vink, 2013), however it is unclear whether this was a valid process or not meaning a judgement cannot be concluded (Figure 4.1).

4.9.2.3 Quality of Quasi-randomised controlled trials

A summary of judgements of quality for the included quasi-RCT studies can be shown in Figure 4.2. Over half the studies (n=5) explicitly reported that blinding did not occur and the remaining four did not provide enough information to make a judgement. There was also lack of accounting for withdrawals and drop-outs, with seven of the nine studies receiving an unclear judgement for this measure. Of the two studies, which did account for this, Casby (1994) received a judgement of low risk, only because the sample size was so small (n=3) meaning it was obvious that drop-outs had not occurred. Ledger (2007) received a high-risk judgement as despite calculating the sample size needed the number of drop outs was high (n=15 out of an initial 60), meaning their tests lacked statistical power. Despite absence of particularly detailed protocols, five studies reported on the outcomes using the measures they identified within the methods section and therefore received judgements of low risk.

Three did not contain enough information within the methods section to know whether selective reporting had occurred and Vanderark (1983) received a high-risk judgement as they failed to identify when residents were tested in the methods section.

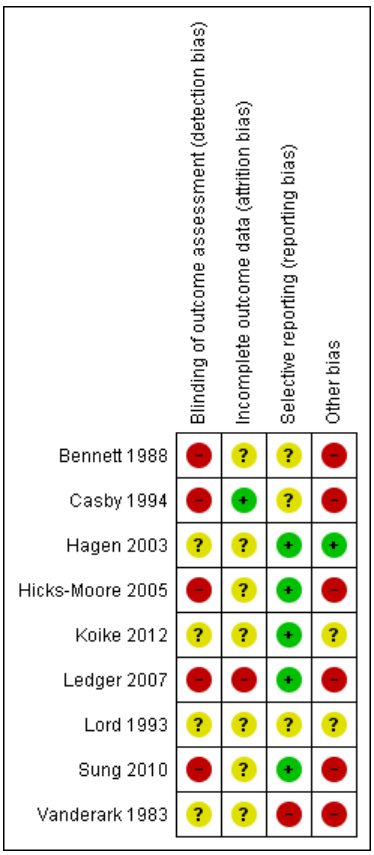


Figure 4.2 Judgements of methodological quality for Quasi-RCT studies. Green corresponds to low risk, yellow to unclear and red to high

In total, six studies contained other sources of bias. Two used questionnaires that were non-validated (Vanderark, 1983; Bennett, 1988) and Casby (1994) only had three residents within his study, with only one taking part in all three conditions meaning findings are subjective.

Ledger (2007) commented that there were discrepancies between findings from the CMAI and those within behavioural charts of residents, however these charts are not mentioned as an outcome measure. Similarly, there is no mention of what behaviours this referred to. Finally Sung (2010) commented the use of RAID within their study was not appropriate for

their study design but was used anyway as they deemed there to be no other method of measuring anxiety.

4.9.2.4 Pre/post-test studies

The twenty (46%) uncontrolled pre/post-test studies represented varying levels of quality (Figure 4.3). Only two studies (Ragneskog, 1996a; Ragneskog, 1996b) explicitly reported that blinding of outcome assessment had taken place. The first assessed different types of music which were not recorded within the videotapes produced for each session. The second implemented blinding (unknown music type) for the standardised outcome assessment measure (the GBS).

Nine studies reported that blinding did not occur, of which seven were observational studies analysed by the same member of the research team. Two were not observational studies but reported that the researcher who delivered the activity also administered the outcome measures pre- and post-test (Smith, 1986; Burrack, 2004). The remaining studies did not explicitly record either way whether blinding did or did not occur.

Evidence of incomplete data were shown for five studies. Two removed residents during the activity as they stopped showing signs of agitation and their data were subsequently not included in analysis (Cohen-Mansfield, 1997; Ziv, 2007). Two reported high levels of participants whose data were not included, Brontons (1996) reported a withdrawal of 27 out of 47 participants and Ragneskog (1996a) featured only five of an initial ten residents as five had their backs to the camera during the filming. Finally, Suzuki (1998) reported only the differences in mean scores with no reporting of scores or outcomes. Only Myskja (2006) received a judgement of low risk as this study appeared to account for all the participants

within the results section. The remaining studies did not account for drop outs or where their data were omitted.

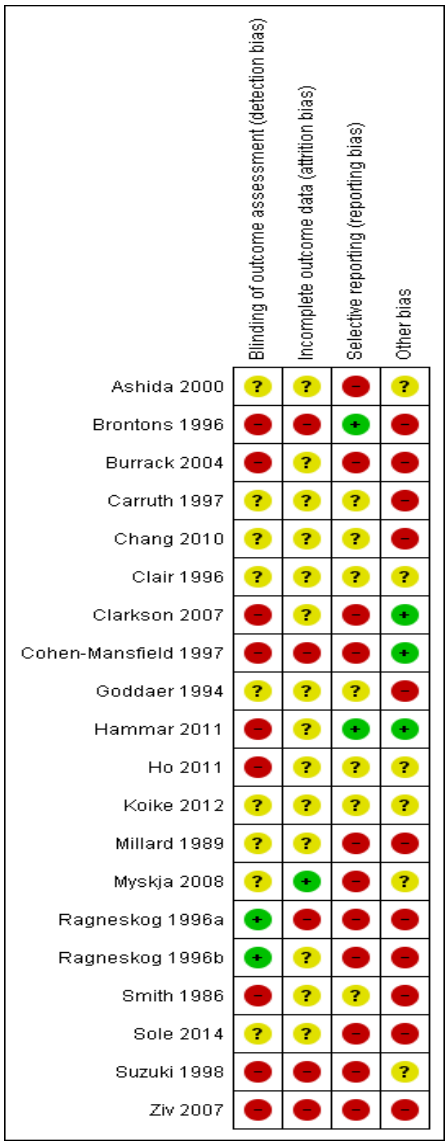


Figure 4.3 Judgements of methodological quality for pre/post-test studies. Green corresponds to low risk, yellow to unclear and red to high

Just two studies contained enough information to be sure that all outcomes were correctly reported within the results section (Brontons, 1996; Hammar, 2011). Over half (n=11) received judgements of high risk for this criterion. Ashida (2000) and Suzuki (1998) made claims within the discussion section that were not reflected within the results, with the first commenting on the relationship between interaction patterns and the second to gender

differences. Myskja (2006) reported results focused on participation levels, however this was not specified as an outcome measure and therefore it is not known how these were measured. Four did not include information within the results section which was specified as an outcome (Millard, 1989; Ragneskog, 1996b; Ziv, 2007; Clarkson, 2009). Ragenskog (1996b) and Burrack (2004) did not report how one of the outcomes (observed behaviours and satisfaction with living environment respectively) would be assessed, nor did they provide statistical values within the results section. Cohen-Mansfield (1997) also provided no statistical values for the CMAI measures obtained within their study, just a sentence to say they were non-significant. The final study (Sole 2014) reports the use of the GENCAT for assessing quality of life, however then report within the results section that several scales were omitted as they 'were not influenced by music therapy'.

Finally, 11 (55% of this study type) studies had risk of other biases. Two of these were related to the outcome measures used, Burrack (2004) used an adapted version of the QoL-AD scale, but it is unknown whether this had reliability and/or validity established and Smith (1986) used subscales from the MMSQ which at the time did not have reliability or validity tested either. Another study (Ziv, 2007) reported inter-rater reliability (IRR) to be higher for positive behaviours during the pilot (between 90-100%) than negative behaviours (80%) but states that individual researchers still observed different residents without any IRR established for the main study. Carruth (1997) reported difficulties recruiting which therefore meant the sample was too small for statistical analysis, with similar issues occurring in Ragneskog (1996a) due to half the residents not being able to be observed due to logistical problems. Discrepancies were found within Millard (1989) who stated ten residents were recruited due to restrictions within the dining room, however then reported

that between 10 and 17 residents took part in the groups. Further to this they also reported within the discussion section that an initial 16 residents were recruited. The remaining three studies all contained potential biases in their methods. Brontons (1996) used nurses to score their outcome measure, however it was not reported whether they were trained in the scales use and Goddaer (1994) used only one researcher to observed 28 residents taking part in the activity. Finally, Chang (2010) states the CMAI was scored by nurses between the hours of 09.00 and 21.00 but that notes were used between 21.00 and 09.00, scoring would likely have been different for these time periods, but data were not presented separately and instead were pooled together.

4.1.9.4 Qualitative studies

None of the qualitative studies received a judgement of grade A, indicating there were no flaws. Two qualitative studies (Gotell, 2007; Hammar, 2010) received a judgement of grade B, meaning they contained some flaws, but these were not likely to impact upon the credibility of the studies. Both did not contain information on selection criteria of participants, these were selected based on recommendations from the head nurse/manager. There was also no evidence of sample size justification for either. Both received ethical approval from the relevant committee, however Hammar (2010) did not provide information based around consent of the caregivers involved with the study. Gotell (2007) failed to provide evidence of reflexivity, although given data were obtained from observations which had been videotaped, the researcher would likely not impact upon these interactions. There was also no evidence of weaknesses and limitations in this study.

Two further studies received a judgement of grade C, meaning they contained flaws which may have impacted upon the credibility of the study. Both studies failed to provide evidence of reflexivity and neither linked the analysis and data with emerging theories or future

directions. They also both failed to recognise the limitations and/or weaknesses of the study. Martin (2004) also did not provide enough background data to contextualise the study, nor show evidence of ethical considerations, there was no reporting of ethical approval, or ethical issues which may have arisen during the study. Chen (2009) did not provide a justification of the sample size, nor evidence of spending time immersed in data during the analysis stage.

4.2.10 Summary of music studies

Music was the most evaluated activity from the included studies within the systematic review (chapter 3, overall synthesis). In total 48 studies evaluated this activity, with the majority of these providing quantitative evidence (n= 44, 92%). A range of study designs were used to assess music, the most popular study design was uncontrolled pre/post-test (n=20, 42%), followed by RCTs (n=16, 33%), quasi-RCTs (8, 17%) and finally descriptive (n=3) (see Table 4.1).

The type of music activity varied across studies and included auditory exposure, music during personal care, structured music therapy programmes and involved both active and passive participation. Most studies evaluated group music activities (n=35, 73%) and there was large variation in the length of programme, from a one-off evaluation to a structured music therapy programme lasting for 28 weeks (see section 4.2.5 (characteristics of the arts activity) and Table 4.1).

The most evaluated outcome measure was changes in behavioural symptoms (n=26, 45%). There were disagreements across studies as to the efficacy of music for behavioural symptoms depending on the type of behaviour assessed. Reduction in agitation levels were observed (see Section 4.2.7.1 and Table 4.1), whilst verbally disruptive behaviours showed less evidence of benefit with some studies showing worsening of this behaviour. Music was

not shown to be more beneficial for behavioural symptoms than other creative and social activities, although it was more effective than usual care comparisons.

Psychological wellbeing was the second most evaluated outcome measure, with specific focus on depression and anxiety. Some effects were observed with respect of depression levels, particularly for those with pre-existing clinical depression. Only one study (of four) focused on anxiety showed a significant impact of the activity post-test. However, studies evaluating both depression and anxiety showed the activity group to have lower scores post-test than the control group, particularly when no other activity was offered which may indicate a reduction in decline, rather than significant improvements.

There was also evidence of impact for measures of cognition, including memory and MMSE scores, the caring process and physical health. Less evidence was shown for improved quality of life when assessed through quantitative measures.

Strength of evidence lay in the number of randomised controlled trials, although judgements of quality were variable for all included study designs (see Section 4.2.9). Moderate quality was displayed within qualitative evidence, with two studies scoring a grade B (few flaws, unlikely to affect credibility) and two receiving grade C (some flaws, may effect credibility). The following section (4.3) provides a narrative descriptive synthesis of studies featuring 'multisensory activities'.

4.3 Narrative descriptive synthesis of studies featuring multisensory activities.

This section provides a narrative descriptive synthesis of studies which featured the activities classified as 'multisensory activities'. Studies were placed within this category when they consisted of arts activities which provided sensory stimulation to residents. This was mainly achieved by changing the aesthetic of the care homes or allowing the residents to experience an environment outside of the care home itself. Studies were also placed within this category where they offered sensory experiences to residents through use of object handling or snozelen rooms (Table 4.3). As with the music subsection (Section 4.2) a descriptive narrative synthesis of the included studies is presented, along with appraisal of methodological quality.

4.3.1 Characteristics of included studies

A total of eight studies (11% of the 71 included within the whole review) were placed within this category (Table 4.3). Published dates spanned 15 years ranging from 2000-2015 (Orsulic-Jeras, 2000; Kincaid, 2003; Cox, 2004; Simard, 2010; Klages, 2011; Chang, 2013; Scott, 2014; Roe, 2015). Only two studies reported the dates during which the study took place, these were October 2006-November 2007 (Chang, 2013) and June-December 2012 (Roe, 2015). A further study reported duration of the study period was nine months, without specifying dates (Orsulic-Jeras, 2000). Seven studies were interventions and one descriptive (Roe, 2015) (Table 4.3). Six intervention studies were quantitative and one a mixed-methods design (Cox, 2004), the descriptive study was qualitative (Roe, 2015). Of the intervention studies, one used an RCT study design (Klages, 2011), one a cluster-RCT (Scott, 2014), two a quasi-RCT design (Orsulic-Jeras, 2000; Cox, 2004), the remaining studies adopted a pre/post-test design (Kincaid, 20 03; Simard, 2010; Chang, 2013) (Table 4.3).

Table 4.3 Characteristics of included studies featuring the arts type multisensory activities

Study ID and type	Location	Participants	Activity Type	Aims	Outcome measures	Findings	Quality score
065 Simard (2010) Intervention Dates not specified Quantitative Uncontrolled pre/post-test	USA EPOCH Senior Living Healthcare Centres	86 in total, 69 (80%) were female and 17 (20%) were male. Mean age was 84.5 (± 7.1), with a range of 68-103 years.	Namaste Care Program The program takes place 7 days a week for around 5 hours a day and is staffed by nursing assistants called Namaste carers.	No aims explicitly stated, however the paper discusses a specially created Namaste Care Program and evaluates its implementation with reference to challenging behaviour, agitation, delirium and anti-psychotic medication use.	Minimum Data Set (MDS) forms were completed prior to the resident enrolling in the programme and then more than 30 days after their enrolment in the programme. MDS forms did not contain any personal information on the residents.	The number of days between first evaluation and enrolment in the Namaste care was on average 44.7 (SD 35.0) and the number of days between enrolment and the second evaluation was 83.6 on average (SD 48.9). No difference in depression scores before and after enrolment (p values and/or significance not reported). Only 8 residents had a score larger than 2 (considered depressed) and 3 of these were not treated for depression. There were also no significant differences between behavioural symptoms before and after enrolment in the whole study population. When only residents who had withdrawal or reduced social interaction were included in the analysis, the Interest subscale of the MDS Challenging Behaviour Profile was significantly decreased following enrolment (3.27, SD 0.30 vs. 2.00, SD 0.47, n= 11, p=0.046) indicating less impairment in social interaction. Scores from the Agitation subscale of the MDS Challenging Behaviour Profile for residents with a CPS score between 1 and 3 decreased following Namaste enrolment, however this was not significant (1.04, SD 1.25 vs. 0.81, SD 1.27, n= 26, p=0.18). Lower levels of agitation were shown for residents with CPS scores between 4 and 6 and these remained unchanged (actually they increased slightly) (0.40, SD 0.69 vs. 0.50, SD 0.89, p values not reported). Total indicators of delirium also decreased non-significantly following enrolment (2.52 SD 1.94 vs. 2.35 SD 1.88, p=.079. When the individual items of this subscale were compared it was found that ratings for 'periods of restlessness' and 'periods of lethargy' were unchanged. When these 2 items were deleted there was a significant decrease of ratings of the remaining delirium indicators following enrolment (2.00 SD 1.53 vs 1.81 SD 1.51, p=.02). Comparison of psychoactive medication administration days did not find any difference in the number of anti-psychotics, antidepressants or hypnotics. However, there was a significant decrease in the days when anti-anxiety medications were administered (0.80 SD 2.18 vs. 0.49 SD 1.79, p=0.35). The differences were caused by 11 residents who were receiving anti-anxiety mediations prior to the Namaste enrolment and only 6 who were receiving such medications after enrolment. The number of residents receiving antidepressants reduced by two (54 vs. 52) and there was a decrease in the number of days when hypnotic medication was administered to one of the residents (4 vs. 1).	GRADE score ? ? ? – Levels of evidence score LII-b, grade B.
076 Chang (2013) Intervention November 2006-October 2007. Quantitative Uncontrolled pre/post-test	Taiwan Nursing Home of St Martin de Porres Hospital in Chiayi.	33 completed the intervention. Ages not reported. 27 (81.8%) were female and 6 (18.2%) were male. Loss to follow-up: 15 participants were lost to follow up, 10 had died, 2 refused to complete at follow up and 3 had been discharged from the nursing home.	Environmental stimulation. Wall murals were created on each floor using acrylic paint. 4 themes were chosen to reflect the life experiences of the residents living there. .	The focus of the study was to enhance the environment of a nursing home by using visual arts aimed at promoting reminiscence. To determine reliability and validity for a scale to assess residents' satisfaction with their living environment (Satisfaction with Living Environment Nursing Home Scale- SLE-NHS) and use this to determine effects of the intervention on satisfaction with the living home environment.	The Satisfaction with Living Environment Nursing Home Scale (SLE-NHS) was developed and validated for use in this study. This 18-item questionnaire measures responses on a Likert-type scale and has three factors: memory recall, pretty and pleasurable and convenience. Scores were obtained at baseline, prior to the installation of the mural and following installation.	T-tests were used to compare pre and post scores on the SLE-NHS subscales. The recalling old memories scale showed significant improvement (mean 1.41 pre vs. -3.47 post, t=13.32, p<0.001). All of the floors showed a significant improvement for this subscale (floor 2, t=5.99, p<0.05; floor 3, t=6.97, p<0.01, floor 5, t=7.29, p<0.001, floor 6, t=10.88, p<0.001). A slight (but non-significant) improvement (3.47 pre and 3.60 post) was observed for the pretty and pleasurable subscale. A significant increase was observed for those participants on the second floor at pre-test (t=13.86, p<0.05). Convenience scores were not significantly different (p>0.05).	GRADE score ? - - - - Levels of evidence score LII-b grade B.

Study ID and type	Location	Participants	Activity Type	Aims	Outcome measures	Findings	Quality score
118. Kincaid (2003) Intervention Dates not specified. Quantitative Uncontrolled pre/post-test	USA A special care unit located within Courtland Terrace Nursing Home, North Carolina.	12 in total (who exhibited exit seeking behaviours from 17 who consented to the study). Of these ten were female (83.3%) and 2 were male (16.7%). Mean age of the participants was 78. Loss to follow-up: not reported.	A mural produced by a local arts class which covered from the floor to the ceiling and the whole door and walls. The home administrator had requested that the colours and themes be in keeping with the existing décor of the home. Residents had an input in the final design of the mural.	To explore the effects of a wall mural on exit-seeking behaviours exhibited by individuals with dementia who were living in a nursing home.	Data were collected in 2 hour blocks for six weeks prior to the installation of a wall mural and then at the same frequency for six weeks following installation of the wall mural. Four different door testing behaviours were observed.	The overall mean score for door testing behaviours prior to the intervention was 55.67 (*df*=11; SD=67.57) and following the intervention was13.42 (*df*=11; SD= 28.41). This represents a significant mean decrease of 42.25 (*t*=2.6; *p*=0.24) All 12 residents showed exit-seeking behaviours prior to the intervention and only 3 remained active following. Type 1 behaviours pre-installation mean average was 35.67 and following 6.17, a mean significant decrease of 29.5 behaviours (t=2.622, *p*=0.024) Type 2 behaviours pre-installation mean of .67 and .17 follow. No significant difference (p=0.5). Type 3 behaviours pre-installation mean average was 3.42 and .75 afterwards, a significant mean decrease of 2.67 (t=2.432, p=0.03). Type 4 behaviours pre-installation mean average was 15.92 and 6.33 following, a mean decrease, that was not significant (p>0.05).	GRADE score ? - - Levels of evidence score LII-b.
125 Scott (2014) Intervention Dates not specified Quantitative Cluster-RCT	Australia 'Aged care facilities' within the South East Queensland area.	Biophilia group (n= 15). Mean age 82.2 years (± 8.9). 10 female (66.7%) 5 male (33.3%). 10 staff members also took part. Reminiscence group (n=10). Mean age 83.3 years (± 8.6). 6 female (60%) 4 (40%) male. 6 staff members also took part. Control group (n=8). Mean age 83.3 (± 5.7). 6 female (75%) 2 male (25.5%).	Sensory enrichment. Comparison of two different wall murals, one inspired by nature (biophilia) and a second inspired by reminiscence. A control group received usual care.	To implement a 'biophilia' environmental stimulation involving placing plants and natural elements within care homes. To assess the benefits of the intervention and compare these with a similar installation which comprised of non-natural factors (a reminiscence intervention). To compare the benefits of both interventions with a control condition which involved no stimulation.	Unvalidated questionnaire for satisfaction with living environment. The Geriatric Depression Scale (GDS) and Geriatric Anxiety Inventory (GAI) collected at Time 1, Time 2 and Time 3. A question was used to assess environmental satisfaction and two items were used to assess satisfaction with opportunities to keep occupied. Wellbeing and engagement were also measured using Likert-type scale of six questions and a staff questionnaire were administered to staff.	No significant differences were shown for baseline measures of depression, anxiety or QoL-AD scores. There were also no significant differences for scores on the satisfaction with environment or satisfaction for opportunities scales. Both those from the biolphilia and reminiscence conditions stated that they were happy with their environments. 90% (n=9) of the staff reported they liked the biophilia installation and 7 (87.5%) from the reminiscence installation. Mixed-MANOVA showed a significant interaction effect where responses differed as a function of group condition and time (T1, T2 and T3) (F=(2, 22)= 2.85, p<0.05, n^2 p= 0.21). Post-Hoc analysis showed there to be a significant difference between the groups at time 2 for social engagement for the biophiilia group (m=2.17) and reminiscence group (m=2.42) compared with the control group (m=2.27). A similar was shown at T3 where social engagement was significantly greater for the biolhilia group (m=2.27) and the reminiscence group (m=2.27) compared with the control group (m=3.20). There was no difference across timepoints for levels of social engagement in the control group (p>.05). Levels of social engagement were significantly different across time for the biophilia condition (p<0.001) and the reminiscence condition (p<.05). There were no significant changes for satisfaction or mood measures (all p>.05), however improvements were observed at T2 and T3 for the biophilia condition. Staff social engagement measures also showed no increase (p>.05). 60% of staff reported they felt the biophilia condition improved memories in participants and all staff in the reminiscence condition reported improvements.	GRADE score + ? ? ? + + Levels of evidence score LI, grade A.
140 Orsulic-Jeras (2000) Intervention 9-month time period, dates not specified. Quantitative	USA Menorah Park Centre for Senior Living, an orthodox Jewish facility. Participants were	25 took part in the whole study (from 44 recruited). 23 females (92%) and two males (8%). 13 residents were assigned to the treatment condition and 12 to the control, matched on their	Montessori group activities. Twice a week involved group, individual and QAR sessions. Group programming sessions lasted 25-45 minutes;	To explore the role of Montessori-based activities in improving engagement levels and reducing the frequency of problem behaviours in residents of a	Outcome measures were carried out at pre-test and then at follow-up nine months later at the end of the programme. The Multidimensional Observation Scale for Elderly Subjects	MMSE and MOSES scores both showed significant effects over time for those taking part in the intervention (F[1, 23]=5.3, p<.03) and (F [1, 23]= 6.0, p<.02) respectively. Initial MMSE scores showed participants to be in the moderate to advanced stages of dementia and MOSES showed moderate ability levels, both showed a steady decline across the study for the sample. No other effects reached significance for any other measure. Pre-test scores for both agitation and depression were low and these remained low for the study. Engagement*:* Non-engagement and self-engagement were rare and therefore not statistically analysed. *Constructive engagement:* mixed 2x3 ANOVA where between group factor was 'group' and the within subjects factor was 'time' (baseline vs. post-test 1, vs. post-test 2). Significant effects were found for group (F (1, 23)= 133, p<0.001) and time (F (2, 46)= 20.9, p<0.001). Group-time interaction was also significant (F[2, 46]= 23.4, p<.001). One-tailed independent t-tests were found groups did not significantly differ at baseline (p>.05) but they did differ significantly at post-test 1 (p<.001) and post-test 2 (p<.001*). Passive engagement* significant main effects were	

Study ID and type	Location	Participants	Activity Type	Aims	Outcome measures	Findings	Quality score
Quasi-RCT	recruited from the special care unit (SCU).	MMSE scores and performance on other cognitive measures.	individual activities 10-30 minutes and QAR (structured reading and discussion) lasted 30-60 minutes. Those in the control group took part in other activities including hand massage, aromatherapy and tai chi. Not reported who carried out the sessions.	special care unit located within a senior living facility.	(MOSES) was used to assess functional ability, the CMAI was used for agitation and the Cornell Scale for Depression (CSD) for depressive symptoms. MMSE scores were also obtained at pre- and post-test.	found for group (F [1, 23]= 29.1, p<.001) and time (F [2, 46]=3.6, p<.04. group-time interaction was not significant (F [2, 46]= 1.4, p<.27). One-tailed independent t-test carried out found the two groups were significantly different from each other at baseline (p<.02) and at post-test 1 (p<.001) and post-test 2 (p<.001). For the control group there was no significant change in passive engagement levels across time (F [1, 12], p<.1) whereas the treatment group did show a significant change across time (F [1, 12]= 8.1, p<.002). A 2 x 2 repeated measures ANOVA was used to analyse passive and engagement measures where the within-subjects factors were "type of programming" (treatment vs. control) and "time" (post-test 1 vs. post-test 2). . Significant types of programming effects were found for both constructive engagement (F [1, 12]= 126, p<.001) and passive engagement (F [1, 12]= 18.8, p<.001). Affect: anger, sadness and anxiety were observed infrequently and therefore not statistically analysed. *Pleasure:* significant main effects were found for group (F[1, 23]= 26.7, p<.001) and time (F [2, 46],= 7.5, p<.002). Group-time interaction was significant (F [2, 46]= 5.7, p<.006). One-tailed independent t-tests showed the two groups were not significantly different to each other at baseline (p<.05) but were significantly different at both post-test 1 (p<.001) and post-test 2 (p<.001). *Passive engagement:* significant main effects were found for group (F [1, 23]= 29.1, p<.001) and time (F [2, 46]=3.6, p<.04. group-time interaction was not significant (F [2, 46]= 1.4, p<.27). one-tailed independent t-test found the two groups were significantly different from each other at baseline (p<.02) and at post-test 1 (p<.001) and post-test 2 (p<.001). For the control group there was no significant change in passive engagement levels across time (F [1, 12], p<.1) whereas the treatment group did show a significant change across time (F [1, 12]= 8.1, p<.002).	
162 Klages (2011) Intervention Dates not specified. Quantitative RCT	Canada An urban not-for-profit residential home located in Ontario Canada, with over 390 residents.	9 in the intervention group and 10 in the control group (n=24). Intervention group: mean age 84 years (SD 6.6). 7 (78%) female and 2 (22%) male. Control group: mean age was 89 years (SD 3.2). 6 females (60%) and 4 males (40%).	Snozelean Relaxed Snozelen sessions to place twice a week for a total of six weeks, with a two day period separating weeks. Preferences of the residents were taken into account and activities included those to stimulate tactile, visual and proprioceptive sensations. Sessions were delivered by the lead author.	To assess the role of Snozelean rooms on the incidence of falls and improvements in balance in older people with dementia living within a care home.	Detailed journals were completed by the facilitators and these recorded balance-enhancing effects such as body posture changes. Positive and negative emotional reactions were also recorded. Four tests were administered to record pre and post intervention balance- the Functional Reach Test, Sharpened Romberg and the Timed Up and Go test. Patient notes were also examined for incidences of falls across the study period.	A spilt MANOVA was carried out to explore changes pre and post intervention between the groups. Within group changes were also analysed using a split-plot analysis. T-tests analysed the changes in falls frequency before, during and after the intervention and Bonferroni corrections were applied to secondary analyses to correct for multiple comparisons. The split=plot MANOVA showed no significant effects of the Snozelen room on balance of participants. Multivariate effect of time was negative from pre to post test (F (4,14)= 1.13, p=0.38) and group was also negative (F (4, 14)= 0.92, p=0.48). Both groups showed small balance improvements over time, however not significantly so. Secondary analysis (incidence of falls) recorded 16 falls in the intervention group compared to 44 in the control group over an 18 week period, however an outlier in the control group was responsible for the majority of these falls (21 out of 44) and these were removed. The intervention group had 5 prior, 7 during and 4 at follow-up and the control had 8 prior, 8 during and 7 at follow up, these changes were not significant at pre, during or post (p=0.29, p=0.47, p=0.47 respectively). The different aspects of the snozelen room and their effects on balance are also discussed.	GRADE score + + - - + - Levels of evidence score LI, grade A.
200 Roe (2015) Descriptive June-December 2012 Qualitative Service Evaluation	UK Manchester Museum and Whitworth Gallery Manchester	8 participants from a supported living facility and nine from a care home (n=17 total). 4 members of staff from the museum and gallery and a supported living project worker. Comments were also provided	Museum and gallery intervention. Programme schedule consisted of monthly sessions. .	To evaluate the feasibility of an arts for health program carried out within museums and galleries for older people who reside within a care home. Particularly to identify the benefits of the program in	Non-participant observations were carried out along with field notes recorded by the research team. Independent content analysis was carried out for each session using the Spradley's framework by two research team members.	*Session and participants:* discusses the structure of the sessions, five of which were carried out as planned and another session which had to be rearranged. *Impact of arts for health programme and sessions:* acts and actors which discussed the structure of the sessions (acts) and those taking part (actors); space and time which evaluates the logistical aspects of carrying out the programme; goals, events, object and activity which presents how engaging with creative activity was beneficial for those taking part; feelings which discusses the positive feelings the participants felt when taking part in the study. *Benefits, impact on wellbeing, feasibility of the sessions and programme:* this category outlines the need for museum staff to be flexible in their approach to the programme. *Feasibility of the programme, sessions and the future:* the gallery and museum staff members expressed their interest in taking part in a similar programme in the future and how they felt this would be achieved.	QA score grade B Levels of evidence score, LIII,. Grade C

Study ID and type	Location	Participants	Activity Type	Aims	Outcome measures	Findings	Quality score
		from an artist and activities coordinator from the care home.		relation to the wellbeing of those taking part.	A final group staff interview was conducted with an interview schedule of 10 questions.		
202 Cox (2004) Intervention Dates not specified Mixed-Methods Quasi-RCT	Australia Rice Village Residential Facility.	24 residents participated, 23 were female (96%) and one male (4%). Ages not reported. Six caregivers and six visitors consented to be interviewed for the qualitative study element.	Each participant took part in each of the activities, the living room (usual care), Snozelen and Garden installations for 16 minute sessions. A total of nine 16-minute observation periods were carried out for each participant (3 of each condition?).	To explore the impact of two types of multisensory environments: Snozelen rooms and a landscape garden on improving the wellbeing of older people and compare these interventions with the usual living room environment within the residential home.	The Affect Rating Scale (ARS) was used to carry out observations with the oparticioants. This contains six different and discrete categories of affect: pleasure, anger, anxiety or fear, sadness, interest and contentment. Immediately prior to entering the installation and at 4 minute intervals during the sessions an observed used to ARS to record participant responses	Wilcoxon signed rank tests were used to examine individual differences prior to and during the intervention. When looking at the interventions individually, we can see that pleasure was increased during all three of the interventions (living room mean 1 before, 44 following, p <0.001, garden mean 0 before, 47 following, p<0.001 and snozelen 0 before, 38 following, p<0.001). interest scores increased only for the living room, mean beforehand 12, following, 31, p<0.01. Contentment significantly reduced for each condition (mean before living room 67, following was 18, p<0.001; mean garden prior 65 and following, 24, p<0.001 and snozelen before 56 and following, 26, p<0.01). Anxiety or fear decreased only for the living room condition (mean prior 14 and following, 4, p<0.05). Qualitative- states nine general themes, however only four are reported for each group. Caregiver group, 4 themes were identified: first impressions, something special, changing patterns, what gets in the way. Relative/visitor group identified 4 themes: changing the patterns, feeling sad about the medication, privacy and meditation, fascination.	

4.3.2 Aims of included studies

Two studies aimed to change the aesthetic of the care home and assessed satisfaction with living environment (Chang, 2013; Scott, 2014) (Table 4.3). The aim of reducing behaviours commonly associated with dementia were reported by two studies, one generically referring to 'problem behaviours' (Orsulic-Jeras, 2000) and the other specifying 'exit-seeking behaviours' (Kincaid, 2003). Two studies specifically referenced an aim to determine the effects of the activity on psychological functioning, namely mood (Cox, 2004; Scott, 2014). Just one study focused on physical improvements, assessing the impact of the activity on falls rates (Klages, 2011). In addition to exploring behaviour, Orsulic-Jeras (2000) also aimed to assess the impact of the activity on engagement levels. Scott (2014) also looked at social interactions between residents and between staff members and residents.

The only descriptive study evaluating this arts type explored feasibility of implementing a museum and gallery-based programme with older people from care homes (Roe, 2015), alongside an account of improvements to wellbeing. No specific aims were reported for one study (Simard, 2010), however background literature related to challenging behaviour, agitation, delirium and anti-psychotic medication, which were indicated to be assessed given the outcome measures used.

4.3.3 Location and setting

Three studies took place in the USA (Orsulic-Jeras, 2000; Kincaid, 2003; Simard, 2010), two in Australia (Cox, 2004; Scott, 2014) and one study each in Taiwan (Chang, 2013), Canada (Klages, 2011) and the UK (Roe, 2015).

The care home was referred to as a 'nursing home' for two studies (Kincaid, 2003; Chang, 2013), two specified residents from a 'residential facility' (Cox, 2004; Klages, 2011) and two included participants from 'senior living' residences (Orsulic-Jeras; 2000; Simard, 2010). The remaining studies stated residents were from 'aged care' facilities (Scott, 2014) and a 'care home and supported living' facility (Roe, 2015). Two studies specified further that residents resided in a special care dementia unit within the homes (Orsulic-Jeras, 2000; Kincaid, 2003).

The location where the arts activity was carried out was consistently reported (Table 4.3). The seven intervention studies stated the activity was carried out within the care home where participants resided (Orsulic-Jeras, 2000; Kincaid, 2003; Cox, 2004; Simard, 2010; Klages, 2011; Chang, 2013; Scott, 2014) and the descriptive study took place within a museum and art gallery (Roe, 2015). This study was the only one the whole of this review (of any arts type) to explicitly describe that the activity was carried out in an environment other than a care home or other care facility (day units, hospital wards etc.).

4.3.4 Populations

The total population for included studies was 249 older adults, with sample sizes ranging from 17 to 86 (mean 31, SD 25). This corresponds to a total of 219 participants who took part in the arts activity, with control groups of eight (Scott, 2014), 10 (Klages 2011) and 12 (Orsulic-Jeras, 2000) (total n= 30) reported for the three studies incorporating a control activity (in the remaining quasi-RCT (Cox, 2004) participants took part in each condition). A total of 28 caregivers (22 staff members and 6 visitors) also provided data for two studies (Cox, 2004; Scott, 2014).

Five studies (Orsulic-Jeras, 2000; Kincaid, 2003; Simard, 2010; Klages, 2011; Scott, 2014) reported the ages of participants with a total mean of 83.5 years (SD 3.2). The age range of

participants was also reported in Roe (2015) as 72-92 years for those attending sessions (72-92 years from the care home residents and 75-88 years for the supported living residents).

The gender of included participants was also reported within all studies, the majority being women (n=199, 80%, men n=50, 20%). This figure was reflected within the gender divide in each individual study (Table 4.3). All studies featured more women than men, this was proportionate for the control and intervention groups for studies that included control groups.

The requirement for residents to have a dementia diagnosis was reported in Kincaid (2003) and Cox (2004). Participants in Orsulic-Jeras (2000) were recruited from a special care dementia unit (SCU) and therefore the majority would presumably have a dementia diagnosis and Simard (2010) reported the 'majority' of residents to have a diagnosis of dementia although others had different medical diagnoses. A further study by Klages (2011) did not specify the need for participants to have dementia but did require 'cognitive deficits' as assessed at baseline via an MMSE cut-off score of 25. Conversely, Chang (2013) assessed participants at baseline using the Short Portable Mental State Questionnaire (SPMSQ) with residents needing to score eight or above to qualify for inclusion. The remaining two studies did not indicate inclusion criteria requiring a dementia diagnosis (or other cognitive impairment), nor did they contain demographic information related to proportions of dementia in their samples (Scott, 2014; Roe, 2015).

4.3.5 Characteristics of the arts activities

Three studies changed the aesthetic of an area within the care home (Kincaid, 2003; Chang, 2013; Scott, 2014); how this was implemented varied between the studies. In the first (Kincaid, 2003) a wall mural was created, which covered the walls from floor to ceiling,

including the door to the home. This was developed by a local artist and input was provided from residents as to the final design, although it was also specified that staff had requested the colour scheme be in keeping with the overall décor of the home. Chang (2013) also created wall murals across different floors of a care home, designed to reflect the earlier lives of the residents who lived on the floor to invoke reminiscence. The final study by Scott (2014) also featured a reminiscence condition which involved not only a wall mural but also included multisensory materials in the form of props including Tables, chairs and books and magazines along with an aroma diffuser (cinnamon scent) and radio which played songs from music from between 1920 and 1950. This was compared with a 'biophilia condition', to differentiate between the effects of reminiscence associated with the activity or the actual sensory enhancement itself. This condition also featured a wall mural and props including a park bench, plants and garden ornaments, an aroma diffuser and sounds of nature (birdsong audio). In the second condition, there was also the opportunity for residents to take part in scrapbooking nature images Table 4.3).

The use of snozelen rooms was evaluated by both Cox (2004) and Klages (2011). As in Scott (2014) the first of these compared the snozelen to a gardening activity (Cox, 2004) although this featured use of a real garden as opposed to a simulated indoor one. Both studies described the snozelen activities as providing visual stimulation (via bubble tubes, colour wheels and projectors), tactile (interacting with different shapes and textures of props in the room), auditory (via exposure to music) and proprioceptive (throwing and kicking balls, swinging in hammocks) often with several senses being stimulated at once.

Two studies evaluated structured programmes, a Namaste programme (Simard, 2010) and a Montessori programme (Orsulic-Jeras, 2000). Although not a snozelen room, the Namaste

programme described in Simard (2010), featured similar multisensory components. The programme used music, aroma diffusers, nature DVDs, touching of stuffed animals and props related to the seasons such as leaves, flowers and snow. In relation to the reminiscence element featured in two of the wall murals (Chang, 2013; Scott, 2014), use of cosmetics which would promote reminiscence (ponds cold cream and old spice) were used during activities of daily living carried out during the sessions. The Montessori programme featured both individual activities such as interacting with everyday objects and group activities which included memory bingo and a reading and discussion activity (Orsulic-Jeras, 2000).

The final study (Roe, 2015) took place within a museum and gallery and offered participants different sensory experiences. These included touring the gallery and museum to view the exhibits (visual), object handling and interacting with live animals (tactile), discussions about the exhibits with staff members and peers, and art-making (proprioceptive) (Table 4.3).

Length of programmes ranged from six weeks to nine months and was reported by four studies (Orsulic-Jeras, 2000; Klages, 2011; Scott, 2014; Roe, 2015). Both studies that created wall murals did not specify the length of time these remained within the home and whether they became a permanent feature or not (Kincaid, 2011; Chang, 2013), the length of time required to construct the murals was two weeks and one year respectively. Simard (2010) evaluated a programme which was established within the residential home, choosing residents who had yet to commence the programme and then assessing again 30 days later. Four studies reported frequency of delivery which ranged from daily (Simard, 2010) to twice weekly (Orsulic-Jeras, 2000; Klages, 2011) and monthly (Roe, 2015). Within the studies that enhanced the surroundings of the care home, it is presumed that access to this was available each day, however this is unclear given these changes were made to communal areas which

may not always be accessed by residents (Kincaid, 2011; Chang, 2013; Scott, 2014). Four of the five studies which delivered specific sessions reported length of these which ranged from 10 minutes (for some individual sessions, reported to range between 10-30 minutes) and five hours (Orsulic-Jeras, 2000; Cox, 2004; Simard, 2010; Roe, 2015). Again, given the wall murals and installations were available constantly it is not clear how often they were accessed by the residents, or for what length of time they engaged (Table 4.3).

Four studies featured control conditions (Orsulic-Jeras, 2000; Cox, 2004; Klages, 2011; Scott, 2014). Of these, two compared the arts activity with usual care (Cox, 2004; Scott, 2014), however Cox (2004) adopted a cross-over design and therefore all the participants took part in both the usual care and activity conditions. Scott (2014) assigned conditions at group level, therefore all those in the control home would have had no changes made to their usual routine. Additional activities were offered for participants in the control conditions of the remaining two studies (Orsulic-Jeras, 2000; Klages, 2011). These activities included hand massage and aromatherapy (Orsulic-Jeras, 2000) and 'activities of interest' such as playing board games, reading magazines and chatting (Klages, 2011) carried out individually with participants in the control conditions (Table 4.2).

4.3.6 Research methods and outcome measures

Two studies measured changes in levels of satisfaction felt by the residents with their living environments (Chang, 2013; Scott, 2014). Both devised their own measures, with Chang (2013) using the study to validate the Satisfaction with Living Environment Nursing Home (SLE-NHS) scale. In addition to satisfaction, Scott (2014) also administered the Geriatric Depression Scale (GDS), Geriatric Anxiety Inventory (GAI), a wellbeing and engagement questionnaire (completed by residents) and a staff questionnaire. Psychological wellbeing was also measured in Cox (2004) who administered the Affect Rating Scale and by Orsulic-

Jeras (2000) via administrating the Cohen-Mansfield Agitation Inventory (CMAI) and Cornell Scale for Depression (CSD) and through observations using the Affect Rating Scale (ARS). Additionally, Simard (2010) used the Minimum Data Set (MDS) to extract information on levels of agitation, depression and psychoactive medications along with information on behavioural symptoms recorded in the 'Challenging Behaviour Profile' of the MDS. Behaviour changes were recorded in both Orsulic-Jeras (2000) and Kincaid (2003) via observations recording levels of engagement and door-testing behaviours respectively. Orsulic-Jeras (2000) used the Myers Research Institute Engagement Scale (MRI-ES) to record levels of engagement and Kincaid (2003) recorded the frequency of four types of door-testing behaviours (ranging from calm-aggressive). Orsulic-Jeras (2000) used the Multidimensional Observation Scale for Elderly Subjects (MOSES) to assess functional ability whereas Klages (2011) was the only study to just assess levels of physical ability, administering four balance tests (Functional Reach Test, Sharpened Romberg, Timed Up and Timed Go), with additional data about falls rates sourced from patient notes. The only descriptive study (Roe, 2015) carried out non-participant observation and kept fieldnotes during the sessions. (Table 4.3).

Data from staff members were provided by Scott (2014) and Roe (2015) obtained through a staff questionnaire (Scott, 2014) and interviews with the staff members implementing the activity and the caregivers facilitating (Roe, 2015). Staff members also completed the outcome measures (MOSES, CMAI and CSD) in Orsulic-Jeras (2000). Only two studies specified who administered the outcome measures, delivered by the researcher (Scott, 2014) and research assistants (Chang, 2013), however in the latter study residents could complete the questionnaires on their own if they did not wish to be interviewed by the assistants.

Data collection time points varied across studies, for two data were collected at baseline, mid-way through the activity and then following (Orsulic-Jeras, 2000; Scott, 2014) and three collected data at baseline and then at follow-up (Simard, 2010; Klages, 2011; Chang, 2013). The observation studies collected data at intervals throughout the study period (Kincaid, 2003; Cox, 2004; Roe, 2015) (Table 4.3).

4.3.7 Quantitative findings

4.3.7.1 Behaviour

Evidence of benefit for this arts type was limited (see Table 4.3 for included studies and findings). Behavioural symptoms were not shown to significantly improve within three studies which assessed this outcome through assessment of agitation levels (Orsulic-Jeras, 2000; Simard, 2010; Scott, 2014). Low levels of agitation at baseline were observed by both Orsulic-Jeras (2000) and Simard (2010) which could explain the lack of impact for this type of activity on behavioural symptoms. Scott (2014) also found no significant effect for improvement between staff and residents. Limited behavioural effects were shown for Kincaid (2003) who found door-testing behaviours decreased during the time the wall mural was present (55.67 (SD 67.57) vs. 13.42 (SD 28.41) t=2.6, p=0.02).

4.3.7.2 Psychological wellbeing

Studies determining the impact on psychological wellbeing also failed to show significant improvements for those taking part (Table 4.3). No significant impact on levels of depression, anxiety or quality of life were observed by Scott (2015). Similarly, Orsulic-Jeras (2000) and Simard (2010) showed no significant changes to depression levels, with the latter study clinically evidencing this through lack of reduction in antidepressant use. Both studies commented on low levels of existing depression and as with the music studies this may have reduced the likelihood of an observable effect (see section 4.2.7). Cox (2004) also found no

significant impact on levels of fear and anxiety for those participating in snozelen, with contentment levels significantly decreasing (mean 24 vs. 56, Wilcoxon value nr, p<0.01), suggesting residents became less happy with the activity as time went on. Further to this, significant reductions in fear and anxiety were observed when the residents were in the living room (mean 14 vs. 4, Wilcoxon value nr, p<0.05) which may indicate participants struggled with the lack of familiarity with the activity (Cox, 2004). This is also reflected in qualitative findings from multisensory activities (Section 4.3.8 below). Simard (2010) did show some clinical evidence of an impact to anxiety, with a reduction in the amount of days that antianxiety medications were prescribed (0.8 pre-test vs. 0.5 post-test, p=0.04), although this was infrequent and only just significant.

4.3.7.3 Cognition, quality of life and socialisation

Significant improvements were shown in both studies determining an impact on pleasure (Orsulic-Jeras, 2000; Cox, 2004). The first study showed changes across time (mean 62 (SD 8) vs 100 (SD13), F(2, 46)= 7.5, p<0.002) for those participating and also evidenced significantly greater scores in comparison to the control group both immediately following the study (100 (SD 13) vs. 67 (SD8) F(1, 12)= 68.2, p<0.001) and at 9-month follow up (100 (13) vs. 67 (SD8) F(1, 12)= 68.2, p<0.001) (Orsulic-Jeras, 2000). The second (Cox, 2004) also demonstrated increased pleasure scores for those participating in the snozelen (mean 0 vs. 38, p<0.001) however the two control conditions also showed increased pleasure scores (Table 4.3) and the snozelen did not impact significantly more than these. Orsulic-Jeras (2000) also found constructive engagement improved in those taking part (F (1, 12)=8, p<0.002).

No significant changes were found with respect of how satisfied residents were with their living environment upon changing the aesthetic of the home (Chang, 2013; Scott, 2015). There was no significant reduction in incidences of falls (Klages, 2011) and Orsulic-Jeras (2000)

showed decreased functional ability ($F(1, 23)= 6.0$, $p<0.02$) over time (determined by MOSES), suggestive that the activity did not prevent normal cognitive decline. Finally, Chang (2013) found there to be improvements to memory recall during the wall mural installation ($t=13.32$, $p<0.001$) (Table 4.3).

For those taking part in the biophilia sensory condition (Scott, 2014) significant improvements in social engagement were found over time ($F2, 18= 13.82$, $p<0.001$). Further analysis showed significant improvements both at time 2 ($t(13)=5.00$, $p<0.001$) and time 3 ($t(9)= 4.26$, $p<0.01$) compared with time 1. Social engagement was also improved for those taking part in the reminiscence sensory condition ($F(2, 14)= 3.62$, $p<0.05$) again with significant improvements shown at both time 2 ($t(7)=4.33$, $p<0.01$) and time 3 ($t(8)=2.48$, $p<0.05$) compared with time 1. There were no significant improvements for engagement between staff members and residents, although it was commented upon how these were high at baseline and therefore may have failed to show an effect due to this (Table 4.3).

4.3.7.4 Improvements to care and physical health

None of the included multisensory studies quantitatively assessed the impact of the activity on behavioural symptoms (Table 4.3).

4.3.8 Qualitative findings

Qualitative findings were provided by two studies (Cox, 2004; Roe, 2015). The first obtained data from caregivers and relatives/visitors, identifying themes from each group (Cox, 2004). Staff reported on the challenges faced with the activity, the difficulties with staffing levels required for the activity and the need for residents to take part in several sessions to be become used to the features of the snozelen. But they also discussed the change of scenery provided by the snozelen and garden, and how these improved residents' affect This was also found in interviews with family members who said they found the garden area relaxing

compared with the usual home environment and described the reactions of residents to the novel environment of the snozelen room and its components. Family members also reported how they felt the activities could form an alternative to the amount of medications often prescribed to residents, in keeping with the reduction in anti-anxiety medications shown in Simard (2010) (Table 4.3).

The second study discussed the sessions and the structure and future feasibility of the programme (Roe, 2015). This study evidenced the success of implementing a museum-based activity with older people, including those from a care home. However, the study also highlighted the possible logistics and barriers which may need to be considered such as staffing levels, the accessibility of venues coupled with the needs of the residents and transporting individuals to and from the location. In addition to exploring the ability to carry out the activity, this study also reported from interviews with staff members who reported on the wellbeing evidenced in the residents taking part through the connections formed with the artefacts they interacted with. This was further evidenced by the findings that residents used positive phrases when referring to the exhibits and artefacts. In addition to the studies mentioned above which quantitatively evidenced engagement, this qualitative evidence also commented on the engagement shown by the residents when they were exposed to different experiences. However, also evidenced was the ability for some residents to find the experience unsettling due to the different location and the need for reassurance should this occur. Unlike Scott (2014) who found there to be no increased engagement of staff members with residents, Roe (2015) reported an improved group dynamic as the sessions progressed, with the residents and staff members interacting more as equals and the staff present at the museum adapting how they interacted with the group.

4.3.9 Quality appraisal of included studies

4.3.9.1 Levels of evidence

Two studies (Klages, 2011; Scott, 2014) provided evidence at level Ib, the highest quality, provided by RCTs (level Ia evidence was not included in this review as it refers to systematic reviews). A further five studies provided evidence at level II, indicative of moderate evidence, two within the first tier level IIa (Orsulic-Jeras, 2000; Cox, 2004) and two the second tier, level IIb (Kincaid, 2003; Simard, 2010; Chang, 2013). One provided the lowest quality of evidence accepted within this review (level III), evidence obtained from descriptive studies (Roe 2015).

4.3.9.2 Quantitative studies

The risk of bias Table for included quantitative studies can be seen in Figure 4.4. Only two studies were randomised controlled trials (Klages, 2011; Scott, 2014). Randomisation was adequately carried out for both these studies via use of a computer programme (Klages, 2011) and a lottery method to assign grouping at cluster level (Scott, 2014). Allocation concealment, however, was not addressed in Scott (2014) and consisted of simply a random number list whereby group assignment occurred when participants returned their questionnaires and were given a number, thereby meaning that concealment would not adequately occur (Klages, 2011).

From the information provided it was evident that four studies had not blinded the scoring of outcome assessments (Orsulic-Jeras, 2000; Kincaid, 2003; Cox, 2004; Klages, 2011). The remaining three studies did not provide information as to whether those administering and scoring outcome assessments were blinded as to condition or group allocation, however it is probable that this did not occur (Simard, 2010; Chang, 2013; Scott, 2014).

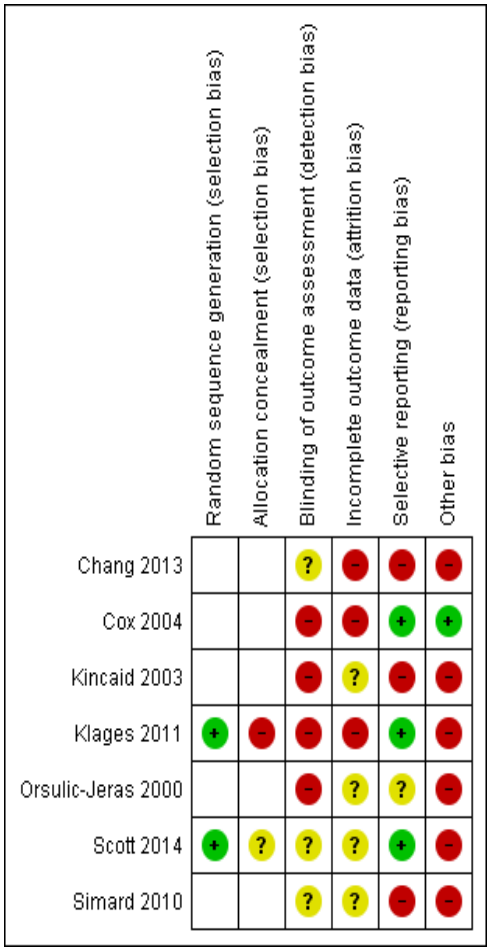


Figure 4.4 judgements on quality of the included quantitative studies which assessed multisensory activities. Green corresponds low risk, red high risk and yellow unclear risk.

Again, four studies did not provide enough information to provide a judgement with respect of incomplete outcome data. The remaining three studies received judgements of high risk. The first (Chang, 2013) provided sample size calculations based upon the initial number recruited, however a drop-out rate of almost a third (15 out of 48) reduced the sample considerably and missing results were excluded from the analysis. Similarly, Klages (2011) reported five out of the sample size of 24 (21%) withdrew and were not included in the analysis which consequently only included 19 participants. Cox (2004) reported there to be 'number of cases with missing data that were excluded from the analysis'.

Judgements of low-risk for selective reporting were given for three studies (Cox, 2004; Klages, 2011; Scott, 2014) with a further study not containing enough of a protocol to make a judgement either way (Orsulic-Jeras, 2000). The remaining studies received judgements of high risk (Kincaid, 2003; Simard, 2010; Chang, 2013). One of these studies referred to collecting qualitative data that was not presented in the methods section (Kincaid, 2003), another reported no aims or outcome measures meaning it is unclear what outcome measures were assessed (Simard, 2010) and the final study refers to significant differences on an outcome of the scale used in the discussion section, however this is not reported within the results section (Chang, 2013).

Only one study (Cox, 2004) did not show any apparent biases in addition to those mentioned. Two made claims in the discussion section not concurrent with the data in the results section, the first referred to semi-structured interviews not mentioned anywhere else in the paper (Chang, 2013) and the second stated participants displayed 'increased interest' but this was not obtained by any outcome measures (Simard, 2010). Methodological flaws were evidenced in a further two studies, one failed to establish reliability or validity for their questionnaire which assessed the satisfaction that older people felt with their living environment (Scott, 2014) and the second used only one observer therefore meaning inter-rater reliability could not be established (Kincaid, 2003). In addition to this, Simard (2010) collected baseline data at vastly different times for residents, ranging from 10-80 days prior to taking part in the programme and participants in Orsulic-Jeras (2000) all took part in different aspects of the Montessori programme without data being grouped accordingly. Finally, Klages (2011) failed to calculate a sample size meaning that following drop-outs a few values were not statistically comparable with the study lacking overall statistical power.

4.3.9.3 Qualitative studies

The qualitative evidence (Roe, 2015) was judged at a level B, meaning it contained few flaws and these were unlikely to impact upon the overall recommendations and conclusions. Minor flaws identified included a lack of rationale for the methods chosen, despite these being in keeping with the aim of the study and well explained hey were not justified. It could be assumed that all those taking part in the programme were eligible for inclusion, however there is no specified inclusion criteria. Finally, there was no evidence of reflexivity from researchers in the study.

4.3.10 Summary of multisensory activities

Overall there was limited evidence of benefit for this arts type (Table 4.3). Most studies did not show an impact on agitation levels, or psychological wellbeing. As found within the studies which focused on music (Section 4.2) there were low levels of agitation and depression at baseline, indicating the included residents did not have clinical need and therefore reducing the likelihood of an observable effect. The requirement of low baseline scores was further evidenced by Scott (2014) who failed to show increased social engagement between staff members and residents, however scores were high initially and therefore showed little room for improvements. Impact was shown for pleasure (Orsulic-Jeras, 2000; Cox, 2004), wellbeing (Roe, 2014), door-testing behaviours (Kincaid, 2013) and memory recall (Chang, 2013).

The ability for this activity type to improve satisfaction with the care home was not shown (Chang, 2013; Scott, 2014) by either study which analysed this outcome. Similarly, both Cox (2004) and Roe (2015) reported some residents struggled with the unfamiliarity of the activity. This was particularly evidenced by Cox (2004) who showed contentment decreased with the amount of time participating with the snozelen and garden activities, an effect not observed when residents were sat within the living room area (familiar environment).

Overall the quality of quantitative evidence for this arts type was lacking (Figure 4.4), which may also offer an explanation as to lack of observable effect. Four out of the seven studies received no judgements of low risk of bias and all received at least one high risk judgement. Qualitative evidence was of a higher standard, however as this was a feasibility study and the only study within this review which utilised a space other than the care home environment, evidence of benefit is still limited (Figure 4.4 and Table 4.3).

This section has focused on studies classified as multisensory activities. The next section (section 4.4) provides a narrative descriptive synthesis of included studies featuring the arts type 'performance'.

4.4 Narrative descriptive synthesis of studies featuring the arts type 'performance'.

This section features studies that evaluated an arts activity which required residents to either actively partake in a performance, or observe a performance carried out by others. All included studies and their characteristics are shown in Table 4.4. As with the previous subsections, a narrative descriptive summary of included studies is presented, along with appraisal of methodological quality.

4.4.1 Characteristics of included studies

A total of six studies (8% of those included within the whole review, n=71) were placed within this category (Table 4.4) published dates ranged from 1998-2014 (Houston, 1998; Palo-Bengtsson, 1998; Heyn, 2003; Noice, 2006; Guzman-Garcia, 2013; Vankova, 2014). Only one study, published over 20 years ago, reported actual dates of data collection (Palo-Bengtsson, 1998) and this occurred between February and May, 1995.

Four were intervention studies reporting quantitative data, of which one reported results from a randomised controlled trial (RCT) (Vankova, 2014), and another featuring a cluster-level randomised controlled trial (Houston, 1998). There were two uncontrolled pre/post-test studies (Heyn, 2003; Noice, 2006). The remaining two studies were descriptive, both reporting qualitative findings (Palo-Bengtsson, 1998; Guzman-Garcia, 2013) obtained via observations and interviews with residents, respectively (Table 4.4)

Table 4.4 Characteristics of included studies featuring performance.

Study ID and type	Location	Participants	Activity Type	Aims	Outcome measures	Findings	Quality score
037. Heyn (2003) Intervention Dates not specified Quantitative Uncontrolled pre/post-test	USA Further information not reported.	13 in total, 12 females (92%) and one male (8%). Mean age 85.7 ± 6.5, range 70-93. Loss to follow-up: None reported.	Multiseneory exercise program. The programme was carried out three times a week for a total of 8 weeks and lasted between 15 and 70 minutes. Comprised of four parts: warm-up, flexibility, strength-training and closure focused on breathing.	To explore the effects of a multisensory exercise program on cognitive functioning (as measured by engagement), behaviour (as measured by mood) and physiological indices (such as blood pressure, resting heart rate and weight) in residents from a nursing home with a diagnosis of moderate to severe Alzheimer's Disease.	The Menorah Park Engagement Scale to assess levels of engagement with the activity. The Caregiver Mood Report, to assess how caregivers felt during the intervention. A video recording was used twice during the intervention to record facial expressions, eye contact and singing. Resting heart rate (RHR), blood pressure and weight.	The MPES showed that 69.2% (n=9) of the participants had engaged in 'more than half' of the activity and 30.8% (n=4) had engaged in 'up to half' of the activity. At the end of the multisensory exercise program, eight residents (61.5%) were classified as showing positive improvements in their overall mood shown by a positive response on "participant looks: 1) happier; 2) calmer and 3) friendlier. 5 residents (38.5%) were reported as showing no significant, or little improvements in their overall mood. RHR showed significant improvements from a mean of 79.69 to 70.92bmp (a mean decrease of 8.77bmp, t= 5.93, DF=12, p=0.002, p<0.01). No significant difference was found in BP and weight. BCRS scores remained sTable, with a cognitive score range from 4.8 to 6.4 (mean= 5.72, SD=.44). The length of the sessions was increased incrementally from 15 to 70 minutes. Residents achieved peak performance at 4 weeks and none had significant changes in their schedules reported.	GRADE score - - - + + - Levels of evidence score LII-b, grade B.
054 Guzman-Garcia (2013) Descriptive Dates not specified. Qualitative Individual interviews	UK Two private care homes located in Newcastle-Upon-Tyne	13 took part in the activity. 3 male (23%) and 10 female (77%). Mean age was 80.5 (SD 6.81). 7 took part in interviews: 5 females (71%) 2 males (28%). Mean age 82.4 (SD 4.86). 9 staff members interviewed. 7 females (71%) and 2 males (28%). Ages not reported.	Latin dancing-Danzon psychomotor intervention. Sessions were carried out bi-weekly for a total of six weeks and lasted for a total of 35 minutes. Sessions were delivered by the author and facilitated by care staff.	To assess the implementation of a Latin ballroom dance intervention for people with dementia who were living in a residential facility.	Each participant giving interviews was interviewed twice and asked to identify the positive and negative aspects of the programme. Results were analysed using grounded theory.	The central category from interviews with staff was identified as 'benefits' which related to residents, staff and family members. This was further condensed into four subcategories: Category A- benefits on the residents who danced, with the following subcategories: behaviour, affective states, mental stimulation, socialising/communication, mobility, reminiscence and physical health. Category B- benefits on the spectator-residents, this related to those who observed the sessions but were not able to take part themselves. Category C- benefits on family members. Category D- benefits on care staff. The central category from interviews with residents was 'enjoyment'. Subcategories identified were: mood, behaviour, socialising, reminiscence, mental stimulation and mobility.	QA score Grade B. Levels of evidence score LIII, grade C.
074 Palo-Bengtsson (1998) Descriptive February-May 1995. Qualitative. Observations.	Sweden A nursing home located in Stockholm, Sweden.	6 participants (4 resided in the care home and 2 were attendees at the day unit). Age range= 76-88. Mean age= 83.3 (± not reported) 4 were female (67%) and 2 were male (33%). Loss to follow-up: 4 lost from 10 participants who were selected to take part (prior to the start of the intervention).	Dancing- social dance sessions. These sessions were already carried out within the home, once a month. 20-50 in each session. Residents attended along with their caregivers. The dance took place between 10.30am and 11.15am (lasting 45 minutes in total). Dance music was performed by a local dance band .	To assess how participants with dementia would respond when taking part in social dance sessions.	Content analysis was performed on qualitative data obtained from video recordings of the sessions using three steps: 1. Define the content, 2. Identify the characteristics, 3. Perform content analysis.	Group category: motor functions. Variables: physical activity, spontaneous activity. Group category: intellectual functions. Variables: orientate themselves in space, recent memory, timing, personal orientation, distant memory, wakefulness, concentration, increased tempo, collected, long-winded, distractibility. Group category: emotional functions. Variables: showing emotions, control of emotional reactions, motivation. Group category: different symptoms common in dementia. Variables: confusion, irritability, anxiety, agony (mental discomfort), reduced mood, restlessness.	Levels of evidence score LIII, grade B.
098. Houston (1998) Intervention Dates not specified Quantitative Cluster-RCT	UK Seven care homes within the UK (one carried out the pilot and the remaining six	61 residents from six homes. 44 were female (72%) and 17 male (28%). Mean age 83.7 (± 6.79) range 63-97. 31 in the intervention	Humour and sing-a-long The researchers sang along with a backing track and danced in a comical fashion and participants were	To ascertain whether there would be an impact on the psychological wellbeing of older people taking part in a 'laughter inducing humorous activity' consisting of a	The general health questionnaire (GHQ-28) the Hospital Anxiety and Depression Scale (HADS). Carried out at baseline and immediately following the intervention.	Baseline t-tests showed no significant differences (p>0.05) for scores on the GHQ and HADS and subscales. *GHQ:* For the GHQ anxiety subscale ANCOVA showed significant correlation between Time 1 and Time 2 F (1, 58)= 2.73, p=<0.001 and significant effect of condition when Time 1 scores were adjusted for F(1, 58)= 15.48, p<0.001. GHQ anxiety mean score intervention 2.11 and control 5.02, F (1, 58)= 15.48, p<0.001. Not significant for somatic symptoms (6.29 intervention vs 6.33 control), p=0.5; severe depression (8.24 vs. 9.38), p=0.43 or social dysfunction (intervention 5.78 vs 3.26 control), p=0.43.	GRADE score ? ? ? ? ? ? Levels of evidence score LI, grade A.

Study ID and type	Location	Participants	Activity Type	Aims	Outcome measures	Findings	Quality score
	took part in the actual study).	condition and 30 in the control.	given props and asked to join in.	sing-a-long and dance activity. Focus on depression and anxiety.		*HADS:* for anxiety significant effects of the covariate F (1, 58)= 45.37, p<0.001 and condition once covariate was adjusted for F(1, 58)= 19.99, p<0.001 (mean intervention 3.15 vs. 6.04 control). For depression effect of covariate was not significant F(1, 58), p=0.17 but effect of condition was, F (1, 58)= 8.89, p=0.004 (mean intervention 3.27 vs. 5.26 control).	
126 Vankova (2014) Intervention. Dates not specified Quantitative RCT	Czech Republic Seven different nursing homes.	Total number participants= 162. Mean age 83 (± 7.98). 149 female (92%) 13 male (8%). Experimental group= 79 participants. Mean age 83 ((± 8.23). 76 female (92%) 3 male (8%). Control group= 83 participants. Mean age 83 ((± 7.87). 73 (92%) female, 10 (8%) male.	Exercise dance EXDASE classes held once a week for a total of three months. Sessions lasted for an hour- 10 minutes warm up, 40 minutes main intervention and 10 minutes cool down. Does not report who delivered the sessions.	To implement an exercise dance intervention for seniors (EXDASE) and evaluate its role on symptoms of depression in older people residing in a nursing home.	A Czech language version of the Geriatric Depression Scale (GDS). Scores were obtained at baseline (one week prior to the intervention) and within one week following the intervention (post-test).	GDS scores were significantly improved following the dance therapy intervention (5.71 mean (± 3.8) versus 5.00 (± 3.3)). In contrast, the control group showed worsening of symptoms (4.86 (± 3.2) pre-test versus 5.27 (± 3.3) post-test), however not significantly so (p=0.81). The group was separated into those taking antidepressants and those not, both groups showed improvements, however only the subgroup not taking antidepressants showed a significant improvement (p=0.008). Generalised linear model: significant for dance therapy (Wilks lambda F=10.58, p=0.001). Controlling for antidepressant usage showed similar results, suggesting changes occurred as a result of the intervention (Wilk lambda F=10.58, p=0.001) rather than antidepressant intake (Wilk lambda F=0.107, p=0.744). Further GLM analysis confirmed that the results occurred as a result of the intervention (Wilks lambda F=10.75, p=0.001) rather than the nursing home (Wilk lambda F=0.26, p=0.621).	GRADE score ? ? + + - + Levels of evidence score LI, grade A.
135 Noice (2006) Intervention Dates not specified.	USA Plymouth Place Continuing Care Facility (La Grange, Illinois), which offers varying support	18 participants. Mean age was 82.3 years (± 5.6), with a range of 72-95 years. 5 were male (27.2%) and 13 females (72.2%).	Drama A course was delivered on Tuesdays and Thursdays for a total of one hour. Emphasis was placed on genuine involvement and communicating thoughts and feelings. These sessions were delivered by a trained drama instructor.	To determine whether a theatrical intervention had bearing on cognition in older people residing within a continuing care facility.	Cognitive battery consisting of word recall and means end problem solving tasks. Psychological wellbeing battery consisting of self-esteem self-reported psychological health and memory.	An initial pre-test was carried out four weeks prior to the intervention and then a second pre-test just before the programme commenced. <u>Cognitive battery</u> A significant MANOVA was computed for the variables of time (within-subjects variable) and cognitive measures (dependent) (p<0.001). Therefore univariate ANOVAs were carried out and showed significant improvements for word recall (F(2, 34)= 4.41, p=0.02), problem solving, (F (2, 34)= 15.46, p<0.001) and set span (F(2, 34)= 6.43, p<0.01). Word recall was significantly increased from pre-test 2 to post-test (t(17)= 2.70, p=0.015) but not between pre-test 1 and pre-test 2 (p=0.59). Problem solving was significantly improved between pre-test 2 and post-test (t (17)= 2.64, p=0.002) and also between pre-test 1 and pre-test 2 (t(17)= 3.01, p=0.01). Marginal increases were observed for working memory from pre-test 1 to pre-test 2 (p=0.70) and pre-test 2 to post-test (p=0.13) but not significantly so. <u>Psychological Functioning</u> A MANOVA showed a significant effect of time (F (2, 14= 4.24, p=0.02) and univariate analyses did not reach significance (p>0.05). T-tests showed no significant self-esteem changes from pre-test 1 to pre-test 2 (p=0.62) and a non-significant improvement from pre-test 2 to post-test (t(17)= 1.63, p=0.12). No significant increases were found for quality of life. The controllability scale consisted of three positive and three negative dimensions about memory decline. Scores for the Present Ability subscale showed a moderate positive correlation with Effort utility scores (r 16=0.48p<.0.05) and a negative correlations with IneviTable Decrement (r(16)=-0.54, p=0.02). Those who believed they were able to improve their memory showed the greatest increase in recall performance r(16=0.52, p=0.028). However, positive correlation was also observed for IneviTable Decrement and problem solving scores r(16)=0.56, p=0.016.	GRADE score ? ? - ? Levels of evidence score LII-b, grade B.

4.4.2 Aims of included studies

Two studies, both randomised controlled trials (Houston, 1998; Vankova, 2014) explicitly stated an aim of assessing impact of the activity on psychological wellbeing, both focusing on depression and Houston (1998) assessing anxiety also. Heyn (2003) outlined the aim of impact on 'behaviour', however further stated this was to be assessed through 'subjective mood changes'. Noice (2006) did not specify an aim, reporting this was a feasibility study based on previous findings with older people living within their own home (Noice et al, 2004). The title of the study referred only to cognition, however it also featured psychological wellbeing measures (see 4.4.6, research methods), despite not stating this as an aim.

Improvements in cognition were reported by two studies, the first of which outlined engagement as a measure (Heyn, 2003) and the second which did not specify a particular measure (Noice, 2006). A multisensory exercise programme (Heyn, 2003) also determined the impact of the activity on physiological changes in residents.

Both descriptive studies focused on dancing activities (Palo-Bengtsson, 1998; Guzman-Garcia, 2013). The first focused on assessing an existing programme and the second implemented a new programme for the study.

4.4.3 Location and setting

Two studies took place within the UK (Houston, 1998; Guzman-Garcia, 2013), two in the USA (Heyn, 2003; Noice, 2006), one in the Czech Republic (Vankova, 2014) and one in Sweden (Palo-Bengtsson, 1998). Data were collected from participants who resided in a total of 17 care homes across the six studies, with half the studies obtaining information from residents from more than one home (Houston, 1998; Guzman-Garcia 2013; Vankova, 2014). Two studies specified 'nursing home residents' (Palo-Bengstton, 1998; Vankova, 2014) two 'care homes' (Houston, 1998; Guzman-Garcia, 2013) one a 'continuing care facility' (Noice, 2006)

and residents from the final study lived within a 'memory care residence' for those with Alzheimer's Disease (Heyn, 2003) (Table 4.4).

Just two studies reported the actual location and environment where the arts activity was carried out. Both stated that sessions were carried out within the care home, with one reporting the activity to take place within a multi-purpose room (Heyn, 2003) and the second within a large entertainment hall (Palo-Bengtsson, 1998).

4.4.4 Populations

Total numbers of participants across the six studies was 273, with a large range in sample size of 6-162 (45.5, SD 55.1) (Table 4.4). The two descriptive studies featured the smallest sample sizes with six (Palo-Bengtsson, 1998) and 13 (Guzman-Garcia, 2013) with additional data provided by seven staff members in the latter study. The largest sample sizes were unsurprisingly reported for the randomised controlled trials with Vankova (2013) reporting 162 in the total sample (79 experimental group vs. 83 control group) and Houston (2014) showing a total sample size of 61 (31 experimental group vs. 30 control group). The remaining uncontrolled pre- test post-test studies featured similar sized samples of 13 (Heyn, 2003) and 18 (Noice, 2006).

All studies reported mean age of residents and in each case this was above the age of 80 years (Table 4.4). The mean age of the six included studies was 83 years, with little variation (SD 1.6, range 80.5-85.7 years). Only one study of two featuring a control group reported separate demographic values for each of the groups (Vankova, 2013). The mean age of the residents in both the intervention (n=76) and control groups (n=79) was matched at 83 years.

Gender of the participants was also consistently reported across the studies with the total number of women a majority 249 (92%) and men, 24 (8%). Again, for Vankova (2013) who

reported separate demographic information for the control and intervention group, numbers of men and women were shown with 92% women in the intervention group (n=76, men 8% n=3) and 88% women in the control group (n=73, men 12% n=10) (Table 4.4).

Half the studies (n=3) did not report whether the residents had a dementia diagnosis, nor did they feature inclusion criteria that would either only include or exclude those with dementia (Houston 1998; Noice, 2006; Vankova, 2014). Two of the studies specified the characteristics of the participants, reporting all had dementia (Palo-Bengtsson, 1998; Guzman-Garcia, 2013), with the latter study reporting the types of dementia in the total sample of 13 (eight had an Alzheimer's disease diagnosis, two had fronto-temporal dementia, one vascular dementia, one mixed dementia and one Parkinson's disease dementia). The final study by Heyn (2003) reported inclusion criteria which specified participants needed to have a documented dementia diagnosis. Despite not reporting specific inclusion criteria, the title of the study as reported by Noice (2006) refers to residents being 'intact'. It is unclear as to what is specifically meant by this term, it does seem probable that it refers to only residents who did not have cognitive deficits. This is further reinforced by the study reporting that residents did not have any 'mental impairment' when screened using the Short PorTable Mental Status Questionnaire (SPMSQ), a ten-item questionnaire which asks questions such as the residents current address, or who is currently president (USA) (Table 4.4).

4.4.5 Characteristics of the arts activities

Characteristics of the individual studies and how they carried out the arts activity are shown in Table 4.4. Both descriptive studies focused on dance, one on ballroom dancing (Palo-Bengtsson, 1998) and the second on Latin dance classes (Guzman-Garcia, 2013). Two studies focused on exercise classes, one implemented an exercise dance class (Vankova, 2014) and the second utilised imagery and storytelling during exercise sessions (Heyn, 2003), however

there is limited evidence within the methods section as to how storytelling was implemented during the activity. The remaining studies featured drama (Noice, 2006) and a 'humorous sing-a-long activity' (Houston, 1998).

All studies were reportedly carried out within a group setting, in keeping with the performance element of the described arts activities. Despite this, only one study provided information as to the group size of the activity (drama) reporting that all 18 residents enrolled onto the programme took part in the drama classes together (Noice, 2006). Consistency of this group size was explained by the fact that residents only participated in the sessions if they could attend each one. A further study did not explicitly state group size, however it seems probable that all the residents attended the sessions (n=13) given that they were recruited from one care home and this was the location where the programme was delivered (Heyn, 2003). Palo-Bengtsson (1998) reported that average attendance at the social dance sessions, which were run monthly, was between 20 and 50 residents, with caregivers also attending. The remaining three studies did not report group size, these were all studies where residents were recruited from more than one care home and therefore unlikely to all have attended the sessions at the same time (Houston, 1998; Guzman-Garcia, 2013; Vankova, 2014).

Only one study featured description of an already established dance programme, held monthly within the care home (Palo-Bengtsson, 1998). These were described as 'social dance sessions' and appeared to not feature residents being taught steps or dances, rather they could turn up and dance freely. There was also a live band which performed music for the dance sessions. The remaining five studies all assessed a programme which was designed and implemented for the study period. Facilitators of these programmes were the authors in two of the studies (Houston, 1998; Guzman-Garcia, 2013) and a trained drama teacher (Noice,

2006) with the remaining two studies (Palo-Bengtsson, 1998; Vankova, 2014) not describing who led the sessions.

4.4.6 Research methods and outcome measures

For measures of depression the Geriatric Depression Scale (GDS) (Vankova, 2014) and Hospital Anxiety and Depression Scale (HADS) (Houston, 1998) were used, with the latter scale also providing information on anxiety measures. In both instances these were administered at baseline and then immediately following the activity to both the intervention and control group. Both were administered to residents by researchers (Houston, 1998) and research assistants (Vankova, 2014) with the second study specifying research assistants were blinded as to group allocation. Heyn (2003) also contained an outcome measure related to 'mood'. This was assessed via the use of a three-item unvalidated questionnaire which was scored by three family members, three caregivers, the activity director and a student assistant. The scale simply asked the panel to record whether the resident 'looked happier', 'is calmer' or 'is friendlier'.

Noice (2006) collected information at an initial pre-test (4 weeks prior to the study), a second pre-test (day before) and then following the activity. A cognitive battery was administered to residents which included a word recall task, listening span, Means End Problem Solving Procedure (MEPS), 'every day practical and social functioning' and Memory Controllability Index. Two psychological wellbeing measures were also included, the Self-Esteem Scale and Self-Reported Psychological Health measure. All tests were administered in a group setting for an hour-long session.

For engagement levels, Heyn (2003) administered the Menorah Park Engagement Scale (MPES) scored by the eight people who completed the mood scale before and following the

activity, with no further information provided as to how many observation periods were utilised, nor how many days these took place for. It is unclear whether these eight people had received training on how to complete the scale, particularly as family members made up three of the group. This study also collected physiological data, blood pressure, resting heart rate and weight. These were measured prior to the study and following by a nurse at the care home.

Both qualitative studies explored the impact of dance activities. The first of these (Palo-Bengtsson, 1998) video recorded social dancing sessions and then analysed these using content analysis to describe incidences of behaviours which were mapped to the elements within the Gottfries-Brane-Steen (GBS) rating scale for symptoms of dementia. Analysed videotapes evidenced how aspects of the GBS were met during the arts activity. These fitted into the categories of motor functions (physical activity and spontaneous activity), intellectual functions (orientating in space, recent memory, timing, personal orientation, distant memory, wakefulness, concentration, increased tempo, collected, long windedness and distractibility), emotional functions (showing emotions, control of emotional reactions and motivation) and the different symptoms common in dementia (confusion, irritability, anxiety, agony (mental discomfort), reduced mood and restlessness The second study (Guzman-Garcia, 2013) carried out semi-structured interviews twice with those taking part (although it is unclear at what points they were interviewed) and then analysed these using grounded theory.

4.4.7 Quantitative findings

4.4.7.1 Behaviour

None of the included performance studies quantitatively assessed the impact of the activity on behavioural symptoms.

4.4.7.2 Psychological wellbeing

Both studies determining the impact of the activity on depression levels showed significant increases (Table 4.4). The first (Vankova, 2014) showed improved GDS scores over time (5.71 (3.84) vs. 5 (3.29) t-test statistic not reported, p=0.005) decreases shown in the control group, although these did not reach significance. Furthermore, dance therapy was shown within a general linear model to significantly impact GDS scores (F= 10.58, p=0.001) indicating an impact across all the homes (no significant effect on depression scores) and irrespective of antidepressant use (again no significant effect). Similarly, Houston (1998) found significantly improved depression scores for the intervention group compared with controls at post-test (mean 3.27 vs. 5.26, F (1, 58)= 8.89, p=0.004) although scores were different midway through the programme, suggesting a cumulative effect. Severe depression measured via the GDS was not significantly changed. Noice (2006) showed there were some changes over time with respect of psychological functioning (individual scores not reported, F2, 14) = 4.24, p=0.02), although univariate analysis failed to reach significance and therefore specific changes could not be pinpointed. Heyn (2003) reported mood improvements, but these are reported descriptively with no statistical analysis. Anxiety changes were observed by Houston (1998) from pre to post-test (F(1, 58)= 45.37, p<0.001 and in comparison to the control group (3.15 vs. 6.04, F(1, 58)= 19.99, p<0.001).

4.4.7.3 Cognition, quality of life and socialisation

Significant cognitive changes were shown across time for Noice (2006) (p<0.001) and Houston (1998), with the second also demonstrating significant differences between groups at post-test (F(1, 58)= 15.48, p<0.001). Noice (2006) also described significant changes to word recall (F(2, 34)= 4.41, p=0.002), problem solving (F(2, 34)= 15.46, p<0.001) and set-span (F(2, 34)= 6.43, p<0.01). Also described within the analysis was the finding that problem solving also

improved between pre-test 1 and pre-test 2, indicating practice effects may be responsible for the overall increase for this measure. However, with the absence of a control group it is not possible to know whether these improvements would have happened without the activity. Heyn (2003) reported no statistically significant changes to measures of cognition. The only study to assess quality of life (Noice, 2006) did not show any significant changes for this outcome (Table 4.4).

4.4.7.4 Improvements to care and physiological changes

Only one study assessed physiological changes and showed a difference in resting heart rate pre and post-test for those taking part (mean decrease of 8.77bpm t=5.93, *p*=0.002) but no other significant physiological changes were observed (Heyn, 2003). No studies assessed the impact on improvements to care (Table 4.4).

4.4.8 Qualitative findings

Guzman-Garcia (2013) interviewed both caregivers and participants and identified the central theme of 'benefits' (Table 4.4). This consisted of four categories, the first of which was benefits to those taking part (category A) with subcategories of behaviour, affect, mental stimulation, socialisation, mobility, reminiscence and physical health. Whilst affect was mainly described in a positive way, there was also discussion from residents as to the negative emotions promoted by the activity, such as being reminded of their decreasing mobility. The second subcategory (category B) identified benefits to other residents observing the sessions, but not taking part due to disability. This was the only study within this review to feature benefits to residents who watched the others partake in an activity but were not able to do so themselves. Finally, category C was benefits to family members and category D benefits to care staff. Improved socialisation was highlighted as a benefits for all categories, between residents, between caregivers and residents and between family members and residents. The

ability to deescalate potentially difficult behaviours from residents was also highlighted as a benefit for staff members. Barriers to carrying out the activity were also discussed within the findings, such as room layout and staff availability. Staff members also reported that whilst beneficial, the increased emotional bond formed with the residents meant it was more painful when they passed away.

The second study (Palo-Bengtsson, 1998) video-recorded social dancing sessions and subjected these to content analysis, mapping evidence to aspects of the GBS scale. Physical and spontaneous activity were said to be well evidenced from the videotapes, along with displays of emotion, control of emotional actions and motivation. These were stated to be demonstrated within the data, however, there is no indication of how this was evidenced i.e. what behaviours had to be evidenced and whether this was consistent for each resident. There was no use of quotes or examples of behaviours to support most of these assumptions. Measures of intellectual functions were also stated to be present, for this measure there were certain criteria reported which indicated residents had met these criteria: orientation in space (capacity to move in the dance hall), recent memory (remembering their place in the dance hall), distant memory (remembering old social patterns, songs and melodies) and being fully awake. There was also evidence of residents being able to keep concentration and change tempo. There was not enough evidence to show personal orientation and the element of long windedness (aspects of the intellectual function criteria on the GBS) was also difficult to evidence. Incidences of symptoms common in dementia were evidenced via certain behavioural symptoms being recorded.

4.4.9 Quality appraisal of included studies

4.4.9.1 Levels of evidence

Levels of evidence scores for the studies were LIb for two (Houston, 1998; Vankova, 2014), this corresponds to the highest level of evidence (as systematic reviews (LIa were not included). None of the studies were judged at LIIa (moderate evidence from well-designed quasi-RCTs) but two (Heyn, 2003; Noice, 2006) were judged at the lower tier of LIIb, reflecting moderate quality evidence from well-designed uncontrolled studies. Finally, two were at LIII (Palo-Bengtsson, 1998; Guzman-Garcia, 2013) the lowest quality eligible for inclusion within the review.

4.4.9.2 Quantitative studies

Judgements on the quality of the included quantitative intervention studies can be found in Figure 4.5. The largest number of responses occurring was that of 'unclear', indicating a general lack of information upon which to make a judgement of high or low quality. Both RCTs scored unclear judgements for both randomisation and allocation concealment with neither containing information as to how randomisation and allocation occurred (Houston, 1998; Vankova, 2014). One study reported that the research assistants who administered the outcome measure were blinded as to the allocation of participants (Vankova, 2014), one reported that blinding of outcome assessment did not occur (Heyn, 2003), and a further two studies did not report any information to make a judgement, however, it seems likely blinding did not occur (Houston, 1998; Noice, 2006). Only one study provided evidence for a judgement of quality to be made regarding incomplete outcome data. Vankova (2014) provided detailed information as to who dropped out and carried out supplementary analysis with those who were withdrawn due to low MMSE scores (therefore failing to meet criteria for inclusion).

Only Noice (2006) provided a detailed enough study protocol to gain a score of low risk, with the others not reporting enough information for a judgement other than unclear. Only one study (Heyn, 2003) featured an evident source of other bias for selective reporting by using measures of mood and engagement that were not assessed for validity.

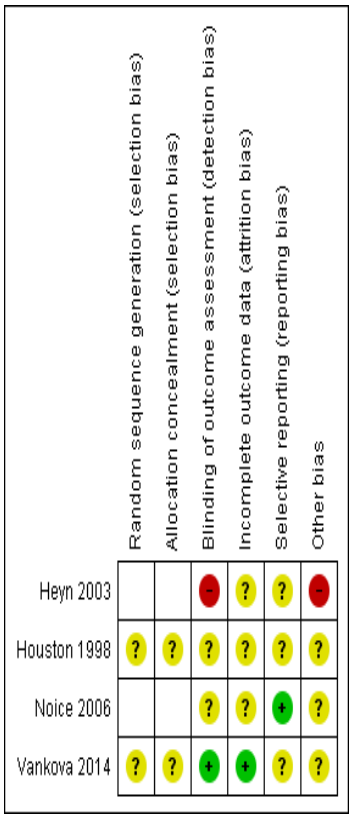


Figure 4.5. Represents quantitative included studies and the judgements on quality. Green corresponds to 'low risk', red to 'high risk' and yellow to 'unclear risk'.

4.4.9.3 Qualitative studies

The two qualitative studies received a judgement of Grade B (few flaws, unlikely to affect the study) (Guzman-Garcia, 2013) and Grade C (some flaws, that may affect the study) (Palo-Bengtsson, 1998). There was an overlap in the sections of the quality appraisal where the studies did not provide information, neither study provided adequate rationale as to why the qualitative method had been chosen, nor the justification for the sample size providing data. In terms of the scope and aim of the study, Palo-Bengtsson (1998) provided a good account

of background, rationale, aims and purposes whereas Guzman-Garcia (2013) failed to report these. Analysis was recorded well within both studies, but they differed in the amount of information provided in the interpretation stage. Guzman-Garcia (2013) gave a good account of the ways in which data were categorised and made good use of data to support their assumptions. A lack of reporting however meant that it was unclear as to how they had collected data and therefore meant that an audit trail would also be unclear when it came to replicate the study. Palo-Bengtsson (1998) provided little information as to how they had made their assumptions, there was a lack of supporting data provided within the results section. Despite there being accounts of behaviours taking place it is unclear how many residents displayed the behaviours, nor how these were determined to have occurred. Furthermore, there was no evidence of immersion with the study data and a lack of information relating to data collection, again meaning it would be unclear whether an audit trail would be possible.

4.4.10 Summary of performance activities

A total of six studies were placed within this category, two qualitative and four quantitative. Activities evaluated were dance, exercise classes, drama and 'humorous sing-a-long'. Overall there was evidence of benefit for depression (Houston, 1998; Vankova, 2014), anxiety (Houston, 1998) and cognition (Houston, 1998; Noice, 2006), with limited evidence for physical health (resting heart rate only (Heyn, 2003)). Of interest is the study carried out by Noice (2006), which was the only study within this entire review (of any arts type) to inform residents that they were not completing the task correctly, or that they were utilising the 'wrong' methods. This may be responsible for a lack of positive impact on quality of life or depression, which was decreased in both Houston (1998) and Vankova (2014).

The two qualitative studies described improvements to affect, behaviour, intellectual functioning, socialisation and physical health (Palo-Bengtsson, 1998; Guzman-Garcia, 2013). The second study also described how the activity could negatively impact residents, particularly with reference to negative emotions associated with physical decline (Guzman-Garcia, 2013). Qualitative evidence from Guzman-Garcia (2013) also showed an influence on those residents observing the sessions, staff members and family members. This study also showed barriers to implementing the activity such as staffing issues and logistics.

Quality of the evidence provided for this activity type was difficult to ascertain, given lack of reporting for many of the domains assessed via GRADE criteria. Only two studies featured control groups (Noice, 2006; Vankova, 2014) and therefore there is a difficulty in understanding how effective this activity type was in comparison to both no activity and other social or creative activities.

This section has provided an overview of studies which evaluated performance-based activities. The following section (4.5) presents a narrative descriptive synthesis of studies which assessed the impact of 'literary spoken and written word' activities.

4.5 Narrative descriptive synthesis of studies featuring 'literary spoken and written word' activities.

Studies within this section evaluated the arts type of 'literary spoken and written word' which included elements of reading or storytelling. Characteristics of the included studies for this section are shown in Table 4.5. As with the previous sections a narrative descriptive synthesis of included studies is presented, along with methodological quality appraisal.

4.5.1 Characteristics of included studies

A total of six studies (8% of the total number) assessed the impact of this arts type (Table 4.5). Published dates of the studies were between 2007 and 2014. All the studies were intervention studies, four were quantitative (Skrajner, 2007; Fritsch, 2009; Philips, 2010; Subramaniam, 2014) and two studies (one with two papers) were mixed methods (Billington, 2013; George, 2014; Houser, 2014). One study reported on a RCT (Subramaniam, 2014) with a further two conducting cluster-level RCTs (Fritsch, 2009; George, 2014; Houser, 2014). Philips (2010) carried out a quasi-RCT and Billington (2013) and Skrajner (2007) were uncontrolled pre/post measures studies.

4.5.2 Aims of included studies

All included studies featured more than one aim (range 2-5), three of these focused on assessing the impact of the activity on engagement levels of those taking part (Skrajner, 2007; Frtisch, 2009; Billington, 2013). Two studies (Skrajner, 2007; Fritsch, 2009) carried out direct observations of those participating to obtain quantitative findings whereas the third conducted interviews with staff members to provide qualitative information as to their perceptions of engagement shown by those taking part (Billington, 2013). Aims of the included studies are shown in Table 4.5.

Table 4.5 Characteristics of included studies featuring spoken and literary word activities

Study ID and type	Location	Participants	Activity Type	Aims	Outcome measures	Findings	Quality score
010. Subramaniam (2014) Intervention Dates of study not specified. Quantitative RCT	UK 14 care homes across North Wales (all but one privately owned) recruited, 11 completed.	23 in total, 11 in the intervention group and 12 in the control group. Intervention group: Range= NR* Mean= 84.5 years (SD 6.7). Control group: Range= NR Mean= 88.3 years (SD 6) Intervention group: 8 females (72.7%) and 3 males 27.3%). Control group: 8 females (66.6%) and 4 males (33.3%).	Life story books. 12 weeks in total. 2 groups, one which looked at participants who created their own life story books compared with a second group who had their books created by relatives. Both groups were presented with the completed books at the end of the programme. Programme delivered by a Clinical psychologist.	To assess the effect of a life review process compared with usual care. To compare the effects of a life story book process compared with books given as gifts. To assess the impact of the life story books on quality of care and the attitudes and knowledge of staff members.	Outcome measures: week 12 (following the programme) and week 18 (six weeks following completion of the programme). The following measures were used: Clinical Dementia Rating Scale (CDR); Quality of Life in Alzheimer's Disease (QoL-AD); Autobiographical Memory Interview- Extended Version (AMI-E); Geriatric Depression Scale (GDS) and Quality of Caregiving Relationship Questionnaire (QCPR). Relatives were also given the QCPR scale and the staff were administered the Approaches to Dementia Questionnaire (ADQ) and staff knowledge of care-recipient questionnaire.	QoL-AD: Life review baseline mean 30.1 (± 8.5), post-intervention mean 36.9 (±6.9). Gift giving group baseline mean 37.5 (±4.7), post-intervention mean 35.5 (±4.7). ANCOVA (F(1 20= 5.11, p=0.035). Six week follow up post-intervention mean life review group 36.1 (±7.8), gift group 38.6 (±3.8) ANCOVA F(1, 20)= 0.08, p=0.77. GDS-12R:Life review baseline mean 4.7 (±3.1), post-intervention mean 4.3 (±3.7), gift giving group baseline mean 2.6 (±1.4) and post intervention mean 2.5 (±1.8). ANCOVA F(1, 20)= 0.93, p=0.34. Six week follow up from post intervention mean life review group 3.5 (±2.7) and gift group 2,7 (±1.7), ANCOVA F(1, 20)= 0.14, p=0.71. AMI total: Life review baseline mean 34.4 (SD 22.0), post-intervention mean 44.5 (SD 28.5), gift group baseline mean 43.2 (19.1), post-intervention mean 34.7 (21.3). ANCOVA F(1,20)= 19.92, p= <0.001* Six week follow up from post-intervention mean life review group 42.0 (SD 23.5) and gift group 42.0 (SD 22.4). ANCOVA F(1,20)= 2.92, p= 0.10. **AMI-E:** *Personal Semantic Schedule Scores (PSS subscale*): Life review baseline mean 31 (SD 19.7), post-intervention mean 36.3 (SD 21.6), gift group baseline mean 36.7 (15.5), post-intervention mean 28.9 (18.3). ANCOVA F(1,20)= 14.01, p= 0.001* Six week follow up from post-intervention mean life review group 35.4 (SD 19.4) and gift group 33.3 (SD 16.6). ANCOVA F(1,20)= 3.98, p= 0.060. *Autobiographical Incident Schedule (AIS):* Life review baseline mean 3.4 (SD 2.8), post-intervention mean 8.2 (SD 8.2), gift group baseline mean 6.5 (4.4), post-intervention mean 5.8 (4.1). ANCOVA F(1,20)= 10.12, p= 0.005* Six week follow up from post-intervention mean life review group 6.6 (SD 5.4) and gift group 8.6 (SD 6.6). ANCOVA F(1,20)= 0.50, p= 0.49 QCPR (participant): Warmth: Life review baseline mean 32.4 (SD 1.0), post-intervention mean 32.3 (SD 2.3), gift group baseline mean 32.2 (1.0), post-intervention mean 31.2 (1.7). ANCOVA F(1,20)= 2.56, p= 0.13 Six week follow up from post-intervention mean life review group 33.5(SD 2.3) and gift group 31.6 (SD 2.1). ANCOVA F(1,20)= 4.51, p= 0.046. Conflict: Life review baseline mean 23.5 (SD 0.8), post-intervention mean 21.5 (SD 2.1.), gift group baseline mean 22.8 (1.7), post-intervention mean 22.3 (1.2). ANCOVA F(1,20)= 0.43, p= 0.52 Six week follow up from post-intervention mean life review group 22.0 (SD 2.1) and gift group 22.5 (SD 3.6). ANCOVA F(1,20)= 1.40, p= 0.25 QCPR (relative): Warmth: Life review baseline mean 34.3 (SD 3.9), post-intervention mean 35.2 (SD 3.7), gift group baseline mean 34.8 (4.6), post-intervention mean 34.5 (4.6). ANCOVA F(1,20)= 0.21, p= 0.65 Six week follow up from post-intervention mean life review group 37.5(SD 3.0) and gift group 37.9 (SD 2.6). ANCOVA F(1,20)= 0.08, p= 0.78. Conflict: Life review baseline mean 21.1 (SD 4.7), post-intervention mean 22.3 (SD 4.1.), gift group baseline mean 23.3 (3.3), post-intervention mean 24.3 (5.2). ANCOVA F(1,20)= 0.120, p= 0.73 Six week follow up from post-intervention mean life review group 26.8 (SD 4.1) and gift group 27.9 (SD 2.2). ANCOVA F(1,20)= 0.03, p= 0.87	GRADE score: + + + + + + Levels of evidence score L1, grade A.
017 Billington (2013) Descriptive Dates not specified Mixed methods Service evaluation.	UK- Wirral, Chester and Liverpool. 3 residential homes carried out the reading program themselves and two groups were carried out in a hospital and one at a day centre.	61 participants and 20 staff members. Ages and Gender NR. Loss to follow-up: 12 participants in total for the whole study (including those not in the care home group). 2 died, 2 left the care home, 1 was discharged from hospital, 1 had gone into a care home after the study began and no baseline data	Reading and storytelling. Get into Reading (GiR) is a literature based intervention. Short stories and/or poetry are read to participants and discussions are encouraged through open-ended questions. Sessions last for up to an hour. Delivered by 'project facilitators'.	To assess whether a shared-reading intervention could impact on behavioural symptoms in older people with dementia. To explore this with relation to what influence the programme had on older people, whether there were changes in dementia symptoms when participating in the programme and staff perceptions of the	Neuropsychiatric Inventory Questionnaire (NPI-Q) and a semi-structured interview schedule carried out with staff members. Collected at baseline and then every four weeks for the duration of the study period.	Average NPI-Q scores for care home 1= baseline 4.07, October (1) 0.45, November (2) 0.43, December (3) 0.0, January (4) 0.17, February (5) 0.17 and March (6) 0.0. Average NPI-Q scores for care home 2= Baseline 8.3, October (1) 8.3, November (2) 0.0, December (3) 0.0, January (4) 0.0, February (5), 0.0 and March (6) 0.0. Average NIP-Q scores for care home 3- Baseline 5.00, October (waiting 0.0), November (waiting) 0.01, December (waiting) 0.00, January (1) 0.0, February (2) 0.0 and March (3) 0.88. Care home 3 adopted a wait-list design which allowed for an ANOVA to be calculated. There were no statistically significant differences between the three phases (F (2, 6)= 0.630, p=0.051 with a sample size of 8. When baseline statistics (x- = 0.39, SD=6.13, n=13) were compared with the average monthly reading statistic (x-= 0.39, SD= 1.42, n=33) a significant effect was observed (F (1, 32= 24.74, p<0.005). Three themes identified for qualitative interviews carried out with staff members- 1. Components of the reading group intervention (literary content, duration, environment); 2. Enjoyment, authenticity and meaningfulness of the reading-group experience, including a renewed sense of identity and 3. Enhancement of listening, memory and attention skills.	GRADE score: ?-+- Levels of evidence III, GRADE B.

Study ID and type	Location	Participants	Activity Type	Aims	Outcome measures	Findings	Quality score
		was collected for 6 of the participants.		levels of engagement of the participants.			
031 George (2014)*tagged with Houser (2014). Intervention June-July 2011. Qualitative Cluster-Randomised pilot study.	USA Landis Homes- a not for profit care retirement community which has 2 nursing care units, one of which participated in this study.	20, 10 intervention group, 10 control group. All female. Age Range Intervention group= NR, Mean= 85.5 years (SD 6.33). 10 participated interviews only conducted with 8. Reasons for this not reported.	Creative storytelling. TimeSlips sessions carried out for one hour twice weekly for six weeks.	Assess the effects of TimeSlips on both residents and staff members within a special care unit.	Thematic analysis. 2 questions asked to residents and 4 asked to staff members. Following the activity.	Main theme- Benefits to Residents. Subthemes- increased creativity, improved quality of life, positively altered behaviour, involvement in meaningful activity. Main theme- Benefits for Staff Members. Subthemes- learning new practices, developing deeper understanding of residents, involvement in meaningful activity, thinking creatively around programmatic challenges. Main theme- Benefits for the Nursing Home Community. Subthemes- nurturing relationships, improved atmosphere.	QA: grade B. Level of evidence: II2, grade B.
032 Houser (2014)* Tagged with George (2014). Intervention June-July 2011. Quantitative Cluster-RCT	USA Landis Homes- a not for profit care retirement community which has 2 nursing care units, one of which participated in this study.	20, 10 in the intervention group, 10 control group Intervention group Range= 73-93 years Mean 85.5 years (no SD reported) Control group Range 68-98 years Mean 84.4 years (no SD reported). Intervention group= all female Control group= 5 females (50%) and 5 males (50%)	Creative Storytelling TimeSlips, a creative storytelling programme was delivered to participants twice a week for a period of six weeks. Sessions lasted for an hour.	To assess whether participation in a TimeSlips program was associated with improved mood and reduced behavioural symptoms in older people residing within a care home environment. Second to this, to assess whether there were any changes in psychotropic medication use as a result of the program.	Outcomes were measured using CareTracker, LH's direct-care data collection tool, which gathers input from caregivers three times daily for each resident. Psychotropic drug prescriptions and dosages were retrieved from LH's electronic medication record. Does not report frequency of this data being obtained, simply reports that data were analysed across the eight month period.	When comparing intervention with control groups with respect to the two primary outcomes of mood and behaviour, there were no statistically significant results. The Hodges-Lehman estimate (a negative sign favours the intervention,and a positive sign favours the control) for the mood score was -1.0, with a 95% confidence interval of (-10.0, 12.0). The Hodges-Lehman estimate for the behaviour score was 1.0, with a 95% confidence interval of (-4.0, 8.0). Out of the secondary outcomes, only the appearance score displayed a non-zero Hodges-Lehman estimate, which was -2.0 with a 95% confidence interval of (-7.0, 2.0). With regard to the psychotropic drug data, although there was some flux in dosages and number of prescriptions, no statistically significant differences were noted within or between groups.	GRADE score: + ? ? ? - + Level of evidence: II-2, grade B
035. Fritsch (2009) Intervention Dates not specified Quantitative Cluster-RCT	USA 20 not-for profit nursing homes, 10 in Milwaukee County, Wisconsin and 10 in Forsyth and Guilford counties, North Carolina. All had a special care dementia unit (SCU).	192 staff members completed surveys. Observations were carried out on 'those taking part in the TimeSlips program'. No further demographic information provided.	Creative storytelling. TimeSlips programme carried out in groups of 10-12 once a week for a total of ten weeks.	To explore the effect of TimeSlips, a creative storytelling intervention, on the 'engagement' of residents who lived within a nursing home. The study also explored whether staff interactions and attitudes improved as a result of the intervention.	Observations between staff and residents using the Quality of Interactions Schedule. Time-sampling methods were four, ten minute observations were carried out within an hour.	<u>Engagement</u> *Disengaged*= TimeSlips facilities- n= 68, ratio .04, control facilities- n= 107, ratio .09, X2 value 24.755, p<.001.*Non-social Engagement*= TimeSlips facilities- n=174, ratio .11, control facilities n=135,ratio .11, X2 value 0.051, p=.822 *Engagement*= TimeSlips facilities- n=1,400, ratio .85, control facilities- n=1,007,ratio .81, X2 value 9.039, p=.003. *Challenging Behaviour*= TimeSlips facilities- n=9, ratio .01, control facilities- n=1,ratio .00, X2 value 4.475, p=.034 <u>Affect</u> *Anger*= TimeSlips facilities- n=6, ratio .00, control facilities- n=1,ratio .00, X2 value 2.368, p=.124 *Fear/anxiety*= TimeSlips facilities- n=39, ratio .02, control facilities- n=11,ratio .01, X2 value 9.195, p=.002 *General alertness*= TimeSlips facilities- n=1,512, ratio .92, control facilities- n=1,111,ratio .89, X2 value 5.535, p=.019. *Other (neutral)*= TimeSlips facilities- n=30, ratio .02, control facilities- n=75,ratio .06, X2 value 35.791, p=.001. *Pleasure*= TimeSlips facilities- n=54, ratio .03, control facilities- n=47,ratio .04, X2 value 0.518, p=.472 *Sadness*= TimeSlips facilities- n=7, ratio .00, control facilities- n=0,ratio .00, X2 value 5.304, p=.0.21.	GRADE score: ? ? - + ? - Levels of evidence score LI, grade A.
044. Skrajner (2007) Intervention Dates not specified. Quantitative	USA 2 settings- 6 participants attended the sessions at an Adult Day Health Centre	6 participants were trained as 'leaders' of the programme. Mean age of the participants was 84.8. Range 75-93.	Creative storytelling RAMP programme- a copy of a story is handed out to a group of participants in an easy to read format. Following the leader	The aim of this study was to determine whether or not individuals in the middle stages of dementia could be trained to lead and	QAR leader assessment form- assess whether the leader effectively implemented the session. The Menorah Park Engagement Scale- an 11 item observational tool that assess	Significantly improved scores were shown for comparisons at Baseline 1 and Treatment for constructive engagement (1.09 vs 1.51, p<0.01) and pleasure (0.27 vs 0.45, p<0.01) and significantly decreased for 'other engagement' (0.70 vs. 0.27, p<0.001) and 'non engagement' (0.36 vs. 0.09, p<0.05). Significantly improved scores were also observed between Baseline 2 and Treatment for 'constructive engagement (0.91 vs. 1.51, p>0.001) and 'pleasure' (0.23 vs. 0.45, p<0.05).	GRADE score ? ? + + Levels of evidence score LII, grade B.

Study ID and type	Location	Participants	Activity Type	Aims	Outcome measures	Findings	Quality score
Uncontrolled pre/post-test	(ADHC- site #1) and 16 attended the sessions at the special care nursing unit (SCU- site #2).	6 participated from the ADHC and all these were female. Mean age was 75 (± 8.4). 16 participated from the SCU. Mean age was 89 (± 6.0). 15 were female (94%) and 1 male (6%).	each participant takes in turn reading their section of the story on the card. Following this, different coloured cards are handed out with facts or discussion on them and a discussion takes place around the cards. In this study, participants were trained to lead the intervention. Residents were trained to deliver the sessions.	deliver a reading activity for those with advanced dementia. The study also explored the satisfaction that those who were trained as leaders felt about their role and the levels of engagement and positive affect of those taking part in the study.	constructive engagement (CE), passive engagement (PE), non-engagement (NE) and other engagement (OE) with the task. Each participant were observed at baseline (baseline 1) for six sessions, then in six sessions during regular activity (baseline 2) programming and then between 6 and 10 sessions during RAMP activities (treatment). A brief satisfaction survey was also administered to those trained as leaders.	Scores significantly decreased for measures of 'other engagement' (0.59 vs 0.27, p<0.001) and 'non engagement' (0.61 vs 0.09, p<0.01) from Baseline 2 to treatment. No significant changes in passive engagement were observed between Baseline 1 and treatment and Baseline 2 and treatment. Increased engagement for both constructive engagement (1.09 vs. 1.51, F(1, 20)= 16.1, p<0.01) and 'other engagement' (0.7 vs. 0.27, F(1, 20)= 27.2, p<0.001) and a decrease in non-engagement (0.36 vs. 0.09, F(1, 20)= 7.6, p<0.05) from baseline to post-test For the QAR measure, for *passing out stories* all of the leaders showed partial adherence for 100% of the activities and all but one leader demonstrated full adherence in at least 80% of the sessions. When it came to *asking someone to read the next section* a total of two leaders showed full adherence for 80% of the sessions and all leaders demonstrated partial adherence in 100% of the activities. Finally, for the *initiating discussion* section, five of the six leaders demonstrated partial adherence in 80% of the sessions and three demonstrated full adherence in at least 80% of the sessions.	
173 Philips (2010) Intervention December 2008-July 2009. Quantitative Quasi-RCT	USA Four nursing homes and two assisted living facilities within central Missouri.	Intervention group (n= 28). Mean age 83.4 years (± 8.0). 23 were female (82%) and 5 were male (18%). Control group (n= 28). Mean age 85.8 (± 6.8). 26 (93%) were female and 2 (7%) were male.	Creative storytelling TimeSlips- an established group story telling programme. This was delivered to the intervention group for a total of six weeks, each session lasted for one hour and group size varied between 6 and 12 participants. Sessions were delivered by the principle investigator and a research nurse, both of whom were trained in delivering the TimeSlips programme.	To assess the role of TimeSlips in relation to communication skills, quality of life, neuropsychiatric symptoms and observed emotions in participants with dementia living within a long-term care facility. It was suggested that TimeSlips would improve cognition and psychological wellbeing.	Observation to record emotion at week 0 (baseline), 1 week post-intervention (Week 7) and 4-weeks post intervention (week 10) for both groups and weeks 3 and 6 for the intervention group. This was measured using the Observed Emotion Rating Scale (OERS). Modified Cumulative Illness Rating Scale. Mini Mental State Examination (MMSE). Cornell Scale for Depression in Dementia (CSDD). Neuropsychiatric Inventory-Nursing Home Version (NPI-NH) Quality of Life in Alzheimer's Disease (QOL-AD) Functional Assessment of Communication Skills (FACS)	A significant group X time interaction was observed for the measures of NPI-NH, social communication and OERS (p<0.05, p<0.001 and p<0.001 respectively). After adjusting for baseline and MMSE scores, OERS pleasure scores were significantly higher at week 3 (p<0.001), week 6 (p<0.001) and weeks 7 (p<0.05) for the intervention group than the control group. Social communication scores were higher for the intervention group at week 7, however not significantly so (p not reported). Week 10 scores were better for the control group compared with the intervention group, but again not significantly so. There was no significant effect for the CSDD scores. Following adjustment for baseline and CIRS scores, however these were significantly lower for the control group at week 10.	GRADE score ? ? + + Levels of evidence score LII-a, grade B.

Three studies featured the aim of assessing the impact of the activity on psychological wellbeing, with two referring to 'mood' (Subramaniam, 2014; Houser, 2014) and one to 'observed emotion' (Philips, 2010). Emotional wellbeing was also qualitatively explored via interviews within one of the mixed-methods studies in relation to both participants and staff members (George, 2014). Improvements to general quality of life were reported within two studies (Philips, 2010; Subramaniam, 2014), both using the same validated questionnaire for this measure. A further two studies aimed to determine whether there was any effect of the activity on neuropsychiatric symptoms, one via questionnaire (Philips, 2010) and the second analysing psychotropic medication use (Houser, 2014). Both mixed methods studies explored whether there were any benefits on behavioural symptoms obtaining both quantitative and qualitative information (Billington, 2013; George, 2014; Houser, 2014).

Two studies aimed to provide an assessment as to whether staff members' knowledge and attitudes of residents was impacted by the activity (Fritsch, 2009; Subramaniam, 2014) and another aimed to determine whether the activity had any effect on communication levels of residents (Philips, 2010). The only study which aimed to assess memory via was carried out by Subramaniam (2014) which may be due to this study evaluating the creation of autobiographical memory books.

Skrajner (2007) trained people with dementia to become leaders of the activity, the only study within the whole of this review (of any arts type) to do so. An assessment via questionnaire was also provided as to how well they implemented the programme and the satisfaction they felt with their new role as a leader. Finally, Subramaniam (2014) featured a group of residents who created their own life story books and a comparison group who received the books as gifts after they had been completed by the families of residents with the aim of assessing

whether taking part in the process of creating the books was beneficial but also whether receipt of the books could promote a positive effect.

4.5.3 Location and setting

Studies originated from just two countries, four (67%) from the USA (Skrajner, 2007; Fristch 2009; Philips, 2010; George, 2014; Houser, 2014) and two (33%) from the UK (Billington, 2013; Subramaniam, 2014). Total number of care homes included in the studies was 40 (range 1-20), with additional data provided by participants from an Adult Day Health Centre (ADHC) by Skrajner (2007) and Billington (2007) and two assisted living facilities in another study (Philips, 2010). Data from participants from two hospital wards was also included in Billington (2013). All but one of the studies (George, 2014; Houser, 2014) featured participants from more than one location (Table 4.5).

The term 'nursing home' was reported in three studies (Skrajner, 2007; Fritsch, 2009; Philips, 2010), one referred to 'care homes' (Subramaniam, 2014), one 'residential units' (Billington, 2013) and the remaining study reported that residents were from a 'care retirement community' (George, 2014; Houser, 2014). Two of the studies, both originating in the USA, specified the included care homes to be not-for-profit facilities (Fritsch, 2009; George 2014; Houser 2014). In contrast Subramaniam (2014) reported of the 14 care homes recruited to the study from within the UK (eleven completed and were included) only one was not privately owned.

The setting where the arts activity was carried out is not fully described for most studies, only one reported that the sessions took place within a private meeting room within one of the care homes and in a shared living space for two of the homes (Philips, 2010), however this study reported collecting data from four homes and so it is unclear where the sessions took

place within the final home. Given that four other studies (Skrajner, 2007; Fristch, 2009; Billington, 2013; Subramaniam, 2014) featured more than one location it is presumed that the sessions took place within each individual home, however this is not explicitly reported, and it is unclear whether the same research team carried out the activity at the homes or whether there was consistency across programmes.

4.5.4 Populations

Data were obtained from a total of 188 older people across the five studies which reported group numbers, giving a mean sample number of 36.4 (SD, 18.5 range 23-61) (Table 4.5). Fritsch (2009) did not report how many people were observed during the sessions, simply stating that observations were made for 'all those taking part'. When taking away control groups of studies which utilised an RCT or cluster-RCT design this leaves a total of 132 residents who provided data after taking part in the activity. Additional data were obtained from a total of 280 care staff members across three studies, 20 in Billington (2013), 68 in Subramaniam (2014) and 192 in Fritsch (2009). Twenty-three relatives also completed questionnaires for Subramaniam (2014).

Mean age was provided by four studies, Fritsch (2009) did not provide any sample information for those taking part and Billington (2013) did not provide demographic information for the 61 residents included in the study. All reported mean ages were above 80 years, with a total mean of 85.7 years (SD 2.4, mean range 82-86.9 years). For the three studies, which featured an intervention and control group (Philips, 2010; Georger 2014; Houser, 2014; Subramaniam, 2014), the mean age of those taking part in the activity (intervention group) was 84.1 (SD 0.25, range 83.4-84.5 years) and for those in the control group was 86.5 years (SD 1.6, range 85.5-88.3 years). Skrajner (2007) featured two groups of residents, those who were chosen to lead the programme and those who took part, with those trained to be leaders (n=6)

showing a slightly lower mean age at 84.4 years (range 75-93 years) compared with the mean age of those taking part from the special care unit (n=16) who had a mean age of 89 (SD 6.0), however those taking part from the adult day centre (n=6) had the lowest mean age at only 75 years (SD 8.4), albeit with a larger deviation.

The same four studies that provided information on the age of participants also provided information as to how many men and women there were within their samples, all four featured more females than males. In total, there were 101 (83%) women and 20 (17%) men. The greatest proportion of females was represented in Skrajner (2007), all six of those trained to be leaders were female and of those taking part in the sessions, all but one (n=15, 94%) were women.

For the studies featuring a control and intervention group the numbers and proportions of men and women within each group were similar. Subramaniam (2014) reported both the intervention and control groups to have eight women (73% and 67% respectively) with three and four men (27% and 33%). Philips (2010) reported 23 women in the intervention group (82%) and 26 in the control group (93%) with five and two males in intervention and control group respectively. The biggest difference in distribution of men and women was shown by George (2014) and Houser (2014) who reported an intervention group size of ten, all of whom were women, however their control group had a 50% split, with five women and five men. It is not explained why this difference was present.

Five studies specified participants required a diagnosis of dementia to be able to participate. The remaining study (Fritsch, 2009) makes preference to people with dementia (PwD) within their justification of their storytelling programme, however no inclusion criteria or demographic information is reported. Residents taking part in Skrajner (2007) were described

as at varying stages of disease with 'some having moderate to advanced dementia' but specifics are not reported. The study did state half of those trained as leaders (n=3) did not have a specific dementia diagnosis, however it is unclear whether this means they were simply recorded as having 'dementia' without identification of a specific condition (i.e. Alzheimer's Disease) or did not have a documented dementia diagnosis at all. Either way, Mini Mental State Examination (MMSE) scores for this group were well below the cut-from 23 (average of 15.3, SD 6.1) which was also specified as an inclusion criterion and matched with those participating in the activity (15.9 average, SD 7.3).

Conversely Subramaniam (2014) excluded residents from taking part if they lacked the 'mental capacity to consent' which may have disadvantaged those in advanced stages of dementia from taking part. Three studies (Skrajner, 2007; Philips, 2010; George, 2014; Houser, 2014) did not require residents to give consent, obtaining consent from a relative or person with Power of Attorney, meaning participants were not excluded on this basis. The requirement of good 'mental capacity' in Subramaniam (2014) may in part be due to the number of self-report measures used within the study (four different questionnaires) and therefore a requirement for individuals to be able to complete such scales.

4.5.5 Characteristics of the arts activities

Three studies focused on storytelling (autobiographical or creative) (Fritsch, 2009; Philips, 2010; Houser, 2014; George, 2014; Subramaniam, 2014) and those which promote the idea of using reading to facilitate discussion (Skrajner, 2009; Billington, 2013).

Three studies (Fritsch, 2009; Philips, 2010; Houser, 2014; George, 2014) evaluated the use of TimeSlips, a creative storytelling programme developed within the USA. Philips (2010) and Houser (2014- tagged with George 2014) carried out the programme for a total of six weeks,

however the activity was only carried out once a week in Philips (2010) and twice weekly for the remaining study (George, 2014; Houser, 2014). Fritsch (2009) reported carrying out the TimeSlips programme weekly, for a total of 10 weeks. All studies described similar methods, which is not surprising given it is an existing programme with guidelines for how it should be delivered. Only Fritsch (2009) did not report on the length of the sessions, with the other studies all reporting duration of the sessions was one hour (Philips, 2010; Houser, 2014; George 2014). Group size of this activity were only recorded by two studies, reporting ranges of 10-12 (Fritsch, 2009) and 6-12 (Philips, 2010).

Subramaniam (2014) also evaluated storytelling, in contrast to the previously mentioned studies these focused on autobiographical memory to produce life story books (Table 4.4). Two groups were included, the first created their own story books during one hour sessions with an average of 12 sessions (range 11-16). Participants were encouraged to recall memories from childhood to the present day which were woven into a book to keep. The second group received usual care and family members provided the information for the life story book fortnightly across 12 weeks, again the completed book was given to the residents to keep.

Two studies therefore focused on reading and discussion (Table 4.5). Unlike the previous studies focused upon storytelling, these activities both gave participants excerpts from stories and poems which were read in turn by the members of the group. Discussion was then prompted using open-ended questions (Billington, 2013) or prompts (Skrajner, 2009). Two of the studies therefore focused on reading and discussion. Both studies provided evaluations of structured programmes, Billington (2013) focused on the Get into Reading (GiR) programme and Skrajner (2009) involved analysis of Resident-Assisted Montessori

Programming 'RAMP'. The latter study was described as a 'storytelling' activity, but mainly involved taking turns reading text from a story, rather than creative storytelling as the TimeSlips studies (Fritsch, 2009; Philips, 2010; Houser, 2014; George, 2014) or autobiographical storytelling (Sumbramiam, 2014). Following reading of this text a discussion around the themes of the story or poem commenced. Session duration of the GiR programme (Billington, 2013) lasted for up to one hour, the study period was six months (month one was used as baseline measure), however frequency of the sessions was not reported. Skrajner (2009) did not provide information on frequency of the sessions and only reported that the duration 'tended to be between 25 and 40 minutes'.

Sessions were delivered by the Principal Investigator in two studies (Philips, 2010; George, 2014; Houser, 2014) and the lead author (a Clinical Psychologist) in another (Subramaniam, 2014). Fritsch (2009) reported the programme was delivered by 'nurse aids, social workers and/or activity directors' who had attended a nine-week training programme to carry out the activity. It is unclear what frequency each of these individuals delivered the sessions, nor whether they delivered them individually or as part of a group. The Get into Reading programme was already established within NHS facilities in Billington (2013) and the programme was described as carried out by 'project facilitators' located within the study sites. Finally, Skrajner (2007) trained people with dementia (PwD) to lead the programme, assessing their level of adherence with training and method of delivery.

Three studies, all of which were RCTs, featured a control condition. Neither of the cluster level RCT's (Fritsch, 2009; George, 2014) described whether differences in care were offered between facilities, nor whether the control group took part in any other form of activity. Finally, Subramaniam (2014) compared the results of those who created life story books with

those who had the story books completed by their relatives and received the completed books as gifts to determine whether the process of creating the books yielded any benefits not shown for the gift giving group.

4.5.6 Research methods and outcome measures

Three studies carried out observations (Table 4.5), one assessing levels of engagement (Skrajner, 2007), one emotional expression (Philips, 2010) and the final study measured both domains (Fritsch, 2009). For assessing engagement, Skrajner (2007) used the Menorah Park Engagement Scale which records incidences of constructive engagement, passive engagement, non-engagement and 'other' engagement. Fritsch (2009) devised their own observation schedule which recorded four different types of engagement: social, non-social, disengaged inactive and challenging behaviour. For assessment of emotional expression, the Observed Emotion Rating Scale (OERS) was used by Philips (2010) focusing on dimensions of pleasure and general alertness. Emotional expression was assessed by Fritsch (2009) using the Geriatric Centre Affect Rating Scale which records levels of pleasure, anger, anxiety or fear, sadness, general alertness and neutral emotion. In addition to these measures Fritsch (2009) also observed interactions that occurred between people with dementia and staff members via the Quality of Interactions Schedule (QIS). In addition to engagement, Skrajner (2007) assessed delivery of the programme via the QAR leader assessment form to determine whether the trained leaders (other residents) adhered to programme delivery.

Observation periods varied across the three studies (Table 4.5). Skranjer (2007) observed residents for 5 minutes per session for six sessions where the activity was carried out and six when normal activity programming was carried out. Philips (2010) observed for ten-minute periods at pre and post-test (week seven, one week after the six-week programme ended) and at three week follow up (week 10). Observed emotion was also assessed mid-way (week

three) and during the last week (week six). Fritsch (2009) used time-sampling methods were four, ten minute observations were carried out within an hour. Four of these hour long observation periods were carried out a day across two weeks. It is unclear which two weeks. Observations were carried out by research assistants (Fritsch, 2009) and members of the research team (Skrajner, 2007; Philips, 2010), with some data being obtained by trained research nurses in Philips (2010).

Symptoms associated with dementia were assessed in three of the studies (Philips, 2010; Billington, 2013; Houser, 2014). Two used the Neuropsychiatric Inventory Questionnaire (NPI-Q) (Philips, 2010; Billington, 2013), with the first of these using the Nursing Home edition of the measure (NPI-NH), in each case these were completed by staff members located within the home. Houser (2014) used patient notes to obtain incidences of psychotropic prescriptions during the study duration.

The Geriatric Depression Scale (GDS) (Subramaniam, 2014) and Cornell Scale for Depression (Philips, 2010) were used for two studies assessing depression, completed by the participants. The same two studies also assessed quality of life, both via the Quality of Life in Alzheimer's Disease questionnaire (QoL-AD), again completed by those taking part with staff also completing this in Philips (2010). Both these studies assessed participants at baseline, following the intervention and then at a follow-up period of 6 weeks' (Subramaniam, 2014) and 4 weeks for (Philips, 2010). Cognitive changes were assessed in Subramaniam (2014) via the Autobiographical Memory Interview Extended Version (AMI-E) (at baseline, immediately following and follow-up). One of the studies assessed the satisfaction of those trained to deliver the programme (Skrajner, 2007) via usage of a non-validated satisfaction survey (Table 4.5).

Houser (2014) was the only study to extract data from notes kept on residents. They used CareTracker, the home's data collection tool, providing information inputted from caregivers three times a day. Data were obtained in relation to mood and behaviour for four months prior to the implementation of timeslips and a four-month period which featured the six-week programme and then compared pre- and post.

In addition to the completion of some of the measures above staff members provided information for other outcomes. Staff were administered the Approaches to Dementia Questionnaire (ADQ) and Quality of Caregiving Relationship Questionnaires (QCPR) to assess whether their knowledge and attitudes of residents had been improved by the activity for the study which explored the use of life story books (Subramaniam, 2014). In Philips (2010) staff completed the Functional Assessment of Communication Skills (FACS) measure to provide information on improvements in communication in residents at baseline, following the programme and follow-up.

Both mixed methods studies (Billington, 2013; George, 2014; Houser, 2014) conducted interviews with staff members to ascertain benefits they had identified with taking part in the activity, both in terms of participants and themselves. Only one of the studies (George, 2014) carried out interviews with residents themselves, asking two questions about their enjoyment of the activity.

4.5.7 Quantitative findings

4.5.7.1 Behaviour

Billington (2013) featured three care homes, with descriptive data presented for two and statistical analysis carried out only on the third. This was explained by the home being the only one to adopt a wait-list design, however a comparison of pre and post-means could have been calculated by the author for the other two. Both studies with descriptive data showed a

trend with NPI-Q scores improving, but this was not found with the third home where statistical comparisons were carried out (Table 4.5). The authors comment how post-hoc comparison of the baseline scores with the average monthly reading statistic for all homes was significant (F (1, 32)= 24.74, p<0.005) suggesting a larger sample size may have yielded a significant effect.

Individuals in facilities carrying out TimeSlips programmes showed higher levels of engagement than those in facilities not carrying out the activity (total n= 1,400 vs. n=1,007, X^2= 9.04, p=0.003) and lower levels of disengagement (n= 68 vs. n=107, X^2=24.76, p<0.01). However, those from the TimeSlips facilities also showed increased levels of challenging behaviour in comparison to the homes not offering the activity (n=9 vs. n=1, X^2= 4.48, p=0.034). These findings were obtained through observational methods and it is not clear how behaviours was assessed (Firtsch, 2009) (Table 4.5). Skrajner (2007) also assessed engagement levels, showing increased engagement for both constructive engagement (1.09 vs. 1.51, F(1, 20)= 16.1, p<0.01) and 'other engagement' (0.7 vs. 0.27, F(1, 20)= 27.2, p<0.001) and a decrease in non-engagement (0.36 vs. 0.09, F(1, 20)= 7.6, p<0.05) from baseline to post-test. AT post-test 2 when the residents with dementia were leading the programme, constructive engagement increased (0.91 vs. 1.51, F(1, 20)= 28.8, p<0.001), and non-engagement (0.61 vs. 0.09, F(1, 20)= 13.3, p<0.01) decreased indicating residents were just as engaged with the programme when it was delivered by the other residents than they were when delivery was by a professional. Adherence to delivery of the programme was also reportedly high (Table 4.5).

4.5.7.2 Psychological wellbeing

There was no significant impact on depression assessed within either Subramaniam (2014) or Philips (2010). There was also no significant difference between the intervention and control

group for mood measures in Houser (2014). There were also no changes in psychotropic medication usage (Houser, 2014).

Fristch (2007) also assessed levels of affect following the activity for the homes offering TimeSlips and compared these to homes that did not offer the activity. He reported significantly greater levels of alertness (occurred 1512 times vs. 1,111 times, X^2= 5.54, p=0.002). However, he also showed greater levels of anxiety/fear (occurred 39 times vs. 11 times, X^2= 9.2, p=0.002 and sadness (occurred 7 times vs. 0 times, X^2= 5.54, p=0.002) within the activity homes. This may show greater emotional expression in those taking part and a reduction in apathy which is further demonstrated by the finding that neutral expression was significantly less observed in those taking part (occurred 30 times vs. 75 times, X^2= 35.79, p=0.001). However, analysis was not carried out across time and there is little information as to what activities were offered by the homes within the control groups.

4.5.7.3 Cognition, quality of life and socialisation

Quality of life was improved at the end of the study duration for those creating life story books and was slightly decreased for those in the gift-giving group (37.5 vs. 35.5) (F(1,20)= 5.11, p=0.035), however there was no significant effects for quality of life at six-week follow-up (Subramaniam, 2014). The second study assessing quality of life (Philips, 2010) showed no significant impact of the activity, rated by both residents and staff members (Table 4.5).

Improvements in memory were observed within Subramaniam (2014), who showed increased scores for the AMI-E in total for the life review group (34.4 (SD 22) vs. 44.5 (SD 28.5)) and a decrease for the gift-giving group (43.2 (SD 19.1) vs. 334.7 (SD 21.3)) at post-test (F(1, 20)= 19.92, p<0.001) Individual sub-scores were also significantly improved (individual values included in Table 4.5). There were no significant changes at six-week follow up for total scale

or subscales, suggesting the impact on memory was only across the duration of the programme.

Two studies assessed the impact of this activity type on social communication, with neither showing any statistically significant changes following the study duration (Philips, 2010; Houser, 2014) (Table 4.5).

4.5.7.4 Improvements to care

There were no significant changes with respect of the quality of caring relationship assessed in Subramaniam (2014). Neither the QCPR completed by the participant nor relative showed any changes prior or following the activity, nor at six-week follow-up. The approaches to dementia questionnaire significantly improved from pre-test to follow-up (73.54 vs. 80.07, $F(2, 73)= 14.31$, $p<0.001$). Staff knowledge of dementia was also shown to improve by significantly less 'don't know' responses both at post-test (5.41 vs. 4.12, $F(2, 115)= 31.65$, $p=0.025$) and six-week follow-up (5.41 vs. 1.78, $F(2, 115)= 31.65$, $p=0.001$). Similarly, the amount of 'incorrect' answers was decreased at six-week follow-up ($F(2, 119)= 24.88$, $p=0.001$). Fristch (2009) also demonstrated increased positive views expressed by care staff in reference to residents ($\beta=3.43$, $p<0.001$). They were also less likely to devalue residents, but not to a significant degree (Table 4.5).

Fritsch (2009) described how different interactions were observed in those from the TimeSlips care homes and the comparison homes. Those from the TimeSlips homes showed greater eye contact ($X^2= 24.27$, $p<0.001$), use of touch ($X^2= 8.7$, $p<0.01$) and verbal interactions ($X^2= 13.85$, $p<0.001$). Despite this, those in the control facilities displayed significantly increased levels of care related touch ($X^2= 8.35$, $p<0.001$ and care-verbal interactions ($X^2= 13.85$, $p<0.01$, the

differences of which may correspond to those in the TimeSlips homes interacting with residents on topics not just related to the care they were receiving.

4.5.8 Qualitative findings

Both qualitative studies primarily obtained data from staff members, with George (2014) asking two questions to residents also. Findings from George (2014) derived three main themes: benefits to residents, benefits to staff members and benefits to the nursing home community. In terms of benefits to residents the subthemes of creativity, quality of life, behavioural alterations and involvement in meaningful activity were reported. Meaningful activity and creativity were also reported as benefits to staff members, along with learning new practices, developing deeper understanding of residents and thinking about challenges creatively. Benefits to the community were nurturing relationships within the home and improving the atmosphere. Benefits outlined by Billington (2013) were less-specific and included practical components of the reading group, enjoyment and meaningfulness and enhancement to listening, memory and attention skills (Table 4.5).

4.5.9 Quality appraisal of included studies

4.5.9.1 Levels of evidence

Three of the studies provided evidence at level I (Fristch, 2009; George, 2014; Subramaniam, 2014), evidence obtained from RCTs and the highest level eligible for inclusion within this review. One provided evidence at level IIa (Philips, 2010) (reflecting moderate quality evidence) and two at level IIb (Skrajner, 2007; Billington, 2013) (again, moderate evidence of a lower quality that level IIa) (Table 4.5).

4.5.9.2 Quantitative studies

Figure 4.6 shows the judgements made on the methodological quality of the included studies. Skrajner (2007), Philips (2010) and Billington (2013) were uncontrolled pre/post studies and as such were not assigned values for random sequence generation or allocation concealment.

In the three RCTs, random allocation was adequately implemented for two studies with Houser (2014) reporting the use of a computer programme to generate the list of homes taking part in the intervention and Subramaniam (2014) using sequential individual-based randomisation which was carried out by a recognised Clinical Trials Unit (The North Wales Organisation for Randomised Trials in Health & Social Care). Fritsch (2009) reported that randomisation occurred, however did not provide any information as to how the random sequence was generated. In terms of allocation concealment, only Subramaniam (2014) adequately explained how allocation was conducted by the trials unit via a web-based system and therefore eliminated the possibility of allocation being guessed. It was probable that allocation concealment occurred, at least at a cluster-level for Houser (2014) however not enough information is provided to make a judgement of low risk. Fritsch (2009) again failed to provide information on how allocation was concealed, however homes within this study were paired based on their characteristics (number of beds, number of residents etc.) which would mean allocation concealment could not be effectively implemented.

Blinding of the person(s) carrying out the outcome assessment was explicitly stated only by Subramaniam (2014) who reported that whilst researchers collected the baseline data, intervention data (weeks 12 and 18) collection and analysis were carried out by assessors who were blind to the participants' allocation. Three studies did not report enough information to make a judgement as to whether analysis of the outcome measures was performed by people blinded to allocation of participants, it would be difficult to blind for these studies as they were uncontrolled, particularly if data were collected at different points and so it would be acceptable to assume that this did not take place (Skrajner, 2007; Philips 2010; Billington, 2013). Houser (2014) explicitly stated that those collecting the data were not blind as to the treatment condition of the participants, however there was no information as to whether the

person carrying out analysis of the data were blinded or not, although again it is likely that blinding did not occur. Finally, Fritsch (2009) reported that the researchers carried out the observations and were involved in the coding and analysis of them and so blinding did not occur.

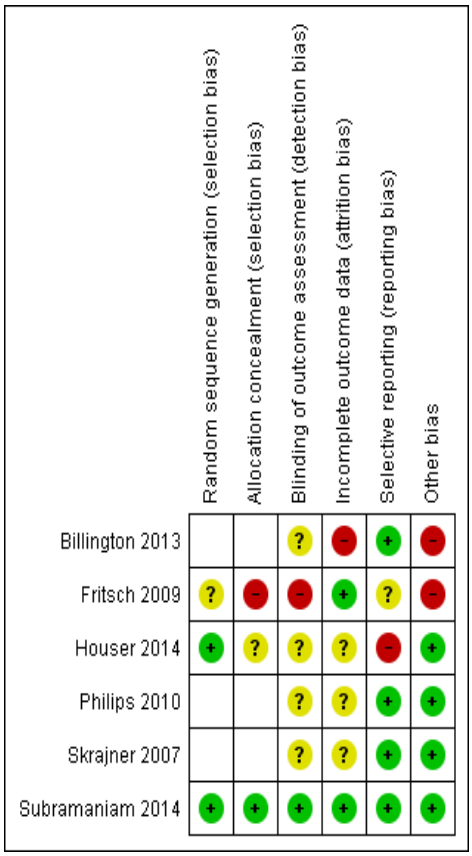


Figure 4.6 Judgements of methodological quality for the arts type of 'literary spoken and written word'. Where green corresponds to a low risk of bias, yellow to an unclear risk and red to a high risk

In terms of incomplete outcome data, both Submanamiam (2014) and Fritsch (2009) reported no data had been removed from the analysis and fully reported results. Only Billington (2013) received a judgement of 'high risk' for this domain, where there was no information at all as to participant numbers within the results section and therefore it is unclear where drop-outs occurred (at which point during the study), from which care home (whether the drop outs evenly split, were they all from one home). Skrajner (2007) did not address the issue of drop outs at all and so it is unclear whether they occurred and what impact they would have. For

Philips (2010) data were presented at baseline for all the participants, regardless of whether they completed the study, however it is unclear whether withdrawal data were utilised at each timepoint and so without this information a judgement cannot be made either way. Finally, Houser (2014) did report there were missing values, however it is not clear whether these were accounted for within the results section or not.

Skrajner (2007), Philips (2010), Billington (2013) and Subramaniam (2014) all contained enough information about the study's outcomes, outcome measures, data collection times and methods to enable a judgement of low risk. One study (Fritsch, 2009) did not give enough information within the study protocol to make a judgement either way and Houser (2014) reported an aim to explore as and when (P.R.N) medication usage, however results for this measure are not given in the results section only a sentence stating there was 'flux in medication use' meaning this was not reported upon adequately.

Two studies were identified as having additional biases which may have influenced the outcome of the study (Fritsch, 2009; Billington, 2013). The first, Fritsch (2009), identified within the results section they opted to use time sampling methods, however this made it difficult to focus observations on the participants meaning they had to present interactions only rather than individual resident-specific data. Billington (2013) presented means for all the care homes that took part within the study (three in total), recording pre-and post-measures for each, but state they only carried out statistical analysis on Care home 3 (which only accounted for nine participants) as this home used a wait-list design meaning an ANOVA was not possible and a descriptive account was given for the other two. However, as means were presented for all the homes, differences between pre-and post-measures could have

been calculated for the other two even if an ANOVA could not be performed given baseline data were collected.

4.5.9.3 Qualitative studies

Qualitative quality appraisal was carried out for the paper which described the qualitative results from the mixed methods study (George, 2014). The study was awarded a Grade B indicating that it possessed some flaws, but these were unlikely to impact upon the transferability, credibility and or overall findings of the study. The study adequately reported upon the purpose, scope and aims of the study and in relation to study design there was relevant information on data collection and evidence of triangulation in respect off the quantitative results from the study being referred to. The only factor which was not met was justification of the use of thematic analysis. Selection criteria was also evident, however there was no sample size justification, a pool of 26 residents met inclusion criteria but only 20 were chosen to participation and it is not clear as to why this figure were chosen. Analysis of data was also well evidenced and there was information as to how the authors had interpreted the data, but they did fail to report a description of the contexts within which data collection occurred. There was also a lack of reflexivity in the paper, with no evidence of this within the results or discussion section. Ethical approval was evidenced, and all aspect of transferability and relevance were met.

4.5.10 Summary of literary spoken and written word activities.

Six studies were placed within this category, all of which originated in the USA and the UK. As with previous activities, there was mixed evidence of benefit for this arts type. Measures of psychological wellbeing, namely depression, failed to show an impact and there was no observed impact on social communication.

There was some evidence for increased engagement with the activity (Skrajner, 2007; Fritsch, 2009) but this was also coupled with increased challenging behaviours (Fritsch, 2009). There was also greater expression of negative affect, however there was also a reduction in neutral affect (Fritsch, 2009), indicative of increased emotional expression. Similarly, studies that evaluated the impact on dementia symptoms also failed to show a significant impact (Philips, 2010; Billington, 2013; Houser, 2014) which may again evidence increased self-expression that could be viewed as disruptive.

Particularly unique to this arts type was the evaluation of how the activity could improve staff performance within the home. Creation of life story books enabled staff members to increase their knowledge of dementia (Sumbramaniam, 2014) and the types of interactions offered along with the TimeSlips programme were less focused on physical care and more on an emotional connection with residents (Fritsch, 2009).

Quality of the included studies was reasonable when evaluating this arts type. One of the RCTs (Subramaniam, 2014) received a judgement of low risk for all domains, with a second receiving only one judgement of high-risk (Houser, 2014). An earlier conducted RCT fared less well (Fritsch, 2009), which could indicate improvements in quality with more recent studies. Selective reporting and other biases also occurred infrequently within the included studies. Finally, the only qualitative study received a judgement of grade B, indicating a high level of confidence can be placed in its findings (George, 2014).

This section has provided a narrative descriptive synthesis of the studies featuring the arts type ‘literary spoken and written word’, along with methodological quality appraisal. The following and final section will present the studies within the category ‘applied arts and crafts’.

4.6 Narrative descriptive synthesis of studies featuring 'applied arts and crafts'

This section focuses upon studies assessing the impact of applied arts and craft activities. The smallest number of studies were placed within this category (n=3). As with the previous sections a narrative descriptive synthesis of included studies is presented, along with judgements of methodological quality. Characteristics of the included studies can be seen in Table 4.6.

4.6.1 Characteristics of included studies

A total of three studies (4% of the 71 included within this whole review, see Table 4.6 for included studies and characteristics) evaluated this arts type and dates ranged from 1997-2011, (Doric-Henry, 1997; La Cour, 2005; De Guzman, 2011). Only one included actual study dates and this took place over twenty years ago, between February and April, 1995 (Doric-Henry, 1997). Two studies were descriptive qualitative studies (La Cour, 2005; De Guzman, 2013) and the third a quantitative intervention study (Doric-Henry, 1997) which adopted a quasi-RCT study design. Doric-Henry (1997) described the study as quantitative and qualitative, however no qualitative data were obtained and analysed, two case studies are presented only.

4.6.2 Aims of included studies

The aim of the intervention study was to assess whether there was any effect of taking part in a pottery activity on levels of psychological functioning, namely self-esteem, depression and anxiety (Doric-Henry, 1997). La Cour (2005) sought to explore 'components of engagement' when taking part in creativity activities, identified via interviews both with those taking part and those delivering the sessions. The final study (de Guzman, 2013) stated the aim was simply to determine the potential for traditional Filipino arts to be used as a therapy with older people living within a care home (Table 4.6).

Table 4.6 Characteristics of included studies featuring applied arts and crafts activities

Study ID and type	Location	Participants	Activity Type	Aims	Outcome measures	Findings	Quality score
097. Doric-Henry (1997) Intervention February 1995-April 1995. Quantitative Intervention group pre-test/post-test with comparison group.	USA Evangelical Home of Saline (church based residential nursing home in Michigan)	40 in total; 20 in the intervention and 20 in the control. Intervention (n=20)- 19 female (95%) and 1 male (5%). Mean age 83.5 (range 50-95). Control (n=20)- 16 female (80%) and 4 male (20%). Mean age 85.9 (range 76-99).	Pottery Sessions were held once a week for eight weeks for an hour. A number of different ceramic processes including wedging, throwing, drying, trimming, bisque-firing, glazing and glaze firing which enabled them to complete a piece of pottery by the end of the course. Sessions were delivered by the researchers.	To explore the role of a pottery intervention on levels of anxiety, self-esteem and depression levels in older adults.	Coopersmith Self-Esteem Inventory. Beck Depression Inventory. State-Trait Anxiety Inventory. Descriptive observational records were also kept throughout the intervention period and a subjective questionnaire about their self-esteem, depression and anxiety was also completed. Outcome measures were administered before and following the course.	The pottery group showed significantly lower depression scores following the intervention than before (mean 4.6 (SD 5.5) pre and 2.1 (SD 2.5) post), alpha level .05. The pottery group also showed significantly improved self-esteem scores following the intervention (72 (SD 13.9) pre and 81.6 (SD 8.4) post), alpha level .05. Trait and State anxiety scores also reduced post-test (28.7 (SD 7.7) vs. 22.6 (SD 5.8)). No changes on any of the outcome measures were observed for the comparison group. Matched pair t-tests showed that those who began the intervention with higher self-esteem scores (mean of 75.2) did not show a significant improvement in their self-esteem scores following the intervention, however those with lower mean scores (60.8) pre-test did (alpha level .05). This was also found for depression scores those with low (mean 1.0) did not show improvements but those with high pre-test scores (mean 8.2) did (alpha level .05). Those with high state (mean 34.3) and trait (mean 34.5) anxiety also showed significant improvements, whereas those with low scores (state anxiety mean 23.0 and trait anxiety 23.5) did not (alpha level .05).	GRADE score ? - ? – Levels of evidence score LII-a, grade B.
102 de Guzman (2011) Descriptive Dates not specified. Qualitative Descriptive	The Philipines 'Group home'.	3 older adults, aged 61, 73 and 86 (mean age 72.2). Loss to follow-up: none, only three interviewed.	Art and Craft-Traditional Filipino Arts (TFA) Puni-making (no detail on what this entails, there are some pictures in the paper).	To explore the potential for traditional Filipino arts as a recreational therapy for older adults.	Phenomenological analysis was used to analyse individual videotaped interviews carried out with the participants.	Themes and Subthemes Me, myself and melancholy- wearing out, walking away and wanting more. Will not let my worth wither- making it through, making it happen and making a difference.	QA score: grade B. Levels of evidence score LIII.
153 La Cour (2005) Intervention Dates not specified Qualitative Descriptive	Sweden A hospital and nursing home which specialises in rehabilitation, palliative and geriatric care	8 residents from the hospital and nursing home and 7 occupational therapy staff members. Demographic data not provided. Loss to follow-up: not reported.	Creative activity The creative activities that participants engaged in included woodwork, pottery, silk painting, soap making, knitting and gardening. Participants either took part individually or within a group. Sessions were delivered by occupational therapy staff members.	To explore the components of engagement with creative activity in the form of occupational therapy for older people who were dealing with a life-threatening illness rom perspectives of participants taking part and those delivering sessions.	The interviews with the older people were carried out either in their own rooms or the occupational therapy department (e.g. in the kitchen area). All of the therapists' interviews were carried out in the occupational therapy department. Each of the interviews were between 30 and 45 minutes in length and were audio tape-recorded and transcribed verbatim.	The core category identified from the analysis was "creating connections to life". Three subcategories were also identified with separate characteristics of engaging in creative activity. The first was "generous receptive environment" which showed the sociocultural setting as necessary for engaging in creative activity. The second was "unfolding creations- an evolving process" which relates to the practical making and process involved with the end products of the creative activity. The third category is "reaching beyond for possible meaning horizons" which related to meaning beyond the current conditions of the body and engagement in the activity.	

4.6.3 Location and setting

The studies took place in different countries (Table 4.6), one in the USA (Doric-Henry, 1997), one in Sweden (La Cour, 2005) and one in the Philippines (de Guzman, 2013). One of the studies specified the site where residents lived as the Evangelical Home of Saline, a church based residential home located within Michigan (Doric-Henry, 1997). One referred to residents from a 'group home', with the specific location within the Philippines not specified (De Guzman, 2013) and a third reported the residents to live within a nursing home and hospital which specialised in rehabilitation, palliative and geriatric care (La Cour, 2005).

Two studies (Doric-Henry, 1997; La Cour, 2005) explicitly stated that the arts activity took place within the care home itself, one did not specify further than this (Doric-Henry, 1997) whilst the second study stated that the sessions took place within the occupational therapy department within the home which featured three rooms, a large room with sofas where residents could get coffee and two smaller rooms designed specifically for arts and crafts (La Cour, 2005).

4.6.4 Populations

A total of 51 older adults were included across the three studies (Table 4.6), 31 of whom participated in the activity and a control group of 20 in Doric-Henry (1997). Additional data were also provided by seven occupational therapists (OTs) in LaCour (2005). Sample sizes were small in the two descriptive studies at three and eight participants (and seven OT staff) respectively (De Guzman, 2011; La Cour, 2005) and a total sample size of 40 (20 in the intervention group and 20 in the control group) shown for the one intervention study (Doric-Henry, 1997).

Two studies reported the gender of participants, with all three participants in de Guzman (2011) reported as women, as were 35 of the total sample of 40 participants (88%) in Doric-Henry (1997). The latter study also reported the proportion of women within both the control and intervention group, with 95% of the intervention group made up of women (n=19) and 80% of the control group (n=16).

The mean age of the participants was only calculated and reported for one study (Doric-Henry, 1997) where it was separately presented for the intervention group and control group (85.5 years (range 50-90 years) and 85.9 years (range 76-99 years) respectively). This allowed for the calculation of a combined total sample mean of 85.7 years (range 50-99 years). De Guzman (2011) did not report a calculated mean age for the group, however reporting separate ages for each of the three participants meant that a mean age of 72.2 years could be determined for the sample (range 61-86). No demographic information was reported for the third study (La Cour, 2005).

None of the studies specified whether participants (all or proportionally) had a dementia diagnosis. However, all three studies featured inclusion criteria which may have excluded those with cognitive impairment from taking part. de Guzman (2011) outlined that participants needed to be 'physically and mentally sTable enough to take part', while La Cour (2005) similarly reported that participants needed to be able to respond 'adequately' to the research questions. Finally, Doric-Henry (1997) based their inclusion criteria on the apparent willingness of participants to take part in the other activities carried out within the home, such as needlework, music, crafts and art, along with their perceived 'physical and emotional needs'.

4.6.5 Characteristics of the arts activities

Two studies assessed an arts activity not currently routinely offered by the homes (Doric-Henry, 1997; de Guzman, 2011), it was explicitly reported in the first of these studies that arts activities of any type were not routinely carried out within the home prior to the research study. The final study assessed an established occupational therapy programme routinely offered within the nursing home and hospital where the study was carried out (La Cour, 2005).

There was almost no information as to the way the arts activity was carried out within de Guzman (2011). The study states that residents took part in Puni-making, a traditional Filipino art form, a few small pictures are provided which outline the process, however there is no description as to what the activity involved or how it was carried out. Both remaining studies outlined art making with Doric-Henry (1997) evaluating a pottery activity and La Cour (2005) assessing a range of creative activities that included woodwork, pottery, silk painting, soap making, knitting and gardening. There were differences in the ways in which these two similar arts activities were carried out, the first study (Doric-Henry, 1997) carried out a series of classes across an eight-week course where residents were taught different techniques to produce pieces of pottery. In the second (La Cour, 2005) residents could choose the activity they wished to take part in and were able to freely take part in their chosen arts type, these sessions were offered up to three times a week. Both studies described sessions lasting for around an hour, one described sessions were delivered by the researcher (Doric-Henry, 1997) and the second by specialist occupational therapy staff members (La Cour, 2005).

The pottery sessions took place in a group environment as the room was large enough to accommodate the entire sample (n=20) (Doric-Henry, 1997). In La Cour (2005) residents are described as able to choose not only their activity but also whether they wished to take part in this individually, or with others as part of a group. For the session evaluated within the

study differences in gender and preferences for group or individual activities are reported, with two males and one female choosing to take part individually, three females took part in group activity and one male took part in both individual and group activities.

4.6.6 Research methods and outcome measures

Doric-Henry (1997) was the only study to administer questionnaires as outcome measures. The Coopersmith Self Esteem Inventory, Beck Depression Inventory and State-Trait Anxiety Inventory were administered at baseline and then immediately following the pottery classes for the intervention group and at the same times for the control group. The study also reports a subjective questionnaire based upon the self-esteem, depression, anxiety and creative experiences from the study were obtained post-test, the results of this are not presented, nor a copy of the included questions.

The two qualitative studies conducted individual interviews. These took place in a 'comfortably furnished' visitation room in de Guzman (2011) and within the occupational therapy department or the residents own rooms in La Cour (2005), it is not clear whether participants chose where their interviews were carried out or whether the location was based on convenience. The interviews with OT staff were carried out in the occupational therapy department.

Interviews were audio-recorded for one study (La Cour, 2005) and videotaped in the second (de Guzman, 2011) and then transcribed. Analysis was carried out via qualitative comparative method (La Cour, 2005) and phenomenological analysis (de Guzman, 2011). In each case the interviews and analysis were carried out by the researchers.

4.6.7 Quantitative findings

4.6.7.1 Behaviour

None of the included applied arts and crafts studies quantitatively assessed the impact of the activity on behavioural symptoms.

4.6.7.2 Psychological wellbeing

Doric-Henry (1997) was the only study of this arts type to report quantitative findings and these all related to psychological wellbeing. T-tests revealed significant improvements in mean scores for measures depression (4.6 vs 2.1), self-esteem (72 vs. 81.6), trait-anxiety and state-anxiety (28.7 vs. 22.6). Actual p values were not reported for the findings, only that the alpha significance level was 0.05. No significant improvements were observed for any of the measures within the control group, however it is worth noting that for all the measures (aside from self-esteem) scores were improved, despite not meeting significance. Comparison between the group means was carried out only at baseline, revealing there to be a significant difference in anxiety scores (it is unclear whether this difference occurred for state or trait anxiety, or both measures), with the intervention group beginning with significantly lower scores than the control group.

In agreement with findings from the previous arts types there was evidence improvements were observed more so in those with existing psychological need. Post-hoc analysis comparing the baseline scores of participants showed those who began with higher self-esteem scores failed to show improvements whereas those with lower scores did. Furthermore, those with 'high' depression scores, state anxiety scores and trait anxiety were said to show significant improvements ($p<0.05$), however mean values for post-test are not reported.

4.6.7.3 Cognition, quality of life and socialisation

None of the included applied arts and crafts studies quantitatively assessed the impact of the activity on either cognition, quality of life or socialisation.

4.6.7.4 Improvements to care and physical improvements

None of the included applied arts and crafts studies quantitatively assessed the impact of the activity on improvements to care or physical health.

4.6.8 Qualitative findings

La Cour (2005) identified a central category of 'creating connections to life' with an additional three subcategories identified. The first of these was entitled 'generous receptive environment' which highlighted the sociocultural factors associated with the programme. These included the freedom to take part in creative activity, support of staff members, the ability for residents to partake in creativity without demands placed upon them and the engagement of therapists which in turn promoted engagement with creativity within the residents. The second category related to the creative process, entitled 'unfolding creations', this included features such as the joy expressed by residents when taking part in the activity and overcoming the initial fears at not knowing how their final products would be shaped. This category also identified the sharing of memories and personal information prompted by the materials and overcoming physical restrictions to create artwork, which in turn facilitated the residents' sense of self by enabling them to produce a creation. Finally, the third category (reaching beyond for possible meaning horizons) represented the residents overcoming physical disability, planning how their creations may evolve and affirming their sense of productivity. Also described was the ways in which this environment facilitated discussions with healthcare workers who would normally not engage.

The second qualitative study (de Guzman, 2011) reported two central categories, the first of which represented the negative emotions and wellbeing expressed by residents. Entitled 'me, myself and melancholy' the subcategories were entitled 'wearing out', 'walking away' and 'wanting more'. The first of these represented how residents felt worn out, not only physically, but in terms of their social and familial roles. The second reflected on the transition from their familial homes to the care home environment and the sense of abandonment they felt from their family members. The final subcategory identified how residents wanted more from the care environment, which they felt catered for their physical but not emotional needs. In direct contrast, the second central category was 'I will not let my worth wither' which represented the positive feelings associated with the activity, with subcategories of 'making it through', 'making it happen' and 'making a difference'. The subcategory of 'making it through' described how residents still had interests and 'making it happen' highlighted how despite residents possessing physical constraints they could take part in the activity which in turn enhanced their self-worth. Finally, 'making a difference' discussed how these feelings of self-worth were further reinforced by being able to acquire new skills and continue learning.

4.6.9 Quality appraisal of included studies

4.6.9.1 Levels of evidence

Levels of evidence scores were low for this category, with one (Doric-Henry, 1997) at LIIb (the lower bracket of 'moderate' evidence) and two (La Cour, 2005; de Guzman, 2013) at level III (the lowest acceptable level within this review).

4.6.9.2 Quantitative studies

The only intervention study within this sub-section (Doric-Henry, 1997) was deemed to be of low quality. It is unclear whether any blinding of outcome assessment took place, however it is probable that this did not occur. The authors state that 'not all participants took part in

each of the sessions' but this is not explored further, it is therefore unclear how many sessions residents attending on average and whether there were participants who dropped out of the programme altogether. The methods section also refers to the use of a subjective mood questionnaire, which was administered following the study period but was not reported within the results section. There are also factors around the sub-analysis of participants scoring high and low on the depression and anxiety scales. Those described as having 'low' depression scores, state anxiety scores and trait anxiety scores did not show significant improvements. A major discrepancy with this finding is a lack of range and/or standard deviation provided for scores on each of the scales. Also, scores are still relatively low when considering the maximum scores allowed on the scales. A score of 1-10 is considered 'normal ups and downs' and not indicative of even slight depression in the Beck Depression Inventory (find reference). Similarly, cut-off scores of higher than the high scorers within this study (state anxiety= 34.3 and trait anxiety 34.5) indicate clinical anxiety. Noted is the fact that the authors do report scores were relative to the sample, but it would be interesting to see if this result had occurred if those with clinical scores indicative of actual clinical depression and/or anxiety (if any) were sub-grouped.

4.6.9.3 Qualitative studies

Both descriptive studies within this subsection (La Cour, 2005; de Guzman, 2013) scored Grade B (indicating few flaws which were unlikely to impact upon the results of the study). Neither of the studies reported evidence of reflexivity and La Cour (2005) failed to adequately provide rationale for their chosen methodology and provided no justification of the sample size chosen. La Cour (2005) adequately reported limitations of the study, whereas these were not reported within Guzman-Garcia (2013). It was difficult to appraise whether La Cour (2005) had adequately met ethical considerations, although the process of ethical approval was

followed and gained from the appropriate body it was unclear whether the interview structure was adequately reported to the residents. Despite these flaws, both studies adequately reported on the scope, purpose and aims of the study and identified clear inclusion criteria for eligible residents. Furthermore, a good account of the analysis carried out was provided in each study, along with a good interpretation of the data and future recommendations and directions.

4.6.10 Summary

This category contained the smallest number of studies, with just three included, two of which were qualitative and one quantitative. Two of the studies (Doric-Henry, 1997; Guzman-Garcia, 2013) reported on a specific craft activity (pottery and puni-making respectively) and the third study provided an account of individuals taking part in different activities of their choosing (LaCour, 2005). Both qualitative studies showed high levels of quality (awarded grade B, the second highest grade and the highest grade awarded within the systematic review). The quantitative study (Doric-Henry, 1997) was of low quality, largely because of poor reporting methods.

There was limited evidence for quantitative improvements in measures of psychological wellbeing, with depression, anxiety and self-esteem improving however sub-analysis indicates this effect may have been dependent upon scores at baseline (Doric-Henry, 1997). Both qualitative studies provided evidence of improved emotional states, along with the sense of achievement felt by residents at completing the activity despite physical constraints. Similar findings were also found in both qualitative studies in respect of increased learning which surprised residents and the improvements in self-esteem promoted by this.

This subsection presented a narrative descriptive synthesis of the studies featuring the arts type ‘applied arts and crafts’. The following subsection (section 4.7) reports an overall summary of each of the five descriptive narrative synthesis sections (one for each arts type) presented within this chapter.

4.7 Overall Chapter 4 Summary

The previous chapter (Chapter 3) presented an overall high-level narrative synthesis of the 71 studies included within the systematic review. This chapter consisted of five subsections focused upon the arts typology developed from the studies within the systematic review. These arts types were: music, multisensory activities, performance, literary spoken and written word and applied arts and crafts. Each section provided a descriptive narrative synthesis of the included studies, along with appraisals of methodological quality.

Music showed evidence of impact on behavioural symptoms, although this varied based on the types of behaviours observed. Music was found to reduce levels of agitation, but this was not found by the studies featuring multisensory activities assessing this outcome. Both studies (Scott, 2014) assessing agitation from the multisensory activities' typology described low baseline levels of agitation in their populations which may account for lack of improvement. Increases in verbally disruptive behaviours were shown by the studies within the music, performance and multisensory activities typologies. Qualitative evidence suggested an increase in self-expression which may be attributed to such behaviours, particularly in studies which required active participation from residents. This was further evidenced particularly by literary spoken and written word studies which showed increased engagement with the activity indicating residents were responsive.

There was also evidence of impact on psychological wellbeing with improvements in depression and anxiety. Studies across the arts types described low levels of baseline scores particularly for depression and anxiety which may have impacted upon evidence of benefit for these measures. This was further evidenced by comparisons with control groups which showed changes post-test (but not across time) indicating arts activities may be beneficial in maintenance of psychological health rather than producing clinically significant

improvements. Studies from the categories of spoken and written word (Fritsch, 2007) and multisensory activities (REF) showed increased levels of anxiety, although the first of these described reductions in apathy also. Increased incidences of VDB may also be reflective of anxiety.

Quality of life measures failed to improve for most of the quantitative studies, although aspects of quality of life were shown through qualitative evidence, particularly in relation to self-expression, creativity, meaning and belonging. Cognitive improvements were demonstrated within the categories of music, performance and literary spoken and written word which would reflect the types of activities featured within these categories and use of cognitive processes. Increased socialisation was demonstrated by the studies which featured music, along with Scott (2014) and Roe (2015) who evaluated multisensory activities. Qualitative evidence also obtained benefits not only for those taking part but also for staff members and the care home community.

Irrespective of arts activity there were variations in the delivery of the activity, suggestive that benefits associated with an arts type may not be applicable to all methods and forms of delivery.

Quality of quantitative studies varied, with RCTs showing lower scores, meaning they were less likely to be at risk of bias in comparison to quasi-RCTs and uncontrolled pre/post-test. This may in part be due to the use of GRADE criteria for appraisal of quality and is discussed further in Chapter 7 (discussion). The music typology (Section 4.2) featured the largest number of studies and therefore showed a large variation in quality. Studies within the performance typology (Section 4.4) were generally poorly reported meaning quality appraisal was difficult as it often warranted scores of ‘unclear’. The multisensory activities typology

(Section 4.3) featured the lowest overall judgements of quality irrespective of study type. The literary spoken and written word typology (Section 4.5) showed a high-quality level of evidence which may be due to the studies within this typology being published more recently (all since 2007). Applied arts and crafts only featured one quantitative study which was of low quality and carried out more than 20 years ago (Doric-Henry, 1997). Across typologies qualitative evidence was found to be of moderate quality, with all studies graded either B (few flaws unlikely to affect the credibility of the study) or grade C (some flaws which may affect the credibility of the study).

This chapter has presented a narrative descriptive synthesis of the included studies from the systematic review based on the arts typology formed. The next two chapters present the methodology and methods (Chapter 5) and findings (Chapter 6) from Study 2, the electronic survey of a stratified randomised national sample of care home managers and follow-up interviews.

Chapter 5- Study 2: Electronic survey of a stratified randomised national sample of care home managers and follow up telephone interviews. Methods.

5.1 Introduction

Chapters 2, 3 and 4 have focused upon Study 1, the systematic review which explores the benefits of arts for health activities based upon current empirical evidence. This chapter (Chapter 5) will provide the methodology and methods for Study 2, the national survey of care home managers and follow-up interviews. Study 2 consisted of two parts, the first was an online survey to ascertain what arts activities are currently offered within a stratified randomised national sample of care homes in England, how they are delivered and potential benefits which may be identified by care home managers. The second part of the study conducted semi-structured telephone interviews with a self-selecting sub-sample of managers from responding care homes to explore further potential benefits and limitations of taking part in arts-based activities.

This chapter will therefore provide an overview of the sampling methods, recruitment of care home managers for telephone interviews, process of quantitative and qualitative data collection and the method of quantitative and qualitative data analysis. This chapter will also provide information as to ethical considerations, the reliability and validity of the survey and public involvement. The results from Study 2 are presented within the next chapter (Chapter 6).

5.2 Objectives

There were three objectives to carrying out this survey:

1. To conduct an online survey of care home managers in England to determine what (if any) arts activities were delivered and map their prevalence.

2. To examine the characteristics surrounding delivery of such activities to gather information including frequency of delivery, method of delivery, funding and what benefits are identified from such activities.

3. To synthesise these results from those obtained from the systematic review (Study 1, results in chapters 3 and 4) to form a typology of arts activities and their perceived benefits (presented within the discussion, chapter 7).

5.3 Research Design

The research design was twofold, the first part utilised a national survey of a stratified randomised sample of care home managers which was administered electronically. The second consisted of semi-structured interviews carried out over the telephone with a self-selecting sub-sample of care home managers.

Surveys are designed to gather information which can then be used to make inferences about the population, or a set group (e.g. a certain profession) (Stoop & Harrison, 2012). For this part of the project the decision was made to administer the surveys electronically. Given the increase in the usage of electronic devices the popularity of electronic surveys is growing. To access electronic surveys participants can be given a URL, either physically (such as in shops and restaurants when people are asked to complete a survey), via email with an embedded URL which they can access directly or via website banners which can be 'clicked on' (Hunter, 2012).

Delivery of electronic surveys have been shown to be beneficial in that they allow for greater reach of hard to reach individuals such as health care professionals (Bruggen, 2009; Holmes, 2009; Hunter, 2012) and require minimal effort and completion time on the part of the participant but still produce high quality data (Gill et al. 2013). Furthermore, they allow an

element of distance between the respondent and the researcher, which may promote more truthful answers than those obtained over the phone or through personally addressed letters (Hunter, 2012).

One of the greatest benefits identified with the usage of web-based surveys is they are time-effective, particularly in comparison to mail-based surveys. Rather than the requirement of each potential participant to be individually contacted via mail, email mailing lists enable many participants to be contacted at a time (Snyder-Halpern, Thompson, & Schaffer, 2000; Fan & Yan, 2010). They also allow for quicker data collection time, with results available upon completion by the participant and they cost less to distribute compared with mail-based surveys (Holmes, 2009; Gill et al. 2013). Web-based electronic surveys also enable completed responses to be immediately available upon completion of the survey and results can be directly exported to statistical packages which not only speeds up the process of analysis, but also reduces the risks associated with human error when inputting large data sets manually (Snyder-Halpern, Thompson & Schaffer, 2000; Gill et al. 2013; Hunter, 2013; McPeake, Bateson & O'Neill 2014). In relation to this, responses obtained via electronic surveys are generally more legible than those which have been completed by hand as they require selection of pre-determined categories and typing of free-text responses (Snyder-Halpern, Thompson & Schaffer, 2000; Holmes, 2009).

Perhaps the most cited problem with internet based surveys is response rates. Despite individual studies showing good response rates (Snyder-Halpern, Thompson & Schaffer, 2000; Gill, 2013), systematic reviews have found that electronic surveys may yield low response rates, particularly amongst health care providers (Fan & Yan, 2010; Cho, Johnson & VanGeest, 2013). It has also been evidenced that recruiting from organisations (such as hospitals, health

care providers) results in low response rates (Swain, 2012). This is further demonstrated by Hunter (2013) who described a 10% response rate from NHS trusts compared with a 74% rate from a web-based forum when contacting midwives for an E-Survey. It has also been shown that whilst our use of technology is increasing, response rates to surveys is declining, possibly due to their increased popularity meaning respondents pick and choose which they are likely to respond to (Cho, Johnson & VanGeest, 2013). In relation to this, individuals are most responsive to surveys where they perceive the topic to be more interesting, or relatable (Fan & Yan, 2010), which may account for some of the low response rates demonstrated for institutional surveys compared with those from forums or patient groups who may initially show more interest in the topic and therefore be more likely to take part.

Whilst some have described the distance which occurs with internet surveys as beneficial by promoting truthful answers (Hunter, 2013), there is also evidence the impersonal way emails are sent in large batches may yield lower response rates compared with contact made with individuals over the phone or through personally addressed letters (Whitehead, 2012). However, such methods are time consuming and not always practical when the aim is to gain information from a large group of individuals.

The way in which the survey is designed and distributed can increase response rates. One such way is making sure the subject title is kept simple, but is designed in a way which appeals to the respondent, without containing the word 'survey' (Edwards et al. 2009; SurveyMonkey, 2016). Sending reminder emails have also been demonstrated to improve response rates to electronic surveys (Cho, Johnson & VanGeest, 2013; Fan & Yan, 2010). The process of delivering reminder emails is quick in comparison to other survey methods (i.e. postal surveys) given that a mailing list of emails can be stored and simply re-sent in minutes (Hunter,

2013). Keeping length of questionnaires short and concise to ensure they aren't time consuming can also increase the number of respondents (Dykema et al. 2013). In addition, the less personal information required for a survey the more likely individuals are to respond (Whitehead, 2007).

As the rate with which we utilise technology increases protective methods are adopted which has led to increased usage of spam filters which may act as a barrier to questionnaires reaching their targets, especially through unsolicited emails (Fan & Yan, 2010). The impact of spam filters can be reduced via utilising similar techniques to improving response rates: sending out smaller batches of emails, providing a good introduction to emails and making sure a descriptive subject line in included with the email (SurveyMonkey, 2016). It is also beneficial to not contain email attachments (i.e. the information sheet or survey itself) within the email as this increases the likelihood of the email being rejected, or not able to be opened (Snyder-Halpern-Halpern, Thompson & Schaffer, 2000). There are also the problems which may occur with using technology. Email addresses may change and are easy to copy down incorrectly (Snyder-Halpern, Thompson & Schaffer, 2000) and internet coverage may dip or be affected by provider based issues (Fan & Yan, 2013).

Semi-structured interviews allow for the creation of questions that allow a narrative to develop, but shape the interview to allow for discussion to be focused on theoretically driven questions (Galetta, 2013). The amount of structure embedded within such interviews can therefore be varied in that interviews which focus on gathering information can consist of more structured questions than those which are more exploratory in nature. However, even when the goal is to gain information they can be shaped and changed as the interview progresses, therefore enabling creation of a topic draft, designed to guide the interviewer

and allow for expansion and change of direction should topics of interest arise (Miles & Gilbert, 2005).

Many of the benefits associated with use of electronic surveys are also applicable to the use of telephone interviews. Use of this method allows for a quicker interview process and reduced costs when compared with the travel required for face-to-face interviews (Wilson, Roe & Wright, 1998; Smith, 2005; Musselwhite et al. 2007; Szolonki & Hoffman, 2013). This requires for the coverage of a large geographical area, including those from other countries. Interviewing via telephone can also enable access to hard to reach professionals and those who work on sites where interviewing face-to-face may be difficult or not possible (Sturges & Hanrahan, 2004; Opdenhakker, 2006).

There have been questions surrounding the quality of data produced with telephone interviews, in comparison to face-to-face interviewing. Particularly in relation to decreased rapport leading to a reduction in discussions of sensitive topics (Opdenakker, 2006; Novick, 2008). However, studies have reported that telephone interviews are capable of producing data of equal quality to that obtained via face-to-face methods (Wilson, Roe & Wright, 1998; Sturges & Hanrahan, 2004; Szolonki & Hoffman, 2013). There is a level of anonymity to telephone interviews which may make discussion of sensitive topics easier (Sturges & Hanrahan, 2004; Novick, 2008). There is also suggestion that interviewees may feel more comfortable with the process of the researcher note taking when interviews are carried out over the phone is this is less intrusive during this method (Smith, 2005; Musselwhite et al. 2007). Even though sensitive questions may still be asked during telephone interviews it has been suggested that they are most beneficial to research questions which focus upon clear

objectives and information gathering (Smith, 2005) and where less emphasis is placed on verbal cues and pauses (Opdenhakker, 2006; Sturges & Hanrahan, 2004).

5.4 Sampling methods, potential samples and national database of care homes

Random via methods such as random number Tables, computer selection programs, or manual selection such as drawing names from a hat ensures all participants stand an equal chance of being selected to take part in a study (Coolican, 2011). However, random sampling does not take characteristics into account which may result in certain populations being over or under-represented which can reduce the overall generalisability of a survey, particularly when members of the public are invited to complete them.

When the participants are from a population of interest (called a *strata*) then stratified random sampling can be used to select participants. Proportional stratified sampling is the method whereby sample sizes are created based upon the proportions of certain groups within a whole population (e.g. men and women). Large populations are divided into smaller groups which allows for samples to be calculated reflective of the population of interest (e.g. if a sample was split into men and women and the male sample contained 40% of the total, then 40% of the sample would be made up of males) therefore allowing for a representative sample (McCormack & Hill, 1997; Johnson & Christensen, 2012). As with many research methods the idea of stratified sampling does not completely negate the possibility of sampling bias occurring, however it minimises the chances of such occurring (Bautista, 2012). In this case, the sample was stratified based upon care home type (nursing care, non-nursing care and dual registered) and geographical location (North West, South West etc.).

The Care Quality Commission (CQC) is the social regulator for England, all facilities providing health care within England must be registered with the body. Their website contains a

directory of all care providers currently registered with the body, which is available for anyone to download (http://www.cqc.org.uk/about-us/transparency/using-cqc-data#directory) (screen shot, Appendix 5). The directory contains all services currently registered with the CQC and therefore consists of almost 50,000 entries. One of the columns records whether the registered facility is a care home and so these were filtered from the original database to form a new database which contained all registered care homes. This was filtered further by removing all the care homes which were reported as providing care for children, young people or adults (identified within columns of the database- screen shot in appendix 5), leaving the care homes which were exclusively for older people.

Table 5.1- stratified sample of included care homes by type and geographical location.

Geographical Location	Dual Registered- Actual number available	Dual registered- Percentage	Nursing- Actual number available	Nursing- percentage	Non-nursing- Actual number available	Non-nursing - Percentage
E. Midlands	30	12%	377	10	521	10
E. England	23	9%	333	8	641	12
London	17	7%	336	9	305	6
N. East	40	16%	239	6	205	4
N. West	33	13%	539	14	813	15
S. East	27	11%	778	20	979	18
S. West	38	15%	496	13	754	14
W. Midlands	20	8%	419	11	565	10
Yorkshire and Humber	22	9%	380	10	586	10
Total	**250**		**3897**		**5368**	

Actual number available shows the potential number of homes and their geographical regions. The percentage of homes located within each region is represented in the percentage column, this corresponds to the proportion of homes the sample was made up of.

A total of 9,515 care homes were identified as providing care for older people and these formed the resulting Care Home database. Care home types were split into those providing nursing care, those who did not provide nursing care (labelled residential) and those who were dual registered. Following stratification of the sample, databases were created for each of the care home types: dual registered homes (n=250, 3% of the total number),

nursing homes (n= 3897, 41% of the total number) and non-nursing residential homes (n= 5368, 56% of the total number). These were then further stratified by location to calculate the proportion of care homes that were located within each region (see Table 5.1). Again, these were separated into databases so that each database contained a care home type and a geographical location (i.e. separate databases for dual registered care homes within the North West, nursing homes in the North West, non-nursing homes in the North West, dual registered care homes in the South East etc.).

5.5 Samples and Recruitment

The potential sample for the survey was based on a previous national postal survey of care home managers in England that used a 20% stratified randomised sample from the CQC database that achieved a 30% response rate (Hays et al 2012; Jasper 2018, personal communication). However, after initial pilot mailing, this was then increased to 40%, given the low percentage of responses obtained from the pilot (see 5.6 reliability and rigour) and evidence that generally response rates from electronic surveys are low, particularly amongst healthcare providers (Swain et al. 2012; Cho, Johnson & VanGeest, 2013; Fan & Yan, 2010). Furthermore, method of recruitment within this survey was via unsolicited emails sent out to care managers who were unknown to the research team and had not been previously contacted, variables which may reduce response rates of electronic surveys (Whitehead, 2007).

Therefore, a 40% total sample size of 3806 care homes was calculated. Due to rounding and calculations a sample of 3817 was calculated, of which 90 (3%) were dual registered care homes, 1527 (41%) were nursing homes and the majority 2200 (56%) were non-nursing homes. These were also calculated so that they represented the geographical spread of the homes (see Table 5.2 below).

Table 5.2 Sample calculation of each care home type and geographical location.

Geographical Location	Dual Registered Percentage represented	Dual registered number invited	Nursing percentage represented	Nursing number invited	Non-nursing percentage represented	Non-nursing number invited
E. Midlands	12%	10	10%	152	10	218
E. Anglia	9%	8	8%	122	12	260
London	7%	6	9%	136	6	130
N. East	16%	14	6%	92	4	136
N. West	13%	12	14%	214	15	326
S. East	11%	10	20%	304	18	390
S. West	15%	14	13%	198	14	304
W. Midlands	8%	8	11%	166	10	218
Yorkshire and Humber	9%	8	10%	152	10	218
Total		**90**		**1527**		**2200**

Percentage represented gives the total number of homes from each geographic region for each care home type. The number invited shows the number that this corresponded to when calculating sample sizes for each care home type and region.

Email addresses were not available from the CQC database and these were gained from the website of the homes contained within the database. Initial emails were sent out at the end of November/beginning of December 2016 to the total potential sample of care home managers. Emails were sent out in sets of around 20 to minimise the likelihood of them being marked as 'spam' or 'junk' and contained a comprehensive subject (subject: research into arts for health in care homes), again to minimises the chances they would be filtered as spam (SurveyMonkey, 2016). Another problem which may arise with distance completed surveys is that they may be passed on to another individual to complete (Whitehead, 2007), to reduce likelihood of this the emails were addresses 'to the care home manager' and it was stated within the email this was the group responses were desired from (see invitation email, Appendix 6).

The invitation email contained brief information about the project, along with an embedded URL which could be accessed if the respondent wished to take part. The URL then took

individuals directly to the survey information sheet. This information sheet outlined the scope of the research project, that participation was voluntary, that responses would be anonymised and the timeframe expected for completion of the survey. Information relating to the telephone interviews was also provided, this informed those taking part that if they so wished they could leave their contact telephone number and a time when it would be convenient to call if they wanted to take part in this aspect of the study. It was explicitly stated that participants could just take part in the survey and they were not obliged to leave their contact details if they did not wish to. Participants were instructed to click

Reminder emails have been shown to increase response rates in electronic surveys (Cho, 2013; Fan & Yan, 2013) and therefore two reminder emails were sent to non-responding participants. One set of reminder emails was sent out after four weeks (end of December 2016) and then a final reminder email was sent out after another four weeks (end of January 2017). The survey remained open for the entire period (i.e. three months) so that it could be accessed any time the email was read by the recipient within this time frame. It has been demonstrated that if respondents see that others have completed a survey then they are more likely to complete the survey themselves (Edwards et al. 2009; Hunter, 2013) and therefore when reminder emails were sent out they contained a 'thank you' to those who had already completed the survey and informed others that they still had the opportunity to take part.

The final question asked whether individuals would be interested in taking part in telephone interviews to provide more information in addition to the survey. Therefore, a self-selecting sub-sample of care home managers provided their name, the name of their care home, a contact telephone number and a time when it would be convenient to call and arrange the

interview. A total of 71 participants (38% of the total number who took part in the survey) left their details. These participants were contacted on their preferred day and time to see if they still wished to take part in the interviews and if they still wished to participate then a suitable time was arranged to call back and carry out the interview.

During the telephone call when interviews were arranged the managers were asked for their email addresses and a participant information sheet (see Appendix 7) was emailed out to them. This explained the process of the interviews, including audio-recording and what would happen in the event of a disclosure being made (contacting the Care Quality Commission (CQC). As it has been demonstrated that non-response to telephone interviews occurs often when the individual has forgotten (Smith, 2005), these emails also served as a reminder to participants of the forthcoming interview.

5.6 Data collection

The survey was hosted via the online hosting platform SurveyMonkey®, accessed via the URL embedded within the invitation email. The survey consisted of 20 questions, split roughly into two sections, the first which asked questions about the demographics characteristics of the home and the second which gathered information as to the arts activities carried out by the home (full copy of the questionnaire, Appendix 8).

Questions one to thirteen focused on the characteristics of the care home. The first three questions enabled the participant to select one option only, given only one response would apply and asked the geographical location (question one), type of care home (dual, non-nursing, nursing- question two) and whether the home was public, private or not-for-profit (question three). Given the likelihood of variation in responses, the next two questions were free-text boxes which enabled actual numbers to be entered asking how many places the

home had in total (question four) and how many residents there currently were within the home (question five). A yes/no response was designed for the next question, which determined whether the home had rooms for respite care. Questions seven to eleven all focused on demographic information about the residents of the homes with relation to dementia diagnosis (question seven), memory problems but no dementia diagnosis (question eight), challenging behaviour (question nine) and depression or low mood (question ten). Question eleven focused on the proportion of residents who had incontinence, pressure sores, risk of falls and multiple medications. Respondents had to select from categories of 25% or less, 26%-50%, 51%-75% and 76% and above for each of these. The following question again consisted of free-text boxes given the expected variation in responses and asked how many staff members were employed within the home, how many of these were nurses and how many volunteers helped at the home. Finally, question thirteen asked how much contact those in the care home had with different health professionals.

The next section (questions fifteen to twenty) assessed the arts activities and the characteristics associated with delivery of such activities. Both question fifteen and sixteen consisted of Tables where respondents could select the arts activity type along with frequency they were delivered (question 15) and who they were delivered by (question 16). A free text box was also featured on both questions so that responses to 'other' could be elaborated upon. The remaining questions (17-20) all consisted of free text boxes and asked participants about funding for activities (question 17), costs associated with carrying out the activity (question 18), whether there were any perceived benefits to residents (question 17), any benefits to staff members (question 18) and any benefits to the care home in general

(question 19). Another free-text box was then provided for respondents to record whether they felt there was anything else of importance that they might not have been asked.

Respondents could skip questions, even though this may lead to missing data, it has been reported that enabling such a feature can reduce the proportion of individuals who complete the survey, as participants are likely to close the survey if they feel a question does not apply to them and they cannot skip it (Gill, 2013). The final question was designed to allow participants to enter their details if they felt they would like to take part in the telephone interview aspect of the study and they could therefore leave their contact number and a day and time it is usually most convenient to call.

When conducting the telephone interviews, respondents were contacted on their agreed date and time and it was confirmed that the correct person was being spoken to. The questions identified on the topic guide (see Appendix 9) were then asked. The topic guide was designed to offer flexibility during the interview process and allow for contextual information to be added to the responses from the survey. The questions were designed to collect information surrounding the arts activities currently delivered by the home and the factors associated with their delivery (such as who they were delivered by, how residents were able to agree to participate). Information was also sought around any impact the managers felt the arts activities had for the residents, care home and staff members, along with any difficulties that may be encountered during the activities. All interviews were audio recorded, with notes made throughout, should there be any technological issues with the audio files.

5.7 Quantitative Data Analysis

The survey was designed to gather provision of arts for health activities within care homes and therefore quantitative data was mainly descriptive. All data were exported from

SurveyMonkey into SPSS Statistics for Windows version 24.0 (released 2016). Descriptive statistics were created from the dataset, along with tests for associations between variables (Chi Square, X^2). A 5% level of significance was accepted. Graphs were created using Excel for Windows (2016 version).

5.8 Qualitative Data Analysis

A described in the data collection section (5.6) all interviews were recorded, and audio-files created. Tessier (2012) suggests a combination of audio files and transcripts is beneficial when preparing to analyse qualitative data as they overcome potential limitations of just using either one or the other. Whilst it is often advised that audio interviews be fully transcribed prior to analysis, Cooperrider and Whitney (1999, pp 25,28) suggest that, for some qualitative data, extraction of 'magic (key) moments' can suffice. Further to this, Bailey (2009) outlines how depth of transcription required is relative to the research question and methodology of data analysis. Given the survey (and qualitative data) were mainly used for information audio files were listened to several times and responses directly related to the questions were transcribed and analysed.

Qualitative Content Analysis (QCA) is the method of coding qualitative data to form themes or patterns which offer descriptions of meaning (Krippendorf, 2013). Content analysis can be used for all types of qualitative data meaning there are no specific rules which must be followed given the varying levels of depth achieved from different methods of data collection (Bengtsson, 2016). The content within the data can be described as both 'manifest', which describes surface content and data and 'latent' which offers underlying meaning and interpretation (Granenheim, Lindgren & Lundman, 2017). Researchers are able to vary their analyses based upon the level of depth sought through their research question and method of data collection (Bengtsson, 2016).

The method of content analysis can be described as either inductive or deductive, with inductive analysis characterised by searching and formulating new patterns and deductive analysis testing existing theory (Krippendorf, 2013; Bengtsson, 2016; Granenheim, Lindgren, Lundman, 2017). However, Armat et al. (2018) argue that content analysis involves processes which conform to both inductive and deductive reasoning. This would be in keeping with the approach adopted when analysing the qualitative data from Study 2. The questions were formulated in part from the responses obtained via the systematic review, therefore assuming a level of prior knowledge which would subscribe to a deductive approach. In addition, due to the undertaking of the systematic review potential benefits were already highlighted. However, codes and categories were formed from the qualitative data alone (i.e. categories were not already formed, and data placed within these) and a survey of care home managers has not previously been carried out, conforming to an inductive process which was adopted for the categorisation of codes.

Given its versatility, content analysis could be used for both the short free-text response questions (questions 17, 18 and 19) and the responses obtained through the telephone interviews. Full responses from the short-text questions were downloaded into a word document and frequencies of occurring words were obtained through reading responses and searching for incidences of commonly occurring words which were then placed into categories. Content analysis allows for an evolving process, the method of forming categories from coded words can lead to searching text again to identify additional codes which may correspond to that category (Krippendorf, 2013; Granenheim, Lindgren, Lundman, 2017). In keeping with this method, further codes were then sought which may correspond to the category and frequency of words calculated (the results Table can be seen in Chapter 6- Tables

6.6, 6.7 and 6.8). These codes and categories were confirmed with another member of the supervisory team (BR) to reach agreement on their formation.

The process for analysing the interview data was like that of the analysis of qualitative questionnaire responses, however quotes were used to provide contextual meaning to identified codes and categories. In keeping with levels of data formed through content analysis the short-text questions could be seen as providing 'manifest' as they evidenced categories of words and the frequency with which they occurred. The qualitative data from the interviews can therefore be described as providing 'latent' data, exploring the meaning behind the frequencies and benefits.

5.9 Ethical considerations

Ethics approval was sought from the Faculty Research Ethics Committee (FREC) in the Faculty of Health and Social Care at Edge Hill University and was granted 26th July 2016 (ethics approval letter, Appendix 10). The study was approved on the basis that a draft interview topic guide be submitted later, as this was to be determined by the responses given to the online survey. The draft topic guide was submitted to FREC in March, 2017 and ethical approval for the interviews was granted on 13th April, 2017 (this approval letter can be found in appendix 11). The University's code of conduct (Edge Hill University, 2012) and degree regulations (Edge Hill University, 2013) were adhered to throughout the course of the research.

Within the information sheet, the telephone interview sheet and when conducting the interviews participants were reminded that their responses would be stored confidentially. Prior to carrying out the project it was anticipated that a list of those consenting to be contacted for the interviews would be kept within a database which was stored on the

University system. However, this was not necessary as all responses were stored individually for each participant on the SurveyMonkey site, with a corresponding respondent number. This allowed for a list of respondent numbers to be created and stored and this could be accessed on the SurveyMonkey site, which was accessed via a login and password. This meant personal data (names and contact numbers) did not need to be transferred to another list and stored separately. When data were exported from the SurveyMonkey site, any identifiable data (IP addresses, contact telephone numbers) were removed prior to analysis as these were not required. The data from all surveys stored on SurveyMonkey is protected via Norton and TRUSTe and encrypted using a Secure Sockets Layer (SSL) (Survey Monkey (no date)). This SSL is reflected within the URL used for data collection: (https://www.surveymonkey.co.uk/r/artsforhealth) by the prefix of https:// rather than http:// where the s corresponds to 'secure' and indicates the presence of the encryption.

The subsequent dataset (which had all identifiable data removed) was stored handled and managed in accordance to the Data Protection Act (1998) and in agreement with the University's data management policy (Edge Hill University, 2012). The University requires all data relating to research projects that do not contain identifiable information be stored within the University repository for a minimum of ten years. Audio files from the interviews were stored until they had been transcribed and were then destroyed. Each file was stored with the respondent number, as participants had the right to withdraw their data for up to two weeks following the interview.

If those receiving the emails were interested in the project, then they could click on the URL which took them to the information sheet explaining the research project and format of the survey. Participants were then given the option to continue with the study, or close the

window and consent was therefore given when the participant clicked 'next' and agreed to take part in the survey. This is a similar process to consent obtained via postal surveys where the returning of the survey gives consent (Whitehead, 2007; Buchanan & Hvizdak, 2009). In cases where consent is implied via return, there needs to be a clear communication of what the survey entails and contact details where questions could be asked, should a participant be unsure (Holmes, 2009). The information sheet therefore offered all required information and contained the contact details for the research team, should any issues arise. Along with a reduction of response rates (Gill, 2013), the ability to skip questions has been highlighted as a potential ethical issue (Buchanan & Hvizdak, 2009) and therefore this option was included within the study.

Prior to the telephone interviews all participants were emailed a participant information sheet which contained information about the research project and the process of carrying out the interviews. Participants were informed that their interviews would be audio recorded, if they agreed to this. Upon phoning the participants to conduct the interview the process was explained again, including the expected length of time of the interview and they were asked if they consented to be recorded.

5.10 Reliability and Validity

Reliability refers to a data tool's ability to obtain results which are replicable whereas validity assesses the ability for the tool to measure and provide evidence for the intended research question (Bolariniva, 2015). Broken down further, internal validity proports that the survey uses questions which enable adequate responses and external validity refers to the generalisability of the survey, that it would be applicable to those beyond the sample.

The questions which formed this survey were based on and informed by the systematic review carried out as Part 1 of this study. The survey questions were therefore constructed to build upon the findings of research with current practice. Questions were assessed by the Faculty Research Ethics Committee (FREC) and through use of public involvement (see section 5.9 below) in to determine that they were realistic and answerable.

Piloting web-based questionnaires is necessary to establish reliability and validity (Wyatt, 2000). A pilot study was carried out with a total of fifteen care homes, five from each care home type (nursing, non-nursing and dual registered). Care homes were selected using random.org, a random number generator was used to select the homes from the databases. The response rate from the pilot study was 20%, corresponding to five of the homes completing the survey. All the questions were completed and the data as expected and therefore no amendments were needed in terms of the composition of the survey. The relatively low response rate of 20% formed the basis for the decision to expand the original sample size calculation of 40% (n=3806) to give a greater chance of responses.

The biggest factor affecting reliability and validity for online surveys is the ability for certain populations to be under-represented due to online accessing issues or not being computer literate (Snyder-Halpern, Thompson & Schaffer, 2000; Wyatt, 2000; Fan & Yan, 2010). This would be unlikely to factor in to this study. Firstly, email is a major method of communication for health care services and it is unlikely there would be issues with homes not internet access. If there were internet server problems, then keeping the survey open for three months and sending reminder emails would allow individuals to complete the survey when these had been resolved. Similarly given the amount of information and communication carried out electronically within the health care sector it is unlikely managers would be unable to use the

email or survey. Every effort was made at sampling stage to ensure that the questionnaire sample would be representative of the care home population. This was achieved via stratified sampling to ensure that populations sampled were reflective of the population i.e. a proportionate number of surveys were sent to those from each care home type and geographical location.

Ten percent of data from both the survey and the interviews were checked via another member of the supervisory team (BR) to establish inter-rater reliability. To provide transparency of reporting for the online survey, the CHERRIES checklist (Eysenbach, 2004) was adhered to determine that all relevant factors had been reported upon.

5.11 Public Involvement

The role of public involvement is to offer a collaborative approach to the design, implementation and delivery of health research, particularly where their skills, knowledge and expertise can help shape the direction of such research (INVOVLE, 2015; Giebel et al. 2017; INVOLVE, 2019). Therefore, rather than members of the public 'taking part' in a research project they are actively contributing to the methods of the research project itself. Collaborating with the public allows researchers to strengthen the ethical validity of their study by improving the experience of participants and ensuring the research topic is relevant and worthwhile (Health Research Authority, 2016).

INVOLVE, the NIHR government programme which sets out recommendations for involve the public in research, outline several ways in which members of the public can be involved with the research process and one of these is through the development of research materials (INVOLVE, 2015; INVOLVE, 2019). Prior to carrying out the survey, a copy was given, via email to those working within the arts and health sectors. Obtaining input from members of the

public through email can be beneficial in ensuring that members of the public who may be unable to attend face-to-face focus groups are still able to contribute to research development (Giebel, 2017). These consisted of a care home manager, arts for health researcher and museum staff member (three in total) who provided feedback based on the questionnaire. Small changes were made to the questionnaire of the basis of this involvement, including adding arts activities which were felt to be of importance and changing the wording of questions to make their purpose clearer.

5.12 Summary

This chapter has provided an overview of the methods used for Study 2, the national survey of a stratified random sample of care homes and managers within England. The aim of the survey was to gather information on arts activities currently being carried out within homes and ascertain any potential limitations or benefits of such activities. Care home managers were invited via email to participate within the survey, which was completed online. This allowed for descriptive data related to arts activity delivery to be recorded and subsequently compared with the results of Study 1, the systematic review (See Chapter 7, Discussion). Managers were given the option to participate in a telephone interview and these responses were audio recorded. The free-text responses in the survey and interviews were then analysed via content analysis to provide themes surrounding arts for health activities delivered in care homes within the UK. The following chapter (Chapter 6) presents the findings from the national survey (Study 2) and these are discussed along with the results from the systematic review in Chapter 7 (discussion).

Chapter 6- Study 2: Electronic survey of a stratified randomised national sample of care home managers and follow up telephone interviews. Results.

6.1 Introduction

Chapter 5 outlined the methods for carrying out the national electronic survey and follow-up semi-structured telephone interviews. This chapter presents the results from the online survey (emailed to care homes within England) and follow-up telephone interviews conducted with a sub-sample of self-selecting care home managers. Descriptive data and associations are provided for quantitative responses, along with content analysis of the qualitative free text questions and subsequent interviews. These are synthesised with the findings from the systematic review within the discussion chapter which follows (Chapter 7).

6.2 Descriptive statistics

6.2.1 Response rates

Invitation emails were sent to a total of 3817 care homes based upon proportional stratified sampling (see Chapter 5 methods section for how these were calculated). Around 10% (382) of the emails elicited bounce-backs from email servers and whilst efforts were made to resend these emails individually (to reduce the likelihood of them being marked as spam), it is not known how many in total reached their intended destination. This means a true sample size calculation of those who received the emails and hence accurate percentage response rate cannot be determined.

A total of 189 responses to the survey were returned, however four were removed as they identified themselves as providing care for populations other than exclusively older adults. If the bounce backs and care homes that did not fulfil inclusion criteria are removed from the target sample (3431), a total sample of 185 responses was achieved, representing 5% of the total adjusted number of invitation emails sent out. Most responses came from the first

mailing (n=116, 63% of total received), 56 came from the second round (30% of total received) and just 14 (7% of total received) from the third round. All 185 responses contained some information (i.e. none were blank) and 128 were complete (69%).

6.2.2 Representativeness of the sample

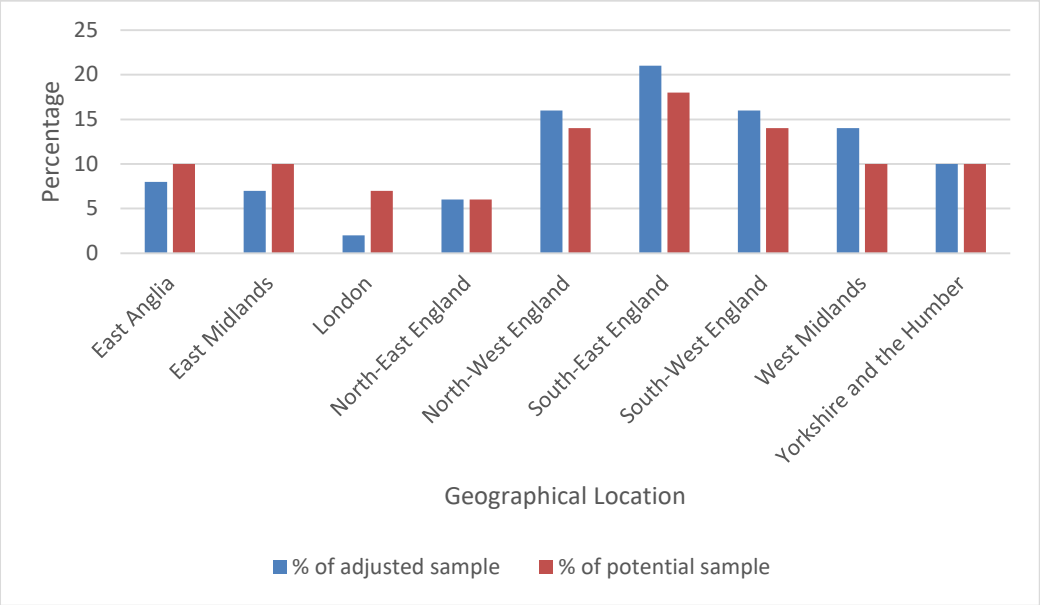
As outlined in the methods chapter (Chapter 5), a stratified sample was used to target emails according to geographical location and care home type to ensure they were reflective of the proportions of care homes within England (representing nursing homes, homes offering residential care for older adults and dual registered care homes, some with and some without a dementia specialty).

The total number of targeted emails and responses for each geographical region is outlined below in Table 6.1 and a visual comparison is shown in Figure 6.1. Percentages were comparable for all the regions, except for London, which corresponded to only 2% (n=3) of the sample, despite 7% (n=272) targeted to the location.

Table 6.1. Geographical location of included care homes responding to the survey along with the total numbers and percentages sent to the potential sample of care homes.

Geographical Location	Number of responses	% of adjusted sample	Number of potential sample sent out	% of potential sample
East Anglia	14	8	390	10
East Midlands	13	7	380	10
London	3	2	272	7
North-East England	11	6	242	6
North-West England	30	16	552	14
South-East England	39	21	704	18
South-West England	30	16	516	14
West Midlands	26	14	392	10
Yorkshire and the Humber	18	10	378	10
No Answer	1	0.5		
Total	**185**		**3817**	

Figure 6.1 Visual comparison of percentage of emails sent to each geographic region and percentage of responses

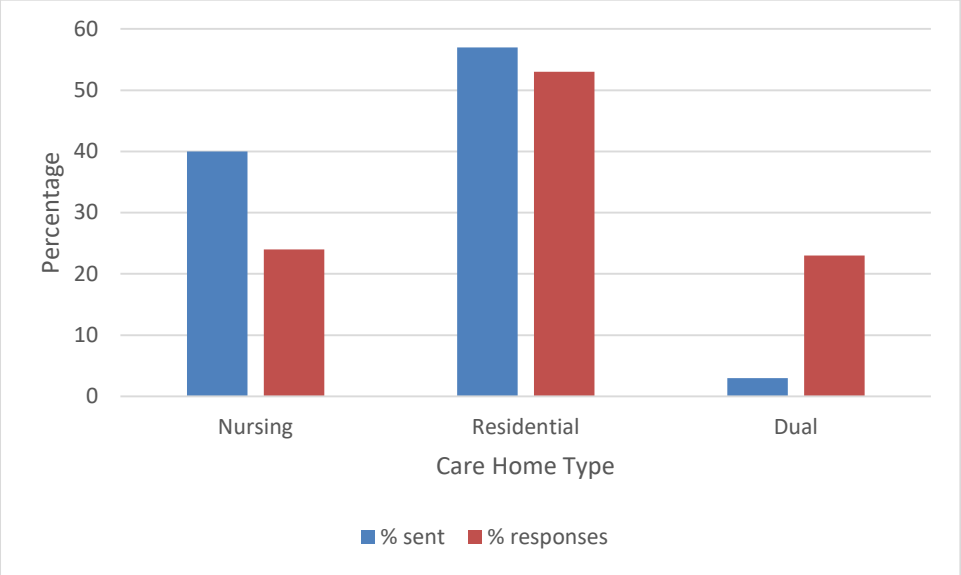


The total care home managers, referred to as respondents by care home type differed from the proportions sent out via the stratified sample (Table 6.2 and Figure 6.2). Only residential homes showed similar proportions of responses sent out (n=2,200, 57%) and those returned (n=92, 50%). Nursing homes were targeted by 40% (n=1527) of the potential sample but only 25% (n=46) of the respondents identified this type. Dual registered homes showed the largest discrepancy, with only 3% (n=90) of the sample targeted to this type with a proportion of 22% (n=41) of responses. Four identified as 'other' and two failed to answer the question.

Table 6.2 Care Home Type and number and percentage of targeted sample with number and percentage of retuned sample

Care Home Type	Number of responses	% of adjusted sample	Number of potential sample sent out	% of potential sample
Residential	92	50	2200	57
Nursing	46	25	1527	40
Dual Registered	41	22	90	3
Other	4	2		
No Answer	2	1		
Total	185		3817	

Figure 6.2 Graph showing the percentage of responses from each care home type compared with the percentage sent out per the stratified sampling.



6.2.3 Care home demographics

The characteristics of the care homes identified within the survey are presented in Table 6.2. All but one of the responses (n=184, 99%) provided the geographical location of the home and whether the home ownership was private, not-for-profit or mixed public and private. Care home type (residential, nursing or dual registered) was also reported by all but two of the responses (n=183, 99%).

Half (92, 50%) the respondents identified their facility as a residential home providing care for older people, 44 (24%) of these reported a dementia care specialty. A quarter of the responses were from managers within nursing homes (n=46, 25%) and 22% (n=31) were from those within dual registered (residential and nursing care) homes. Of those which identified their home as 'other' (n=4, 2%), these were stated as: 'residential older adults with mental health', 'specialist care with challenging behaviour', 'nursing care older people with end of life care' and 'neurological rehabilitation and dementia care'.

Table 6.3 Characteristics of the Responding Care Homes

Characteristic	Number of Responses and (%)
Geographical Location	
East Anglia	14 (8)
East Midlands	13 (7)
London	3 (2)
North-East England	11 (4)
North-West England	30 (16)
South-East England	39 (21)
South-West England	30 (16)
West Midlands	26 (14)
Yorkshire and the Humber	18 (10)
No Answer	1 (0.5)
Care Home Type	
Residential care older people	48 (26)
Residential care older people with dementia care specialty	44 (24)
Nursing care older people	27 (15)
Nursing care with dementia specialty	19 (10)
Dual registered care older people	22 (12)
Dual registered with dementia care specialty	19 (10)
Other	4 (2)
No response	2 (1)
Ownership	
Private	116 (63)
Not-for-profit	39 (21)
Mixed public and private	29 (16)
No response	1 (0.5)

Number of responses to each characteristic is given with the percentage (%) of total responses to the question. Total responses n=185.

Over half the homes were privately owned (n=116, 63%), 39 (21%) were not-for-profit facilities and 29 (16%) were jointly owned, both public and privately.

The number of reported places available varied greatly between the homes from 10-114 (reported by 179 homes, 97%). The average number of places was 41 (SD 19.5). This was closely matched by the average number of residents currently living within the homes, recorded by 178 respondents with an average of 40 (SD 17) and range of 10-94. Similarly, the number of staff employed within the homes varied from 12 to 160 with an average number of staff calculated as 58 (SD 28). Ninety-seven (52%) of respondents reported that volunteers helped within their homes.

There was no association between the geographical location of the care home and the number of places available) nor between geographical location and whether the care home was publicly or privately owned. There was an association between the care home type and the geographical location of the homes, however this was expected given the proportions identified within the sampling calculations (X^2 (1, n=179) = 96.6, p<0.001) and is reflective of the overall spread of care homes throughout England.

Of the homes that identified the number of places (n=176, 96% of total responses) and the number of residents currently within the home 62 (35%) were operating at full capacity. An additional 65 (37%) homes reported that more than 90% of the places within their homes were currently filled, with 27 of these indicating they had only one less resident than the total number of places. Seventeen (9%) of the remaining homes reported 50% capacity or above (1 at 50%, 3 at 60-69%, 13 at 70-79%) and only three reported low capacity levels (25%, 20% and 17%).

6.6.4 Resident demographics

The reported demographics of the residents living within the care homes is shown in Table 6.3. Care home managers were asked to report the proportions of residents within their home they felt suffered with incontinence, pressure sores, multiple medication needs or were at risk of falls. Estimates of proportions of residents with pressure sores was very low (97%, n=157 identified 25% or less residents). Incontinence rates were high, most homes (n= 126, 77%) reporting proportions of 50% or higher (see Table 6.3).

There was an association between the proportions of residents with incontinence and care home type (X^2(18, n=165) = 50.6, df 18, p=<0.001), not surprisingly nursing care homes were most likely to report high proportions of residents with incontinence, whereas residential

homes for older people with dementia were more likely to report moderate levels and residential homes for older people (not dementia specific) in general reported low proportions (Table 6.4). There were similar associations between the proportion of residents who were at risk of falls and care home type (X^2 (18, n= 162) = 34.8, p=0.01), with those from the nursing homes more likely to report high proportions of residents as being at risk of falls and residential homes reporting lower proportions. There was no association between the proportions of pressure sores, nor those with older people identified as being on multiple medications and care home type.

Table 6.4 Demographics of the residents from the included care homes

What Proportion of Residents	Proportion Identified (%)				
	25% or less N (%)	26-50% N (%)	51-75% N (%)	76% and above N (%)	Total responses question (%)
Incontinence	9 (5)	30 (18)	62 (38)	64 (39)	165 (89)
Pressure sores	157 (97)	3 (2)	2 (1)	0	162 (88)
Risk of falls	39 (24)	66 (41)	38 (23)	19 (11)	162 (88)
Multiple medication needs	2 (1)	20 (12)	43 (27)	96 (60)	161 (88)
Have a confirmed diagnosis of dementia?	49 (29)	40 (24)	39 (23)	41 (24)	169 (91)
Have memory problems but no confirmed dementia diagnosis?	90 (53)	42 (25)	24 (14)	13 (8)	169 (91)
Would you see as having challenging behaviour?	142 (85)	16 (10)	7 (4)	3 (2)	168 (91)
Would you see as having depression/low mood?	123 (73)	36 (21)	8 (5)	2 (1)	169 (91)

Total response given for each question with percentage (%) in brackets.

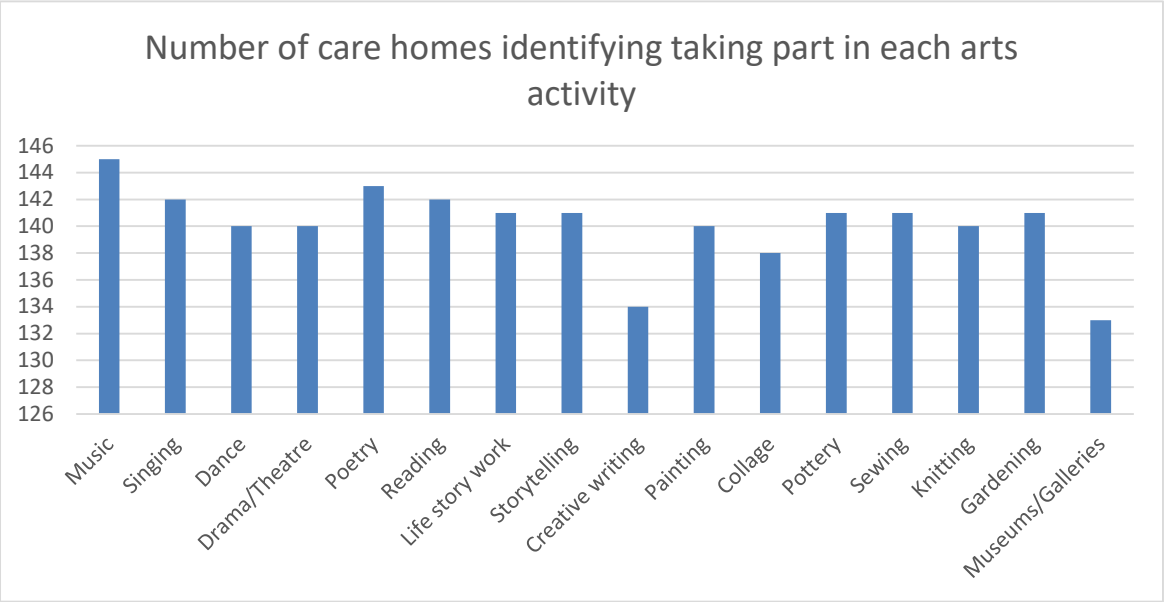
Proportions of residents with dementia were almost equal across the categories, with only 22 (37%) homes reporting more than half their residents had memory problems but no dementia diagnosis. Low levels of challenging behaviour and mood problems were reported, with most homes identifying this to be a problem in a quarter of their residents or less (n= 142 (85%) and n=123 (73%) respectively). Not surprisingly the proportion of those with confirmed dementia diagnoses was associated with the type of care home (X^2 (18, n=169) = 83, p<0.001), those with a dementia specialty were more likely to report high proportions of residents with

dementia. The proportion of residents with memory problems, challenging behaviour and depression/low mood was not associated with care home type.

6.3 Characteristics of the arts activity

A total of 29 homes (16%) indicated they received some additional funding for carrying out arts based activities, however 11 of these specified this was obtained through fundraising carried out by the homes itself and not from an external body. The number of homes taking part in each arts activity is displayed in Figure 6.3 below and frequency of delivery by arts type is reported in Table 6.5.

Figure 6.3. Graph showing the number of homes partaking in each arts activity.



Music was the most popular activity, with a total of 143 homes (77%) indicating they delivered some musical activity. This was also delivered with the highest frequency levels, a total of 98 (69% of those partaking in music) carried out music at least once a day, with the majority (n=67, 47% of those partaking in music) reporting activity levels of more than once a day. None of the homes indicated that music was delivered less frequently than once a month. Singing was also popular amongst the homes, identified as an activity by 142 homes, with 85%

(n=121) reporting this took place weekly or more frequently. Other frequently offered activities reportedly being carried out weekly or more frequently included reading (78% of those reporting the activity, n=111), life story work (61% n=86), painting (56%, n=79) and dance (50%, n=64).

Table 6.5 Frequency with which each arts activity was delivered

Frequency	Arts Type															
	Music	Singing	Dance	Drama or Theatre	Poetry	Reading	Life story work	Story-telling	Creative writing	Painting	Collage	Pottery	Sewing	Knitting	Gardening	Engaging with museums or galleries
More than once a day	68	36	7	0	0	32	13	6	1	6	1	0	0	4	2	0
Once a day	32	23	7	1	4	35	12	11	1	3	2	0	5	11	2	1
Every other day	10	10	7	0	4	12	8	8	0	9	6	1	3	3	3	0
Twice a week	17	18	20	2	6	7	16	8	6	12	4	1	3	14	9	2
Weekly	15	34	30	11	24	25	37	26	7	49	30	2	14	22	18	8
Fortnightly	2	5	7	9	15	7	6	12	2	20	23	3	6	19	11	6
Monthly	1	13	18	34	18	9	28	21	15	18	28	3	13	11	34	22
Less often	0	3	44	84	72	15	21	49	104	23	44	131	97	56	62	95
Total number	145	142	140	141	143	142	141	141	136	140	138	141	141	140	141	134

Pottery was recorded as an activity carried out by a total of 141 (76%) homes, but delivered with the least frequency, with 93% (n =131) of homes which carried it out reporting this activity occurred less than once a month. Other infrequently delivered activities by care homes included creative writing (delivered less than once a month, 76%, n= 104), engagement with museums (occurring less than once a month, 71%, n= 95), sewing (delivered less than once a month by 69%, n= 97) and drama (delivered less than once a month by 60%, n=84).

There was no significant association between geographical location and frequency of delivery for any of the arts activities nor between frequency and care home type for any of the arts types.

The survey also reported on staff members responsible for carrying out each activity (Table 6.6). For each arts type, activity coordinators were identified as carrying out the activity the most. Life story work was almost equally as likely to be carried out by care staff members as activity coordinators (n=56 (46%) and n=62 (51%) respectively). Reading and music were also often carried out by care staff (n=36 (31%) and n=40 (30%) respectively). Staff external to the care home were reported in relatively high numbers for delivering drama (n=37, (37%)) and pottery (n=19 (28%)) in comparison to other activities. Volunteers were reported in low numbers for all activities.

Table 6.6 Staff members responsible for carrying out each arts activity

Activity	Staff Member				
	Activity Coordinator	Care Staff	Staff Not Employed Within the Home	Volunteers	Total
Music	86 (64)	40 (30)	7 (5)	2 (1)	135
Dance	74 (67)	23 (21)	11 (10)	2 (1)	110
Drama	55 (55)	6 (6)	37 (37)	2 (2)	100
Poetry	65 (68)	16 (17)	9 (9)	6 (6)	96
Reading	69 (60)	36 (31)	2 (2)	8 (7)	115
Life Story Work	62 (51)	56 (46)	2 (2)	2 (2)	122
Storytelling	65 (68)	25 (26)	1 (1)	5 (5)	96
Creative Writing	52 (64)	10 (12)	13 (16)	6 (7)	81
Painting	94 (80)	19 (16)	3 (3)	3 (3)	119
Collage	80 (77)	13 (13)	8 (8)	3 (3)	104
Pottery	42 (62)	5 (7)	19 (28)	2 (3)	68
Sewing	62 (73)	16 (19)	5 (6)	2 (2)	85
Knitting	71 (74)	17 (18)	4 (4)	4 (4)	96
Gardening	79 (68)	26 (22)	5 (4)	6 (5)	116
Engaging with Museums	55 (65)	13 (15)	14 (16)	3 (4)	85

6.4 Benefits of the arts activities identified by the survey

Questions 17, 18 and 19 were free text responses that asked the respondents to identify whether they felt arts activities were beneficial for residents, care staff and their care home. In a similar way to the full content analysis carried out on the telephone interviews, these were sorted using condensed codes which made up categories. The frequencies of words related to each category were then calculated.

6.4.1 Benefits to Residents

The Table below (Table 6.7) identifies the frequency of words from question 17 which asked care home mangers to outline what benefits they felt arts activities gave to residents within their care home. This question produced the greatest response from the three free-text questions around benefits of the arts.

Words associated with the category of psychological health and quality of life were the most frequently reported benefits, occurring 77 (24%) and 76 (24%) times respectively within the answers. Terms most frequently referred to general aspects of psychological health, most often ‘wellbeing’ and ‘mood’, with lesser frequencies of specific terms including ‘happiness’, ‘(improvements to) depression’, ‘(improved) loneliness’ and ‘(relief of) boredom’. Within the category of quality of life phrases frequently related to increased worth, such as ‘achievement’, ‘self-esteem’, ‘purpose’, ‘value’, self-worth’ and ‘confidence’. Secondary to this was enabling residents to maintain a sense of self through ‘expression’, ‘choice’ and ‘inclusion’ (Table 6.7).

The category of cognition also featured a high frequency of terms (55, 17%), the majority of which focused on engagement and stimulation of residents. Lesser used terms related to specific areas of cognitive functioning including memory, coordination and concentration.

The remaining categories indicated via the question were increased socialisation (43, 14%), which featured several similar words and phrases outlining increases in residents interacting with staff and one another, enjoyment and pleasure (27, 8%), physical benefits (29, 9%) and behaviour (16, 5%).

Table 6.7 Identified Benefits to Residents from the Free Text Response (Question 17)

Category	Condensed Codes	Frequency
Socialisation	"Social engagement"	1
	"Contact"	1
	"Building Friendships"	1
	"Social/Socialisation/Socialising"	8
	"Social inclusion"	1
	"Social Interaction/Interaction"	13
	"Engaging/connecting with community"	2
	"Social skills"	4
	"Helping one another participate"	1
	"Enhanced social circle"	1
	"Promotes communication"	3
	"Prevents from feeling isolated/Reduces isolation"	3
	"Getting together to chat"	1
	"Family participation increased"	1
	"Part of/belonging to a group"	2
		Total: 43
Psychological health	"Mental wellbeing/mental health"	3
	"Happy/Happiness/Happier"	9
	"Smiling/Smile"	3
	"Relaxed/relaxation"	4
	"(enhancing/lifting/improving) Mood"	15
	"Wellbeing"	25
	"(reduced) depression"	5
	"(reduced) boredom"	4
	"(reduced) loneliness"	1
	"(reduced) helplessness"	1
	"General mental state"	1
	"Psychological (improvements)"	2
	"(healthy) mind"	1
	"Contentment"	1
	"Morale"	1
	"Feeling positive"	1
		Total: 77
Quality of Life	"Purpose"	7
	"Value/Valued"	7
	"Self-worth/Worth"	8
	"Respect/Respected"	1
	"Confidence"	7
	"Achievement/(what you have) Made/Accomplishment"	9
	"Belonging"	3
	"Involved/Included"	9
	"(feeling) Worthwhile"	1
	"Dignity"	1
	"Self-Esteem"	6
	"Creativity/creative"	2
	"Usefulness"	1
	"Express/Expression"	4
	"Choice"	3
	"Fulfilment"	1
	"Quality of Life"	6
		Total: 76
Cognition	"Coordination"	2
	"Alert"	5
	"Engaged/Engagement"	15

Category	Condensed Codes	Frequency
	"Reminiscence" "Active brain/Mentally active/Active mind" "Concentration" "Memory" "Stimulation" "Interest" "Cognitive/Cognition/Cognitive functioning"	1 5 3 5 15 1 3 Total: 55
Enjoyment/pleasure	"Enjoy/Enjoyment" "Fun" "Laughter/Laughing" "Look forward (to activities)" "Pleasure" "Contentment" "Joy"	11 5 2 4 2 2 1 Total: 27
Physical	"Balance" "Keep (physically) active/Active body" "(improved) mobility" "Motor skills" "Better fluid intake/Food intake" "Less medication" "General health/Health needs" "Keeping fit" "Physical benefits"	1 8 4 3 2 1 7 2 1 Total: 29
Behaviour	"Behaviour" "(show) Less anxiety" "(show) Less agitation" "(show) Less aggression" "Calmer" "Relaxed" "(manage) Challenging behaviour"	3 3 4 1 1 2 2 Total: 16

6.4.2 Benefits to Staff Members

Fewer benefits to staff members were reported compared with the number of benefits to residents. Responses to this question fitted into three distinct categories: connecting with residents, enjoyment and job satisfaction (Table 6.8). The greatest frequency of words was attributed to the ability for staff members to connect with residents (n= 67, 54%) through improved socialisation and sharing of experiences and memories. In parallel with a benefit for residents, enjoyment was also frequently mentioned as a staff benefit (n= 30, 24%). Finally,

job satisfaction (n= 26, 21%) was outlined as being improved when taking part in the arts, most frequently mentioned was a sense of 'fulfilment' experienced by staff.

Table 6.8 Identified Benefits to Staff Members from the Free Text Response (Question 18)

Category	Condensed Codes	Frequency
Connecting with residents	"Interaction" "Social/Socialising/Socialisation" "Chat" "(improved) Communication" "Engage/Engaged (with residents) "Quality time" "(good/positive/strong/close) Relationships" "Get to know (residents/service users" "Exchange life experiences" "Learning (about residents) "Bond/Bonding" "Share/Sharing" "Connect (to residents) "Talk" "Involved"	14 7 1 3 7 2 5 7 1 5 4 4 1 3 3 Total: 67
Job Satisfaction	"Fulfilled/Fulfilment/Fulfilling" "(improved) Morale" "Development" "Job satisfaction" "Variety" "Diversity" "Less pressure" "Confidence" "Stops boredom" "Something other than care work" "Improving (care)" "Diversity (to their role)	7 2 2 5 1 1 1 2 1 1 2 1 Total: 26
Enjoyment	"Fun" "Creative" "Engaging (with activities)" "Wellbeing" "Enjoy/Enjoyment" "Laughter"	9 1 3 2 14 1 Total: Total 30

6.4.3 Benefits to the care home

The final question surrounding potential benefits asked whether the respondents felt there was any benefit to their care home in general when taking part in the arts. Responses to this question were sorted into three categories: the atmosphere of the home, linking with the

outside community and quality (see Table 6.8). Improvements to the atmosphere of the home was mentioned most frequently within the responses (30, 24%). Arts activities were shown to make the overall atmosphere within the homes 'happier' and 'nicer'. Participation was also described as a way of linking the home with the outside community (n= 14, 25%), particularly with residents' 'families' and 'visitors' but also mentioned were links with the 'community' and 'engagement with local artists'. Finally, a few of the responses outlined 'quality' (n=11, 20%) including improvements to the 'reputation' of the home.

Table 6.9 Identified Benefits to the Care Home from the Free Text Response (Question 19)

Category	Condensed Code	Frequency
Atmosphere of the Home	"Atmosphere" "Make the home their own" "Feeling of home" "Fun" "Happy/Happier/Happiness" "Content/Contented/Contentment" "Display (artwork)" "Boosts morale" "Contentment" "Brighter corridors" "Nicer/Nice" "Relaxed"	18 2 1 3 26 3 3 3 2 1 5 2 Total: 30
Quality	"High quality service" "CQC" "Reputation" "Marketing tool" "Sells beds" "Good occupancy" "(activities) attractive to potential clients"	1 2 4 1 1 1 1 Total: 11
Linking with outside community	"Community" "Family (involvement)" "Relatives" "Visitors" "Engaged with local artists"	5 2 3 3 1 Total: 14

6.4 Telephone Interviews

This part of the results section focuses on analysis of the qualitative data obtained via telephone interviews. A total of 71 care home managers left their details within their survey responses. The care home managers were then contacted in the order their survey responses

were returned. Upon contacting these respondents, ten indicated they would take part in the interviews. All were female, three were from the North West, two each from the South East, West Midlands and South West and one from Yorkshire.

Transcripts from telephone interviews were analysed using content analysis. As outlined within the methods Chapter 5 (See Chapter 5, 5.7 qualitative analysis), each interview was recorded and listened to, with meaningful quotes from the responses to questions transcribed. Following this the text was read and reread, before codes were assigned. These codes were then further sorted into categories. Five categories were identified as: psychological wellbeing, adaptability and inclusivity, improving the caring relationship, communication and barriers. These made up the overarching theme of 'impact of arts for health in care homes'. Each category and corresponding code can be seen in the diagram (Figure 6.4) below.

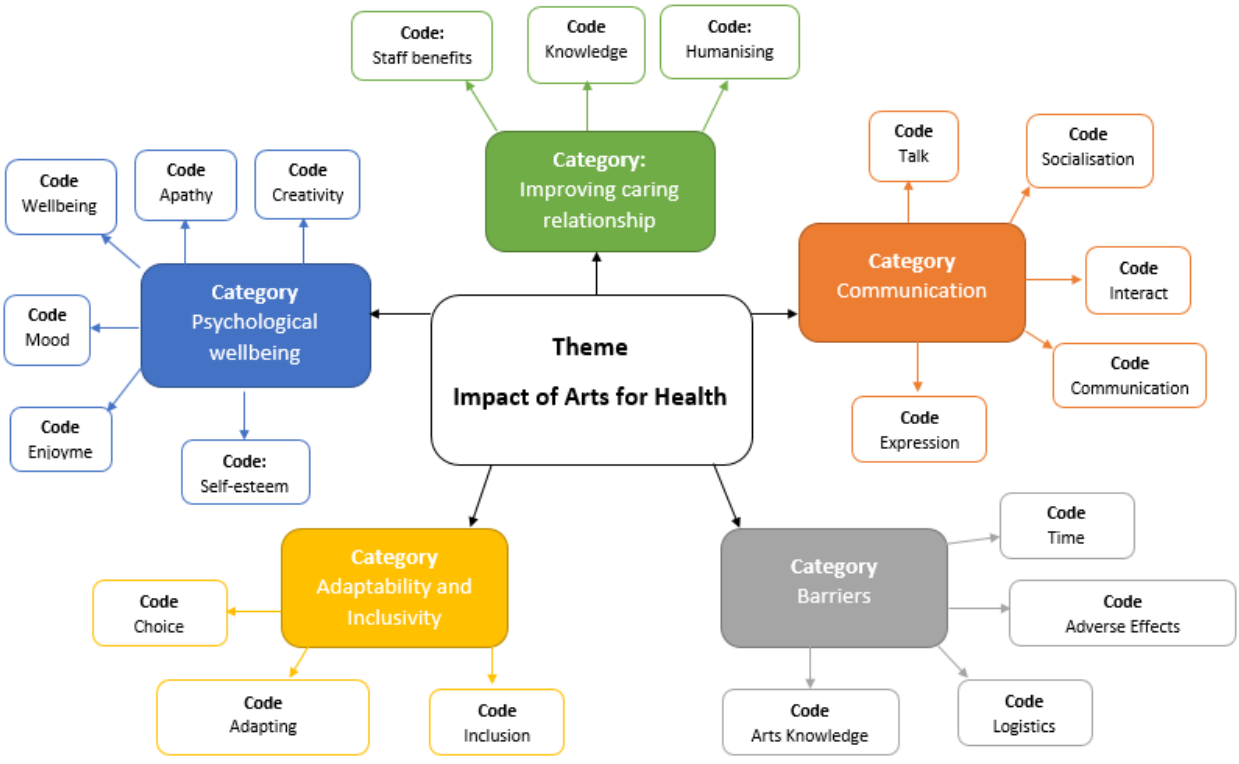


Figure 6.4 overarching theme 'impact of arts for health' with corresponding categories and codes

6.4.1 Psychological wellbeing

Improvements in psychological health and functioning of the residents were mentioned frequently. Reference was made to the ways in which taking part in the activity improved the overall mood of the residents in the moment they were actively participating or engaging with the arts.

> *"You can see the difference in residents from the beginning of the activity to halfway through the activity. You can see the change in mood and activity levels." (R3).*

Also discussed was the ability for the arts activities to challenge residents, by allowing them to partake in something different from the norm. The managers spoke of the routines adopted by the home and how this could cause the residents to become apathetic with carrying out the same activities every day. Taking part in the arts was seen as a novel activity which peaked the interest of those living within the home and impacted upon their energy levels and wellbeing.

> *"… most of our residents are in their 90s, four are over 100 so by the time they get up and get showered and dressed, or bathed and dressed they say they are shattered and want a snooze, which is fine, but if something is going on they are suddenly interested and want to take part." (R2).*

> *"It breaks up the daily routine. I know myself if I stay indoors…. watch tv and things… it'll drive you stir crazy. It's stimulation which has to have a knock-on effect on both your physical and mental wellbeing." (R8)*

Along with improvements to mood and reductions in apathy, there was enjoyment at participating in arts for health. This was discussed with reference to improvements in self-worth; art allowed those taking part to gain a sense of taking part in something worthwhile which in turn enhanced their self-esteem.

> *"It's the pleasure that they get from creating something as well. These people that feel that they're useless and can't really do anything that can then produce (art)…." (R4)*

Also described was a sense of pride and achievement when participation in arts activities resulted in a finished product. This was evidenced when the pieces of art were displayed for visitors, staff members and other residents of the home to observe.

> *"The pride, the self-esteem, the changes were astronomical. Especially when we framed (their) pictures and put them on the wall. People that had been withdrawn and very quiet, to realise that people thought they were good at something." (R1)*

6.4.2 Communication

The various ways in which participation could aid communication were also discussed. Production of art was described as offering a way for residents to express themselves, particularly for those with advanced dementia or other illnesses which may prevent traditional means of communication from taking place.

> *"Sometimes with dementia especially, the residents can become quite catatonic and withdraw into their own little worlds. Or, all they're doing is shouting because they're frustrated about something. So, it helps the staff to find out what is going on. Because they (residents) have got another way of expressing a feeling." (R1)*

Further to this, when residents took part in arts activities they could personalise the end products, allowing for self-expression, which may be hindered by inabilities to verbalise. One respondent discussed how cake decorating had allowed residents to express their personalities.

> *"I know it isn't art therapy, but something as simple as baking cakes and decorating the cakes… just to individualise the cakes…You can tell who has created them. Sometimes people who you think are far down the dementia trail… who you haven't spoken to in a conversational way, it's just been yes and no answers, or little words, sometimes they can express something in what they're doing." (R1).*

Residents were also noted to socialise more with one another, particularly when the activity took place within a group session. Furthermore, older people were also described as interacting more with staff members. It was discussed how some of the residents liked to

'keep to themselves' and didn't engage much with other members of the home or staff members, yet when they were partaking in activities they socialised more.

> *"I think it encourages them to interact, it encourages them to participate, it also gives them the chance to verbalise…basically the whole social conduct of it all is imperative" (R8)*

> *"It's the chatting, it's the way that they engage and they talk to us about things. These are people that very rarely talk. So, we've noticed a big change in that." (R1)*

6.4.3 Improvements to the caring relationship

It was discussed how, secondary to promoting communication, staff members could improve their knowledge of the residents they were caring for when jointly participating in arts activities. This was particularly true of activities that were noted to promote discussions of residents' lives outside of the home (past and present).

> *"Recently, we have been doing a lot of life story things… talking to the residents about childhood. What colours. What textures (remind them of their childhood). Kind of a therapy, we talk to them about the past, about their childhood, what they would associate." (R1)*

> *"It's surprising when you're doing a session what comes out of that, if you've got the right people and you're listening, you can pick things up can't you… what they've done and what they enjoy doing. It's having that skill and being able to pick it up, it's not just a job." (R4)*

Consequently, improvements in communication and staff members' knowledge of the residents resulted in them seeing them as human beings with interests, likes and dislikes. The caring profession was described as one which often promoted those carrying out specific tasks, such as personal care and medication distribution, to become detached from the service users as human beings. Art was described as a way in which carers gained knowledge of the residents other than that which was task-oriented.

> *"It (art) forms a link, sometimes somebody could have been caring for somebody for 6-12 months and not really know anything about them. They might know how many sugars they have in their tea, they know what time they like to get up, so they don't really engage in things." (R1)*

This was further described as being beneficial to staff members as it resulted in a higher level of job satisfaction. Aside from simply gaining knowledge of their service users they were also able to learn from them. There was a sense of enjoyment facilitated by the discussions.

> *"They (staff) gain a lot of information from the residents when they interact, you learn from each other and it's good therapy for them. They get pleasure out of it as well." (R6)*

6.4.4 Adaptability and inclusivity

Managers discussed the need for activities to be tailored to the individual needs of the residents. This was particularly related to both physical and mental limitations perceived within the care home residents. However, in contrast to this being presented as a reason for not joining in arts for health activities, interviews focused on the ways in which activities were adapted such that they were enjoyed by all:

> *"Our activities are designed across the spectrum. We do have clients who are fully competent but we also have those who are mentally frail too. So, we would gear up different types of activities for different people." (R2).*

> *"We would try activities with everybody. We try and look at… for example we went out and it was wheelchair friendly, so we took the ones in wheelchairs who might not always be able to go. This week when we go out it is not wheelchair friendly, so we would take residents who are able bodied. I think residents know their own limits anyway to be honest." (R4)*

When residents were described as very unwell, or disabled, there was a description of how activities became more passive, requiring less active participation to allow those who may struggle to involve themselves physically or mentally with the activity to take part:

> *"A lot of the activities are non-practical, such as listening to music. As we have a lot of residents with high complex needs here." (R6)*

> *"One resident enjoyed going to the theatre and was no longer able to do so because she was frail… so we had a group reading plays to offer as a sort of compromise for her." (R3)*

There were also descriptions of residents that the managers felt would be unable to partake in group activities, however these were able to be offered other forms of art via one-to-one activities.

> *"Not everybody would be eligible for taking part. For example, we have those who are receiving palliative care so they wouldn't be able to join in. We might do something quite gentle with those people, such as at Christmas when we had carol singers in, they sang to everybody and then they went off separately and sang to her (person receiving palliative care)" (R9).*

> *"We have a resident who lost her husband a year ago and who doesn't want to participate. We would go into her room and offer her one-to-one activities." (R8)*

Within the interviews, managers spoke of the care home residents with a great deal of respect. Overwhelmingly residents were given choice as to how they participated in arts based activities. The input from residents with relation to activity selection and delivery varied greatly between the homes, with some indicating service users were consulted on every aspect of the activities, from the planning of activities to how they were delivered and with what frequencies.

> *"We meet with residents once a month at a staff and residents group. We talk to them about activities….ask them what they enjoyed, what they weren't so keen on, what they would change and we ask them for their ideas. Whatever they come up with we try and include in the programme." (R3)*

All those interviewed reported that residents could self-select to join in activities, sessions were always optional and no one was made to take part. Care was also taken to ensure that residents were given the choice to 'opt-out' of activities. Whilst there was great recognition surrounding the potential benefits from partaking in the arts, many interviewees also discussed the importance of residents who did not wish to take part. This formed the basis for where sessions were carried out, with many managers stating they tried to avoid communal areas so that exposure was kept to those who wished to take part.

"We very much value arts activities. However, I recognise that some people actually don't want them. And that's important to recognise. The CQC would probably say that everyone needs them and I don't agree. I think that, there are a small number who like to sit in their own space and not be actively involved in something that may distress them or challenge them." (R7).

6.4.5 Barriers

One of the main barriers to undertaking arts activities was the perception of not having staff with arts based expertise to carry out or facilitate such sessions. Art was perceived as a specialist subject and discussions focused upon managers expressing a desire to carry out more activities but feeling they lacked the skills to do so themselves.

"I would find it very difficult to sit down and do artwork with somebody because I don't know where to start, I don't know how to draw. That's how I came up with the idea of the scrapbooks, with the different materials, because I can stick things in a book, but I can't draw…. there isn't any kind of training for staff to help them break down that barrier. I would willingly go somewhere and learn about art therapy" (R1)

"We don't do a lot of art as in drawing pictures…. Because we haven't got anybody who is competent enough with technique and things like that..." (R4)

The delivery of arts based interventions was therefore seen as a specialised activity which was to be carried out separately to normal caring tasks. Within most of the homes arts activities were delivered by activities coordinators, or by volunteers and the frequency and delivery of the arts was based upon the availability of such staff. Those with in-house activities coordinators and a high level of voluntary support therefore reported the greatest proportion of arts provision.

"We do loads of activities with our residents, I've got three activities coordinators so that's 60 hours a week." (R4)

"We have two activities coordinators at the moment, they carry out activities Monday to Friday and we have lots of external bodies that come in…. including a singer" (R6)

Homes without such staff members discussed how they struggled with the need for care staff to appropriately allocate time within the working day to carry out other tasks and therefore felt they needed to fit in arts activities around this.

> *"There would be more benefits to staff members if they had more time to sit and create. The difficulty is we are a very busy home…. in an ideal world, we would spend all day doing arts and crafts but that isn't really the main purpose of why we are here." (R2)*

The final barrier was the logistics of carrying out the activities. This was not related to activities which took place within the home, most homes expressed they had dedicated areas where arts activities took place (although individual descriptions of these areas were not provided), but was associated to arts exposure away from the home. This mainly centred around the physical requirements of the residents within the home. The survey reported high levels of physical disability within the homes and this was described within the interviews as posing a problem when it came to leave the home for outings.

> *"I think a lot of residents, they always ask can we go out. That creates a logistical difficulty because most of them would require a wheelchair. Wheelchair taxis will generally only take one wheelchair which makes them very expensive. Minibuses are also expensive and often only take two wheelchairs." (R3)*

6.5 Summary

This chapter presented the results from the quantitative survey along with the responses to free text questions and the telephone interviews conducted with care home managers from a stratified sample of care homes within England.

A total of 185 (5%) care homes/ care home managers responded to the online survey, representing nursing homes, homes offering residential care for older adults and dual registered care homes, some with and some without a dementia specialty. Responses included homes which were privately owned, not-for-profit and dually owned. Most of the homes reported operating at capacity (or near to capacity). Demographic data were

provided with respect of the residents from the care home which reported high medication needs, but relatively low levels of challenging behaviour or mood disorders.

Several different arts types were reportedly offered within the responding homes and frequency of delivery also varied for each arts type, with music carried out both by most homes and with the highest frequency. Most activities were carried out by a dedicated activity coordinator.

Benefits to residents, staff and the home itself were reported through free-text responses within the survey. Benefits to residents included the categories of psychological health, quality of life, cognition, socialisation, enjoyment/pleasure, physical health benefits and behaviour. Those identified to staff members were connecting with residents, job satisfaction and enjoyment. With respect of benefits to the home itself managers reported improving the atmosphere of the home, improving quality and linking with the outside community.

Follow-up semi-structured interviews were conducted with a total of ten care home managers. These were analysed, and the following themes were formed: psychological wellbeing, adaptability and inclusivity, improving the caring relationship, communication and barriers.

These results are discussed within the following chapter (Chapter 7), along with the findings from the systematic review. Results from both studies are synthesised and discussed in relation to policy, research and practice into the potential benefits of arts for health delivery in care homes. Following this Chapter 8 will make overall conclusions and recommendations for policy and practice.

Chapter 7- Discussion

7.1 Introduction

Presented within this thesis are two studies, the first (Study 1) was a systematic review of arts for health studies to determine the impact and benefits of such activities on the health, wellbeing and quality of life of older adults in care homes. The second (Study 2) was an electronic survey of a stratified randomised national sample of care home managers and follow up telephone interviews with a self-selecting sub-sample of respondents.

The following chapter discusses and synthesises the findings from both studies along with current literature and arts for health policy and legislation. The limitations of both studies will also be discussed, along with directions for future research and implications for policy and practice. The final chapter (Chapter 8) outlines recommendations for policy and practice along with overall conclusions from findings of both studies.

7.2 Summary of Study 1 and Study 2

Within Study 1 studies were eligible for inclusion if they evaluated an arts activity and included study population were older adults (with a mean age of 60 or over) who resided within a care home. In addition, studies had to feature a comparison condition and an outcome related to health, wellbeing or quality of life. Both quantitative and qualitative studies were eligible for inclusion (see chapter 2 for specific inclusion criteria). Potentially relevant studies were screened for eligibility using a bespoke screening tool, with 71 studies meeting the criteria for inclusion. Included studies underwent data extraction and quality appraisal (see Chapter 2 for a detailed description of these processes) and the results of the systematic review are presented in Chapter 3 (overall narrative synthesis) and Chapter 4 (narrative synthesis by arts type). The main outcomes assessed by the studies within the review were behavioural changes (most often associated with dementia) (n=32, 45%), psychological wellbeing (n=24,

34%), cognition (n=10, 14%), improvements to the caring process (n=9, 13%) and quality of life (n=8, 11%). Lesser reported outcomes included feasibility of carrying out arts for health activities (n=6, 8%), physical health (n=4), satisfaction with living environment (i.e. the care home) (n=2), socialisation (n=2) and neuropsychiatric symptoms associated with dementia (n=2). A typology of arts activities was also formed which consisted of: music, performance, spoken and written literary word, multisensory activities and applied arts and crafts.

Study 2 was an electronic survey within a stratified randomised sample of care home managers with follow-up interviews conducted with a self-selecting sub-sample of respondents. The objectives of the electronic survey were to ascertain what (if any) arts activities were delivered within care homes in England and map their prevalence and determine the characteristics surrounding delivery of such activities along with identified impact. Follow-up interviews were conducted with self-selecting sub-sample of responding care home managers to provide contextual information to that obtained through the survey responses. A total of 185 homes responded to the survey (5% of the total mailed out n= 3,817) and all responding homes carried out arts-based activities (characteristics of both the care homes and the arts activities is provided in Chapter 6). Ten care home managers were interviewed, and their data analysed using content analysis. Five categories of impact and benefits were identified: psychological wellbeing, improving caring relationships, communication, adaptability and inclusivity and barriers (see Chapter 6, Figure 6.3 for a diagram of themes and codes).

Both studies provide a new insight into Arts for Health, to date there has been no systematic review conducted to assess the impact of arts for health activities within the care home population. A description of arts activities currently delivered within care homes in England

has also not been conducted prior to this study, nor has there been interviews with care home managers to establish their experience whether arts may be beneficial to their residents, staff and home environment. The following chapter discusses and synthesises findings from both studies along with current literature and arts for health policy and legislation.

7.3 Existing Evidence Relating to Arts for Health

7.3.1 Quality of Existing Evidence

Interest in arts for health is increasing based upon the need for modern health care to encompass psychological and social needs (Clift, 2012; Clift and Camic, 2015; Crossick and Kaszynska, 2016; Public Health England, 2016). Both the World Health Organisation (2015) and NICE (2013; 2015) recognise participation of 'worthwhile' activities as important for healthy ageing with arts offering a way of participating in such activities. The option for older adults to partake in worthwhile activities is also recommended by the CQC within care homes (CQC 2017). Increased interest is also supported by the systematic review in study 1 where a total of 57 (80%) included studies were published since the year 2000.

However, whilst there has been increasing levels of research carried out, discrepancies in quality of research related to arts for health have been noted extensively within the literature. Factors cited for lower quality include a lack of description of the arts activity, methodological weaknesses, absence of control groups and lack of self-report outcome measures (Vink, Bruisnma & Scholten, 2009; Mederios & Basting, 2013; National Endowment for the Arts, 2013; Cowl & Gaugler, 2014; Noice, Noice & Kramer, 2014; Young et al. 2016; APPG, 2017). Such problems can also lead to the failure to include some arts for health studies within systematic reviews, which often adopt measures of quality for inclusion criteria (Mederios & Basting, 2014).

Quality appraisal of the studies included in study 1 support previous findings suggestive of varying quality of arts for health research. To ensure a wide range of research the systematic review (study 1, results in chapters 3 and 4) included empirical studies of any design provided they met the inclusion criteria (outlined briefly in 7.1 above and detailed in Chapter 2). For fairness of comparison, included studies were grouped according to study design (RCT, quasi-experimental, pre/post-test and descriptive) and appraised for quality. Risk of bias was correlated with study design, with randomised controlled trials overall showing lowest risk, followed by quasi-RCTs and finally uncontrolled pre/post-test designs.

Of the RCTs included within the review, all but one (Houston, 1998) were published since the year 2000 and just under half within a five-year period (2011-2016), suggesting a move towards more robust study designs. Only two RCTs (Cooke 2010; Subramaniam, 2014) obtained low risk of bias for all domains assessed by GRADE. Just eight studies received low risk of bias judgements for allocation concealment and seven for blinding (see Figure 3.2 in Chapter 3, Section 3.9), however it could be suggested that for arts for health interventions these factors are less influential for the study quality. Within their systematic review, Medeiros and Basting (2014) outline that arts for health research should not be judged solely on how well it adheres to the design of a randomised controlled trial. In addition, Spector et al. (2000) outline the difficulties with allocation concealment and blinding when conducting arts for health research.

Whilst RCTs generally showed less bias than other study designs, the quality of any study design (RCT, quasi-RCT and pre/post-test) was not correlated with publication date (see Chapter 3, Section 3.9 for a detailed description of research quality appraisal and Figures 3.2, 3.3 and 3.4). It could be argued that the use of GRADE criteria impacted upon the quality

scores obtained by quasi-RCT's and pre/post-test designs, given it is traditionally used for RCT appraisal only (see Section 7.6, limitations). However, despite this, risk of bias was most often determined due to selective reporting and incomplete outcome data which arguably should not appear in any research, irrespective of study design. Given the difficulty in obtaining data through traditional RCTs in support of arts for health, (Spector, 2000; Mederios & Basting, 2013) this systematic review would confirm there needs to be better quality studies of other research designs to add to the body of evidence.

The importance of good quality qualitative research in capturing the subtle and personal aspects of wellbeing has been stressed within reviews and reports focused upon arts for health (Medeiros & Basting, 2013; Crossick & Kaszynska, 2016). The All-Party Parliamentary Report into arts for health recognises that qualitative evidence may be particularly beneficial for capturing 'existentially significant differences' rather than focusing on statistically significant differences relied upon by quantitative methods (APPG, 2017, p 131). However, little quality appraisal of qualitative studies is shown in existing systematic reviews related to arts for health. Studies within this systematic review were assessed using a quality appraisal checklist developed by Walsh and Downe (2006) and used previously by Cooke, Mills and Lavender (2010) (see Chapter 2 for an overview of quality appraisal methods for qualitative studies). Grading then took place using a system developed by Downe et al. (2009, based on the work by Lincoln & Guba, 1985). Overall the qualitative evidence was good, although none were judged at grade A (free from any biases). Five of the studies (La Cour 2005; Gotell, 2007; deGuzman, 2013; Guzman-Garcia, 2013; Roe, 2015) upon quality appraisal were classified as Grade B for risk of bias, meaning they possessed some flaws which were unlikely to affect the certainty of their findings (see Chapter 2 study 1 methods, for an overview of the qualitative grading criteria). However, three of the studies were judged at grade C (Palo-Bengtsson, 1998;

Martin, 2004; Chen, 2009), indicating the flaws identified may affect the certainty of findings (see Chapters 3 and 4 for detailed quality appraisal of qualitative studies).

There is recognition of the need for greater follow-up in arts for health research to longitudinally determine the effects of arts for health (Vink, Bruinsma & Scholten, 2009; Noice, Noice & Kramer, 2013; Crossick and Kaszynska, 2016; Cowl & Gaugler, 2014; Young et.al, 2016; APPG, 2017). This is reinforced by this systematic review, only two of the studies assessed existing arts programmes (Palo-Bengtsson, 1998; Myskja, 2007). The remaining studies implemented programmes for short periods of time with little analysis beyond the study duration.

7.3.2 Scope of Existing Evidence

Within the systematic review 68% (n=48) of the included studies focused on music interventions. Consistent with findings from other systematic reviews (Beard, 2012; Cowl & Gaulger, 2014) this makes comparison of different arts for health activities difficult. The reliance on music studies also makes it difficult for positive findings from the systematic review to conclude evidence of benefit for all types of arts activity. Arts therapies garner more evidence than arts for health activities as they adhere to stricter protocols which lend themselves to traditionally viewed higher quality study designs such as randomised controlled trials (APPG, 2017). In addition, music is the most popular art therapy and therefore assessments of its efficacy occur more frequently than other art forms (Beard, 2012). A systematic review exploring the benefits of arts for health in older adults with dementia by Cowl and Gaulger (2014) found that music was popular in studies where individuals showed greater cognitive decline because of dementia. This occurred because music activities required less active participation compared with other activities such as painting and dance

sessions. This could be the reason for a greater number of music studies within this systematic review as the population of older adults in care homes show a greater incidence of dementia.

Reflective of the large proportion of music-based studies within the systematic review (Study 1), music activities were delivered by the care homes (within Study 2) more so than any other activity and with the greatest frequency. However, there were disparities between the frequency of other arts activities carried out within the homes (from Study 2) and the numbers of studies focused on such activities within the systematic review (Study 1). Specifically reading, life story work, painting and dance were carried out frequently within the homes. The systematic review included some studies which assessed the use of reading (Clair, 1996; Orsulic-Jeras, 2000; Skrajner, 2007; Cooke et al. 2010; Billington, 2013). In general, whilst widely discussed in the survey and interviews, applied arts and crafts were scarce within existing literature (included in Study 1). Just one study (La Cour, 2006) featured painting in any form and as with studies which focused upon dance, the activity was not the main arts type evaluated by the study but rather a feature of a larger activity programme. Beard (2012) evaluated arts and crafts within a systematic review of art therapies in dementia care and recognised that often such activities are evaluated through assessment of ability and self-worth, meaning they may be less represented in empirical studies which often feature validated outcome assessments.

Crossick and Kaszynska (2016) discuss the popularity of social dance sessions for those with dementia, yet these were the main activity in only two of the studies included within the systematic review (Palo-Bengtsson, 1998 and Guzman-Garcia, 2013). Similarly, a previous systematic review found that whilst there were studies evaluating the use of both dance and 'visual arts' these lacked overall quality mainly due to poor assessment of outcomes and

reliance on anecdotal evidence (Beard, 2009). As discussed, the number of studies focused on music is partly due to art therapy protocols leading to more robust study designs and therefore greater likelihood of publication (APPG, 2017).

The differences between impact and benefits identified from the systematic review and the national survey may in part be explained by the way in which outcomes are determined. Within the systematic review, only seventeen studies featured outcome measures completed by the residents themselves. There was a high use of observational methodologies which were often scored by members of the research team. Whilst observations carried out by those who do not know the residents may be traditionally preferred for removing biases, they may in this case make determining benefits difficult given the team would be blind as to the usual behaviour of the residents, or indeed other factors which may influence their behaviours. Existing arts for health research has not addressed whether discrepancies exist between self-report and observer-reported measures and this would warrant further study. Additionally, the use of standardised outcome tools whilst beneficial in terms of quality of evidence are not always relevant given the topics covered which may not be relevant to care home populations (Crossick & Kaszynska, 2016; Meredios & Basting, 2013). Benefits identified from the survey and interviews (Study 2) such as improved 'happiness', 'achievements' and 'self-worth' are difficult to determine via observational methods or by those who do not have existing knowledge of the residents and this may explain their scarcity within the systematic review.

Finally, there was a large discrepancy in the gender of the participants within the literature. Three-quarters of the total number of participants for all studies were female, however this figure is typical of the proportions of females and males within care home populations.

Most recent UK demographics (ONS, 2014) report there to be 2.8 females for each male and therefore the systematic review would be reflective of the older adult care home population. A recent study also confirmed there to be more females than males within older adult residential care (Windle et al. 2018). However, also discussed within the care home demographic is the narrowing of the gender gap between males and females, particularly within the oldest-old (ONS, 2014). If this trend continues it may be useful to have more research with older males to ascertain whether any identified benefits or adverse effects are shown for both genders. This concurs with the findings of Bell et al (2016, p108) and Gandy et al (2017) on their study of active ageing and community activities groups, older men and suitable activities need to become the focus of the future in terms of impacts on their health and wellbeing.

7.4 Arts for Health Typology

Previous arts typologies have shown subtle differences in the way in which arts activities are classified and defined. Davies (2011) in a study based on perceived engagement with art activity, formed categories of 'performing arts', 'visual arts', 'design and craft', 'community and cultural events', 'literature' and 'online digital and electronic arts'. A broader typology is offered by Arts Council England (2015) within their classification of arts activities as: 'performance arts', 'visual arts', 'combined arts' and 'recreational activities'. Studies within this systematic review were separated according to the specific activity being carried out. The systematic review was designed to be inclusive of all activities and therefore no distinction was made between method of delivery of the arts activity (i.e. active or passive participation, group or individual) nor whether the activity was part of a structured arts therapy programme or not. This is in keeping with previous typologies (Davies, 2011; Arts Council England, 2015) who have formed their typologies based on the arts type alone rather than differentiating

between delivery or participation levels. The categories were defined as 'music', 'multi-sensory', 'spoken and written word', 'performance' and 'applied arts and crafts'. Each of these categories will be discussed separately within the following subsections.

7.4.1 Music

The first and most evidenced arts type was that of music. The categories presented within the review from Study 1 overlap with those proposed by previous systematic reviews, which tend to categorise the arts activity itself rather than the way in which the activity delivered or engaged with. A separate 'music' category is always included within such reviews (Beard, 2009; Cowl & Gaugler, 2014; Noice, Noice & Kramer, 2014; Medeiros & Basting, 2014), possibly due to the large evidence base related to this arts type. Such categories will obviously be driven by findings from the review, where papers could be grouped together according to their arts similarities.

Most papers within the systematic review were focused on assessing music activities and in Study 2 music was reportedly carried out with the greatest frequency. Music activity could be classified within Davies (2011) or Arts Council England (2014) categories of 'performance' and therefore could have been placed within the 'performance' category of this typology. However, given the number of studies which featured varying levels of engagement with music, from active participation to passive participation it was decided that music be defined in its own category. The large number of music studies was also reflected in survey responses within Study 2 where music was an activity carried out by most of the homes and with the greatest frequency. Responses from Study 2 also indicate there were often varying levels of participation within the same arts sessions as adaptions and allowances were made for those with lower physical and/or cognitive abilities (this is further discussed in Section 7.5, barriers and adaptions).

Method of delivery for music activities differed across the studies within the systematic review (Study 1) which would be expected given the large number of included studies. Participation ranged from actively playing instruments to listening to CD's. This may account for the popularity of music delivery in care homes (Study 2). Respondents described the time required to carry out arts activities, and therefore listening to music would enable residents to engage with music without the need for structured delivery as other activities may require.

7.4.2 Multi-Sensory Activities

Given the focus of this thesis was older adults a lack of 'community and cultural' arts category as shown in Davies (2011) is explained by the content of activities within the included studies from the systematic review. Roe et al. (2015) was the only included study which evaluated an activity that took place in the wider community. The study highlighted the difficulties related to the logistics of such an activity. Responses from telephone interviews with care home managers in Study 2 reinforce the concerns of activities which take place away from the home (discussed further in section 7.4- barriers and adaptions). Rather than a category focused upon the community, the category of 'multi-sensory experiences' was included within this typology. This decision was based on the inclusion of studies (Orsulic-Jeras, 2000; Kincaid, 2003; Cox, 2004; Simard, 2010; Klages, 2011; Chang, 2013; Scott, 2014) which would reflect an experience normally achieved by visiting a museum, gallery or other cultural location, but without older adults having to leave the care home environment. A similar grouping of studies is shown in Medeiros & Basting (2014) which included studies with similar delivery to those within this systematic review.

It could be argued that some of the studies placed within the multi-sensory experiences category could be classified as 'visual arts', however those incorporating visual arts offered more immersion than simply viewing art and either 'concealed' undesirable features of the

home (Kincaid, 2003; Chang, 2013) or involved sensory stimulation such as fragrances (Scott, 2014). Secondly, activities which were included within this review, may not be included in more traditional 'art therapy' reviews as this review is the first to focus solely on the care home population. Such activities included snozelen, (Cox, 2004; Klages, 2011), a Namaste care programme (Simmard, 2010) and Montessori programme (Orsulic-Jeras, 2000) all of which were included as they provided tactile stimulation and visual engagement. Recent Arts Council England (2017) initiatives would support this category with the development of a travelling theatre company which visits care homes and allows participants to engage with activities which would traditionally be delivered the wider community.

Activities of this type were not widely discussed within the survey responses. Only 85 of the homes (46%) reported engaging with museums, mainly due to the barriers associated with such activities (discussed in 7.4). The aesthetic of the external grounds of the homes were discussed more so than the internal grounds and the opportunity for the residents to take part in gardening activity. Such activity types were not evaluated by studies included within the systematic review and may therefore warrant further study to ascertain potential benefits and impact.

7.4.3 Literary Spoken and Written Word

The second distinct category formed from the systematic review was 'literary spoken and written word'. This included activities related to life story work and storytelling. Such activities appear to have not been included within many previous systematic reviews yet were frequently discussed within the interviews carried out with care home managers. Within Young, Camic and Tischler (2015) literary arts were discussed together, but none of the studies featured creation of life story books or life story work which produced collage. Such activities were commonplace within the care home responses, particularly when used to

gather information about the residents. This may account for the absence of this category in previous systematic reviews which have not exclusively focused on care home populations.

Interestingly, activities of this type were delivered by care staff more frequently than any other type (see chapter 6, Table 6.5). As reported by Young Camic and Tischlet (2015), this is indicative of the potential for activities such as life story work and storytelling to enable care homes to gather information from residents in a creative way (discussed further in 7.4.4 below).

7.4.4. Performance

As with both the definition offered by Davies (2011) and that from Arts Council England (2015) there is the presence of a category related to performance. Within Beard (2012) and Noice, Noice & Kramer (2014) two distinct categories were formed for 'dance' and 'drama' whereas in the present review a category of 'performance' was formed.

The performance category was formed based on six studies (Houston, 1998; Palo-Bengtsson, 1998; Heyn, 2003; Noice, 2006; Guzman-Garcia, 2013; Vankova, 2014) which required the residents to actively partake in an activity and therefore performing. Whereas previous reviews have featured dance and drama as separate categories (Beard, 2012; Noice, Noice & Kramer, 2014) these included studies where older people were passively watching dance or drama performances. Therefore, within this category the performance element was synonymous amongst the studies rather than the *genre* of activity.

Within the survey both drama and dance were relatively popular activities (identified as an activity by 100 (54%) and 110 homes (59%) respectively) but varied in frequency of delivery. Dance was delivered more often than drama and may overlap with the delivery of music activities. Drama was more likely to be seen as a 'specialist' activity which may reflect the

lower proportion of homes stating they participated in this arts type. This is further highlighted by the fact that drama was more likely to be delivered by external staff (from specialist companies) more so than any other of the arts activities.

7.4.5 Applied Arts and Crafts

The only other category, formed by the literature within this systematic review, which overlapped with previous typologies was that of 'applied arts and crafts'. Within previous systematic reviews, art activity such as painting and drawing is classified as 'visual arts' (Beard, 2009) and this category also exists within Arts Council England (2015) definition of the arts. The decision to form the category of 'arts and crafts' was again driven by the focus of the studies within the systematic review. The three studies which formed this category involved residents actively engaging with arts materials in order to create something (Doric-Henry, 1997; La Cour, 2005; de Guzman, 2011) and therefore differed from 'visual arts'.

This category featured the least amount of studies (just three from a total of 71) and yet activities included within it were delivered often within care homes, particularly were painting (an activity which took place in 119 (64%) homes) and collage (reported by 104 (56%) homes). This further reflects the preference within the literature for traditional 'art therapy' (APPG, 2017), particularly music, and less focus on therapeutic activity such as arts and crafts.

7.5 Impact of Arts for Health Activities

The focus of the studies included within the systematic review were: behaviour, psychological wellbeing, quality of life, cognition, socialisation, improvements to care and physical improvements. Overall themes from the survey responses and interviews with care home managers were psychological wellbeing (including wellbeing, apathy, mood, creativity, enjoyment and self-esteem), improvements to the caring relationship (staff benefits, knowledge and humanising) and communication (including socialisation, communication and

expression). This section will discuss the systematic review and survey findings to determine specific impact and benefits related to arts for health.

7.5.1 Behaviour

The most frequently cited outcome measures from the studies in the systematic review were changes in behaviour, mainly behavioural symptoms associated with dementia including agitation and verbally disruptive behaviours (VDB). Previous reviews have highlighted the efficacy of arts for health as a behavioural intervention, particularly in individuals with dementia (O'Connor et al. 2009a; Beard, 2012; Cowl & Gaugler, 2014). This systematic review would offer some support for this assumption, particularly with respect of agitated behaviours which reduced over time for several studies (Goddaer, 1994; Hicks-Moore, 2007; Garland, 2007; Ledger, 2007; Ho, 2011 & Vink, 2013).

Modifications to behaviour of residents was infrequently reported as a benefit within the national survey and was not discussed within the telephone interviews. Benefits associated with behaviours identified by both the survey and telephone interviews often related to changes which impacted upon the resident rather than care staff and tie in with psychological improvements such as looking less anxious, less agitated and less aggressive.' The most frequently reported behavioural impact was shown for improvements in agitation and reinforces the findings of the systematic review and previous reviews that have shown reductions in such symptoms (O'Connor et al. 2009a; Beard, 2012; Cowl & Gaugler, 2014).

Less evidence was demonstrated by the review with respect to verbally disruptive behaviours, three studies failed to show an improvement (Svansdottir, 2006; Garland, 2007; Nair, 2011) and two showed some benefits but with limited analysis (Cohen-Mansfield, 1997; Hagen, 2003). This was mirrored in the survey responses and telephone interviews which did not

report improvements to verbally disruptive behaviours as a potential benefit. Reporting of such improvements may reflect the change in focus of arts for health from treatment and control of symptoms towards improvements in wellbeing and psychological health (Medeiros & Basting, 2014; APPG, 2017).

7.5.2 Psychological Wellbeing & Quality of Life

Within the systematic review, the most cited quantitative measures of mood were changes in depression levels. Of the studies which explored this outcome, just over half (a total of nine studies) showed a positive impact of arts activities compared with eight studies who did not show any changes. Despite high reported incidences of depression in older people residing in care homes (Atkins et al. 2013; Ferreira, Dias & Fernandes, 2015; Age UK, 2016; Thraves, 2016), depression was not reported as being frequently prevalent in the population of care home residents within the national survey, nor was it reported frequently as being impacted upon by the arts. Given depression in older adults is often mediated by loss (Flood & Philips, 2007; Myskja & Nord, 2008; Basting, 2009; DeGuzman, 2009; Age UK, 2015) and feeling less useful (Age UK, 2015; ONS, 2015b; ONS, 2018) the fact that all the responding care homes offered participation in the arts may account for this.

Two of the studies included within the systematic review showed improvements to residents with pre-existing clinical depression (Suzuki, 2006; Liu, 2010), suggestive that mood is improved most in those with baseline scores indicative of clinical depression from the outset. This was further evidenced by findings that individuals with low depression scores (indicating no clinical depression) did not show improvements in their scores when taking part in the arts (Orsulic-Jeras, 2000). This may also account for depression being infrequently reported as a benefit within the survey responses and telephone interviews given relatively low levels of depression were reported. A recent systematic review of the impact of music on wellbeing

(Daykin et al. 2018) commented on the need for future research to include participants who had low baseline levels of wellbeing, otherwise significant changes are not observed.

The systematic review found good evidence of benefit for anxiety in particular, with a number of studies showing significant changes (Houston, 1998; Svansdottir, 2006; Guetin, 2009; Sung, 2010; Mohammadi, 2011). Interestingly anxiety was always assessed through self-report measures which as discussed earlier (in section 7.2.2) may reflect participants' perceived benefits which may differ from those which have been observed. Again several studies commented on low baseline levels of anxiety within residents which may have reduced the likelihood of statistically significant changes. For both depression and anxiety comparisons with control groups evidenced maintenance of wellbeing and psychological health suggestive of prevention rather than cure.

Descriptors of benefits to residents from the survey and telephone interviews focused less on specific mood disorders, such as reductions in anxiety and depression and were more associated with the impact of the activity on general descriptions of wellbeing which overlapped with quality of life concepts. Managers discussed factors such as self-expression, self-worth and feelings of inclusion. Such concepts are difficult to capture within clinical outcomes (Mediros & Basting, 2014; Crossick & Kaszynska, 2016). This is reinforced by the fact that studies within the systematic review that quantitatively measured quality of life showed limited improvements for this outcome (Bennett, 1998; Hagen, 2003; Hilliard, 2004; Philips, 2010; Chang, 2013; Subramaniam, 2014). However, they could be overlapped with studies within the systematic review which featured general mood questionnaires (Lord, 1993; Suzuki, 2006; Hammar, 2010; Sole, 2014) and particularly those that adopted qualitative methods (Martin, 2004; LaCour, 2005; DeGuzman, 2009; LaCour, 2010; Billington,

2013; Houser, 2014). Such factors were recently explored within a longitudinal evaluation of a visual arts programme as described as 'in the moment' benefits which occur within the residents whilst partaking in the experience of the arts (Windle et al. 2018). This further demonstrates the importance of both quantitative and qualitative evidence and mixed methods approaches.

Whilst arts activities may be beneficial to older people, there was limited evidence that concludes they were more beneficial than other forms of activity (such as talking with family members or taking part in relaxation). This would be supported by the findings within the systematic review (Study 1). Perhaps the largest differences between arts activities and their comparators, however, are that the arts offer the opportunity for creativity. As with certain aspects of wellbeing, this would be difficult to quantitatively measure as an outcome, but qualitative findings again highlighted this as a benefit or impact (La Cour, 2005; Chen, 2009; Billington, 2013; Houser, 2014). The ability for residents to be creative and engage with the arts was discussed within the survey responses. This reflects existing recommendations that research focused on arts for health move away from clinical outcomes to evaluating more personal experiences (Cowl & Gaulger, 2014; Crossick & Kaszynska, 2015; APPG, 2017).

7.5.3 Cognition and Socialisation

The systematic review reflects not only an increase in interest of arts for health, but also how the focus of such studies has changed. Studies which focused purely on the ability to improve cognition were carried out in the 80s and 90s (Smith, 1986, Clair, 1999) whereas more recent studies assessing cognition did so whilst evaluating other outcomes such as affect (Suzuki, 1998; Noice, 2006; Philips, 2010; Subramaniam, 2014) and behaviour (Hagen, 2003; Koike, 2012). In contrast with a previous review evaluating the impact of arts for health activities for

older adults (Cowl and Gaugler, 2014) this systematic review showed improvements in different aspects of cognition because of participating in the arts.

Cognitive benefits outlined within the survey related to how engaged with the activity the residents were and included 'engagement', 'stimulation' and remaining 'mentally active'. Engagement was also found to be improved by all studies which assessed this outcome within the systematic review (Orsulic-Jeras, 2000; Holmes, 2006; Skrajner, 2007; Fritsch, 2009). This would reinforce the recommendations that older people partake in activities that are 'worthwhile' (Basting, 2009; WHO, 2015; NICE, 2015), engagement would outline the residents saw the worth in taking part and wanted to maintain their participation.

Previous reviews focused on arts participation and older people have highlighted the benefits of arts for health in improving socialisation in older people (Beard, 2012; Cowl & Gaugler, 2014). Despite being discussed within the survey responses, there were few studies within this systematic review which focused on socialisation and social engagement. Improvements to socialisation were mainly discussed within descriptive studies (Gotell, 2007; Hammar, 2010; Billington, 2013; Guzman-Garcia, 2013; Roe, 2015) and included improved communication between residents and staff members.

Improved socialisation was also discussed within the national survey responses, again reflecting the priorities of the home. The CQC recommends that steps should be taken within care homes to maintain social relationships with 'people that matter' and 'avoid social isolation' (CQC, 2017, code R1.4). The care home managers discussed how arts allowed the residents to maintain links within the community, which was also shown in Roe et al. (2015) where the activity took place externally to the home.

7.5.4 Improvements to care and physical improvements

Perceived benefits to the caring process were reported differently within both the systematic review and the national survey. Several studies within the systematic review assessed quality of care with respect of improvements to nutritional intake (Ragneskog, 1996a; Ragneskog, 1996b) and reducing restiveness to care (Gotell, 2007; Hammar, 2010; Houser, 2014), neither of which were mentioned within the survey responses or telephone interviews. The most discussed impact or benefit of arts for health activities for staff members within the survey responses and telephone interviews was the ability for such activities to aid communication between care staff and residents. Qualitative studies within the systematic review also found this to be a benefit, along with the ability for such activities to prompt sharing of memories and experiences (Martin, 2004; Chen, 2009; Billington, 2013; Guzman-Garcia, 2013; Roe, 2015).

A previous systematic review assessing older adults and arts participation by Young, Camic and Tischler (2016) showed that literary arts were particularly beneficial to communicating basic needs between residents and care givers. Art was noted by the care home managers as aiding communication both within residents in the home and between residents and staff members. One way in which arts activities were used was to gather information about those living within the home to adopt a person-centred approach for future care and activities. This is reflected by the survey which showed life story work was often carried out by care staff as an information tool. The use of arts, specifically activities such as creation of memory books and collages, were beneficial in allowing information about residents to be obtained in a creative way, rather than relying on filling in forms or asking relatives. This in turn enabled future social and arts activities to be tailored to the residents for them to participate in activities they enjoyed. Within the systematic review (Study 1) only two studies explored the

role of the arts in aiding knowledge of residents (Fritsch, 2009 and Philips, 2010), both of which showed improvements. This may also be demonstrated within the second study also with life story work mainly carried out by care staff rather than activities coordinators or other staff members. This reinforces the use of arts-based activity as a way of gathering information and improving staff knowledge.

Only three studies within the systematic review assess changes to physical health and all showed limited evidence of benefit (Hagen, 2003; Heyn, 2003; Koike, 2012). Similar findings were reflected in the survey responses, where general physical benefits were discussed but not with a high frequency. This may reflect a change in research to focus more upon the psychological benefits of arts participation. Older people in care homes have discussed how psychological factors impact more upon their quality of life than physical ones do (Hoe et.al 2006; Hoe et al. 2009; Mortimer & Green, 2016), therefore priority should be given to such outcomes.

7.6 Adverse Effects

Guidelines into arts for health stipulate that activities need to be 'meaningful' for older adults for them to be beneficial (NICE, 2015; APPG, 2017). The CQC also outlines that residents should be encouraged to participate in activities that are 'socially and culturally relevant and appropriate to them' (CQC, 2017, code R1.3), recommendations which are also discussed in the recent All Parliamentary Report (2017). Further to this, preference has been demonstrated as important for taking part in arts for health activities (O'Neill, 2011; Kolanowski et al. 2011). Preferences was discussed in only 12 (17%) of the studies within the systematic review. Despite stating music was preferred, a total of seven studies gathered information from caregivers or family members and therefore in a similar way to proxy outcome measures the accuracy of how preferred this music would be cannot be guaranteed.

Such preferences may account for a lack of positive findings for multisensory activities. Of note were two studies which changed the aesthetic of parts of the care home completely and both showed limited improvements (Chang, 2013; Scott et al. 2014). The décor of the home preferred by one resident may not have been like that of another and therefore whilst some may have preferred the environment following the intervention, some may not.

Overall the small number of studies which indicated adverse effects echo findings from Cowl and Gaugler's 2014 systematic review of arts activities and older people, in that while there may not be dramatic improvements to wellbeing, there is minimal evidence of harm from such activities. This was largely endorsed by the survey responses and telephone interviews which rarely cited any risk of harm they felt may occur to residents engaging in arts for health activities. The only mention of potential harm was discussed in relation to painful memories which may be uncovered when taking part in memory-based creative activity (such as autobiographical memory books).

Interestingly, the systematic review included studies suggestive of an increase in the amount of 'disruptive behaviours' such as increased verbalisations. However, there is evidence that these 'negative findings' may reflect increases in communicative ability and reductions in apathy. Of the studies which reported increased verbalisations (Clarkson, 2007; Cooke, 2010; Nair, 2011) none commented upon what types of verbalisations increased. A fourth study (Ziv, 2007) was the only one to examine verbalisation type and this showed overall residents produced more positive expressions as opposed to negative. Similar anecdotal evidence was reported for two additional studies (Millard, 1989 and Sole, 2014). A decrease in self-reported ratings of quality of life outlined in Sole (2014) may also be attributed to this increased self-awareness, as residents may become more aware of their emotions. Such findings were

echoed in the national survey responses, which described residents of the homes as being able to express themselves clearer, communicate with others more and remain engaged with the activity.

7.7 Barriers and Adaptions

The potential barriers to carrying out arts activities in care homes were similarly identified by both Study 1 and Study 2. Within the systematic review studies which described such difficulties reported staff requirements, physical disability of residents and the potential for individuals to become distressed by certain aspects of the activity (Cox, 2004; Billington, 2013; Roe, 2015). These were all echoed by the responses from the care home managers within the telephone interviews. Study 2 featured one barrier to carrying out such activities which was not identified by the systematic review and that was the perception of specialist knowledge required for carrying out arts-based activities. This likely did not feature within the systematic review as many of the interventions were carried out by members of the research team and were specifically designed for the purposes of the study. This relates to the overall picture of lack of training for care home staff within England, particularly for holistic care (Age UK, 2019). Current literature also suggests many care homes within England are working at capacity (Kings Fund, 2019) and this was reflected within the national survey (Study 2) which could impact on the provision of arts for health delivery.

The discrepancies between 'arts therapy' and 'arts for health' is often discussed within the literature (Stickley, 2012; Beard, 2012; Crossick and Kaszynska, 2016). The telephone interviews particularly referenced this, with managers preceding answers with phrases such as 'I know it's not art therapy, but…' (outlined in Chapter 6- Study 2 results). This was further evidenced by the idea that carrying out arts for health activities required specialist knowledge of art-based methods. Along with the large evidence base for music, this may in

part explain why music is so frequently delivered as it may be viewed as an activity which can be carried out by those who do not have specific arts training. In addition to this there was a desire for further training the enable staff members to carry out meaningful arts activities but expressions of frustration at a lack of information surrounding such training. The recent inquiry into arts for health (APPG, 2017) reinforced this idea, recommending occupational therapists and arts for health practitioners liaise with care homes to establish the most effective ways of delivering such activities.

It was highlighted within the interviews how residents often wished to take part in activities which took place away from the home (Chapter 6- Study 2 results). Managers discussed how they tried to give residents access to such outings, but this was logistically difficult for them. There was a high level of need in terms of wheelchairs and medical assistance which in turn meant many staff members would need to accompany the residents. This was echoed in the only study within the systematic review (Roe et al. 2015) where the difficulties related to taking older people from care homes to museums and galleries were discussed. The studies identified within the systematic review which focused upon changing the aesthetic of the home (Kincaid, 2003; Chang, 2013; Scott, 2014) may offer a compromise to the difficulties when engaging with museums and galleries. These studies provided the residents with access to visual and tactile experiences which they would experience from visits to cultural establishments. However, as discussed this promotes difficulties in preferred activities if personal preferences are not taken into account. The Arts Council England (2017) project which delivers performances within care homes may be more likely to offer a compromise for older adults who are unable to leave the care home for logistic reasons as residents could choose to engage or not depending on their preferences.

Within the systematic review there was one study (Doric-Henry, 1997) which explored the use of pottery making with older people within a care facility. The study concluded that whilst beneficial to the residents, carrying out such an activity was difficult given the lack of facilities within the homes for glazing the completed works. Within the survey, pottery making was identified by the fewest care homes was delivered with the lowest frequency. This is likely due to the logistics of such an activity, as highlighted by Doric-Henry (1997) which described the length of time needed for carrying out the creative artwork and specialist equipment such as kilns.

Finally, the amount and type of activities that could be carried out by the homes was limited by the amount of funding received by the homes specifically for carrying out arts. Only 16% (n=30) of the homes reported receiving funding for arts-based activities and when asked for additional information, eleven stated this was obtained through the homes own fundraising efforts, meaning only 10% (n=19) received funding from an outside source. In addition, care homes which specifically paid activity coordinators reported the highest level of arts activity. Funding was scarcely mentioned within the systematic review; few studies reported on how the arts activity was funded. This is representative of the lack of programme evaluation seen within arts for health research (APPG, 2017). Despite there being little funding available to the homes, care home managers did not express this as being a barrier and mainly carried out activities that they could within their budget, however small this was.

Despite the barriers to engagement faced by the homes (identified by care home managers in Study 2) there was a sense of all residents being eligible to partake in activities. The managers were asked if they felt there were any grounds for residents being unable to participate in the activities offered and all said no. Generally, levels of participation within

the activities became more passive in nature as the level of disability (either physical or mental) increased.

Qualitative evidence from Guzman-Garcia (2013) described how physical disability was a barrier identified by residents themselves, but how this was overcome to maintain participation in the activity. Wider research into wellbeing has also found that whilst physical health is important, many older adults do not see disability as a barrier to living their lives (Mortimer & Green, 2016). Some studies (Bennett, 1988; Houston, 1998; Suzuki, 1998; Burrack, 2004; Chen, 2009; Chang, 2013; Vankova, 2014) within the review excluded participants if they were deemed to have certain 'cognitive disabilities' or 'physical impairments' which the researchers felt excluded them from being eligible to participate, likely due to the difficulty in obtaining outcome measures from such individuals. The evidence from the national survey would indicate that no individual should be excluded from arts participation and that care homes are accommodating to the diverse needs of residents.

7.8 Implications for Practice

NICE guidelines (2013) state that those residing in care homes should be able to partake in 'meaningful activity' and that such activities should meet with the preferences and abilities of those taking part. This was further discussed within the All-Parliamentary Report (APPG, 2017) where the importance of preferred and meaningful activity was highlighted particularly for older people with care homes. All care homes responding to the national survey within Study 2 indicated that they carried out some form of arts activity and these were carried out frequently and matched to the preferences of residents. Further to this the care homes clearly described the ways in which art was indeed providing 'meaningful' activity through the benefits outlined within the survey and the telephone responses. Whilst difficult given the

large numbers within care homes, the ability for residents to have input into which activities are offered and carried out is important for providing activity which is classed as 'meaningful'. Several studies within the review highlighted how preference of residents was vital for providing benefit and minimising adverse effects of the activity (Gerdner, 2000; Kolanowski, 2011; Nair, 2011).

Continuity of arts availability is also important. Within the systematic review many of the studies focused on assessment at the beginning and end of the programme, only two assessed the impact of the programme for one year (Ledger, 2007; Lin, 2010) and none revisited residents for follow-up of longer than a month following the study duration. This is echoed by previous systematic reviews (Vink, Bruinsma and Sholten, 2011; Beard 2012; Cowl and Gaugler, 2014) which have suggested that benefits of arts participation are only assessed within the short-term.

The systematic review collected information as to how the arts activity was delivered but this was not assessed in great detail by the national survey. Of interest was a study by Noice (2006) which was one of the few studies within the systematic review that reported adverse effects of the activity. This was the only study which explicitly stated that residents were told when they were 'wrong' and were 'corrected'. Whilst this on its own would not be sufficient to conclude that this delivery was what caused the negative effects to residents it would tie in with the interviews within Study 2 which discussed how art enabled residents to self-express and improve self-esteem.

Furthermore, whilst studies which provided environmental changes highlighted the benefits of exploring novel surroundings there were also descriptions of when this was potentially disruptive to residents and this should be considered when adopting such methods (Cox 2004;

Billington, 2013; Roe et al. 2015). As discussed, there is evidence for the arts to increase vocalisations (not always negative) and so the opportunity for self-expression must also be considered when arts sessions are being delivered.

The evidence within both studies, but the systematic review in particular indicates that arts-based interventions may be particularly beneficial to those with existing clinical symptoms, specifically anxiety and depression. This was particularly evidenced by both Suzuki (2006) and Liu (2010) who described populations with existing clinical depression, who showed improvements when taking part in arts activities, along with Simard (2010) who found only those with low levels of social function to improve. Orsulic-Jeras (2000) described how low depression scores at baseline may explain a lack of effect within this study, given that residents already functioned well psychologically and had already reached a 'ceiling' in scores on outcome measures. The arts may be particularly beneficial for individuals with certain mental health needs within the care home population.

7.9 Limitations

The biggest limitation of Study 2 overall was the lack of a control group within the national survey due to lack of responses from care homes which do not routinely offer arts activities. This meant that comparisons could not be drawn between homes which did offer arts for health activities and those that did not. However, finding homes which do not offer at least some form of arts for health activity would prove difficult, given CQC, NICE guidelines and All-Party Parliamentary Group on Arts for Health (NICE, 2015; APPG, 2017; CQC 2017) recommend the ability for care home residents to take part in meaningful activity. In addition to this, the absence of an option within the survey indicating that the homes did not take part in activities meant that it is unclear whether those who did not respond to the question

related to activities carried out within the home (question number 13) were saying they did not carry out those activities, or simply did not respond to the question.

A second limitation of the national survey was the poor response rate. The survey was mailed out to 3817 of care homes and just 185 responded, giving a response rate of just 5% (n=185). It is not clear why the response rate was so low, given that those who were interviewed spoke highly of the use of arts for health within their settings. An important factor in the completion of online surveys is whether the respondent is interested in the topic (Fan & Yan, 2010; Saleh & Bista, 2017) and therefore responses may have been obtained only from those with a genuine interest in arts for health which could question the generalisability of the findings. Surveys administered via email show the lowest response rates (Fincham, 2008) and therefore conducting the survey in this way limits the number of response that would be obtained. In addition, surveys from health professionals are reported as yielding poor response rates that those of other professionals (Swain, 2012; Cho, Johnson & VanGeest, 2013; Hunter, 2013). The lack of response rate makes it difficult to draw conclusions upon whether delivery of arts-based activities was impacted on by factors such as geographical location and whether there was a difference in those from private and public care homes.

Although a rigorous way of assessing bias, the use of GRADE for all included studies may also be a limitation to the systematic review within this study. GRADE is a tool developed for use with Randomised Controlled Trial (RCT) study designs only and therefore required modification for use with quasi-experimental and pre-post-test designs.

A further limitation of the systematic review is the inclusion only of papers published in English. This is perhaps reflected within the country the included studies originated, with over half the papers sourced from English-speaking countries (33% from the USA, 10% from the

UK and 10% from Australia). This may therefore limit the generalisability of the findings made within the review. Most studies within the systematic review included individuals with a dementia diagnosis and this may also factor into the generalisability of findings to the care home population. However, it is likely to reflect the high proportions of individuals with cognitive impairment and dementia living within care homes (Age UK, 2017).

7.10 Future Research

As discussed, significant findings are most often described by studies with populations who display baseline scores indicative of lower functioning (such as depression, anxiety, low levels of wellbeing). It would obviously be recommended that such populations would therefore provide evidence of whether arts for health interventions are beneficial for those with clinical need. However, more difficulty would lie in establishing whether arts for health activities can reduce the pattern of decline in wellbeing shown with advancing age, or maintain a sense of wellbeing. In addition, rather than focusing purely on clinical outcomes, there is a need to explore personal aspects of wellbeing in more detail (Cowl & Gaugler, 2014; Medeiros & Basting, 2014; Crossick & Kaszynska, 2016). Recommendations set out by the All-Party Parliamentary Group (APPG, 2017) outline the need for health and wellbeing outcomes to form part of future knowledge in arts for health.

In keeping with findings of Noice, Noice and Kramer (2013) and Young et.al (2016), there was a lack of longitudinal data reflected within the studies in the review. Indeed, many of the interventions were described as providing limited benefits, only during the activity and for a short while after. Again, the All-Party Parliamentary Inquiry into arts for health recognised the shortcomings in good-quality evaluation of existing programmes being delivered within care homes (APPG, 2017) along with a greater need for good-quality longitudinal studies to ascertain impact and benefit.

Whilst there is a clear bias for music studies within the literature, Medeiros & Basting (2014) argue the importance of evaluating the method of delivery of the arts activity in order to assess effectiveness, rather than the type of arts used. Given the wide range of arts activities utilised within care homes (as determined by the survey responses within study 2) it would be beneficial to have wider evidence of benefit from a range of different arts types to broaden the evidence base related to arts for health. Furthermore, it would be advantageous for future research to compare methods of delivery which adhere to arts therapy requirements in comparison to those which offer alternative methods, or less structured forms, of arts for health delivery.

This thesis related to the needs of individuals who reside within care homes. Interestingly a recent study (Windle et al. 2018) showed no difference in outcomes between those within care homes, those who resided within the community or individuals living within their own homes. This would warrant further investigation to ascertain whether the benefits of arts for health are significantly different within those from care homes compared with other populations of older adults. It would also be beneficial to determine whether arts activities are better suited to individuals or populations in care homes compared with those within the community depending on diagnosis, health status or preferred activity (see section 7.5 for a discussion around preferred activity).

Inclusion of a meta-analysis was not possible within this systematic review given the differences between the studies, not only due to the number of different outcome measures used, but the timings with which they were administered. This would be beneficial to future systematic review, particularly those which focus on one arts activity or one outcome measure to allow for increased likelihood of homogeneity among findings. Additionally, as

the wealth of qualitative evidence increases, meta-synthesis of findings from this area would also contribute to knowledge of how beneficial arts for health are.

The aim of the national survey within this study was to provide an overview of the arts activities currently offered by care homes, along with information as to their frequency of delivery and whom they were led by. Most papers included within the systematic review assessed arts activities that were delivered specifically for the purposes of the study and therefore may not be reflective of real-life activities carried out within homes. Further evidence of such activities is therefore warranted. In addition, any surveys which are carried out to assess the use of arts for health activities may wish to look at alternative methods of administration to enhance response rates and adherence to the activities delivered.

Although the systematic review (Study 1) featured a range of different arts types, there was no inclusion of studies which focus upon the use of digital methods of arts delivery. Whilst existing studies have indeed evaluated the use of digital technology within care homes, these have focused on general iPad usage rather than for exposure to arts-based activities (Jones, Kay, Upton & Upton, 2013; Evans, Bray & Evans, 2017). As the use of digital technologies within care homes increases it may be beneficial for future research into arts for health to evaluate the use of such devices as a method of arts for health delivery. Such research could offer an alternative experience of museum and gallery environments and therefore provide an alternative to the barriers identified by both Study 1 and Study 2 with respect of providing such visits for older adults living within care homes.

Arts for health research should incorporate the impact on wider social circles rather than just that of the participant (Medeiros & Basting, 2014). Given the benefits identified with respect of socialisation (outlined in section 7.4.3) future research should incorporate family members

and friends who are in regular contact with the residents. It would also be of interest to explore the benefits to staff members and the overall attitudes and culture within care homes. Within the survey responses benefits to staff members were reported less so than benefits to residents or the care home in general. However, the categories of job satisfaction and enjoyment were reported, and these topics would benefit from further exploration.

7.11 Summary

This thesis presented two studies, the first, Study 1, sourced, appraised and synthesised existing evidence of benefit for arts for health activities carried out within care homes. The second, Study 2, conducted a national survey of care homes within England to ascertain which activities were undertaken within homes and the perceived benefits of carrying out such activities.

Disparities between benefits ascertained by the two studies show that research is often focused upon the control and mediation of clinical symptoms, such as anxiety, depression and 'disruptive behaviours' however, such impacts and benefits are not viewed with the same levels of importance within the care homes themselves. More important were general aspects of health and wellbeing, including the ability for participants to foster their creativity.

Despite there being variation in the quality and types of evidence there were areas where arts for health was shown to impact the population of older adults living within care homes. The arts were demonstrated to impact on agitation levels, general aspects of psychological health, aid communication and socialisation and encourage engagement and creativity.

The following and final chapter (chapter 8) presents overall conclusions arising from each of the studies, their objectives and synthesis of key findings. Chapter 8 will also state the overall implications for practice, policy and further research which can be drawn from this research.

Chapter 8- Conclusions and recommendations for policy, practice and future research.

8.1 Introduction

Following on from Chapter 7 which provides a discussion synthesising the results from Study 1 and Study 2 this chapter (Chapter 8) presents overall conclusions which can be drawn from the two studies. The chapter also provides implications for current arts for health policy along with recommendations for practice and further research.

8.2 Summary of Study 1 and Study 2

Two studies were carried out to assess the impact of arts for health activities on the health, wellbeing and quality of life of older people in care homes. Study 1 (Chapter 2-methods, Chapter 3-overall descriptive narrative synthesis, Chapter 4- descriptive narrative synthesis by arts type) consisted of a systematic review and descriptive narrative synthesis of published studies evaluating the benefits and impact of arts for health for older people residing in care homes.

The objectives of Study 1 were:

1. To identify published empirical studies focused upon arts for health and the impact to older adults with respect of health, wellbeing and quality of life.

2. To create a typology of arts for health activities and their perceived benefits.

The second study, Study 2 (Chapter 5- methods, Chapter 6- results) was an electronic national survey of a stratified randomised sample of care home managers within the UK, with telephone interviews conducted with a self-selecting sub-sample of responding managers.

The objectives for Study 2 were:

1. To conduct an online survey of care home managers in England to determine what (if any) activities were delivered and map their prevalence.

2. To carry out telephone interviews with a self-selecting sample of care home managers to add further contextual information.

3. To examine the characteristics surrounding delivery of such activities to gather information including frequency and method of delivery, funding and associated benefits.

The final objective was to synthesise the results from Study 1 and Study 2 and form a typology of arts activities and their perceived benefits, this is presented in the Chapter 7.

8.3 Conclusions

The findings from Study 1 and Study 2 allow for the following conclusions to be made on the impact of arts for health activities for older people residing in care homes:

- There is an emerging body of published research evaluating the impact of arts for health in older people living within care homes. In Study 1, a total of 80% (n=57) of the included studies were published since the year 2000, with 39% (n=28) published between 2010 and 2015. Study 2 further confirmed the interest in the use of arts for health with older people residing in care homes. Whilst only representing a small sample size (n= 184), all the responding care home managers reported they carried out arts for health activities within their homes.
- Published empirical arts for health studies vary greatly in quality. Randomised controlled trials showed higher levels of quality than Quasi-RCTs or pre/post-test studies, although only two were scored at low risk of bias for all domains. Pre/post-

test studies showed the lowest levels of quality, with many domains (assessed via GRADE criteria) unable to be judged due to poor reporting. There was little evidence that quality of quantitative studies improved with publication date, suggesting that more robust study methods are needed for all study designs to place confidence in findings of benefit and impact.

- Existing qualitative evidence was appraised to be of moderate quality (all judged Grade B or C), with none being given the highest (Grade A) or lowest (Grade D) judgements of quality. Qualitative studies were more recent, with all being published after 1998. As with quantitative studies there was little evidence that quality improved with publication date.
- Most studies within the systematic review (Study 1) evaluated music and this activity was reportedly carried out by the largest number of care homes and with the greatest frequency (Study 2). However, other activities often delivered by the home such as painting, collage and knitting (applied arts and crafts) were not frequently evaluated by existing published literature.
- Evidence from Study 1 and Study 2 indicates arts for health are beneficial to psychological wellbeing, agitated behaviours, cognition, socialisation and improving the caring process. However, included studies within the systematic review (Study 1) consisted of small sample sizes which impacted upon the amount of statistically significant findings.
- In addition, studies within the systematic review (Study 1) outlined low baseline levels for measures such as depression, anxiety and agitation which likely contributed to lack of observable impact. Study 2 also reported low baseline levels of depression and

behavioural symptoms. Therefore, focus should lie in establishing the impact of arts for health activities to maintain levels of health, wellbeing and quality of life.

- The scope of arts for health's impact may not be accurately evidenced due to the evaluations used. Few studies included in the systematic review (Study 1) that assessed the impact of arts for health on quality of life showed evidence of benefit or impact. However, qualitative evidence from Study 1 and Study 2 reported quality of life to be impacted upon by arts for health. Standardised quantitative quality of life measures may therefore not be sufficient in detecting impact to quality of life in older people in care homes. This reinforces the need for mixed methods studies to capture all benefits associated with arts for health.
- Many studies included in Study 1 that featured comparison with other activities failed to show arts activities to be more beneficial than the comparator. However, none of these qualitatively assessed the activity, nor did they feature assessment of outcomes unique to arts for health such as creativity and opportunity for self-expression. Such aspects were frequently cited within published qualitative evidence (Study 1) and the responses from the care home managers (Study 2).
- Adverse effects were not frequently reported by either Study 1 or Study 2. Therefore, whilst there is a lack of conclusive evidence of arts for health being beneficial for older adults in care homes, they are not likely to be harmful.
- Within Study 2 only ten percent (n=19) of the responding care homes reported they received external funding for carrying out arts for health activities, most fundraised within the homes or set aside internal budgets. Other barriers to delivery of arts for health activities within care homes included the perception that specialist knowledge was required and the logistics of delivering such activities.

8.4 Implications for policy

The implications for current international and national arts for health policy are:

- The World Health Organisation's Active Ageing strategy (WHO, 2015) stipulates older adults should be offered the opportunity to participate in activities which are worthwhile and meaningful. Such activities are also recommended by NICE (2015) and CQC (2017) guidelines, along with the All-Party Parliamentary report into arts for health delivery (APPG, 2017). Responses from care home managers within Study 2 would evidence this to be the case within care homes in England as all indicated they carried out a range of activities with no difference in delivery of activities between different types of home or by geographical location.
- Guidelines for activities delivered in care homes state such activities should be 'meaningful' and 'preferred' for them to be beneficial (NICE, 2015; WHO, 2015; CQC, 2017; APPG, 2017). Preferences were reportedly considered by those delivering activities within the care homes (Study 2) but preferred activities were rarely evaluated in published studies (Study 1). When preferred activities were evaluated they were shown to have greater impact.
- The All-Party Parliamentary Inquiry (2017) also highlighted the need for communication between arts for health practitioners and care homes. Within Study 2 care home managers identified training needs and the need for greater competencies in arts for health activities which reinforces this suggestion.
- The inquiry into arts for health delivery within the UK also described current arts for health evidence to be 'unevenly distributed, of varying quality and sometimes inaccessible' (APPG, 2017, p10). The findings from Study 1 would agree with this statement.

8.5 Recommendations for practice

The following recommendations for practice can be drawn from this research:

- In keeping with existing policy recommendations (NICE, 2015; CQC, 2017; APPG, 2017) arts for health activities offered to older people within care homes should be matched to their own preferences to provide an experience that is 'meaningful and worthwhile'. Older people should be consulted on their preferences and the delivery of such activities.
- Findings from both Study 1 and Study 2 showed that most arts for health activities within care homes were delivered by activities coordinators and external staff members. However, there is evidence that such activities can be beneficial for staff, their knowledge and relationship with residents. This should be considered when implementing arts for health activities within care homes.
- Whilst adverse effects were infrequently discussed, those delivering arts for health activities should be mindful of them. Care should be taken when carrying out activities that could cause distress, such as those which offer sensory stimulation, or may promote memory recall.
- Arts for health activities should promote the opportunity for older adults to foster their creativity.
- Most evidence suggested arts for health offers greatest benefit 'in the moment' and this should be considered when anticipating the timing and delivery of such activities in order to maximise their potential benefits.

8.6 Recommendations for future research

The following recommendations for future research can be made based on the findings from both Study 1 and Study 2:

- The All-Party Parliamentary Report (2017) outlines the need for future research to longitudinally evaluate the effects of arts for health activities for older people residing in care homes. There was a lack of longitudinal methods existing in published research included in the systematic review (Study 1) and would therefore agree with the need for longitudinal research. Such methods would also enable changes in decline and maintenance of health, wellbeing and quality of life in older adults residing in care homes, rather than just focusing on short-term improvements.
- Furthermore, only 10% (n=7) of the studies within the systematic review (Study 1) originated within the UK. Given the unique structure of long-term care for older adults within the UK further research would be beneficial in ascertaining specific impact and benefits within this population.
- Randomised controlled intervention studies adopting mixed methods designs are warranted to contribute to the growing field of arts for health evidence. Such studies should include adequate, justified sample sizes to ensure they are able to effectively assess the impact of arts for health activities for older people in care homes.
- Where homogeneity of outcomes allows it would be beneficial for future systematic reviews to feature a meta-analysis of findings from included studies, particularly around outcomes related to depression, anxiety, agitation and cognition. Furthermore, a meta-synthesis of qualitative evidence would allow for greater confidence to be placed in the findings from these studies.

- Published empirical studies failed to quantitatively demonstrate an effect on quality of life, despite qualitative evidence reporting this outcome. It would be beneficial for a specific quality of life outcome measure which reflects the unique needs of care home residents to be developed in order to assess this outcome.
- Future research should focus on assessing the impact of activities less represented within the literature but frequently carried out within homes, such as applied arts and crafts. In addition, given the wide range in method of arts delivered it would be beneficial to evaluate this aspect of arts for health.
- It would also be useful to explore whether experiences and benefits of taking part in the arts are different for older adults in care homes and those within the community, or whether they are universal.

8.7 Original contribution to knowledge

Both studies within this thesis offer original contributions to knowledge and these will be discussed briefly. The first, Study 1 is the first systematic review assessing the impact of arts for health activities for older people residing in care homes. Furthermore, it featured comparison of different study designs to determine the quality of such evidence and how this varies between randomised controlled trials, quasi-RCTs, pre/post-test studies and qualitative studies which has not taken place in previous arts for health systematic reviews focused on older adults. Findings from Study 1 also allowed for a unique typology of arts for health activities for this population to be formed and the benefits associated with such activities.

The second study, Study 2, was the first electronic national survey of a stratified randomised sample of care home managers in England to determine what activities were delivered within care homes and characteristics surrounding delivery of such activities. This allowed for

evaluation of arts activities delivered in different care home types and geographical locations to determine whether any differences were observed based on these factors. The study is also the only one to date that interviewed care home managers within England to ask them about their perceptions of arts for health activities and how they can impact on care home residents, staff members and the care home environment in general.

References

Age UK. 2015 *Agenda for Later Life 2015: a great place to grow older* [online]. Available from: http://www.ageuk.org.uk/professional-resources-home/policy/agenda-for-later-life/ [accessed 07/08/2019]

AGE UK. 2016. *Later Life in the United Kingdom* [online]. Available from: http://www.ageuk.org.uk/Documents/EN-GB/Factsheets/Later_Life_UK_factsheet.pdf?dtrk=true [accessed 31/03/2019]

Age UK. 2018. *Care Homes: finding, choosing and paying for a care home.* [online]. Available from: https://www.ageuk.org.uk/globalassets/age-uk/documents/information-guides/ageukig06_care_homes_inf.pdf [accessed 07/08/2019].

Age UK. 2019. Briefing: Health and Care of Older People in England 2019. Available from: https://www.ageuk.org.uk/globalassets/age-uk/documents/reports-and-publications/reports-and-briefings/health--wellbeing/age_uk_briefing_state_of_health_and_care_of_older_people_july2019.pdf Accessed 10/01/2020.

Atkins, J., Naismith, S.L., Luscombe, G.M. & Hickie, I.B. 2013. Psychological distress and quality of life in older persons: relative contributions of fixed and modifiable risk factors. *BMC Psychiatry.* 13: 249.

All Parliamentary Group (APPG) on Arts, Health and Wellbeing. 2017. *Inquiry Report. Creative Health: The Arts for Health and Wellbeing.* [online]. Available from: https://www.culturehealthandwellbeing.org.uk/appg-inquiry/ [accessed 07/08/2019]

Allen, J. 2008. *Older People and Wellbeing.* London: Institute for Public Policy and Research.

Anderson, D., Banerjee, S., Barker, A., Connelly, P., Junaid, O., Series, H.S. & Seymour, J. 2011. *The Need to Tackle Age Discrimination in Mental Health- A Compendium of Evidence* [online]. Avavilable from: http://www.rcpsych.ac.uk/pdf/Royal%20College%20of%20Psychiatrists%20-%20The%20Need%20to%20Tackle%20Age%20Discrimination%20in%20Mental%20Health%20Services%20-%20Oct09.pdf [accessed 04/04/2019].

Anderson, R. & Schiedermayer, D. (2003). 'The art of medicine through the humanities: an overview of a one-month humanities elective for fourth year students', *Medical Education*, 37, 6, 560-562.

Armat, M.R., Assarroudi, A., Rad, M. & Heydari, A. 2018. Inductive and deductive: ambiguous labels in qualitative content analysis. *The Qualitative Report.* 23(1): article 16.

Arts Council England 2015. *Cultural Activities, Artforms and Wellbeing.* Manchester: Arts Council England.

Ashida, S. 2000. The effect of reminiscence music therapy sessions on changes in depressive symptoms in elderly persons with dementia. *Journal of Music Therapy.* 37(3), 170-182.

Bailey, J. 2008. First steps in qualitative data analysis: transcribing. *Family Practice Advance Access.* Pp 127-131.

Balarinwa, O.A. 2015. Principles and methods of validity and reliability testing of questionnaires used in social and health science research. *Niger Postgraduate Medical Journal.* 22, pp. 195-201.

Basting, A. 2003. Exploring the creative potential of people with Alzheimer's disease and related dementia: Dare to imagine, pp353-367 In Ronch JL, Goldfield JG. Mental Wellness in Aging: Strengths-Based Approaches. Baltimore: Health Professions Press.

Basting, A. 2009. Part One: Understanding our fears about dementia. In *Forget Memory: Creating Better Lives for People with Dementia*. Pp 7-30. Baltimore: Johns Hopkins University Press.

Beard, R.L., 2012. Art therapies and dementia care: a systematic review. *Dementia*. 11 (5), pp 633-656.

Belfrage, C. & Hauf, F. 2016. *The Gentle Art of Retroduction: Critical Realism, Cultural Political Economy and Critical Grounded Theory.* Organization Studies.

Bell A, Gandy R & Roe B (2016/17) Benefits and impacts of Active Lives groups for older people living in the community. Research, Policy and Planning 32(2), 97-112

Bengtsson, M. 1026. How to plan and perform qualitative study using content analysis. *Nursing Plus Open.* 2, pp. 8-14.

Bennett, S.L. & Maas, F. 1988. The effects of music-based life review on the life satisfaction and ego integrity of elderly people. *British Journal of Occupational Therapy.* 51 (12), 433-436.

Billington, J., Carroll, J., Davis, P R., Healey, C & Kinderman, P. 2013. A literature-based intervention for older people living with dementia. *Perspectives in Public Health.* 133(3), 165-173

British Medical Association. 2011. *Science and Education. The Psychological and Social Needs to Patients*. London: BMA Marketing and Publications Department.

British Medical Association. 2011. *Science and Education. The Psychological and Social Needs to Patients*. London: BMA Marketing and Publications Department.

Brontons, M. & Pickett-Cooper, P.K. 1998. The effects of music therapy intervention on agitation behaviors of Alzheimer's Disease Patients. *Journal of Music Therapy.* 33 (1), 2-18.

Bruggem, E. 2009. A critical comparison of offline focus groupsn, online focus groups and e-Delphi. *International Journal of Market Research*. 51, pp.363-381.

Burrack, O.R., Jefferson, P. & Libow, L.S., 2004. Individualized music: a route to improving the quality of life for long-term care residents. *Activities, Adaption and Aging.* 27 (1), 63-76.

Calman, K., Downie, R. (1996). 'Why arts courses for medical curricula?', The *Lancet,* 34, pp. 1499-1500.

Cameron, M., Crane, N., Ings, R. And Taylor, K. 2013. Promoting well-being through creativity: how arts and public health can learn from one another. *Perspectives on Public Health.* 133 (1), pp. 52-59.

Carruth, E.K. 1997. The effects of singing and the spaced retrieval technique on improving face-name recognition in nursing home residents with memory loss. *Journal of Music Therapy.* 34 (3), 165-186.

Casby, J.A. & Holm, M.B. 1994. The effect of music on repetitive disruptive vocalizations of persons with dementia. *The American Journal of Occupational Therapy.* 48 (10), 883-889.

Centres for Disease Control and Prevention. 2013. *Well-being Concepts* [online]. Available from: http://www.cdc.gov/hrqol/wellbeing.htm [Accessed 31/03/2016]

Cesario, S., Morin, K., Santa-Donato, A. Evaluating the level of evidence in qualitative research. Journal of Obstetrics*. Gynaecology and Neonatal Nursing*. 31, pp. 531–537.

Chang, C-H., Lu, M-S., Lin T-E. & Chen, C-H. 2013. The effectiveness of visual art on environment in nursing home. *Journal of Nursing Scholarship.* 45 (2), 107.

Chang, F Y, Huang, H C., Lin, K C. & Lin, L C. 2010. The effect of a music programme during lunchtime on the problem behaviour of the older residents with dementia at an institution in Taiwan. *Journal of Clinical Nursing.* 19 (7-8), 939-948.

Chen, S-L., Lin, H-C. & Jane, S-W. 2009. Perceptions of a group music therapy among elderly nursing home residents in Taiwan. Complementary Therapies in Medicine. 17 (4), 190-195.

Cho, Y. I., Johnson, T. P., & VanGeest, J. B. (2013). Enhancing Surveys of Health Care Professionals: A Meta-Analysis of Techniques to Improve Response. Evaluation & the Health Professions, 36(3), 382–407

Clair, A.A. 1996. The effect of singing on alert responses in persons with late stage dementia. *Journal of Music Therapy.* 23 (4), 243-247.

Clarkson, K.A., Cassidy, K-L. And Eskes, G.A., 2007. Singing soothes: music concerts for the management of agitation in older adults with dementia. *Canadian Journal of Geriatrics.* 10 (3), 80-83.

Clift, S. And Camic, P.M. 2015. Introduction to the field of creative arts, wellbeing and health: achievements and current challenges. In S. Clift and P.M. Camic, *Oxford Textbook of International Perspectives on Practice, Policy and Research* pp. 235-244. Oxford: Oxford University Press.

Cochrane, A.L. 1972. *Effectiveness and Efficiency: random reflections on health services.* London: Nuffield Hospital Provincial Trust.

Cohen, G.D., 2006. Research on creativity and aging: the positive impact of the arts on health illness*. Generations*. 30 (1), pp 7-15.

Cohen-Mansfield, J. And Werner, P., 1997. Management of verbally disruptive behaviors in nursing home residents. *Journal of Gerontology: Medical Sciences.* 52a (6), M369-M377.

Cohen-Mansfield, J., Dakheel-Ali, M., Marx, M.S., Thein, K. & Reginer, N.G. 2015. Which unmet needs contribute to behaviour problems in persons with advanced dementia? *Psychiatry Research.* 228(1), pp59-64.

Cooke, A., Mills, T.A. And Lavender, T. 2010. Informed and uninformed decision making'—Women's reasoning, experiences and perceptions with regard to advanced maternal age and delayed childbearing: A meta-synthesis. *International Journal of Nursing Studies.* 10, pp. 1317-1329.

Cooke, M., Moyle, W., Shum, D., Harrison, S & Murfield, J. 2010a. A randomized controlled trial exploring the effect of music on quality of life and depression in older people with dementia. Journal of Health Psychology, 15 (5), 765-767.

Cooke, M., Moyle, W., Shum, D., Harrison, S & Murfield, J. 2010b. A randomized controlled trial exploring the effect of music on agitated behaviours and anxiety in older people with dementia. Aging and Mental Health, 14 (8), 905-916.

Cowl, A.L. And Gaugler, J.E., 2014. Efficacy of creative arts therapy in treatment of Alzheimer's disease and dementia: A Systematic Literature Review. *Activities, Adaption and Aging.* 38 (4), pp 281-330.

Cox, H., Burns, I. & Savage, S. 2004. Multisensory environments for leisure: promoting well-being in nursing home residents with dementia. *Journal of Gerontological Nursing.* 30 (2), 37-45.

Critical Appraisal Skills Programme (Casp). 1999*. Ten questions to help you make sense of qualitative research*. Oxford: Clinical Appraisal Skills Programme.

Crossick, G. And Kaszynska, P. 2016. *Understanding the Value of Arts and Culture* [online]. Available from http://www.ahrc.ac.uk/documents/publications/cultural-value-project-final-report/ [accessed 02/04/2019].

Collier, A. 1994. Critical Realism: An introduction to Roy Bhaskar's Philosophy. Verso.

Coolican, H. 2011. *Research Methods and Statistics in Psychology.* East Sussex, UK: Psychology Press.

Curtis, A., O'brien, M., Gibson, L., Maden, M., Gallagher, W. And Roe, B. 2015. *The impact of arts for health on the health status, well-being and quality of life in older adults who reside in care homes- protocol* [online]. Available from: http://www.crd.york.ac.uk/PROSPERO/. Registration number CRD42015026264.

Davidson, S. And Rossall, P. 2014. *Evidence Review: Loneliness in Later Life* [online]. Available from: http://www.ageuk.org.uk/Documents/EN-GB/For-professionals/Research/Age%20UK%20Evidence%20Review%20on%20Loneliness%20July%202014.pdf?dtrk=true [accessed 02/04/2019].

Daykin, N., Mansfield, L., Meads, C., Julier, G., Tomlinson, A., Payne, A., Duffy, L.G., Lane, J., D'Innocenzo, D., Burnett, A., Kay, T., Dolan, P., Testoni, S. & Victor, C. 2017. What works well for wellbeing? A systematic review of wellbeing outcomes for music and singing in adults. *Perspectives in Public Health.* 138(1), pp. 39-46.

de GUZMAN, A.B., SATUITO, J.B.C., SATUMBA, M.A.E., SEGUI, D.R.A., SERQUINA, F.E.C., SERRANO, L.J.P. and SEVILLA, M.D. 2011. Filipino arts among elders in institutionalized care settings. *Educational Gerontology*. 37 (3), pp 248-261.

De Medeiros, K & Basting, A. 2013. "Shall I compare thee to a dose of Donepezil?": cultural arts interventions in dementia care research. *The Gerontologist.* 54(3), pp. 344-353.

Department Of Health. 2011. *No Health Without Mental Health*: *A Cross-Government Mental Health Strategy for People of All Ages.* London: Department of Health.

Division Of Mental Health And Prevention Of Substance Abuse. 1997. WHOQOL Measuring Quality of Life [online]. Available from: http://www.who.int/mental_health/media/68.pdf [accessed 31/03/2019]

Doric-Henry, L. 1997. Pottery as Art Therapy with elderly nursing home residents. *Art Therapy.* 14 (3), 163-171.

Downie, R. Medical humanities some uses and problems. *Journal of the Royal College of Physicians.* 46, pp. 288-294.

Downe, S., Finlayson, K., Walsh, D. And Lavender, T. 2009. Weighing up and balancing out': a meta-synthesis of barriers to antenatal care for marginalised women in high-income countries. *BJOG- An International Journal of Obstetrics and Gynaecology.* 116 (4), pp. 518-529.

Dykema, J., Jones, N.R., Piche, T. & Stevenson, J. 2013. Surveying clinicians by web: current issues in design and administration. *Evaluation and the Health Professions.* 36, pp. 352-381.

Edwards PJ, Roberts I, Clarke MJ, DiGuiseppi C, Wentz R, Kwan I, Cooper R, Felix LM, Pratap S. 2009. Methods to increase response to postal and electronic questionnaires. *Cochrane Database of Systematic Reviews.* Version 3.

Eden, J., Levit, L., Berg, A. And Morton, S. 2011. *Finding What Works in Healthcare. Standards for Systematic Reviews.* Washington, USA: The National Academies Press.

Egger, M., Davey Smith, G. And Altman, D.G. 2005. *Systematic reviews in health care. Meta-analysis in context.* London, BMJ Publishing.

Egger, M., Davey Smith, G.D. And O'rourke, K. 2008. Rationale, potential and promise of systematic reviews. In: M. EGGER., G. D. SMITH. And D. ALTMAN eds. *Systematic Reviews in Healthcare: Meta-Analysis in Context*. London: BMJ Publishing. Pp 3-22.

Fan, W. & Yan, Z. (2010) Factors affecting response rates of the Web survey: A systematic review. Computers in Human Behavior, 26, 132-139.

Ferreira, A.R., Dias, C.C. & Fernandes, L. 2016. Needs in nursing homes and their relation with cognitive function decline, behavioural and psychological symptoms. *Frontiers in Aging Neurosciences.* 8: 72.

Flood, M & Phillips, K.D., 2007. Creativity in older adults: A Plethora of Possibilities. Issues in Mental Health Nursing. 28 (4), pp 389-411.

Fritsch, T., Kwak, J., Grant, S., Lang, J., Montgomery, R.R. & Basting, A.D. 2009. Impact of TimeSlips, a creative expression intervention program, on nursing home residents with dementia and their caregivers. The Gerontologist. 49(1), 117-127.

Fujiawara, D. And Mackerron, G. 2015. *Cultural Activities, Artforms and Wellbeing* [online]. Available from: http://www.artscouncil.org.uk/cultural-activities-artforms-and%C2%A0wellbeing [accessed 28/03/2019].

Galetta, A. 2013. *Mastering the Semi-Structured Interview and Beyond: From Research Design to Analysis and Publication.* New York: New York University Press.

Gandy R, Bell A, McClelland R & Roe B. (2017) Evaluating the Delivery, Impact and Cost-Benefits of an Active Lives Programme for Older People Living in the Community. Journal of Primary Health Care Research and Development. 18, 122-134.

Garland, K., Beer, E., Eppingstall, B. & O'Connor, D.W. 2007. A comparison of two treatments of agitated behaviour in nursing home residents with dementia: simulated family presence and preferred music. American Association for Geriatric Psychiatry. 15 (6), 514-521.

George, D R. & Houser, W S. 2014. "I'm a storyteller!" Exploring the benefits of a TimeSlips creative expression program at a nursing home. *American Journal of Alzheimer's Disease and Other Dementias.* 29(8) 678-684.

Gerdner, L. 2000. Effects of individualized versus classical 'relaxation' music on the frequency of agitation in elderly persons with Alzheimer's disease and related disorders. *International Psychogeriatrics.* 12 (1), 49-65.

Glasziou P, Irwig L, Bain C, Colditz G (2001) *Systematic Reviews in Health Care. A Practical Guide*. Cambridge: Cambridge University Press.

Glenton, C., Lewin, S., Carlsen, B., Colvin, C., Mumthe-Kaas, H., Noyes, J. And Rashiain, A. 2013. *Assessing the certainty of findings from qualitative evidence synthesis: the CerQual approach* [online]. Available from: http://gradepro.org/Rome2013/22acerqual.pdf [accessed 31/03/2016].

Goddaer, J. & Abraham, I.L. 1994. Effects of relaxing music on agitation during meals among nursing home residents with severe cognitive impairment. *Archives of Psychiatric Nursing.* 8 (3), 150-158.

Gotell, E., Brown, S. & Ekman, S-L. 2009. The influence of caregiver singing and background music on vocally expressed emotions and moods in dementia care: a qualitative analysis. *International Journal of Nursing Studies.* 46(4), pp. 422-430.

Graneheim, U.H., Lindgern, B-M. & Lundman, B. 2017. Methodological challenges in qualitative content analysis: a discussion. *Nurse Education Today.* 56, pp. 29-34.

Greer, N., Fleuriet, K.J. And Cantu, A.G. 2012. Acrylic Rx: a program evaluation of a professionally taught painting class among older Americans. *Arts and Health.* 4(3), pp. 1-12.

Guetin, S., Portet, F., Picot, M.C., Pommie, C., Messaoudi, M., Djabelkir, L., Olsen, A.L., Cano, M.M., Lecourt, E. & Touchon, J., 2009. Effect of music therapy on anxiety and depression in patients with Alzheimer's type Dementia: randomised, controlled study. *Dementia.* 28(1), 36-46.

Guyatt, G., Oxman, A.D., Alk, E.A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falckytter., Glasziou, P., Debeer, H., Jaeschke, R., Rindi, D., Meerpohi, J., Dahm, P. 7 Schunemann, H.J. 2011. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*. 64 (4), pp 383-394.

Guzman-Garcia, A., Mukaetova-Ladinska, E. & James, I. 2012. Introducing a Latin ballroom dance class to people with dementia living in care homes, benefits and concerns: A pilot study. *Dementia.* 12 (5), 523-535.

Hagen, B., Armstrong-Esther, C. & Sandilands, M., 2003. On a happier note: validation of musical exercise for older persons in long-term care settings. *International Journal of Nursing Studies.* 40, 347-357.

Ham, C., Imison, C., Goodwin, N., Dixon, A. And South, P. 2011. *Where Next for NHS Reforms? The Case for Integrated Care.* London: The Kings Fund.

Hammar, L.M., Emami, A., Engstrom, G & Gotell, E. 2010. Finding the key to communion- Caregivers' experience of 'music therapeutic caregiving' in dementia care: A qualitative analysis. *Dementia.* 10 (1), 98-111.

Hammar, L.M., Emami, A., Gotell, E. & Engstrom, G. 2011. The impact of caregivers' singing on expressions of emotion and resistance during morning care situations inn persons with dementia: an intervention in dementia care. *Journal of Clinical Nursing.* 20 (7-8), 969-978.

Hancock, G.A., Woods, B., Challis, D., And Orrell, M., 2006. The needs of older people with dementia in residential care*. International Journal of Geriatric Psychiatry*. 21 (1), pp 43-49.

Hays R, Clarkson P, Tucker S, Challis D (2012) Healthcare support services for care home residents. Nursing Older People24, 10, 26-30.

Health Research Authority (partnership with INVOLVE). 2016. Public Involvement in Research and Research Ethics Committee Review.

Heyn, P. 2008. The effect of a multisensory exercise program on engagement, behaviour, and selected physiological indexes in persons with dementia. *American Journal of Alzheimer's Disease and other Dementias.* 18 (4), 247-251.

Hicks-Moore, S.L. And Robinson, B.A. 2007. Favorite music and hand massage: two interventions to decrease agitation in residents with dementia. *Dementia.* 24 (4), 195-196.

Hicks-Moore, S.L., 2005. Relaxing music at mealtime in nursing homes. *Journal of Gerontoloical Nursing.* 31 (12), 26-32.

Higgins J.P.T. And Green S. 2011 *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* (updated March 2011). [online] The Cochrane Collaboration. Available from www.cochrane-handbook.org [Accessed 04/04/2019].

Higgins, J.P.T., Altman, D.G., Gotzsche, P.C., Juni, P., Moher, D., Oxman, A.D., Savovic, J., Schultz, K.F., Weeks, L. & Sterne, J.A.C. 2011. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ.* 343, d5928.

Hilliard, R.E., 2004. A post-hoc analysis of music therapy services for residents in nursing homes receiving hospice care. *Journal of Music Therapy.* 41 (4), 40-41.

Ho, S-Y., Lai, H-L., Jeng, S-Y., Tang, C-W., Sung, H-C. & Chen, P-W. 2011. The effects of researcher-composed music at mealtime on agitation in nursing home residents with dementia. *Archives of Psychiatric Nursing.* 25 (6), e49-e55.

Hoe, J., Hancock, G., Livingston, G And Orrell, M. 2006. Quality of life of people with dementia in residential care homes. *British Journal of Psychiatry*. 188 (5), pp 460-464.

Hoe, J., Hancock, G., Livingstone, G., Woods, B., Challis, D. And Orrell, M., 2009. Changes in the quality of life of people with dementia living in care homes. *Alzheimer's disease and Associated Disorders*. 23 (3), pp 285-290.

Holmes, C., Knights, A., Dean, C., Hodkinson, S. & Hopkins, V. 2006. Keep music live: music and alleviation of apathy in dementia subjects. International Psychogeriatrics. 18 (4), 623-630.

Hong, I.S. And Choi, M.J., 2011. Songwriting oriented activities improve the cognitive functions of the aged with dementia. *The Arts in Psychotherapy.* 38, 221-228.

Houser, W.S. & George, D. R. 2013. Impact of TimeSlips creative expression program on behavioural symptoms and psychotropic medication use in persons with dementia in long-term care: a cluster-randomized pilot study. American Journal of Geriatric Psychiatry. 22 (4), 337-340.

Houston, D.M., Mckee, K.J., Carroll, L. & Marsh, H., 1998. Using humour to promote psychological wellbeing in residential homes for older people. *Aging and Mental Health.* 2(4), 328-332.

Houston, S and Mcgill, A. 2016. English National Ballet, Dance for Parkinson's: An Investigative Study 2 A report on a three-year mixed methods research study [online]. Available from: http://www.ballet.org.uk/media/filer_public/2016/01/20/enb_research-report_final_2.pdf [accessed 01/04/2019]

Human Rights Watch. 2019. Unmet Needs: Improper Social Care Assessment for Older People in England. Available from: https://www.hrw.org/report/2019/01/09/unmet-needs/improper-social-care-assessments-older-people-england. Accessed 10/01/2020.

Hunter, L. 2012. Challenging the reported disadvantages of e-questionnaires and addressing methodological issues of online data collection. Nurse Researcher. 20, pp. 11-20.

Hsieh, H-F. & Shannon, S.E. Three approaches to qualitative content analysis. *Qualitative Health Research.* 15 (9), pp. 1277-1288.

INVOLVE 2015. Public Involvement in Research, Values and Principles Framework. Eastleigh: INVOLVE.

INVOLVE. 2019. What is public involvement in research? [online]. Available from: https://www.invo.org.uk/find-out-more/what-is-public-involvement-in-research-2/ [accessed 05/08/2019].

Jasper R. 2018. Personal communication.

Jessop, B. & Kino, K. 2019. Critical Realism, Symptomology and the Pedagogy of Crisis. In B. Jessop & K. Kino (Eds). *The Pedagogy of Economical, Political and Social Crises.* Routledge: Oxon, UK.

Johnson, B., & Christensen, L. (2012*). Educational Research* (4th ed.). Los Angeles, CA: Sage

Khan, K., Kunz, R., Kleijnen, J., Antes, G. 2011. *Systematic Reviews to Support Evidence Based Medicine* (2nd Edition). London: Hodder Arnold.

Kincaid, C. And Peacock, J.R., 2003. The effect of a wall mural on decreasing four types of door-testing behaviors. *Journal of Applied Gerontology.* 22 (1), pp. 76-88.

Kings Fund. 2019. What's your problem social care? Available from: https://www.kingsfund.org.uk/publications/whats-your-problem-social-care. Accessed 10.01.2020.

Kjorstad, M. & Solem, M-B. 2017. Critical Realism for Welfare Professionals. Routledge.

Klages, K., Zecevic, A., Organge, J.B. And Hobson, S., 2011. Potential of Snoezelen room multisensory stimulation to improve balance in individuals with dementia: a feasibility randomized controlled trial. *Clinical Rehabilitation.* 25(7), 607-616.

Koike, Y., Hoshitani, M., Tabata, Y., Seki, K., Nishimura, R & Kano, Y., 2012. Effects of Vibroacoustic therapy on elderly nursing home residents with depression. *Journal of Physical Therapy Science.* 24 (3), 291-294.

Kolanowski, A., Litaker, M., Buettner, L., Moeller, J And Costa Jr, P.T., 2011. A Randomized Clinical Trial of Theory-Based Activities for the Behavioral Symptoms of Dementia in Nursing Home Residents. *Journal of the American Geriatrics Society*. 59 (6), pp 1032-1041.

Krippendorf, K. 2013. *Content Analysis: an introduction to its methodology.* California: Sage.

La Cour, K., Josephsson, S. And Luborsky, M., 2005. Creating connections to life during life-threatening illness: creative activity experienced by elderly people and occupational therapists. *Scandinavian Journal of Occupational Therapy.* 12 (3), 98-109.

Ledger, A.J. & Baker, F.A. 2007. An investigation of long-term effects of group music therapy on agitation levels of people with Alzheimer's disease. *Aging and Mental Health.* 11 (3), 330-338.

Lewin, S., Glenton, C., Munthe-Kaas, H., Carlsen, B., Colvin, C.J., Golmezoglu, M., Noyes, J., Booth, A., Garside, A. And Rashidian, A. 2015. Using Qualitative Evidence in Decision Making for Health and Social Interventions: An Approach to Assess Confidence in Findings from Qualitative Evidence Syntheses (GRADE-CERQual). *PLoS Med 12*(10): e1001895.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gotzsche, P.C., Loannidis, J.P. Clarke, M., Devereaux, P.J., Kleiknen, J. And Moher, D. 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *British Medical Journal* doi: 10.1136/bmj.b2700

Lin, Y., Chu, H., Yang, C-Y., Chen, C-H., Chen, S-G., Chang, H-J., Hsieh, C-J., Chou, K-R. 2010. Effectiveness of group music intervention against agitated behaviour in elderly persons with dementia. *International Journal of Geriatric Psychiatry.* 26 (7), pp. 670-678.

Lincoln, Y.S. And Guba, E.G. *Naturalistic Enquiry.* London: Sage Publications.

Liu, X., Niu, X., Feng, Q. & Liu, Y. 2014. Effects of five-element music therapy on elderly people with seasonal affective disorder in Chinese nursing home. Journal of Traditional Chinese Medicine. 34 (2), pp. 159-161.

Lord, T.R. And Garner, J.E. 1993. Effects of music on Alzheimer patients. *Perceptual and Motor Skills.* 76, 451-455.

Marmot, M. Fair society, healthy lives: The Marmot Review: strategic review of health inequalities in England post-2010. [online]. Available from: https://www.gov.uk/dfid-research-outputs/fair-society-healthy-lives-the-marmot-review-strategic-review-of-health-inequalities-in-england-post-2010#citation. [accessed 09/08/2019].

Martin, L.S., Mcdowell, S., O'leary, J., Hagens, C., Bonas, B., Waxman, S., Abraham, B., Morden, P., Chong, A., Harris, C. And Rozon, L., 2004. An observational study of the impact of vibrotactile rhythm playing on seniors in long term care. *Activities, Adaption and Aging,* 28, 57-71.

Mays, N. And Pope, C. 2000. Qualitative research in health care: assessing quality in qualitative research. *British Medical Journal*. Pp. 320:50–52

McCormack, B. & Hill, L. 1997. *Conducting a Survey: The SPSS workbook.* UK: Cengage Learning Business Press.

Mcdermott, O., Orrelle, M And Ridder, H.M., 2014. The importance of music for people with dementia: the perspectives of people with dementia, family carers, staff and music therapists. *Aging and Mental Health*. 18 (6), pp 706-716..

Miles & P. Gilbert (Eds.). 2005. *A Handbook of Research Methods for Clinical and Health Psychology.* New York, USA: Oxford University Press.

Millard, K A O., Smith & Smith, J.M. 1989. The influence of group singing therapy on the behavior of Alzheimer's Disease patients. *Journal of Music Therapy.* 26 (2), 58-70.

Mohammadi, A.Z., Shahabi, T. And Panah, F.M. 2011. An evaluation of the effect of group music therapy on stress, anxiety and depression levels in nursing home residents. *Canadian Journal of Music Therapy.* 17(1), 55-68.
Moher, D., Liberati, A. Ytetzlaff, J. And Altman, D.G. 2009. Preferred reporting items for systematic reviews and meta-analysis. The PRISMA statement. *PLoS Med*, 6 (6) e1000097

Mortimer, J. And Green, M. 2015. *Briefing: The Health and Care of Older People In England 2015* [online]. Available from: http://www.cpa.org.uk/cpa/docs/AgeUK-Briefing-TheHealthandCareofOlderPeopleinEngland-2015.pdf [accessed 07/08/2019]

Mowlah, A., Niblett, V., Blackburn, J., And Harris, M. 2014. *The Value of Arts and Culture to People in Our Society. An Evidence Review* [online]. Available from: http://www.artscouncil.org.uk/publication/value-arts-and-culture-people-and-society [accessed 01/04/2019].

Murphy, E., Dingwall, R., Greatbatch, D. 1998. *Qualitative research methods in health technology assessment: a review of the literature*. York: Department of Health.

Musselwhite, K., Cuff, L., McGregor, L. & King, K. 2007. The telephone interview is an effective method of data collection in clinical nursing research: a discussion paper. *International Journal of Nursing Studies.* 44(6), pp. 1064-1070.

Myskja, A & Nord, P.G. 2008. "The day the music died": a pilot study on music and depression in a nursing home. *Nordic Journal of Music Therapy*. 17 (1), pp 30-40.

Nair, B.K., Heim, C., Krishnan, C., D'este, C., Marley, J. And Attia, J. 2011. The effect of baroque music on behavioural disturbances in patients with dementia. *Australasian Journal on Ageing*. 30 (1), pp 11-15.

National Institute for Clinical Excellence (NICE). 2013. *Mental Wellbeing of Older People in Care Homes.* [online]. Available from: https://www.nice.org.uk/guidance/qs50/resources/mental-wellbeing-of-older-people-in-care-homes-pdf-2098720457413 [accessed 07/08/2019].

National Institute for Health and Care Excellence (NICE). 2015. *Older People In Care Homes* [online]. Available from: https://www.nice.org.uk/advice/lgb25/chapter/introduction [accessed 20/03/2019].

National Institute for Health Care An Excellence. 2012. *The Guidelines Manual* [online] Available from: https://www.nice.org.uk/article/pmg6/chapter/6-reviewing-the-evidence [accessed 02-04-2016]

New Economics Foundation (NEF). 2008. Five Ways to Wellbeing [online]. Available from: http://www.fivewaystowellbeing.org/ [accessed 07/08/2019].

Noice, H. & Noice, T., 2006. A theatrical intervention to improve cognition in intact residents of long term care facilities. *Clinical Gerontologist.* 29 93), pp. 59-76.

Noice, T., Noice, H. & Kramer, A.F. 2014. Participatory arts for older adults: a review of benefits and challenges. *The Gerontologist.* 54(5), pp. 741-753.

Novick, G. 2008. Is there a bias against telephone interviews in qualitative research? *Research Nursing Health.* 31(4), pp. 391-398.

Noyes, J., Popay, J., Pearson, A., Hannes, K And Booth, A. 2011. Qualitative research and Cochrane reviews. In J.P.T. Higgins and S. Green (Eds). 2011 *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* (updated March 2011). [online] The Cochrane Collaboration. Available from www.cochrane-handbook.org [Accessed 04/04/2016].

O'Neill, D., 2009. The art of medicine: the art of the demographic dividend. The Lancet. 377 (9780), pp 1828-1829.

O'connor, D.W., Ames, D., Gardner, B. & King, M., 2009a. Psychosocial treatments of behavior symptoms in dementia: a systematic review of reports meeting quality standards. *International Psychogeriatrics.* 21 (2), pp. 225-240.

O'connor, D.W., Ames, D., Gardner, B. & King, M., 2009b. Psychosocial treatments of psychological symptoms in dementia: a systematic review of reports meeting quality standards. *International Psychogeriatrics.* 21 (2), pp. 241-251

Office For National Statistics, 2015. *Past and projected data from the period and cohort life tables, 2014-based, UK, 1981-2064* [online]. Available from: http://www.ons.gov.uk/ons/dcp171778_426798.pdf

Office for National Statistics. 2018*. Living longer: how our population is changing and why it matters* [online]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/ageing/articles/livinglongerhowourpopulationischangingandwhyitmatters/2018-08-13 [accessed 07/08/2019]

Opdenakker, Raymond (2006, August). Advantages and Disadvantages of Four Interview Techniques in Qualitative Research. *Qualitative Social Research*, 7(4), Art. 11

Orsulic-Jeras, S., Schneider, N.M & Camp, C.J., 2000. Special feature: Montessori-based activities for long-term care residents with dementia. *Topics in Geriatric Rehabilitation.* 16 (1), 78-91.

Palo, Bengtsson, L., Winblad, B. & Ekman, S.L. 1998. Social dancing: a way to support intellectual, emotional and motor functions in persons with dementia. *Journal of Psychiatric and Mental Health Nursing.* 5, 545-554.

Petticrew, M And Roberts. H. *Systematic Reviews in the Social Sciences. A Practical Guide.* Oxford: Blackwell Publishing.

Popay, J., Roberts, H., Sowden, A., Petticrew M., Acru, L., Rodgers, M., Britten, N., Roen, K And Duffy. S. 2006. Guidance on the conduct of narrative synthesis in systematic reviews. ESRC Methods Programme.

Popay, J., Rogers, A. 1998. Rationale and standards for the systematic review of qualitative literature in health services research. *Qualitative Health Research.* 8, pp. 341–352.

Pope, C., Mays, N. And Popay, J. 2007. *Synthesising Qualitative and Quantitative Health Research: A Guide to Methods*. Berkshire: McGraw-Hill Publishing

Public Health England. 2016. *Arts for Health and Wellbeing- An Evaluation Framework.* London: Public Health England.

Ragneskog, H., Kihlgren, M., Karlsson, I. & Norberg, A., 1996a. Dinner Music for Demented Patients. *Clinical Nursing Research.* 5, 3, 262-282.

Ragneskog, H., Braane, G., Karlsson, K & Kihlgren, M. 1996b. Influence of dinner music on food intake and symptoms common in dementia. *Scandinavian Journal of Caring Sciences.* 10, 11-17.

Remington, R. 2002. Calming music and hand massage with agitated elderly. *Nursing Research.* 51 (5), 317-323.

Ridder, H M O., Stige, B., Qvale, L.G. & Gold, C., 2013. Individual music therapy for agitation in dementia: an exploratory randomized controlled trial. Aging and Mental Health. 17 (6), 667-678.

Robinson, K.A., Saldanha, I.J And Mckay, N.A. 2011. Framework for determining research gaps during systematic reviews. *Methods Future Needs Report*. 2, report no 11- EHC043-EF

Roe, B. 2007. Chapter 2- Key considerations when undertaking a systematic review: bladder training for the management of urinary incontinence. In C.Webb and B.Roe (Eds) *Reviewing Research Evidence for Nursing Practice: Systematic Reviews* pp.9-21. Oxford: Wiley-Blackwell.

Roe, B., McCormick, S., Lucas, T., Gallagher, W., Winn, A. & Elkin, S. 2014. Coffee, cake & culture: Evaluation of an art for health programme for older people in the community. *Dementia.* 15 (4), 539-559.

Royal Society Of Public Health. 2013. *Arts Health and Wellbeing Beyond the Millennium: how far have we come and where do we want to go?* London: Working Group on Arts, Health and Wellbeing.

Royal Society of Public Health. 2013. *Arts Health and Wellbeing Beyond the Millennium: how far have we come and where do we want to go?* London: Working Group on Arts, Health and Wellbeing.

Ryan, R. 2013. *Cochrane consumers and communication review group: data synthesis and analysis* [online]. Available from: http://cccrg.cochrane.org/sites/cccrg.cochrane.org/files/uploads/AnalysisRestyled.pdf [accessed 04/04/2016]

Sackett,D.L., Straus, S.E., Richardson, W.S., Rosenberg, W And Haynes, R.B. 2000. *Evidence based medicine: How to practice and teach*. Edinburgh, Churchill Livingstone

Scott, T.L., Masser, B.M. And Pachana, N.A., 2014. Multisensory installations in residential aged-care facilities: increasing novelty and encouraging social engagement through modest environmental changes. *Journal of Gerontological Nursing*. 40 (9), pp 20-31

Scottish Government Social Research Team. 2013. *Healthy Attendance? The Impact of Cultural Engagement and Sports Participation on Health and Satisfaction with Life in Scotland* [online]. Available from: http://www.gov.scot/Resource/0043/00430649.pdf [accessed 20/03/2019]

Shaw, C., Mcnamara, R., Abrams, K., Cannings-John, R., Hood, K., Longo, M., Myles, S., O'mahony, S., Roe, B. And Williams, K. 2009. Systematic review of respite care in the frail elderly. *Health Technology Assessment: Executive Summaries* [online]. Available from http://www.ncbi.nlm.nih.gov/books/NBK56857/ [Accessed 31/03/2015].

Simard, J. & Volicer, L., 2010. Effects of Namaste care on residents who do not benefit from usual activities. *American Journal of Alzheimer's Disease and Other Dementias.* 25 (1), 46-50.

Skrajner, M.J. AND Camp, C.J. 2007. Resident-Assisted Montessori Programming (RAMP): use of a small group reading activity run by persons with dementia in adult day health care and long-term care settings. American Journal of Alzheimer's Disease and Other Dementias. 22 (1), 27-36.

Smith, E.M. 2005. Telephone interviewing in healthcare research: a summary of the evidence. 12, pp.32-41.

Smith, G.H. 1986. A comparison of the effects of three treatment interventions on cognitive functioning of Alzheimer patients. *Music Therapy.* 6A (1), 41-56.

Snyder-Halper, R., Thompson, C. & Schaffer, J. 2000. Comparison of mailed vs. Internet applications of the Delphi technique in clinical informatics research. Proceedings / AMIA ... Annual Symposium. AMIA Symposium, pp. 809-813.

Szlonki, G. & Hoffman, D. 2013. Online, Face-to-Face and Telephone Surveys – Comparing Different Sampling Methods in Wine Consumer Research. *Wine Economics and Policy.* pp. 1-15.

Social Care Institute For Excellence. 2012. *At a Glance 60: Preventing Loneliness and Social Isolation among Older People* [online]. Available from: http://www.scie.org.uk/publications/ataglance/ataglance60.asp [accessed 27/03/2019]

Sole, C., Mercadal-Brontons, M & Galati, A. 2014. Effects of group music therapy on quality of life, affect and participation in people with varying levels of dementia. *Journal of Music Therapy.*

Spencer, L., Ritchie, J., And Lewis, J. 2003. *Quality in qualitative evaluation: a framework for assessing research evidence*. Government Chief Social Researcher's Office, Occasional Papers Series 2.

Stallings, J. 2010. Collage as a Therapeutic Modality for Reminiscence in Patients with Dementia. *Art Therapy,* 27 (3), 136-140.

Staricoff, R.L. And Clift, M. 2011. *Arts and Music in Healthcare: An Overview of the Medical Literature: 2004-2011* [online]. Available from: http://www.lahf.org.uk/sites/default/files/Chelsea%20and%20Westminster%20Literature%20Review%20Staricoff%20and%20Clift%20FINAL.pdf [accessed 19/03/2019].

Stickley, T. 2012. *Qualitative Research in Arts and Mental Health: Context, Meanings and Evidence.* Herefordshire, UK : PCCS Books.

Stoop I., Harrison E. (2012) Classification of Surveys. In: Gideon L. (eds) Handbook of Survey Methodology for the Social Sciences. Springer, New York, NY

Sturges, J.E. & Hanraha, K.J. 2004. Comparing telephone and face-to-face qualitative interviewing: a research note. *Qualitative Research.* 4(1), pp.107-118.

Subramaniam, P., Woods, B. & Whitaker, C. 2014. Life review and life story books for people with mild to moderate dementia: a randomised controlled trial. *Aging and Mental Health.* 18 (3), 363-375.

Sung, H-C., Chang, A M and Lee W-L. 2010. A preferred music listening intervention to reduce anxiety in older adults with dementia in nursing homes. Journal of Clinical Nursing. 19 (7-8), 1056-1064.

Sung, H-C., Chang, S-M., Lee, W-L. & Lee, M-S. 2006. The effects of group music with movement intervention on agitated behaviours of institutionalized elders with dementia in Taiwan. Complementary Therapies in Medicine. 14 (2), 113-119.

Suzuki, A. 1998. The effects of music therapy on mood and congruent memory of elderly adults with depressive symptoms. *Music Therapy Perspectives.* 16 (2), pp. 75-80.

Svansdottir, H, B., & Snaedal, J. 2006. Music therapy in moderate and severe dementia of Alzheimer's type: a case-control study. *International Psychogeriatrics.* 18 (4), 613-621.

Thomson, G. 2015. *Key Issues for the 2015 Parliament* [online]. Available from: https://www.parliament.uk/business/publications/research/key-issues-parliament-2015/ [accessed 07/08/2019]

Thraves, L. 2016. *Fix Dementia Care: NHS and Care Homes.* [online]. Available from: https://www.alzheimers.org.uk/sites/default/files/migrate/downloads/fix_dementia_care_nhs_and_care_homes_report.pdf [accessed 07/08/2019].

Ullan, A.M., Belver, M.H., Badia, M., Moreno, C., Garrido, E., Gomez-Isla, J., Gonzalez-Ingelmo, E., Delgado, J., Serrano, I., Herrero, C., Manzanera, P. And Tejedor, L., 2013. Contributions of an artistic educational program for older people with early dementia: An exploratory qualitative study. *Dementia*. 12 (4), pp 425-446.

United Nations Population Fund. 2012. *Ageing in the 21st Century: A Celebration and a Challenge* [online] Available from: http://www.unfpa.org/sites/default/files/pub-pdf/Ageing%20report.pdf [accessed 07/08/2019].

United Nations, Department of Economic and Social Affairs, Population Division (2013). World Population Ageing 2013.

Vanderark, S., Newman, I. & Bell, S. 1983. The effects of music participation on quality of life in the elderly. *Music Therapy.* 3 (1), 71-81.

Vankova, H., Holmerova, I., Machacova, K., Volicer, L., Veleta, P. & Celko, A.M., 2014. The effect of dance on depressive symptoms in nursing home residents. *Journal of the American Medical Directors Association (JAMDA).* 8 (15), 582-587.

Vella-Burrows, T. 2015. Chapter 29- The arts and older people: a global perspective. In S. Clift and P.M. Camic (Eds) *Oxford Textbook of International Perspectives on Practice, Policy and Research.* Pp. 235-244. Oxford: Oxford University Press.

Vink, A.C., Bruinsma, M.S. And Scholten, R.J.P.M., (2011). Music therapy for people with dementia. *Cochrane Database and Systematic Reviews*. Issue 4. Chichester: John Wiley & Sons.

Vink, A.C., Zuidersma, M., Boersma, F., De Jonge, P., Zuidema, S.U. & Slaets, J.P.J. 2013. The effect of music therapy compared with general recreational activities in reducing agitation in people with dementia: a randomised controlled trial. *International Journal of Geriatric Psychiatry.* 28 (10), 1031-1038.

Walsh, D and Downe, S. 2006. Appraising the quality of qualitative research. *Midwifery.* 22(2), pp.108-119.

Webb, C. And Roe, B. 2007. Introduction to Systematic Reviews. In C.Webb and B.Roe (Eds) *Reviewing Research Evidence for Nursing Practice: Systematic Reviews.* pxi Oxford: Wiley-Blackwell.

Whitehead, L.C. (2007) Methodological and ethical issues in internet-mediated research in the field of health: an integrated review of the literature. *Social Science and Medicine.* 65: 4, 782–791.

World Health Organisation, 2014. *Facts about ageing* [online]. Available from: http://www.who.int/ageing/about/facts/en/ [Accessed 07/08/2019]

World Health Organisation. 2006. *Constitution of the World Health Organisation (amendments). Basic Documents*. 45th Edition. Geneva: World Health Organisation.

World Health Organisation. 2012. Chapter 1: Introduction, dementia a public health issue. pp 5-9. *In Dementia: A Public Health Priority*. Geneva: World Health Organisation.

World Health Organisation. 2015. *World Report on Ageing and Health.* Switzerland: World Health Organisation Press.

World Health Organisation. 2015. *World Report on Ageing and Health.* Switzerland: World Health Organisation Press.

Wyatt, J.C. 2000. When to use Web-Based Surveys. *Journal of the American Medical Information Association.* 7 (4), pp.426-429.

Yardley, L. 2000. Dilemmas in qualitative health research. *Psychology and Health.* 15 (2), pp. 215-228.

Yasamy, M.T., Dua, T., Harper, M & Saxena, S. 2013. *Mental Health of Older Adults: Addressing a Worldwide Health Concern* [online]. Available from: http://www.who.int/mental_health/world-mental-health-day/WHO_paper_wmhd_2013.pdf [accessed 07/08/2019].

Young, R., Camic, P.M & Tischler, V. 2015. The impact of community-based arts and health interventions on cognition in people with dementia: a systematic literature review. *Ageing and Mental Health* [online]. Available from: http://www.tandfonline.com/doi/full/10.1080/13607863.2015.1011080#.VQGOfDSsVKc [accessed 29/03/2019].

Ziv, N., Granot, A., Hai, S., Dassa, A. & Halmov, I. 2007. The effect of background stimulative music on behaviour in Alzheimer's patients. Journal of Music Therapy. 44 (4), 329-343.

Appendix 1

Example Search String

31. exp DANCE THERAPY/ OR exp MUSIC THERAPY/ OR exp ART THERAPY/
32. PERFORMING ARTS/
33. SINGING/
34. exp STORYTELLING/ OR exp READING/
35. POETRY/
36. REMINISCENCE THERAPY/
37. "memory box*".ti,ab
38. "sew*".ti,ab
39. "knit*".ti,ab
40. "creative art*".ti,ab
41. "paint*".ti,ab
42. "performing art*".ti,ab
43. (arts AND health).ti,ab
44. (arts AND humanities).ti,ab
45. "life stor*".ti,ab
46. "life narrative*".ti,ab
47. "life review*".ti,ab
48. theatre*.ti,ab
49. "visual art*".ti,ab
50. ((arts AND well-being)).ti,ab
51. 31 OR 32 OR 33 OR 34 OR 35 OR 36 OR 37 OR 38 OR 39 OR 40 OR 41 OR 42 OR 43 OR 44 OR 45 OR 46 OR 47 OR 48 OR 49 OR 50
52. exp NURSING HOMES/
53. "nurs* home*".ti,ab
54. "care* home*".ti,ab
55. (residential adj3 home*).ti,ab
56. 52 OR 53 OR 54 OR 55
57. 51 AND 56
58. 57 [Limit to: (Language English) and (Age Groups Middle Aged: 45-64 years or Aged~ 65+ years)

Appendix 2

Article

Dementia 2018, Vol. 17(6) 645–669

DOI: 10.1177/1471301217740960

journals.sagepub.com/home/dem

Systematic review of the impact of arts for health activities on health, wellbeing and quality of life of older people living in care homes

Amy Curtis, Lucy Gibson, Mary O'Brien and Brenda Roe
Edge Hill University, UK

Abstract
Care home populations frequently feature older people who often experience poor physical health and cognitive difficulties, along with vulnerability to psychological and social stressors. To date there has been no systematic review which focuses on the impact of arts for health activities to the care home population. Evidence was sourced from several databases and 71 studies were deemed eligible for inclusion in this review. These studies underwent data extraction and quality appraisal and the findings associated with health, wellbeing and quality of life are presented within this paper.

Keywords
older adults, arts for health, care home, systematic review, wellbeing, quality of life

Introduction

As the population of older adults increases, it is likely that the number who require long- term care will also rise (WHO, 2014). Care homes are facilities offering varying levels of support to those who require nursing care or assistance carrying out activities of daily living, with differing levels offered depending upon need and care home type. In many countries of the world, care home facilities (such as nursing homes, residential homes, supported or assisted living or aged care homes) are operated by local councils, voluntary agencies

Corresponding author:

Amy Curtis, Edge Hill University, St Helens Road, Ormskirk, Lancashire, UK.
Email: curtisa@edgehill.ac.uk

or are privately owned. Within the UK they must all be registered with the Care Quality Commission (Age UK, 2013; NICE, 2013).

Care home populations are diverse and residents often have chronic illnesses with many possessing more than one long-term health condition (Clift & Camic, 2015; Ham, Imison, Goodwin, Dixon, & South, 2011; O'Neill, 2009). A decrease in physical health can be a predictor for individuals moving into long-term care facilities (Age UK, 2015). In addition to physical health conditions, around 70% of individuals who reside within a care home may have dementia and those without a formal diagnosis will often have some degree of memory problems or cognitive difficulties (Alzheimer's Society, 2014). Due to the physical and social changes associated with living within a care home environment, residents are also susceptible to depression (Age UK, 2016; Anderson et al., 2011; NICE, 2015) and loneliness (Cowl and Gaugler, 2014; Social Care Institute for Excellence, 2012).

There is recognition that care homes often adequately respond to the physical needs of residents; however, there is also suggestion that long-term care should incorporate a more holistic approach to meet psychological and sociological needs (New Economics Foundation, 2008; NICE, 2013; WHO, 2014, 2015). It has been demonstrated that not meeting such needs can increase incidences of anxiety, low mood and lead to behavioural disturbances in older

people residing in care homes (deGuzman et al., 2011; Hancock, Woods, Challis, & Orrell, 2006). Quality of life is a quality indicator for care homes (Department of Health, 2014) along with the recommendation that older people in care homes are able to partake in 'meaningful activities' (NICE, 2015).

Research suggests residents can maintain quality of life, health and wellbeing by participation in arts-based activities. Benefits to older adults have been reported in terms of behaviour, psychological functioning, cognition, social relationships and overall health (Cowl and Gaugler, 2014; O'Connor, Ames, Gardner, & King, 2009a; O'Connor, Ames, Gardner, & King, 2009b, Kolanowski, Litaker, Buettner, Moeller, & Costa, 2011; Vink, Bruinsma, and Sholten, 2011). Along with these benefits, the ability to continue learning throughout the lifespan and foster creativity have also been highlighted (Age UK, 2015; Clift and Camic, 2015; Cowl and Gaugler, 2014; Vella-Burrows, 2015; Young, Camic, and Tischler, 2016).

Methods

Research aim

Despite recognition of emerging benefits to older people, to date, there has been no systematic review focused exclusively on evaluating the evidence of arts for health activities for older people who reside in care homes. The aim of this systematic review was to locate, analyse, appraise and synthesise existing empirical studies which have evaluated the impact of arts for health activities with older people who reside in care homes. The outcome measures focused upon were health, wellbeing and quality of life.

Registration of the review

Prior to conducting the review, a protocol was developed and registered with the University's Faculty Research Ethics Committee and the International Prospective Register of Systematic Reviews (PROSPERO, Curtis et al., 2015, CRD42015026264). PRISMA guidelines were followed for the methods and reporting of the review (Liberatiet al., 2009, Moher, Liberati, Ytetzlaff, & Altman, 2009) as well as the PICOS framework (Robinson, Saldanha, & Mckay, 2011).

Inclusion criteria

Published papers of empirical studies of arts for health activities with older people living in care homes were eligible for inclusion in this review. Arts for health activities could take place within or external to the care home (i.e. studies where residents visited arts establishments or where activities were carried out within a separate venue). The term 'care home' referred to nursing homes, dual registered care homes, aged care facilities and assisted living facilities. Study populations had a mean age of 60 years or above. Arts for health activities included those classified as performing arts, creative arts, visual arts and/or recreational activities, as defined by Arts Council England (2015). The arts activity could include active or passive engagement and be delivered individually or as a group. Study designs were qualitative, quantitative or mixed methods. For quantitative studies, there had to be a clear comparison between usual care or a comparison activity. However, uncontrolled pre–post/test studies were also eligible for inclusion where clear pre- and post-measures were reported.

Exclusion criteria

Studies were excluded where those taking part in the activity lived within their own homes, unless there was a proportion of residents who resided in a care home whose data could be extracted and included separately. Studies that reported on the effects of 'reminiscence' were also excluded unless they involved aspects of creativity such as 'life story work' or 'musical based reminiscence'. Non-empirical studies were also excluded, along with those not published in English.

Search strategy

No date restrictions were applied to the search; key databases were searched from inception to December 2016, with continual updates obtained monthly. Databases included PsychInfo, AHMED, MEDLINE, EMBASE and CINAHL. The Cochrane Library, Joanna Briggs Foundation and Web of Science Arts and Humanities and Social Sciences Citation Index were also searched. Additional hand-searching of reference lists from systematic reviews and relevant papers was also conducted. It was specified that the returned articles were available in English (where the database search allowed for this specification). Search terms included arts activities: 'dance therapy', 'music therapy', 'art therapy', 'singing', 'reading', 'poetry', 'reminis- cence therapy', 'memory box', 'sew', 'knit', 'creative art(s)', 'paint(ing), 'performing art', 'arts and health', 'arts and humanities', 'life story', 'life narrative', 'life review', 'theatre', 'visual art', and 'arts and wellbeing' combined with terms related to care homes: 'nursing home', 'residential home' and 'care home' (Literature search strategy available from the lead author and online supplementary table S1 – Literature Search Strategy). Search terms were piloted prior to use.

Search outcome

A total of 1091 potentially relevant studies were identified (see Figure 1). Following removal of duplicates, 790 papers were screened for eligibility at title and abstract level with 624 removed that did not meet inclusion criteria. The resulting 166 papers underwent full-text screening independently by two reviewers using a bespoke inclusion/ exclusion form and 74 met the inclusion criteria. The 74 papers represented a total of 71 studies: 59 (83%) quantitative, eight qualitative (11%) and four (6%) mixed methods (see Figure 1).

Data extraction

Data from the included studies were then extracted using a bespoke data extraction form based on the PICOS framework (Robinson et al., 2011) to extract all relevant data by the lead reviewer/author. All data extraction forms and pdfs of included studies were distributed amongst the review team and data extraction independently checked for errors or disagreements and consensus reached.

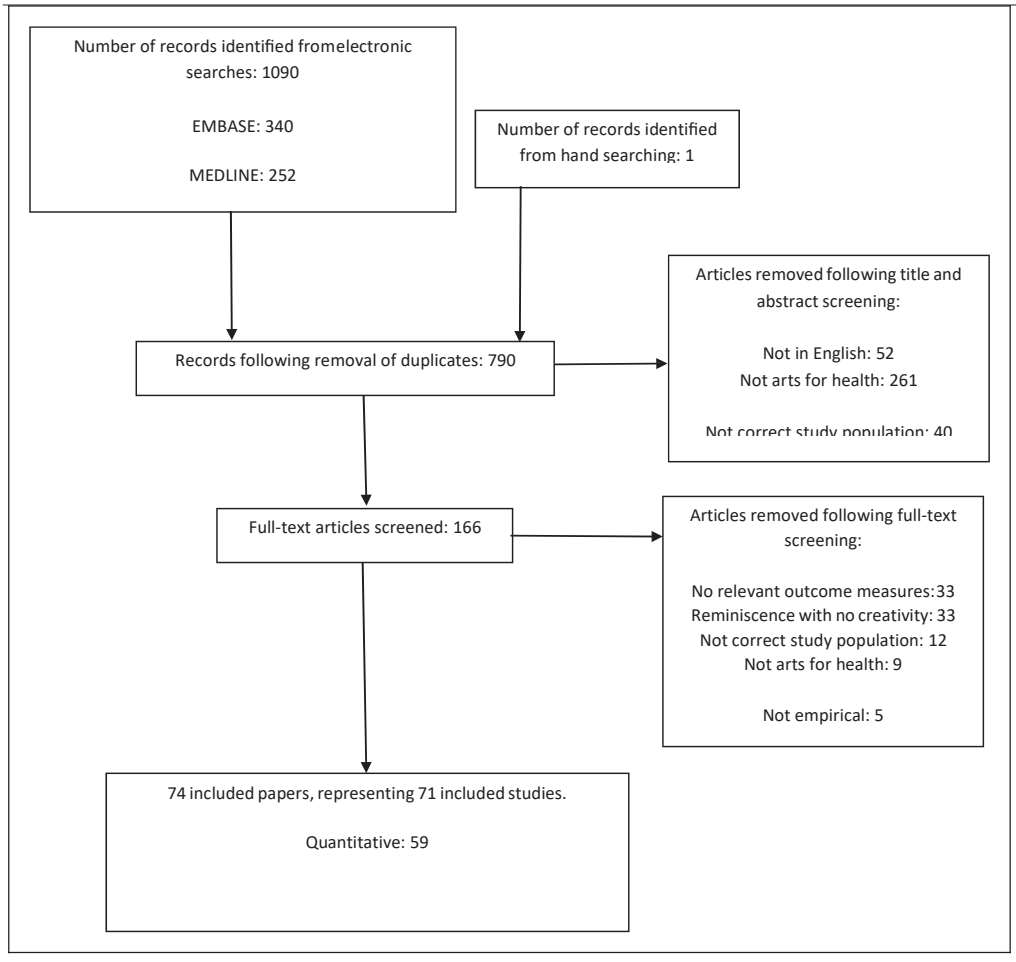


Figure 1. PRISMA flow chart of included studies.

Box 1 Criteria based on GRADE (Higgins and Green, 2011) to assess the risk of bias in quantitative studies.

- Evidence adequate randomisation took place
- Evidence of allocation concealment
- Evidence of blinding of outcome assessment – given the nature of the arts activities it would not be possible to blind those taking part and therefore this was judged to have occurred if the person scoring the outcome assessment was blind to group allocation.
- Complete/incomplete outcome data
- Selective reporting
- Any other apparent risk of bias

Box 2 Quality appraisal criteria for judgement and scoring of included qualitative studies. See Cooke et al. (2010a) and Downe et al. (2009).

- Whether the scope, purpose and aims were presented and clear
- If the study design was appropriate and well explained
- Sampling size adequacy and characteristics
- Whether there were data and evidence to support the author's interpretation of the results
- Evidence of reflexivity from the author

A – No, or few flaws. The study credibility, transferability, dependability and confirmability are high.

B – Some flaws, unlikely to affect the credibility, transferability, dependability and/or confirmability of the study.

C – Some flaws that may affect the credibility, transferability, dependability and/or confirmability of the study.

Quality appraisal

Quality appraisal of included studies was also undertaken by the lead reviewer and independently assessed and verified by another reviewer with agreement reached.

The Cochrane tool GRADE (Higgins and Green, 2011) was used for all randomised controlled trials (RCTs). Appraisal of quasi-RCT and uncontrolled pre/post-test studies was also undertaken using these criteria. A score of low risk (þ), high risk (–) or unclear risk (?) of bias was determined for each of the factors with a rationale explaining each decision, in keeping with the standards for appraising quality in a systematic review (Guyatt et al., 2011; Higgins and Green, 2011; Higgins et al., 2011) (see Box 1). Risk of bias graphs were created for quantitative studies using Review Manager 5.3 (RevMan).

A tool developed by Walsh and Downe (2006), based on Lincoln and Guba (1985), was used to appraise and score qualitative studies (Cooke, Mills, & Lavender, 2010a; Downe, Finlayson, Walsh, & Lavender, 2009) (see Box 2).

All studies were allocated a level of evidence score: I results based on a RCT, IIa results based on well-designed controlled study without randomisation, IIb evidence from a well-designed quasi-experimental design or III evidence from well-designed descriptive studies (NICE guidelines, 2012).

Data synthesis

Given the number and heterogeneity of included studies, their differing study designs, methods and outcome measures, data are presented in a narrative descriptive summary. This provides an overview and high level comparison of the different studies, their charac- teristics, key findings in relation to health, wellbeing and quality of life and their quality appraisal.

Findings

Characteristics of included studies

A table of included studies is available online, supplementary table 2 – table of included studies. Of the 71 included studies publication dates ranged from 1983 to 2015 with the majority published from 2000 onward. There was an increase in publications per decade, with four from the 1980s (4%), 14 from the 1990s (20%), 29 (41%) between 2000 and 2010 and 28 (39%) from the five-year period, 2010–2015. Study dates were only reported by 13 (18%) studies (Chang, Lu, Lin, & Chen, 2013; Cooke et al., 2010a; Doric-Henry, 1997; Guetin et al., 2009; Houser, 2014; Lin, 2010; Liu, Niu, Feng, & Liu, 2014; Myskja & Nord, 2008; Palo-Bengtsson, Winblad, & Ekman, 1998; Philips, 2010; Ragneskog, Kihlgren, Karlsson, & Norberg, 1996b; Ridder, Stige, Qvale, & Gold, 2013; Roe et al., 2016) and these ranged from 'spring' 1991 to December 2012.

Sixty-three (89%) studies were intervention studies: 59 were quantitative (83%) and four mixed methods (6%). Eight studies were qualitative (11%), all of which were descriptive. A total of 27 quantitative studies used uncontrolled pre/post-test designs (38%), 23 (32%) reported on RCTs, 12 quasi-experimental designs (17%) and one being a retrospective cohort study.

Each of the studies was broadly classified via arts type and consisted of the following: 48 (68%) involved activities classified as 'music', eight (11%) as multi-sensory activities, six (8%) as 'spoken and written word', six (8%) as performance and three as applied arts and crafts.

Location and setting

A third of studies took place within the USA (24, 33%). Second most cited countries were the UK (seven, 10%), Australia (seven, 10%) and Taiwan (seven, 10%). Six were carried out in Sweden (8%), five within Canada (7%) and two in Israel. The remaining studies were carried out in Belgium, Iceland, Norway, France, Holland, Spain, China, Japan, South Korea, Iran, Czech Republic and the Philippines, with one study originating from each country. Only one study featured data from more than one country, with participants included from 14 care homes in total, 10 from Norway and four from Denmark (Ridder et al., 2013).

The term 'nursing home' was used by 29 of the studies (41%), 11 (15%) referred to 'care homes', seven (10%) to 'long-term care facilities' and five (7%) reported residents lived

within 'residential homes'. Lesser featured terms were 'dementia specialty facility' (n = 3), 'residential facility' (n = 2), 'senior living residence' (n 2), 'residential units' (n 1), 'care facility' (n 1), 'care retirement community' (n 1), 'continuing care facility' (n 1), 'memory care residence' (n 1) and 'supported living facility' (n 1). Five studies did not specify a term for where participants resided.

Most studies (49, 69%) reported the arts activity took place within the care home itself. There was just one study which carried out an activity external to the home, where residents visited a gallery and museum (Roe et al., 2016). The remaining 21 (30%) studies did not explicitly state where the arts activity took place, although, it is implied from the description of the activities that they were carried out within the home environment.

Aims of the included studies

Most studies aimed to assess the impact of the activity on more than one outcome measure. The most popular being the impact of the arts activity on health behaviour symptoms reported by just under half of the studies (n 32, 45%). Many of these studies focused on determining whether there were reductions in agitated behaviours (n 17 = 24%). In addition, four studies aimed to assess the impact on verbally disruptive behaviours (VDB), five on 'behavioural symptoms associated with dementia', two exploring 'irritability and restlessness' and one focusing specifically on 'exit-seeking behaviours' exhibited by those who resided upon a ward within a care home. Five studies explored positive behaviours, aiming to determine levels of engagement with the arts activity.

The second most reported aim was assessing the impact of the arts activity on measures of psychological wellbeing, cited by a total of 24 studies (34%). Specific aims included the effects on depression, anxiety and general wellbeing measures.

Assessments of cognition were reported by 10 studies (14%), nine (13%) focused on the influence of the arts activity on the care process and eight (11%) explored quality of life . Lesser reported outcomes included feasibility of carrying out arts for health activities (n = 6, 8%), physical health (n = 4), satisfaction with living environment (i.e. the care home) (n = 2), socialisation (n = 2) and neuropsychiatric symptoms associated with dementia (n = 2).

Characteristics of included populations

The total number of older adults included was 2086 (reported for a total of 70 out of the 71 studies, 99%). There was large variation in sample size, from 3 to 117. Just over 70% of the studies (n 51) reported the mean age of those involved with the arts activity (or data which allowed for the calculation of a mean age) and across these the mean age of residents was 84 years, with a range between 69 and 87.5 years. Forty-four of these 51 studies (86% of those which reported a mean age) recorded their sample as aged 80 years or above and a further six studies reported a mean age above 70 years (Casby & Holm, 1994; Garland, 2006; Ho et al., 2011; Hong & Choi, 2011; Kincaid & Peacock, 2003; Sung, Chang, Lee, & Lee, 2007). This left only one study which reported a mean age of less than 70 years; however, this was very close at 69.4 years (Mohammadi, 2011).

The gender of those taking part was reported by 62 studies (87%), with women making up three quarters of the total sample across the studies (75%, n 2068), and men represent- ing 25% of the total (n 508). Only three studies featured more men than women (Mohammadi, 2011; Sung et al., 2006; Sung, Chang, & Lee, 2010), two reported equal

numbers of men and women (Hilliard, 2004; Ragneskog, Braane, Karlsson, & Kihlgren, 1996a) and two reported on data obtained from women only (Carruth, 1997; deGuzman, 2009).

Just under half the studies specified within inclusion criteria that residents required a dementia diagnosis to participate (n 35, 48%). Whilst not specifying a dementia diagnosis, the requirement for participants to display 'cognitive difficulties' was specified by Klages, Zecevic, Organge, and Hobson (2011). In addition, Simard and Volicer (2010) reported most participants to have 'severe' Cognitive Performance Scale scores. In addition, Martin et al. (2004) explicitly stated no residents were excluded from taking part based on cognitive abilities.

In contrast, nine studies featured inclusion criteria which may have prevented those with dementia (or other forms of cognitive impairment) from partaking (Bennett & Maas, 1988; Burrack, Jefferson, & Libow, 2004; Chang et al., 2013; Chen, Lin, & Jane, 2009; Hagen, 2004; Houston, Mckee, Carroll, & Marsh, 1998; Scott, Masser, & Pachana, 2014; Suzuki, 1998; Vankova et al., 2014). Both Hagen (2003) and Houston et al. (1998) excluded partic- ipants who had 'severe cognitive impairment'. AQ: The reference "Hagen (2004)" is cited in the text but is not listed in the references list. Please either delete in-text citation or provide full reference details following journal style.] Bennett and Maas (1988) excluded people with 'mental dysfunction or confusion', and Chang et al. (2013) included only those with 'good cognitive functioning'. Suzuki (1998) specified similar criteria outlining that residents required 'adequate reality orientation' to complete the outcome measures. Mini Mental State Examination (MMSE) cut-off scores were adopted as inclusion criteria by Burrack et al. (2004), Chen et al. (2009) and Vankova et al. (2014), with Scott et al. (2014) also assessing MMSE scores, although the cut-off was not reported.

Outcome measures

Of the studies which reported changes in behaviour as an outcome measure, 16 (23%) used the Cohen–Mansfield Agitation Inventory (CMAI). Five further studies used observational methods to assess behavioural changes, of which four devised their own schedules and one used an existing tool, the Behavioural Pathology in Alzheimer's Disease Rating Scale (Svansdottir & Snaedal, 2006). Two studies assessing engagement used the Menorah Park Engagement Scale (Heyn, 2003; Skrajner & Camp, 2007), another used the Myers-Research Institute Engagement Scale (Orsulic-Jeras, Schneider, & Camp, 2000).

The most commonly assessed area of psychological wellbeing was depression as measured by the Geriatric Depression Scale (Cooke et al., 2010a; Guetin et al., 2009; Scott et al., 2014; Subramaniam, Woods, & Whitaker, 2014; Vankova et al., 2014) and Cornell Scale for Depression (Ashida, 2000; Clarkson, 2007; Orsulic-Jeras et al., 2000; Philips, 2010). Lesser utilised measures were the Montgomery Asberg Rating Scale (Myskja, 2008), Gottfries–Brane–Steen Scale (Ragneskog et al., 1996a), Self-Rating Depression Scale (Liu, 2014), Dementia Mood Assessment Scale (Koike et al., 2012) and Beck Depression Inventory (Doric-Henry, 1997). Levels of anxiety were measured using the Rating Anxiety in Depression tool (Cooke, Moyle, Shum, Harrison, & Murfield, 2010b; Sung et al., 2010), Hamilton Scale for Anxiety (Guetin et al., 2009) and the Geriatric Anxiety Inventory (Scott et al., 2014). Combined measures of both anxiety and depression were obtained via the Depression and Anxiety Stress Scale (Mohammadi, 2011) and the Hospital Anxiety Depression Scale (Houston et al., 1998). In addition, the use of psychotropic medications,

often used to moderate anxiety and agitation, was collected in five studies (Brontons, 2009; Clarkson, 2007; Houser, 2014; Ridder et al., 2013; Simard & Volicer, 2010). Assessment of general mood and affect was assessed almost exclusively via observational methods using the Observed Emotion Rating Scale (Hammar, Emami, Gotell, & Engstrom, 2011; Philips, 2010), the Philadelphia Geriatric Affect Rating Scale (Fritsch et al., 2009) and the Affect Rating Scale (Cox, Burns, & Savage, 2004). Un-validated measures of mood were reported in three studies (Heyn, 2003; Lord & Garner, 1993; Sole, Mercadal-Brontons, & Galati, 2014).

Quality of life measures were assessed via the Quality of Life in Alzheimer's Disease Scale (Burrack et al., 2004), Dementia Quality of Life Scale (Cooke et al., 2010a), Alzheimer's Disease Rated Quality of Life, the Life Satisfaction Index-A (Bennett & Maas, 1988) and an un-validated scale created by Vanderark, Newman, and Bell (1983). Resident's satisfaction with their living environment was measured in two studies, each devising their own measure (Chang et al., 2013; Scott et al., 2014) but only the first study showed validation of the questionnaire within their study (Satisfaction with Living Environment Nursing Home).

Cognition was evaluated via scoring of the MMSE (Hong & Choi, 2011; Smith, 1986), Memory Retrieval Test (Suzuki, 1998), Cognitive Assessment Scale (Hagen, Armstrong-Esther, & Sandilands, 2003), Functional Assessment of Communication Skills (Philips, 2010) and Autobiographical Memory Interview (Subramaniam et al., 2014). One study used face–name recognition (Carruth, 1997) and another assessed 'alert responses' (Clair, 1996). Finally, Noice and Noice (2006) administered a range of measures which assessed word recall, listening span and the Means End Problem Solving Procedure.

Physical health was measured by balance and joint movement range (Hagen et al., 2003), oxygen saturation, temperature and pulse (Koike et al., 2012) and balance tests (Klages et al., 2011). Overall functioning was assessed via use of the Multidimensional Observational Scale of Elderly Subjects by Orsulic-Jeras et al. (2000).

Measures of behavioural changes and agitation were therefore always obtained via observations of residents, along with quality of care. General expression of 'emotion' and 'affect' was mostly obtained through observational methods, except for Koike et al. (2012). Residents always completed quality of life measures and depression outcome measures for all but two of the studies (Myskja, 2008; Orsulic-Jeras et al., 2000) both of whom used care staff to complete perceived depression. Similarly, anxiety was self-reported for all but one of the studies (Sung et al., 2006).

Quantitative findings

Behaviour and symptoms associated with dementia

Significant changes over time were reported by seven of the studies which utilised the CMAI (Garland, Beer, Eppingstall, & O'Connor, 2007; Goddaer & Abraham, 1994; Hicks-Moore & Robinson, 2007; Ho et al., 2011; Ledger & Baker, 2007; Lin et al., 2010; Vink et al., 2013). Despite this, four of the five studies that featured a comparison group taking part in other activities failed to show a significant difference between the two groups (Hicks-Moore & Robinson, 2007; Ledger & Baker, 2007; Remington, 2002; Vink et al., 2013). In contrast, two studies showed significant differences between the scores obtained between the com- parison and control group when the comparison group took part in 'normal daily activities' (Lin et al., 2010; Sung et al., 2006), suggesting that arts activities were more effective than

normal daily activities. Further evidencing this, Garland et al. (2007) showed both the arts activity (music) and comparison activity (simulated family presence) to significantly reduce agitation, but not one more so than the other. Despite this, both were more effective than the third control condition which involved no activity.

In respect of other behavioural symptoms, both Hagen et al. (2003) and Svansdottir and Snaedal (2006) showed changes over time for both groups (those taking part in the arts activity and the control group). No significant differences were found between the two groups for the first study (Svansdottir & Snaedal, 2006) and the second did not provide a comparison (Hagen et al., 2003). Houser (2014) and Orsulic-Jeras et al. (2000) failed to show significant effects of the arts activity. In contrast, Brontons (1998) did show the control group to have significantly increased behavioural disturbances at post-test; however, these scores fluctuated throughout the study for both groups and did not change over time.

Studies specifically focused on VDBs produced mixed findings. Cohen-Mansfield and Werner (1997) showed decreases over time, but not more so for the music group compared with the control group. Casby and Holm (1994) also suggested decreases over time but with only three participants and did not report any statistical analyses. Increases in verbal disruptions were evidenced in three studies (Clarkson, Cassidy, & Eskes, 2007; Cooke et al., 2010a; Nair et al., 2011); however, none explicitly stated what types of verbalisations increased. Ziv, Granot, Hai, Dassa, and Halmov (2007) showed a significant increase in positive verbalisations whereas negative verbalisations decreased. Similar anecdotal evidence was provided by both Millard and Smith (1989) and Sole et al. (2014). Further to this, four studies provided evidence of increased engagement with the activity (Fritsch et al., 2009; Holmes, Knights, Dean, Hodkinson, & Hopkins, 2006; Orsulic-Jeras et al., 2000; Skrajner & Camp, 2007). Again, descriptive accounts were provided by Heyn (2003) and Millard and Smith (1989) but no statistical evidence was provided.

None of the studies which specifically focused on dementia symptomatology showed improvements (Billington, Carroll, Davis, Healey, & Kinderman, 2013; Houser, 2014; Philips, 2010). The only study to assess door testing (Kincaid & Peacock, 2003) showed a statistically significant fall in exit-seeking behaviours.

Psychological wellbeing

Nine studies analysing depression showed significantly reduced scores over time (Doric-Henry, 1997; Guetin et al., 2009; Houston et al., 1998; Koike et al., 2012; Liu et al., 2014; Mohammadi, Shahabi, & Panah, 2011; Myskja, 2008; Ragneskog et al., 1996a; Vankova et al., 2014). However, eight failed to show any significant impact of the activity on depression levels (Clarkson et al., 2007; Cooke et al., 2010a; Houser, 2014; Orsulic-Jeras et al., 2000; Philips, 2010; Scott et al., 2014; Simard & Volicer, 2010; Subramaniam et al., 2014).

All but one of the studies (Cooke et al., 2010a) which assessed changes in anxiety showed a significant reduction in related scores over time (Guetin et al., 2009; Houston et al., 1998; Mohammadi et al., 2011; Sung et al., 2010; Svansdottir & Snaedal, 2006). Additionally, the four studies which featured a comparison group all showed significantly lower anxiety scores at follow-up for those who had taken part in the arts activity (Guetin et al., 2009; Houston et al., 1998; Sung et al., 2010; Svansdottir & Snaedal, 2006).

Significantly improved scores over time were also reported for general mood measures (Hammar, Emami, Engstrom, & Gotell, 2010; Lord & Garner, 1993; Sole et al., 2014; Suzuki, 1998) with evidence that the arts activity group improved significantly more than a comparison group (Lord & Garner, 1993). Noice and Noice (2006) failed to show improvements related to self-esteem or general psychological health. Increases in expression of positive emotions were also observed to significantly improve in Hammar et al. (2010), although this finding was not shown by Fritsch et al. (2009) or Suzuki (1998). However, Suzuki (1998) demonstrated a significant reduction in expressions of negative affect indicative that whilst positive emotions were not more apparent, there were reduced expressions of negative emotions. Fritsch et al. (2009) found that whilst those taking part in the arts activity displayed more levels of anger and fear, they also displayed less neutral affect.

Of the studies assessing medication changes, two reported that there was not enough data to analyse any trends (Brontons, 1998; Clarkson et al., 2007) and a further study found no significant changes (Houston, 2014). In contrast, two studies showed an impact of the arts activity, one via increased prescriptions of antipsychotic medications for the control group (Ridder et al., 2013), and the second through decreased anti-anxiety prescriptions through- out but no changes in antidepressants (Simard & Volicer, 2010).

Quality of life and general wellbeing

Evidence of benefit on general quality of life and wellbeing measures was limited, with improvements being reported by six studies (Bennett & Maas, 1988; Chang et al., 2013; Hagen et al., 2003; Hilliard, 2004; Philips, 2010; Subramaniam et al., 2014). Even these findings were tentative, Hilliard (2004) assessed length of life and determined that those taking part in music therapy lived longer. However, this failed to account for the fact that these individuals may have exhibited less advanced disease than those not taking part and therefore may have had greater life expectancy related to disease stage irrespective of the arts activity. Chang et al. (2013) also reported positive findings for only one measure of satisfaction, the living environment. This was associated with recalling pleasant memories which may indicate improvements in memory rather than quality of life. Sole et al. (2014) actually reported worsening of quality of life scores, which occurred for both the interven- tion and control group, although not to a significant degree. Both Hagen et al. (2003) and Subramaniam et al. (2014) showed improvements for those who took part in the arts activity at post-test; however, this was not sustained at follow-up, where both groups showed com- parative improvements.

Cognition

With respect to cognition, both studies using the MMSE showed improvements over time for those taking part in the music activity (Hong & Choi, 2011; Smith, 1986). Scores were also significantly improved compared with the control group for the RCT (Hong & Choi, 2011). Both Hagen et al. (2003) and Subramaniam et al. (2014) showed improvements in cognitive assessment and memory scores, respectively, over time; however, this was not compared with other activities. Similarly, Clair (1996) showed alert responses to have increased for the music condition significantly more than the silent condition, but not significantly more so than for the reading condition. Unpleasant memories decreased for those taking part in music reminiscence but the percentage of pleasant memories did not

increase (Suzuki, 1998). Working memory did not significantly improve when taking part in drama in Noice and Noice (2006), although word recall and problem solving did. Social communication was not shown to significantly increase (Philips, 2010) and there were very limited findings for face–name recognition (Carruth, 1997), where 'four participants showed an increase, three did not' with no statistical comparisons, possibly due to the limited sample size.

Quality of care and physical health

Three quantitative studies reported the aim of assessing quality of care. Two of these explored similar outcomes (Ragneskog et al., 1996a, 1996b) focusing on nutritional intake during mealtimes. The first showed nurses fed the residents significantly more during the music conditions and residents spent more time with them during meals. The second showed residents consumed larger portions of food. Hammar et al. (2011) was the only study to focus on restiveness to care. It was found that significantly less pulling away, grabbing objects and behaviour adduction were observed during the Music Therapeutic Caregiving condition as opposed to the normal morning care routines.

Significant physical health improvements were only shown by Koike et al. (2012) who found pulse and temperature to have improved. Balance (Hagen et al., 2003; Klages et al., 2011) and overall physical 'functioning' (Orsulic-Jeras et al., 2000) did not improve due to the arts activity.

Qualitative finding

Qualitative findings overlapped with quantitative findings in terms of identified themes and benefits. All qualitative studies outlined improvements to wellbeing. Specified by several of these was the ability for the arts activity to improve a sense of self in residents (Billington et al., 2013; Chen et al., 2009; DeGuzman, 2011; La Cour, 2005; Martin et al., 2004). These five studies also reported on the participation of residents, despite physical barriers which may exist. Creativity and the opportunity to take part in meaningful activity were reported by Billington et al. (2013), Chen et al. (2009), Houser (2014) and La Cour, Josephsson and Luborsky (2005). Cognitive improvements were reported by three studies including benefits to listening, memory and attention (Billington et al., 2013) and concentration (Palo- Bengtsson et al., 1998).

Difficulties with the usual care process were reported by studies which conducted interviews with staff members (Gotell, 2007; Hammar et al., 2010; Houser, 2014). This was contrasted with providing care during and following the arts activity which was reported as being more unified and cooperative. The ability for staff members to use their creativity from the sessions for problem solving during work was also highlighted by both Billington et al. (2013) and Guzman-Garcia (2013) along with general improvements to the overall atmosphere of the care home (Houser, 2014). Improved communication between the residents and staff members was also evidenced (Gotell, 2007; Hammar et al., 2010; Roe et al., 2016).

Increased socialisation was also reported in terms of interacting with other residents and between the care staff and residents (Billington et al., 2013; Chen et al., 2009; Guzman-Garcia, 2013; Martin et al., 2004; Roe et al., 2016). Sharing of memories prompted

by the arts activity was highlighted in all five of these studies to aide and promote these interactions.

The feasibility of carrying out the arts activity was described within Billington et al. (2013), Cox et al. (2004) and Roe et al. (2016). Aspects of the programme that were dis- cussed included the process of physically situating residents at the arts activity, staff require- ments, physical barriers to participation and carrying out the arts activity itself. Two studies that featured changes in environments, Cox et al. (2004, Snoezelen and garden activity) and Roe (2016, museum and gallery) described the benefits of exploring novel surroundings that differed from the care home. However, both also contained accounts of times when this was distressing for certain residents.

Quality of included studies

Levels of evidence scores were assigned to each of the studies. Twenty-three (32%) studies provided level I evidence based on an RCT, 11 (15%) provided level IIa results based on well-designed controlled study without randomisation and most, 29 (41%) level IIb based on well-designed quasi-experimental studies. The remaining eight studies (11%) corresponded to an evidence score of level III, well-designed descriptive studies.

RCTs

The summary of RCT judgements of quality can be found in Figure 2. Twelve of the RCTs provided adequate accounts of randomisation to obtain a low risk score for this measure and nine of these also gave detail as to methods of allocation concealment. Blinding of outcome assessment was reported for a total of four of the studies, with nine reporting no blinding to have taken place. The remaining studies did not provide enough information to make a judgement either way. Nine studies received judgements of low risk for incomplete data, just two receiving high-risk judgements, and the remaining studies did not contain enough data to make a judgement. Fourteen appeared to be free of selective reporting, four showed selective reporting and the remaining studies did not provide enough information. Finally, eight studies possessed other biases which may affect the credibility of the studies.

Quasi-RCTs

Summary of quasi-RCT judgements is presented in Figure 3. Two of the quasi-RCTs demonstrated blinding of the individual scoring the outcome assessment, five stated this did not happen (therefore judged as high risk) and four did not contain enough information to make a judgement. Eight were unclear as to whether there was any missing outcome data, two were judged as high risk and just one was deemed to be low risk. In terms of selective reporting, there was only evidence of this occurring within one study, a further six were judged as low risk and it was not clear whether this had taken place in the remaining two studies. Two demonstrated risks of other biases, six were deemed low risk and two did not provide enough information.

Uncontrolled pre/post-test

Summary of uncontrolled pre/post-test judgements is shown in Figure 4. Evidence of blinding of outcome assessments was shown for just two studies; this explicitly did not

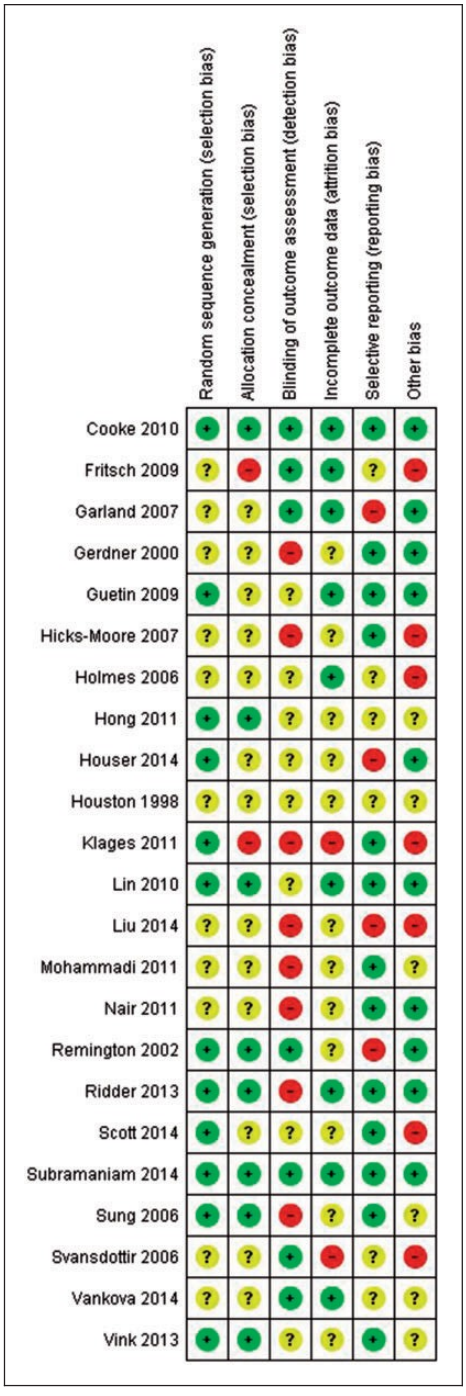


Figure 2. Risk of bias judgement for randomised controlled trials.

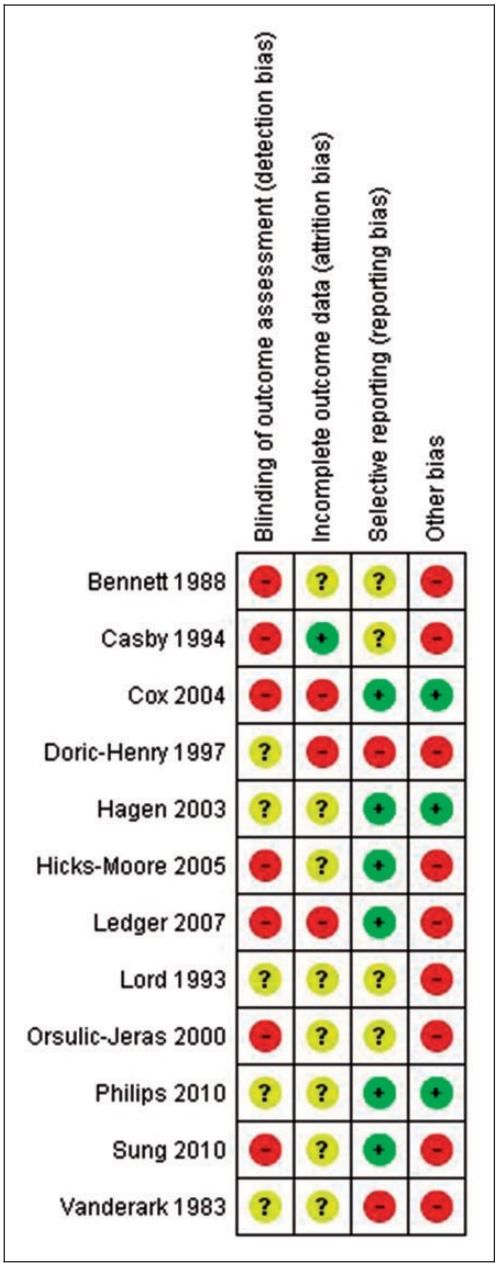


Figure 3. Risk of bias judgments for quasi-randomised controlled trials.

occur for 10 studies, and there was insufficient information to form a judgement about the others. Incomplete data were evidenced for seven studies, with 18 unclear and just one judged as having a low risk of bias. Selective reporting was also apparent in 14 studies, with just five receiving a judgement of low risk for this category. Finally, most of the studies

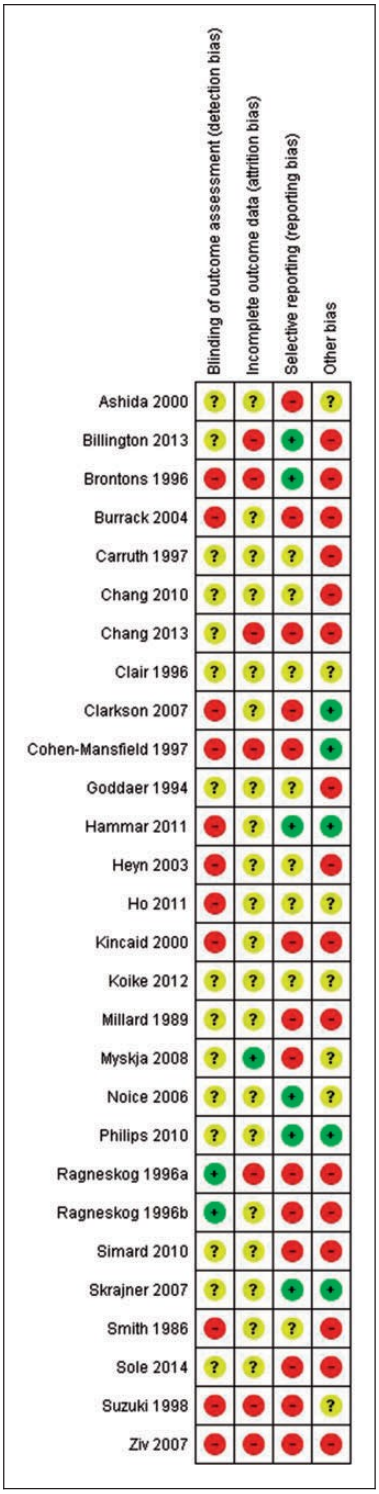


Figure 4. Risk of bias judgments for uncontrolled pre/post-test studies.

(15 in total) showed evidence of other biases, with only five low-risk judgements for this category.

Qualitative

Overall quality of the qualitative studies was judged to be good, with five studies (deGuzman, 2011; Gotell, 2007; Guzman-Garcia, 2013; La Cour et al., 2005; Roe et al., 2016) classified as Grade B for risk of bias and three judged at grade C (Chen et al., 2009; Martin et al., 2004; Palo-Bengtsson et al., 1998).

Discussion

The interest in arts activities for older people in care homes is illustrated by the increase in publications over time. Eighty per cent of the papers sourced within this review had been published since the year 2000, with 41% published between 2010 and 2015 alone. A higher proportion of studies focused upon music, accounting for almost 70% of the papers mean- ing that much of the evidence relates to this specific arts type. Future research would therefore benefit from explorations of other arts for health activities to achieve a greater evidence of potential benefits associated with older people living in care homes.

The focus of studies has also shifted over time. Apart from Hong and Choi (2011), studies with the aim of purely assessing cognitive changes were carried out in 1986 (Smith) and 1996 (Clair). Whilst subsequent studies have also explored the role of cognition, these also aimed to explore other areas such as quality of life or psychological wellbeing. Another (Philips, 2010) determined changes in social communication therefore focused more on socialisation. Similarly, the measures of quality of care within earlier studies (Ragneskog et al., 1996a, 1996b) provide evidence in regards to food consumption or nutritional intake, whereas latter studies explored the relationship between the carer and residents, along with restiveness to care (Gotell, 2007; Hammar et al., 2011; Houser, 2014).

Evidence was sourced from a range of study designs and there were disparities in quality across the studies. Unsurprisingly RCTs fared well in terms of quality, with other study designs showing a lower quality. This may partially be due to use of GRADE criteria for all study designs. A good standard of evidence was reflected in the scores obtained via qualitative evidence, particularly in those carried out within the last 10 years.

Overall there was mixed evidence of benefit for arts for health activities within care home populations. It may be suggested that negative findings may still be attributed to benefits. There was evidence that increased verbalisations were associated with increases in positive behaviours (Millard & Smith, 1989; Sole et al., 2014; Ziv et al., 2007), such as joining in with the arts activity which consequently reduced apathy in residents. Studies reporting on increased verbal disruptions did not discuss whether these were in respect of negative behaviours indicative of agitation or anxiety, or, whether this represents increased participation (Clarkson et al., 2007; Cooke et al., 2010a; Nair et al., 2011). Further evidence of the latter is provided by increased engagement with the arts activity (Fritsch et al., 2009; Holmes et al., 2006; Orsulic-Jeras et al., 2000; Skrajner & Camp, 2007) and increased expressions of positive affect (Hammar et al., 2010) along with a decrease in negative affect (Suzuki, 1998). Whilst Fritsch et al. (2009) showed increased incidences of fear and anxiety, there were also reductions in neutral affect, indicative that residents could express their emotions, even if these were negative, therefore improving levels of communication. Furthermore, qualitative

evidence showed a unique outcome regarding increased sense of self, which was not cap- tured by any quantitative studies (Billington et al., 2013; Chen et al., 2009; DeGuzman, 2011; La Cour, 2010; Martin et al., 2004). A decrease in self-reported ratings of quality of life outlined in Sole et al. (2014) may also be attributed to this increased self-awareness, again suggesting residents may become more aware of their emotions. Interestingly Noice and Noice (2006) failed to quantitatively capture increases in self-awareness and was the only study within this review which stated that residents were told when they were making mistakes when acting and therefore may have felt discouraged. Further analysis on the aspects of arts for health activities would be useful in determining what content works best.

There is also evidence that arts for health activities may be particularly beneficial for those with existing clinical need, such as pre-existing depression and anxiety. Both Liu (2014) and Suzuki (1998) described populations with clinical depression who showed improvements when taking part in arts activities. In direct contrast, Orsulic-Jeras et al. (2000) comment on low depression scores at baseline may explain a lack of effect within this study, given that residents already functioned well psychologically. A sub-analysis car- ried out by Simard and Volicer (2010) showed that improvements to behaviour were only evidenced in patients with low social interaction levels at baseline, which further evidences the fact that low pre-existing functional levels may be needed to show improvements.

Despite positive findings over time for those taking part in arts activities, there was limited evidence for arts activities to foster improvements more than comparison activities. Many quantitative studies exploring the effects of the arts activity on behaviour and psychological wellbeing showed improvements over time, but not compared with another activity. However, there was evidence that taking part in such arts activities improved outcomes for those in care homes more so than when no activities were offered. Further illustration is provided by qualitative findings that would be difficult to capture quantitatively, the most relevant being the opportunity for creativity (Billington et al., 2013; Chen et al., 2009; Houser, 2014; La Cour et al., 2005), which is unlikely to improve taking part in other activities. Improved socialisation with both peers and staff members was also captured by qualitative evidence (Billington et al., 2013; Chen et al., 2009; Guzman-Garcia, 2013; Martin et al., 2004; Roe, 2016) meaning both types of evidence are needed to ascertain the true benefits associated with the arts.

Limitations of the review

This review was confined to studies published in English which therefore may have impacted upon the countries evidence was sourced from. This may therefore mean not all cultural practices or countries being represented. A meta-analysis was also not able to be performed given the differences and heterogeneity of the included studies and this should be under- taken where possible in the future when evidence from similar studies and outcomes becomes available.

Conclusion

International evidence emerging to date for arts for health activities for older people in care homes whether within or external to the home looks promising and can benefit residents' health, wellbeing and quality of life. This review identifies evidence for including arts

activities within care homes and which activities may ascertain the greatest benefits. However, more research is required, particularly in relation to lesser evidenced arts activities such as crafts. Larger mixed methods complex intervention studies with qualitative elements embedded that adhere to international recognised standards for conduct and reporting such as CONSORT (Moher, 1998) and the MRC Framework for complex interventions/studies (Moore et al., 2015) are warranted. Where possible future meta-synthesis of qualitative evidence and meta-analysis of quantitative outcomes should be conducted.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Supplementary Material

Supplementary material for this article is available online.

References

Age UK. (2013). Care homes: Finding the right care home. Retrieved from http://www.ageuk.org.uk/ Documents/EN-GB/Information-guides/AgeUKIG06_Care_homes_inf.pdf?dtrk=true

Age UK. (2015). Agenda for later life 2015: A great place to grow older. Retrieved from http://www.ageuk.org.uk/professional-resources-home/policy/agenda-for-later-life/

Age UK. (2016). Later life in the United Kingdom. Retrieved from http://www.ageuk.org.uk/Documents/EN-GB/Factsheets/Later_Life_UK_factsheet.pdf?dtrk=true

Alzheimer's Society. (2014). Dementia 2014: Opportunity for change. Retrieved from http://www.alzheimers.org.uk/site/scripts/download_info.php?fileID=2317

Anderson, D., Banerjee, S., Barker, A., Connelly, P., Junaid, O., Series, H. S., & Seymour, J. (2011). The need to tackle age discrimination in mental health – A compendium of evidence. Retrieved from http://www.rcpsych.ac.uk/pdf/Royal%20College%20of%20Psychiatrists%20-%20The%20 Need%20to%20Tackle%20Age%20Discrimination%20in%20Mental%20Health%20Services% 20-%20Oct09.pdf

Arts Council England. (2015). *Cultural activities, artforms and wellbeing*. Manchester, UK: Arts Council England.

Ashida, S. (2000). The effect of reminiscence music therapy sessions on changes in depressive symptoms in elderly persons with dementia. *Journal of Music Therapy*, *37*(3), 170–182.

Bennett, S. L., & Maas, F. (1988). The effects of music-based life review on the life satisfaction and ego integrity of elderly people. *British Journal of Occupational Therapy*, *51*(12), 433–436.

Billington, J., Carroll, J., Davis, P. R., Healey, C., & Kinderman, P. (2013). A literature-based intervention for older people living with dementia. *Perspectives in Public Health*, *133*(3), 165–173.

Brontons, M., & Pickett-Cooper, P. K. (1998). The effects of music therapy intervention on agitation behaviors of Alzheimer's disease patients. *Journal of Music Therapy*, *33*(1), 2–18.

Burrack, O. R., Jefferson, P., & Libow, L. S. (2004). Individualized music: A route to improving the quality of life for long-term care residents. *Activities, Adaption and Aging*, *27*(1), 63–76.

Carruth, E. K. (1997). The effects of singing and the spaced retrieval technique on improving face- name recognition in nursing home residents with memory loss. *Journal of Music Therapy*, *34*(3), 165–186.

Casby, J. A., & Holm, M. B. (1994). The effect of music on repetitive disruptive vocalizations of persons with dementia. *The American Journal of Occupational Therapy*, *48*(10), 883–889.

Chang, C.-H., Lu, M.-S., Lin, T.-E., & Chen, C.-H. (2013). The effectiveness of visual art on envi- ronment in nursing home. *Journal of Nursing Scholarship*, *45*(2), 107.

Chen, S.-L., Lin, H.-C., & Jane, S.-W. (2009). Perceptions of a group music therapy among elderly nursing home residents in Taiwan. *Complementary Therapies in Medicine*, *17*(4), 190–195.

Clair, A. A. (1996). The effect of singing on alert responses in persons with late stage dementia. *Journal of Music Therapy*, *23*(4), 243–247.

Clarkson, K. A., Cassidy, K.-L., & Eskes, G. A. (2007). Singing soothes: Music concerts for the management of agitation in older adults with dementia. *Canadian Journal of Geriatrics*, *10*(3), 80–83.

Clift, S., & Camic, P. M. (2015). Introduction to the field of creative arts, wellbeing and health:

Achievements and current challenges. In S. Clift & P. M. Camic (Eds.), *Oxford textbook of international perspectives on practice, policy and research* (pp. 235–244). Oxford, UK: Oxford University Press.

Cohen-Mansfield, J., & Werner, P. (1997). Management of verbally disruptive behaviors in nursing home residents. *Journal of Gerontology: Medical Sciences*, *52a*(6), M369–M377.

Cooke, A., Mills, T. A., & Lavender, T. (2010a). 'Informed and uninformed decision making'

–Women's reasoning, experiences and perceptions with regard to advanced maternal age and delayed childbearing: A meta-synthesis. *International Journal of Nursing Studies*, *10*, 1317–1329.

Cooke, M., Moyle, W., Shum, D., Harrison, S., & Murfield, J. (2010b). A randomized controlled trial exploring the effect of music on quality of life and depression in older people with dementia. *Journal of Health Psychology*, *15*(5), 765–767.

Cowl, A. L., & Gaugler, J. E. (2014). Efficacy of creative arts therapy in treatment of Alzheimer's disease and dementia: A systematic literature review. *Activities, Adaption and Aging*, *38*, 281–330.

Cox, H., Burns, I., & Savage, S. (2004). Multisensory environments for leisure: Promoting well-being in nursing home residents with dementia. *Journal of Gerontological Nursing*, *30*, 37–45.

Curtis, A., O'Brien, M., Gibson, L., Maden, M., Gallagher, W., & Roe, B. (2015). The impact of arts for health on the health status, well-being and quality of life in older adults who reside in care

homes – protocol. Retrieved from http://www.crd.york.ac.uk/PROSPERO/ Registration number CRD42015026264.

Deguzman, A. B., Satuito, J. C. B., Satumba, M. A. E., Segui, R. A., Serquina, F. E. C., Serrano, L. J. O., & Sevilla, M. D. (2011). Filipino arts among elders in institutionalized care settings. *Educational Gerontology*, *37*, 248–261.

Department of Health. (2014). Adult social care outcomes framework. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/375431/ASCOF_15-16.pdf

Doric-Henry, L. (1997). Pottery as art therapy with elderly nursing home residents. *Art Therapy*, *14*, 163–171.

Downe, S., Finlayson, K., Walsh, D., & Lavender, T. (2009). Weighing up and balancing out': A meta-synthesis of barriers to antenatal care for marginalised women in high-income countries. *BJOG – An International Journal of Obstetrics and Gynaecology*, *116*, 518–529.

Fritsch, T., Kwak, J., Grant, S., Lang, J., Montgomery, R. R., & Basting, A. D. (2009). Impact of TimeSlips, a creative expression intervention program, on nursing home residents with dementia and their caregivers. *The Gerontologist*, *49*, 117–127.

Garland, K., Beer, E., Eppingstall, B., & O'Connor, D. W. (2007). A comparison of two treatments of agitated behaviour in nursing home residents with dementia: Simulated family presence and preferred music. *American Association for Geriatric Psychiatry*, *15*, 514–521.

Goddaer, J., & Abraham, I. L. (1994). Effects of relaxing music on agitation during meals among nursing home residents with severe cognitive impairment. *Archives of Psychiatric Nursing*, *8*, 150–158.

Gotell, E., Brown, S., & Ekman, S.-L. (2007). The influence of caregiver singing and background music on vocally expressed emotions and moods in dementia care: A qualitative analysis. *International Journal of Nursing Studies*, *46*, 422–430.

Guetin, S., Portet, F., Picot, M. C., Pommie, C., Messaoudi, M., Djabelkir, L., ... Touchon, J. (2009). Effect of music therapy on anxiety and depression in patients with Alzheimer's type dementia: Randomised, controlled study. *Dementia*, *28*, 36–46.

Guyatt, G., Oxman, A. D., Alk, E. A., Kunz, R., Vist, G., Brozek, J., .. . Schunemann, H. J. (2011). GRADE guidelines: 1. Introduction- GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, *64*, 383–394.

Guzman-Garcia, A., Mukaetova-Ladinska, E., & James, I. (2013). Introducing a Latin ballroom dance class to people with dementia living in care homes, benefits and concerns: A pilot study. *Dementia*, *12*, 523–535.

Hagen, B., Armstrong-Esther, C., & Sandilands, M. (2003). On a happier note: Validation of musical exercise for older persons in long-term care settings. *International Journal of Nursing Studies*, *40*, 347–357.

Ham, C., Imison, C., Goodwin, N., Dixon, A., & South, P. (2011). *Where next for NHS reforms? The case for integrated care*. London, UK: The Kings Fund.

Hammar, L. M., Emami, A., Engstrom, G., & Gotell, E. (2010). Finding the key to communion – Caregivers' experience of 'music therapeutic caregiving' in dementia care: A qualitative analysis. *Dementia*, *10*, 98–111.

Hammar, L. M., Emami, A., Gotell, E., & Engstrom, G. (2011). The impact of caregivers' singing on expressions of emotion and resistance during morning care situations inn persons with dementia: An intervention in dementia care. *Journal of Clinical Nursing*, *20*, 969–978.

Hancock, G. A., Woods, B., Challis, D., & Orrell, M. (2006). The needs of older people with dementia in residential care. *International Journal of Geriatric Psychiatry*, *21*, 43–49.

Heyn, P. (2003). The effect of a multisensory exercise program on engagement, behaviour, and selected physiological indexes in persons with dementia. *American Journal of Alzheimer's Disease and Other Dementias*, *18*, 247–251.

Hicks-Moore, S. L., & Robinson, B. A. (2007). Favorite music and hand massage: Two interventions to decrease agitation in residents with dementia. *Dementia*, *24*, 195–196.

Higgins J. P. T., Green. S. (eds). *Cochrane Handbook for Systematic Reviews of Interventions* Version

5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from http://handbook.cochrane.org.

Higgins, J. P. T., Altman, D. G., Gotzsche, P. C., Juni, P., Moher, D., Oxman, A. D., .. . Sterne,

J. A. C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials.

BMJ, *343*, d5928.

Hilliard, R. E. (2004). A post-hoc analysis of music therapy services for residents in nursing homes receiving hospice care. *Journal of Music Therapy*, *41*, 40–41.

Ho, S.-Y., Lai, H.-L., Jeng, S.-Y., Tang, C.-W., Sung, H.-C., & Chen, P.-W. (2011). The effects of researcher-composed music at mealtime on agitation in nursing home residents with dementia. *Archives of Psychiatric Nursing*, *25*, e49–e55.

Holmes, C., Knights, A., Dean, C., Hodkinson, S., & Hopkins, V. (2006). Keep music live: Music and alleviation of apathy in dementia subjects. *International Psychogeriatrics*, *18*, 623–630.

Hong, I. S., & Choi, M. J. (2011). Songwriting oriented activities improve the cognitive functions of the aged with dementia. *The Arts in Psychotherapy*, *38*, 221–228.

Houser, W. S., & George, D. R. (2014). Impact of TimeSlips creative expression program on behav- ioural symptoms and psychotropic medication use in persons with dementia in long-term care: A cluster-randomized pilot study. *American Journal of Geriatric Psychiatry*, *22*, 337–340.

Houston, D. M., Mckee, K. J., Carroll, L., & Marsh, H. (1998). Using humour to promote psychological wellbeing in residential homes for older people. *Aging and Mental Health*, *2*, 328–332.

Kincaid, C., & Peacock, J. R. (2003). The effect of a wall mural on decreasing four types of door- testing behaviors. *Journal of Applied Gerontology*, *22*, 76–88.

Klages, K., Zecevic, A., Organge, J. B., & Hobson, S. (2011). Potential of Snoezelen room multisen- sory stimulation to improve balance in individuals with dementia: A feasibility randomized con- trolled trial. *Clinical Rehabilitation*, *25*, 607–616.

Koike, Y., Hoshitani, M., Tabata, Y., Seki, K., Nishimura, R., & Kano, Y. (2012). Effects of vibroa- coustic therapy on elderly nursing home residents with depression. *Journal of Physical Therapy Science*, *24*, 291–294.

Kolanowski, A., Litaker, M., Buettner, L., Moeller, J., & Costa, P. T. Jr (2011). A randomized clinical trial of theory-based activities for the behavioral symptoms of dementia in nursing home residents. *Journal of the American Geriatrics Society*, *59*, 1032–1041.

La Cour, K., Josephsson, S., & Luborsky, M. (2005). Creating connections to life during life- threatening illness: Creative activity experienced by elderly people and occupational therapists. *Scandinavian Journal of Occupational Therapy*, *12*, 98–109.

Ledger, A. J., & Baker, F. A. (2007). An investigation of long-term effects of group music therapy on agitation levels of people with Alzheimer's disease. *Aging and Mental Health*, *11*, 330–338.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Loannidis, J. P., .. . Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *British Medical Journal.* Retrieved from https://doi.org/10.1136/bmj.b2700

Lin, Y., Chu, H., Yang, C.-Y., Chen, C.-H., Chen, S.-G., Chang, H.-J., .. . Chou, K.-R. (2010).

Effectiveness of group music intervention against agitated behaviour in elderly persons with demen- tia. *International Journal of Geriatric Psychiatry*, *26*, 670–678.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic enquiry*. London, UK: Sage Publications.

Liu, X., Niu, X., Feng, Q., & Liu, Y. (2014). Effects of five-element music therapy on elderly people with seasonal affective disorder in Chinese nursing home. *Journal of Traditional Chinese Medicine*, *34*, 159–161.

Lord, T. R., & Garner, J. E. (1993). Effects of music on Alzheimer patients. *Perceptual and Motor Skills*, *76*, 451–455.

Martin, L. S., McDowell, S., O'Leary, J., Hagens, C., Bonas, B., Waxman, S., … Rozon, L. (2004). An observational study of the impact of vibrotactile rhythm playing on seniors in long term care. *Activities, Adaption and Aging*, *28*, 57–71.

Millard, K. A., & Smith, J. M. (1989). The influence of group singing therapy on the behavior of Alzheimer's Disease patients. *Journal of Music Therapy*, *26*, 58–70.

Mohammadi, A. Z., Shahabi, T., & Panah, F. M. (2011). An evaluation of the effect of group music therapy on stress, anxiety and depression levels in nursing home residents. *Canadian Journal of Music Therapy*, *17*, 55–68.

Moher, D. (1998). CONSORT: An evolving tool to help improve the quality of reports of randomised controlled trials. Consolidated standards of reporting trials. *JAMDA*, *279*, 1489–1491.

Moher, D., Liberati, A., Ytetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analysis. The PRISMA statement. *PLoS Med*, *6*, e1000097.

Moore, G., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D. & Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *British Medical Journal, 350*: h1258.

Myskja, A., & Nord, P. G. (2008). "The day the music died": A pilot study on music and depression in a nursing home. *Nordic Journal of Music Therapy*, *17*, 30–40.

Nair, B., Heim, C., Krishnan, C., D'este, C., Marley, J., & Attia, J. (2011). The effect of Baroque music on behavioural disturbances in patients with dementia. *Australian Journal of Ageing*, *1*, 11–15.

National Institute for Health and Care Excellence. (2012). The guidelines manual. Retrieved from https://www.nice.org.uk/article/pmg6/chapter/6-reviewing-the-evidence

National Institute for Health and Care Excellence. (2013). Dementia: Independence and wellbeing.

Retrieved from https://www.nice.org.uk/guidance/qs30

National Institute for Health and Care Excellence. (2015). Older people in care homes. Retrieved from https://www.nice.org.uk/advice/lgb25/chapter/introduction

New Economics Foundation (NEF). (2008). Five ways to wellbeing. Retrieved from http://www. fivewaystowellbeing.org/.

Noice, H., & Noice, T. (2006). A theatrical intervention to improve cognition in intact residents of long term care facilities. *Clinical Gerontologist*, *29*, 59–76.

O'Connor, D. W., Ames, D., Gardner, B., & King, M. (2009a). Psychosocial treatments of behavior symptoms in dementia: A systematic review of reports meeting quality standards. *International Psychogeriatrics*, *21*, 225–240.

O'Connor, D. W., Ames, D., Gardner, B., & King, M. (2009b). Psychosocial treatments of psycho- logical symptoms in dementia: A systematic review of reports meeting quality standards. *International Psychogeriatrics*, *21*, 241–251.

O'Neill, D. (2009). The art of medicine: The art of the demographic dividend. *The Lancet*, *377*, 1828–1829.

Orsulic-Jeras, S., Schneider, N. M., & Camp, C. J. (2000). Special feature: Montessori-based activities for long-term care residents with dementia. *Topics in Geriatric Rehabilitation*, *16*, 78–91.

Palo-Bengtsson, L., Winblad, B., & Ekman, S. L. (1998). Social dancing: A way to support intellectual, emotional and motor functions in persons with dementia. *Journal of Psychiatric and Mental Health Nursing*, *5*, 545–554.

Phillips, L., Reid-Arndt, S.A. and Pak, Y. (2010). Effects of a creative expression intervention on emotions, communication, and quality of life in persons with dementia. *Nursing Research. 6*, 417–425.

Ragneskog, H., Braane, G., Karlsson, K., & Kihlgren, M. (1996a). Influence of dinner music on food intake and symptoms common in dementia. *Scandinavian Journal of Caring Sciences*, *10*, 11–17.

Ragneskog, H., Kihlgren, M., Karlsson, I., & Norberg, A. (1996b). Dinner music for demented patients. *Clinical Nursing Research*, *5*, 262–282.

Remington, R. (2002). Calming music and hand massage with agitated elderly. *Nursing Research*, *51*, 317–323.

Ridder, H. M. O., Stige, B., Qvale, L. G., & Gold, C. (2013). Individual music therapy for agitation in dementia: An exploratory randomized controlled trial. *Aging and Mental Health*, *17*, 667–678.

Robinson, K. A., Saldanha, I. J., & Mckay, N. A. (2011). *Framework for determining research gaps during systematic reviews*. *Methods Future Needs Report* 2, report no 11-EHC043-EF. Baltimore: The Johns Hopkins University Evidence-based Practice Centre

Roe, B., McCormick, S., Lucas, T., Gallagher, W., Winn, A., & Elkin, S. (2016). Coffee, cake & culture: Evaluation of an art for health programme for older people in the community. *Dementia*, *15*, 539–559.

Scott, T. L., Masser, B. M., & Pachana, N. A. (2014). Multisensory installations in residential aged- care facilities. *Journal of Gerontological Nursing*, *40*, 20–31.

Simard, J., & Volicer, L. (2010). Effects of Namaste care on residents who do not benefit from usual activities. *American Journal of Alzheimer's Disease and Other Dementias*, *25*, 46–50.

Skrajner, M. J., & Camp, C. J. (2007). Resident-Assisted Montessori Programming (RAMP): Use of a small group reading activity run by persons with dementia in adult day health care and long-term care settings. *American Journal of Alzheimer's Disease and Other Dementias*, *22*, 27–36.

Smith, G. H. (1986). A comparison of the effects of three treatment interventions on cognitive func- tioning of Alzheimer patients. *Music Therapy*, *6A*, 41–56.

Social Care Institute for Excellence. (2012). At a glance 60: Preventing loneliness and social isolation among older people. Retrieved from http://www.scie.org.uk/publications/ataglance/ata glance60.asp

Sole, C., Mercadal-Brontons, M., & Galati, A. (2014). Effects of group music therapy on quality of life, affect and participation in people with varying levels of dementia. *Journal of Music Therapy*, *51*, 103–125.

Subramaniam, P., Woods, B., & Whitaker, C. (2014). Life review and life story books for people with mild to moderate dementia: A randomised controlled trial. *Aging and Mental Health*, *18*, 363–375. Sung, H.-C., Chang, A. M., & Lee, W.-L. (2010). A preferred music listening intervention to reduce anxiety in older adults with dementia in nursing homes. *Journal of Clinical Nursing*, *19*, 1056–1064.

Sung, H.-C., Chang, S.-M., Lee, W.-L., & Lee, M.-S. (2006). The effects of group music with move- ment intervention on agitated behaviours of institutionalized elders with dementia in Taiwan. *Complementary Therapies in Medicine*, *14*, 113–119.

Suzuki, A. (1998). The effects of music therapy on mood and congruent memory of elderly adults with depressive symptoms. *Music Therapy Perspectives*, *16*, 75–80.

Svansdottir, H. B., & Snaedal, J. (2006). Music therapy in moderate and severe dementia of Alzheimer's type: A case-control study. *International Psychogeriatrics*, *18*, 613–621.

Vanderark, S., Newman, I., & Bell, S. (1983). The effects of music participation on quality of life in the elderly. *Music Therapy*, *3*, 71–81.

Vankova, H., Holmerova, I., Machacova, K., Volicer, L., Veleta, P., & Celko, A. M. (2014). The effect of dance on depressive symptoms in nursing home residents. *Journal of the American Medical Directors Association (JAMDA)*, *15*, 582–587.

Vella-Burrows, T. (2015). Chapter 29 – The arts and older people: a global perspective. In S. Clift & P. M. Camic (Eds.), *Oxford textbook of international perspectives on practice, policy and research* (pp. 235–244). Oxford, UK: Oxford University Press.

Vink, A. C., Bruinsma, M. S., & Scholten, R. J. P. M. (2011). *Music therapy for people with dementia. Cochrane database and systematic reviews* (Issue 4). Chichester, UK: John Wiley & Sons.

Vink, A. C., Zuidersma, M., Boersma, F., De Jonge, P., Zuidema, S. U., & Slaets, J. P. J. (2013). The effect of music therapy compared with general recreational activities in reducing agitation in people with dementia: A randomised controlled trial. *International Journal of Geriatric Psychiatry*, *28*, 1031–1038.

Walsh, D., & Downe, S. (2006). Appraising the quality of qualitative research. *Midwifery*, *22*, 108–119.

World Health Organisation. (2014). Facts about ageing. Retrieved from http://www.who.int/ageing/about/facts/en//.

World Health Organisation. (2015). *World report on ageing and health*. Geneva, Switzerland: World Health Organisation Press.

Young, R., Camic, P. M., & Tischler, V. (2016). The impact of community-based arts and health interventions on cognition in people with dementia: A systematic literature review. *Ageing and Mental Health*, *20*, 337–351. Retrieved from http://www.tandfonline.com/doi/full/10.1080/13607863.2015.1011080#.VQGOfDSsVKc

Ziv, N., Granot, A., Hai, S., Dassa, A., & Halmov, I. (2007). The effect of background stimulative music on behaviour in Alzheimer's patients. *Journal of Music Therapy*, *44*, 329–343.

Appendix 3

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
010. Subramaniam (2014) Intervention Dates of study not specified.	UK 14 care homes across North Wales (all but one privately owned) recruited, 11 completed.	23 in total, 11 in the intervention group and 12 in the control group. Intervention group: Range= NR* Mean= 84.5 years (SD 6.7). Control group: Range= NR Mean= 88.3 years (SD 6) Intervention group: 8 females (72.7%) and 3 males 27.3%). Control group: 8 females (66.6%) and 4 males (33.3%). Loss to follow up: 24 in total, 12 randomised to each group. One lost to follow up from the intervention group (died).	Life story books. 12 weeks in total. 2 groups, one which looked at participants who created their own life story books compared with a second group who had their books created by relatives. Both groups were presented with the completed books at the end of the programme. Programme delivered by a Clinical psychologist.	To assess the effect of a life review process compared with usual care. To compare the effects of a life story book process compared with books given as gifts. To assess the impact of the life story books on quality of care and the attitudes and knowledge of staff members.	Outcome measures: week 12 (following the programme) and week 18 (six weeks following completion of the programme). The following measures were used: Clinical Dementia Rating Scale (CDR); Quality of Life in Alzheimer's Disease (QoL-AD); Autobiographical Memory Interview- Extended Version (AMI-E); Geriatric Depression Scale (GDS) and Quality of Caregiving Relationship Questionnaire (QCPR). Relatives were also given the QCPR scale and the staff were administered the Approaches to Dementia Questionnaire (ADQ) and staff knowledge of care-recipient questionnaire.	**QoL-AD:** Life review baseline mean 30.1 (± 8.5), post-intervention mean 36.9 (±6.9). Gift giving group baseline mean 37.5 (±4.7), post-intervention mean 35.5 (±4.7). ANCOVA (F(1 20= 5.11, p=0.035). Six week follow up post-intervention mean life review group 36.1 (±7.8), gift group 38.6 (±3.8) ANCOVA F(1, 20)= 0.08, p=0.77.**GDS-12R**:Life review baseline mean 4.7 (±3.1), post-intervention mean 4.3 (±3.7), gift giving group baseline mean 2.6 (±1.4) and post intervention mean 2.5 (±1.8). ANCOVA F(1, 20)= 0.93, p=0.34. Six week follow up from post intervention mean life review group 3.5 (±2.7) and gift group 2,7 (±1.7), ANCOVA F(1, 20)= 0.14, p=0.71. **AMI-E:** *Personal Semantic Schedule Scores (PSS subscale*): Life review baseline mean 31 (SD 19.7), post-intervention mean 36.3 (SD 21.6), gift group baseline mean 36.7 (15.5), post-intervention mean 28.9 (18.3). ANCOVA F(1,20)= 14.01, p= 0.001* Six week follow up from post-intervention mean life review group 35.4 (SD 19.4) and gift group 33.3 (SD 16.6). ANCOVA F(1,20)= 3.98, p= 0.060. *Autobiographical Incident Schedule (AIS):* Life review baseline mean 3.4 (SD 2.8), post-intervention mean 8.2 (SD 8.2), gift group baseline mean 6.5 (4.4), post-intervention mean 5.8 (4.1). ANCOVA F(1,20)= 10.12, p= 0.005* Six week follow up from post-intervention mean life review group 6.6 (SD 5.4) and gift group 8.6 (SD 6.6). ANCOVA F(1,20)= 0.50, p= 0.49**AMI total**:Life review baseline mean 34.4 (SD 22.0), post-intervention mean 44.5 (SD 28.5), gift group baseline mean 43.2 (19.1), post-intervention mean 34.7 (21.3). ANCOVA F(1,20)= 19.92, p= <0.001* Six week follow up from post-intervention mean life review group 42.0 (SD 23.5) and gift group 42.0 (SD 22.4). ANCOVA F(1,20)= 2.92, p= 0.10.**QCPR (participant):** Warmth: Life review baseline mean 32.4 (SD 1.0), post-intervention mean 32.3 (SD 2.3), gift group baseline mean 32.2 (1.0), post-intervention mean 31.2 (1.7). ANCOVA F(1,20)= 2.56, p= 0.13 Six week follow up from post-intervention mean life review group 33.5(SD 2.3) and gift group 31.6 (SD 2.1). ANCOVA F(1,20)= 4.51, p= 0.046. Conflict: Life review baseline mean 23.5 (SD 0.8), post-intervention mean 21.5 (SD 2.1.), gift group baseline mean 22.8 (1.7), post-intervention mean 22.3 (1.2). ANCOVA F(1,20)= 0.43, p= 0.52 Six week follow up from post-intervention mean life review group 22.0 (SD 2.1) and gift group 22.5 (SD 3.6). ANCOVA F(1,20)= 1.40, p= 0.25.**QCPR (relative):** Warmth: Life review baseline mean 34.3 (SD 3.9), post-intervention mean 35.2 (SD 3.7), gift group baseline mean 34.8 (4.6), post-intervention mean 34.5 (4.6). ANCOVA F(1,20)= 0.21, p= 0.65 Six week follow up from post-intervention mean life review group 37.5(SD 3.0) and gift group 37.9 (SD 2.6). ANCOVA F(1,20)= 0.08, p= 0.78. Conflict: Life review baseline mean 21.1 (SD 4.7), post-intervention mean 22.3 (SD 4.1.), gift group baseline mean 23.3 (3.3), post-intervention mean 24.3 (5.2). ANCOVA F(1,20)= 0.120, p= 0.73 Six week follow up from post-intervention mean life review group 26.8 (SD 4.1) and gift group 27.9 (SD 2.2). ANCOVA F(1,20)= 0.03, p= 0.87).**QCPR total:** mean score baseline (time 1) 56.83 (SD 5.49), following life review (time 2) 58.17 (SD 7.44) and six week follow up (time 3) 65.13 (SD 5.40) ANOVA F(2, 39)= 19.37, p < 0.001. T1 v. T2 p > 0.99 T1 v T3 p < 0.001 T2 v T3 p < 0.001. QCPR warmth: time 1 34.57 (SD4.18) time 2 34.87 (SD 4.07) time 3 37.74 (SD 2.73) ANOVA F(2, 39) =7.16, p=0.003 T1 v T2 p > 0.99 T1 v T3 p=0.013 T1 v T2 p < 0.003. QCPR conflict: time 1 22.26 (SD 4.11) time 2 23.30 (SD 4.74) time 3 27.39 (SD 3.23) ANOVA F(2, 35)= 16.21, p < 0.001 T1 v T2 p=0.76 T1 v T3 p < 0.001 T1 V T2 p= 0.005. **ADQ total** (n= 46): Time 1 73.54 (SD 10.50), Time 2 73.74 (SD 9.23), Time 3 80.07 (SD 8.65). ANOVA F(2, 74) = 14.31, p <0.001. T1 v T2 p>0.99, T1 v T3 p < 0.001, T2 v T3 p < 0.001. ADQ hopefulness (n=46): Time 1 26.72 (SD 5.16), Time 2 27.00 (SD 4.88), Time 3 31.09 (SD 5.55). ANOVA F(2, 84) = 19.38, p <0.001. T1 v T2 p>0.99, T1 v T3 p < 0.001, T2 v T3 p < 0.001. ADQ person-centred (n=46): Time 1 46.83 (SD 6.82), Time 2 46.74 (SD 6.06), Time 3 48.99 (SD 4.65). ANOVA F(2, 74) = 3.92, p = 0.035. T1 v T2 p>0.99, T1 v T3 p = 0.11, T2 v T3 p = 0.001. Knowledge: correct (n=68): Time 1 5.93 (SD 3.77), Time 2 6.28 (SD 4.14), Time 3 8.79 (SD 5.31). ANOVA F(2, 120) = 14.31, p <0.001. T1 v T2 p>0.99, T1 v T3 p < 0.001, T2 v T3 p < 0.001. **Knowledge: don't know*** (n=68): Time 1 5.41 (SD 3.86), Time 2 4.12 (SD 3.33), Time 3 1.78 (SD 2.25). ANOVA F(2, 115) = 31.65, p <0.001. T1 v T2	Creation of life story books (LSB) either by life review involving the PwD or by relatives not involving the PwD has benefits for PwD, relatives and care staff. Life review is a process that requires supervision and training.	GRADE score: + + + + + + Levels of evidence score L1, grade A.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
						p=0.025, T1 v T3 p < 0.001, T2 v T3 p < 0.001. Knowledge: incorrect* (n=68): Time 1 2.01 (SD 2.37), Time 2 1.54 (SD 1.83), Time 3 0.32 (SD 0.68). ANOVA F(2, 119) = 24.88, p <0.001. T1 v T2 p=27, T1 v T3 p < 0.001, T2 v T3 p < 0.001.		
017. Billington (2013). Descriptive Dates not specified	UK- Wirral, Chester and Liverpool. 3 residential homes carried out the reading program themselves and two groups were carried out in a hospital and one at a day centre.	61 participants and 20 staff members. Ages and Gender NR. Loss to follow-up: 12 participants in total for the whole study (including those not in the care home group). 2 died, 2 left the care home, 1 was discharged from hospital, 1 had gone into a care home after the study began and no baseline data was collected for 6 of the participants.	Reading and storytelling. Get into Reading (GiR) is a literature based intervention. Short stories and/or poetry are read to participants and discussions are encouraged through open-ended questions. Sessions last for up to an hour. Delivered by 'project facilitators'.	To assess whether a shared-reading intervention could impact on behavioural symptoms in older people with dementia. To explore this with relation to what influence the programme had on older people, whether there were changes in dementia symptoms when participating in the programme and staff perceptions of the levels of engagement of the participants.	Neuropsychiatric Inventory Questionnaire (NPI-Q) and a semi-structured interview schedule carried out with staff members. Collected at baseline and then every four weeks for the duration of the study period.	Average NPI-Q scores for care home 1= baseline 4.07, October (1) 0.45, November (2) 0.43, December (3) 0.0, January (4) 0.17, February (5) 0.17 and March (6) 0.0. Average NPI-Q scores for care home 2= Baseline 8.3, October (1) 8.3, November (2) 0.0, December (3) 0.0, January (4) 0.0, February (5), 0.0 and March (6) 0.0. Average NIP-Q scores for care home 3- Baseline 5.00, October (waiting 0.0), November (waiting) 0.01, December (waiting) 0.00, January (1) 0.0, February (2) 0.0 and March (3) 0.88. Care home 3 adopted a wait-list design which allowed for an ANOVA to be calculated. There were no statistically significant differences between the three phases (F (2, 6)= 0.630, p=0.051 with a sample size of 8. When baseline statistics (x- = 0.39, SD=6.13, n=13) were compared with the average monthly reading statistic (x-= 0.39, SD= 1.42, n=33) a significant effect was observed (F (1, 32= 24.74, p<0.005). Three themes identified for qualitative interviews carried out with staff members- 1. Components of the reading group intervention (literary content, duration, environment); 2. Enjoyment, authenticity and meaningfulness of the reading-group experience, including a renewed sense of identity and 3. Enhancement of listening, memory and attention skills.	This study looked at three different care homes who took part in a reading programme. Descriptive changes in NPI-Q scores are presented for two of the homes and statistical analysis performed on the third. Changes were observed for the homes for the study period, however the one home which had statistical analysis presented did not show a significant difference across the study periods. The aspects of the programme were qualitatively explored through interviews with staff members who discussed the characteristics of the intervention, and its impact on residents.	Levels of evidence III, GRADE B.
031 George (2014) Intervention June-July 2011.	USA Landis Homes- a not for profit care retirement community which has 2 nursing care units, one of which participated in this study.	20, 10 intervention group, 10 control group. All female. Demographics only provided for the intervention group: Range= NR Mean= 85.5 years (SD 6.33). Loss to follow up: 10 participated but interviews were only conducted with 8. Reasons for this not reported.	Creative storytelling. TimeSlips sessions carried out for one hour twice weekly for six weeks.	Assess the effects of TimeSlips on both residents and staff members within a special care unit.	Thematic analysis. 2 questions asked to residents and 4 asked to staff members.	Main theme- Benefits to Residents. Subthemes- increased creativity, improved quality of life, positively altered behaviour, involvement in meaningful activity. Main theme- Benefits for Staff Members. Subthemes- learning new practices, developing deeper understanding of residents, involvement in meaningful activity, thinking creatively around programmatic challenges. Main theme- Benefits for the Nursing Home Community. Subthemes- nurturing relationships, improved atmosphere.	This study was the first to qualitatively explore the benefits of the TimeSlips programme. Themes were identified that highlight the benefits of taking part for residents, staff members and the whole of the nursing home. N.B- this paper is tagged with 032 which provides the quantitative counterpart of the study.	Risk of bias score Grade B, some flaws unlikely to affect the credibility, transferability, dependability and/or confirmability of the study.
032 Houser (2014) Intervention June-July 2011	USA Landis Homes- a not for profit care retirement community which has 2 nursing care units, one of which participated in this study	20, 10 in the intervention group, 10 control group Intervention group Range= 73-93 years Mean 85.5 years (no SD reported) Control group Range 68-98 years Mean 84.4 years (no SD reported). Intervention group= all female. Control group= 5 females (50%) and 5 males (50%) Loss to follow-up: missing variables are reported but no further information.	Creative Storytelling TimeSlips, a creative storytelling programme was delivered to participants twice a week for a period of six weeks. Sessions lasted for an hour.	To assess whether participation in a TimeSlips program was associated with improved mood and reduced behavioural symptoms in older people residing within a care home environment. Second to this, to assess whether there were any changes in psychotropic medication use as a result of the program.	Outcomes were measured using CareTracker, LH's direct-care data collection tool, which gathers input from caregivers three times daily for each resident. Psychotropic drug prescriptions and dosages were retrieved from LH's electronic medication record. Does not report frequency of this data being obtained, simply reports that data were analysed across the eight-month period.	When comparing intervention with control groups with respect to the two primary outcomes of mood and behaviour, there were no statistically significant results. The Hodges-Lehman estimate (a negative sign favours the intervention,and a positive sign favours the control) for the mood score was -1.0, with a 95% confidence interval of (-10.0, 12.0). The Hodges-Lehman estimate for the behaviour score was 1.0, with a 95% confidence interval of (-4.0, 8.0). Out of the secondary outcomes, only the appearance score displayed a non-zero Hodges-Lehman estimate, which was -2.0 with a 95% confidence interval of (-7.0, 2.0). With regard to the psychotropic drug data, although there was some flux in dosages and number of prescriptions, no statistically significant differences were noted within or between groups.	This study was the first to use quantitative methods to clarify how the practice of TimeSlips may be beneficial for participants and the nursing community at large. Creative expression programs engage people with dementia and professional caregivers in meaningful relationship-based activity that focuses on remaining strengths and capabilities. The authors highlight results also raise the question of whether the practice of TimeSlips may be similarly efficacious within other patient care settings for older adults (eg, rehabilitation clinics, home-based care, etc).	GRADE score: + ? ? ? - + Level of evidence: II-2, grade B

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
034. Garland (2007) Intervention Dates not specified	Australia Participants were residents at a total of 9 nursing homes within Australia.	30 participants in total. Mean age of the participants was 79 years, range 66-93. 19 (63%) were female and 11 male (37%). Loss to follow-up: None reported.	Music- preferred and stimulated family presence. 15 minute audiotapes of simulated family presence and preferred music (obtained from family members) were compared with a neutral audiotape. Interventions were applied once a day for three days during weeks 2, 3 and 4 and participants were randomly assigned to a group (A B C) and rotated through treatments. This was compared with usual care.	The aim of the study was to compare two interventions; preferred music and stimulated family presence to determine which, if any, was more effective at reducing behavioural disturbances in individuals with dementia who resided within a care home. A comparison group receiving usual care was also included.	Trained researchers observed the participants and recorded whether or not behaviours were present or absent during two minute intervals prior, during and following exposure to the 15 minute tapes, each observation period lasted for a total of 45 minutes. Behaviours were categorised into physically aggressive agitation, physically nonaggressive agitation, verbally aggressive agitation and verbally nonaggressive agitation but varied amongst the participants.	2880 target behaviours were observed in total for this phase. 3.8% were physically aggressive, 64.8% were physically non-aggressive and the remainder were non-aggressive and aggressive verbal behaviours. Physically aggressive behaviours were not included in the analysis. Actual counts of physically agitated behaviours decreased by 30% for the simulated family presence, 25% for music and 15% for placebo. Actual counts of verbally agitated behaviours decreased by 33% for the simulated family presence, 18% for music and 29% for placebo. Time*Treatment: (F12, 18 =2.62, p=0.032).Separate analysis of each of the types of agitation uncovered significant time-treatment interactions for both physical agitation (F6, 174=2.42, p=0.029) and verbal agitation (F3.71,107.42 =3.37, p=0.014). ***Difference in physical mean behaviour counts before and during treatments***: simulated presence vs. music difference in means, 0.31, F 0.77, p 0.388. Simulated presence vs. placebo difference in means 0.84, F 8.29, p 0.007. Simulated presence vs. usual care difference in means 1.12, F 10.23, p 0.003. Music vs. placebo difference in means 0.53, F 3.06, p 0.091. Music vs. usual care difference in means 0.81, F 4.67, p 0.039. Placebo vs. usual care difference in means 0.28, F 2.42, p 0.369. ***Difference in mean physical behaviour counts during and after treatments*:** simulated presence vs. music difference in means, -0.26, F 0.51, p 0.483. Simulated presence vs. placebo difference in means -0.34, F 1.13, p 0.296. Simulated presence vs. usual care difference in means -0.44, F 1.46, p 0.237. Music vs. placebo difference in means -0.09, F 0.07, p 0.790. Music vs. usual care difference in means -0.19, F 0.27, p 0.605. Placebo vs. usual care difference in means -0.10, F 0.15, p 0.705.***Difference in mean verbal behaviour before and during treatments****:* simulated presence vs. music difference in means, 0.43, F 1.44, p 0.239. Simulated presence vs. placebo difference in means 0.23, F 0.46, p 0.504. Simulated presence vs. usual care difference in means 0.96, F 4.78, p 0.037. Music vs. placebo difference in means -0.19, F 0.57, p 0.458. Music vs. usual care difference in means 0.53, F 2.86, p 0.101. Placebo vs. usual care difference in means 0.73, F 5.18, p 0.030. ***Difference in mean verbal behaviour during and after treatments***: simulated presence vs. music difference in means, -0.39, F 3.59, p 0.068. Simulated presence vs. placebo difference in means 0.26, F 1.18, p 0.287. Simulated presence vs. usual care difference in means -0.11, F 0.21, p 0.654. Music vs. placebo difference in means 0.66, F 8.29, p 0.007. Music vs. usual care difference in means 0.29, F 2.46, p 0.127. Placebo vs. usual care difference in means -0.37, F 2.52, p 0.124.***Magnitude of change****:* of the 30 participants, 11 experienced a fall of 50% or more in physical and/or verbal agitation while simulated presence tapes were played and 15 experienced such a fall while listening to preferred music.	The fact that participants responded to the placebo tape showed that even simple technology can improve the lives of those who exhibit behavioural manifestations of dementia. Both the music and simulated family presence tapes were modestly, but variably effective at reducing behavioural agitation, however the music tapes were easier to make as family members had difficulty in recalling memories sufficient for the simulated family presence tape. However, some of the participants showed increased agitation, particularly with the simulated family presence intervention.	GRADE score: ? ? + + - + Level of evidence II:2, grade B.
035. Fritsch (2009) Intervention Dates not specified	USA 20 not-for profit nursing homes, 10 in Milwaukee County, Wisconsin and 10 in Forsyth and Guilford counties, North Carolina. All had a special care dementia unit (SCU).	192 staff members completed surveys. Observations were carried out on 'those taking part in the TimeSlips program'. No further demographic information provided. Loss to follow-up: None reported.	Creative storytelling. TimeSlips programme carried out in groups of 10-12 once a week for a total of ten weeks.	To explore the effect of TimeSlips, a creative storytelling intervention, on the 'engagement' of residents who lived within a nursing home. The study also explored whether staff interactions and attitudes improved as a result of the intervention.	To explore the effect of TimeSlips, a creative storytelling intervention, on the 'engagement' of residents who lived within a nursing home. The study also explored whether staff interactions and attitudes improved as a result of the intervention.	**Engagement** *Disengaged*= TimeSlips facilities- n= 68, ratio .04, control facilities- n= 107, ratio .09, X2 value 24.755, p<.001.*Non-social Engagement*= TimeSlips facilities- n=174, ratio .11, control facilities n=135,ratio .11, X2 value 0.051, p=.822 *Engagement*= TimeSlips facilities- n=1,400, ratio .85, control facilities- n=1,007,ratio .81, X2 value 9.039, p=.003. *Challenging Behaviour*= TimeSlips facilities- n=9, ratio .01, control facilities- n=1,ratio .00, X2 value 4.475, p=.034 **Affect** *Anger*= TimeSlips facilities- n=6, ratio .00, control facilities- n=1,ratio .00, X2 value 2.368, p=.124 *Fear/anxiety*= TimeSlips facilities- n=39, ratio .02, control facilities- n=11,ratio .01, X2 value 9.195, p=.002 *General alertness*= TimeSlips facilities- n=1,512, ratio .92, control facilities- n=1,111,ratio .89, X2 value 5.535, p=.019. *Other (neutral)*= TimeSlips facilities- n=30, ratio .02, control facilities- n=75,ratio .06, X2 value 35.791, p=.001. *Pleasure*= TimeSlips facilities- n=54, ratio .03, control facilities- n=47,ratio .04, X2 value 0.518, p=.472 *Sadness*= TimeSlips facilities- n=7, ratio .00, control facilities- n=0,ratio .00, X2 value 5.304, p=.0.21.	There were higher levels of engagement in the TimeSlips facilities, and higher levels of disengagement in the control facilities. However, there was also more challenging behaviours in the TimeSlips group. There was also more general alertness, fear or anxiety and sadness within the TimeSlips group, those within the control groups exhibited more neutral affect. TimeSlips staff engaged in a greater total of interactions than those in the control group.	GRADE score: ? ? - + ? - Levels of evidence score LI, grade A.
037. Heyn (2003) Intervention Dates not specified	USA Further information not reported.	13 in total, 12 females (92%) and one male (8%). Mean age 85.7 ± 6.5, range 70-93. Loss to follow-up: None reported.	Multiseneory exercise program. The programme was carried out three times a week for a total of 8 weeks and lasted between 15 and 70 minutes.	To explore the effects of a multisensory exercise program on cognitive functioning (as measured by engagement), behaviour (as	The Menorah Park Engagement Scale to assess levels of engagement with the activity. The Caregiver Mood Report, to assess how caregivers felt during the intervention.	The MPES showed that 69.2% (n=9) of the participants had engaged in 'more than half' of the activity and 30.8% (n=4) had engaged in 'up to half' of the activity. At the end of the multisensory exercise program, eight residents (61.5%) were classified as showing positive improvements in their overall mood shown by a positive response on "participant looks: 1) happier; 2) calmer and 3) friendlier. 5 residents (38.5%) were reported as showing no significant, or little improvements in their overall mood. RHR showed significant improvements from a mean of 79.69 to 70.92bmp (a mean decrease of 8.77bmp, t= 5.93, DF=12, p=0.002, p<0.01). No significant difference was found in BP and weight.	Findings suggest multisensory exercise approaches may decrease RHR, increase exercise engagement and preserve function in individuals with Alzheimer's Disease.	GRADE score - - - + + - Levels of evidence score LII-b, grade B.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
			Comprised of four parts: warm-up, flexibility, strength-training and closure focused on breathing.	measured by mood) and physiological indices (such as blood pressure, resting heart rate and weight) in residents from a nursing home with a diagnosis of moderate to severe Alzheimer's Disease.	A video recording was used twice during the intervention to record facial expressions, eye contact and singing. Resting heart rate (RHR), blood pressure and weight.	BCRS scores remained stable, with a cognitive score range from 4.8 to 6.4 (mean= 5.72, SD=.44). The length of the sessions was increased incrementally from 15 to 70 minutes. Residents achieved peak performance at 4 weeks and none had significant changes in their schedules reported.		
040 Hagen (2003) Intervention Dates not specified.	UK Three 30 bedded wards within long term care facilities.	20 from each ward. Each ward was assigned a condition (Exercise Therapy (ET), Occupational Therapy (OT) or Control (C)) Demographic information not reported. Loss to follow-up: none reported.	Musical exercise ET group: sessions lasting 40 minutes, three times a week for a total of 10 weeks. The programme involved movement to music from the 20s-40s and involved mild stretching and aerobics. The OT group: lasted for an hour, three times a week for 10 weeks. Involved crafts, social activities and ADLs.	To assess the immediate impact of a ten-week musical exercise therapy intervention on physical, emotional, behavioural and cognitive functioning in older people and to evaluate these changes at a 10-week follow up timeframe.	Cognitive Assessment Scale (CAS) Behaviour Rating Scale (BRS) Overall Dependency Scores (ODS) Life Satisfaction Index (LSI) Physical Assessment Administered at baseline, posttest 1 (following ten weeks of treatment) and at posttest 2 (10 week follow-up).	**CAS scores:** showed a significant interaction between group and time of assessment (F 2, 144= 19.6, p<0.001). Scores for the C group decreased at both post-tests whereas the OT and ET groups showed improved scores at first follow-up with slight declines at 10 week follow up. **BRS:** significant interaction between group and time of assessment (F (4, 114)= 4.11, p<0.019. The C group showed increased behavioural disturbances at both follow-ups, whereas the ET and OT groups showed improvements immediately following which were further increased at 10 week follow-up. **ODS:** significant interaction between group and time of assessment (F (4, 114)= 19.38, p<0.001. **LSI scores**: significant group and time interaction (F (4, 114)= 4.11, p<0.019. Score remained the same for the C and OT group, but rose immediately following the intervention for the ET group before falling at 10 week follow up.	This study explored the potential benefits of a music exercise group with both a control group who took part in no additional activities, and a second control group which took part in occupational therapy. Psychological measures of cognition, behaviour and dependency were improved for both the music therapy and occupational therapy group immediately following the programme and then again 10 weeks after the intervention. The music group also observed initial improvements at the first follow up that were not sustained to the 10 week follow-up. Taking part in both music exercise and occupational therapy were beneficial to the psychological and physical functioning of older people.	Levels of evidence score LII-a, grade B.
041. Nair (2011). Intervention. Dates not specified.	Australia A dementia specific aged care facility located in Newcastle, Australia. Two units were selected for the study, Unit A which provided high-level care and Unit B which offered low-level care.	75 in total: 38 from Unit A and 37 from Unit B. Unit A- high level care Mean age 85.8 (± 5.37). 29 (75%) were female and 9 (25%) were male. Unit B- low level care Mean age 81.7 (±8.52). 28 were female (75%) and 9 (25%) were male. Loss to follow-up: none reported.	Music- listening to Baroque music. A selection of Baroque music was played between 3 and 7pm within the common room areas, loud enough so that it could be heard within the room but not in the participants'' bedrooms. Residents could come and go from the common room area as they pleased.	This study tested the effects of listening to Baroque music on the behavioural disturbances exhibited by older people who lived within an aged care facility for people with dementia.	Behaviour chart documenting physical aggressiveness, verbal abuse, agitation, wandering and inappropriate sexual advances. This was completed at the end of the afternoon shift (3-11pm) and following the night shift (11pm-8am) by nurses who were working that shift.	Music played High level ward (afternoon) 2.39 (SD 2.104). Music played High level ward (night) 0.42 (SD 0.783). Music played Low level ward (afternoon) 1.16 (SD 1.451). Music played Low level ward night 0.98 (SD 1.351). No music High level ward (afternoon) 1.76 (SD 1.927). No music High level ward (night) 0.5 (SD 0.906). No music Low level ward (afternoon) 1.14 (SD 1.398). No music Low level ward (night) 0.54 (SD 1.13). The lower care unit had 0.3 fewer episodes per week than higher (p=0.001), women had 0.2 more episodes per week than men (p=0.04) and afternoon shifts had one more episode per week than night shifts (p<0.001). Weeks where music was played showed significantly more episodes than observation weeks (0.23 more episodes per weeks, p=0.004). When multivariate analysis was carried out Sex and unit type did not remain significant (p=0.39 and p=0.26 respectively). Shift did remain significant, with one extra episode of behavioural disturbance per week in the afternoon shift (p<0.0001). Music also remained significant with 0.23 extra episodes per week compared with no music (p=0.01).	The study found that Baroque music increased behavioural disturbances in those with dementia. This was shown by an increase in the afternoon shift, however it did not continue on to the evening shift. Anecdotally, nursing staff had said that some residents were agitated by the music and requested for it to be turned off. It may be that music needs to be individualised and targeted to a person's history to be beneficial.	GRADE score ? ? - + + + Levels of evidence score LI, grade A.
043. Casby (1994) Intervention Dates not specified	USA A long term care setting in Washington State	Three participants in total. Participant 1: 87 year's old, female. Participant 2: 77 year's old, female. Participant 3: 69 year's old, male. Loss to follow-up: none, only three recruited.	Music- listening to music. Three phases: Phase A- baseline, no music and no headsets. Phase B- Intervention listening to relaxing classical music (Pachelbel's *Canon in D)*	To determine whether listening to music had an effect on the disruptive vocalisations of three individuals with dementia who resided within a long term care facility.	Data were collected during the times of day when instances of disruptive vocalisations were normally the greatest. For each phase data were collected during two 10 minute sessions each day over four days for a total of 12 observations days for subjects 2 and 3 (who took	Subject 1: Phase A mean 15 (± 19), Phase B mean 0.5 (± 9), Phase C 2.9 (± 5) Phase A mean 0. Subject 2: Phase A mean 45 (± 51), Phase C 15 (± 15), Phase A 12 (± 16). Subject 3- Phase A mean 43 (± 48), Phase B mean 32 (± 52), Phase A 1 (± 4). Celeration lines were presented for each of the subjects, which showed the number of disruptive vocalisations that were observed over a 10 minute period. Celeration lines show the variability of the data, the more varied the data points are (i.e. they do not show a distinct pattern) then the less valid the final results are. In this case, the data did not show a stable pattern at baseline. For subject's 2 and 3 there was an accelerating trend at baseline, which suggests that verbally disruptive behaviours rose over the four days at baseline level. However, for subject 1 there was a decrease in verbally disruptive behaviours, as shown by a decline in the data points at baseline.	This study looked at the effects of both favourite music and classical (non-preferred) music, using 3 phases. Phase A- baseline, phase B- classical and phase C- favourite. Subject 1 took part in all, subject 2 took part in A and C (favourite music) and subject 3 took part in A and B (classical). Music reduced the amount of disruptive vocalisations in two of the residents, but not in the third. Subject 3, who did not show a significant difference in the treatment	GRADE score ? - - - ? + Levels of evidence score, LII, grade B.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
			Phase C- Intervention listening to favourite music chosen by the participants.		part in phases ACA and ABA respectively) and 16 for subject one (who took part in phases ABCA).	For subject 3 there was no significant difference ($p > 0.05$) in the intervention (phase B- classical music) and all points fell below the celeration line. For subject 2 all data points during the intervention phase (phase C- favourite music) fell below the celeration line and this was significant ($p < 0.05$). For subject 1, data points fell below the celeration line for both the classical music phase and the favourite music phase, and this was significant ($p<0.05$).	condition, was exposed only to the classical music condition and therefore may have shown different results had they been exposed to the favourite music.	
044. Skrajner (2007) Intervention Dates not specified.	USA 2 settings- 6 participants attended the sessions at an Adult Day Health Centre (ADHC- site #1) and 16 attended the sessions at the special care nursing unit (SCU- site #2).	6 participants were trained as 'leaders' of the programme. Mean age of the participants was 84.8. Range 75-93. 6 participated from the ADHC and all these were female. Mean age was 75 (± 8.4). 16 participated from the SCU. Mean age was 89 (± 6.0). 15 were female (94%) and 1 male (6%). Loss to follow-up: none reported.	Creative storytelling RAMP programme- a copy of a story is handed out to a group of participants in an easy to read format. Following the leader each participant takes in turn reading their section of the story on the card. Following this, different coloured cards are handed out with facts or discussion on them and a discussion takes place around the cards. In this study, participants were trained to lead the intervention. Residents were trained to deliver the sessions.	The aim of this study was to determine whether or not individuals in the middle stages of dementia could be trained to lead and deliver a reading activity for those with advanced dementia. The study also explored the satisfaction that those who were trained as leaders felt about their role and the levels of engagement and positive affect of those taking part in the study.	QAR leader assessment form- assess whether the leader effectively implemented the session. The Menorah Park Engagement Scale- an 11 item observational tool that assess constructive engagement (CE), passive engagement (PE), non-engagement (NE) and other engagement (OE) with the task. Each participant were observed at baseline (baseline 1) for six sessions, then in six sessions during regular activity (baseline 2) programming and then between 6 and 10 sessions during RAMP activities (treatment). A brief satisfaction survey was also administered to those trained as leaders.	Significantly improved scores were shown for comparisons at Baseline 1 and Treatment for constructive engagement (1.09 vs 1.51, $p<0.01$) and pleasure (0.27 vs 0.45, $p<0.01$) and significantly decreased for 'other engagement' (0.70 vs. 0.27, $p<0.001$) and 'non engagement' (0.36 vs. 0.09, $p<0.05$). Significantly improved scores were also observed between Baseline 2 and Treatment for 'constructive engagement (0.91 vs. 1.51, $p>0.001$) and 'pleasure' (0.23 vs. 0.45, $p<0.05$). Scores significantly decreased for measures of 'other engagement' (0.59 vs 0.27, $p<0.001$) and 'non engagement' (0.61 vs 0.09, $p<0.01$) from Baseline 2 to treatment. No significant changes in passive engagement were observed between Baseline 1 and treatment and Baseline 2 and treatment. For the QAR measure, for *passing out stories* all of the leaders showed partial adherence for 100% of the activities and all but one leader demonstrated full adherence in at least 80% of the sessions. When it came to *asking someone to read the next section* a total of two leaders showed full adherence for 80% of the sessions and all leaders demonstrated partial adherence in 100% of the activities. Finally, for the *initiating discussion* section, five of the six leaders demonstrated partial adherence in 80% of the sessions and three demonstrated full adherence in at least 80% of the sessions.	People with dementia are able to effectively train to lead the RAMP programme to other residents from a care home environment. People with dementia also showed greater engagement with the RAMP based activities in comparison to regular activities carried out within the home.	GRADE score ? ? + + Levels of evidence score LII, grade B.
046. Ridder (2013) Intervention Data were collected in three 15-week periods during Autumn 2010, Spring 2011 and Autumn 2011.	Denmark and Norway. 14 nursing homes in total: 10 in Norway and 4 in Denmark.	42 participants. 21 in each group. 29 (69%) were female and 13 were male (31%). Age data was only provided for 26% of participants, who had a mean age of 81, range 66-96 years. Loss to follow-up: 45 participants were assessed for eligibility, 3 were excluded prior to baseline: 1 was hospitalised and 2 were excluded as their music therapist left their care home. At week 7 data collection- 1 participant was lost (from the group allocated to music therapy first) collection due to a disagreement on	Music therapy. Participants had received referrals from their homes to take part in music therapy. All completed a music therapy protocol carried out with a music therapist employed within their home. Sessions were delivered by a music therapist.	To carry out individual music therapy with individuals who had moderate to severe dementia and lived within a care home and determine its effects on frequency and disruptiveness of agitation. The study also aimed to explore the effects of the intervention on psychotropic medication use and quality of life of residents taking part.	Agitation was assessed using the CMAI-fr (frequency) and CMAI-di (disruptiveness) at baseline, following 7 weeks (crossover) and following 14 weeks. Quality of Life was assessed using the Alzheimer's Disease Rated Quality of Life (ADRQL). These scales were adapted so that the respondent only needed to focus on the previous week. Anxiolytic medication prescriptions were also assessed at baseline and week 14.	During standard care the frequency of agitation slightly increased (0.46) whereas it decreased during music therapy (-2.96), the difference was therefore -3.41 which shows a small effect size (-0.21) which was not significant ($p=0.38$). Agitation disruptiveness increased during standard care (3.26) and decreased during music therapy (-3.51) and this difference reached significance ($p=0.027$). Sensitivity analysis using LOCF confirmed the results, the difference in perceived agitation disruptiveness remained significant ($p=0.03$). Per-protocol analysis (excluding those who received less than 8 music therapy sessions) showed similar effect sizes and significance ($p=0.02$). Quality of life showed a decrease during standard care (-5.88) and an increase during music therapy (10.42), however this did not reach statistical significance ($p=0.44$). This analysis also showed that those who received music therapy later may have had greater improvements (by 8 points from 20.81 to 12.59) then those who received music therapy first (who had a smaller reduction from 15.71 to 15.65). The patterns of change during standard care did not seem to depend on the sequence in which it was received. Little change during standard care was seen in agitation frequency and quality of life and there was a tendency of deterioration in agitation disruptiveness which suggests carry-over effects were small. Medications were recorded at baseline and week 14. For this period there was no change in 48% of the participants. Increases in psychotropic medication occurred significantly more often during standard care than during music therapy (McNemar's $X2=5.14$, $df=1$, $p=0.02$). It also remained significant in the per-protocol analysis excluding participants who received fewer than eight sessions ($p= 0.02$).	This study was a randomised controlled trial that explored the role of a music intervention in manifestations of agitation in older people who lived within a care home and had dementia. This was compared with their usual care. The study found that there were significant reductions in disruptive agitation scores in those taking part in the music therapy compared with controls (as measured by the CMAI). Furthermore, there were no increases in psychotropic medications during the time when the music therapy was being administered, compared with significant increases during the times not engaging with the music therapy.	GRADE score + + + + + - Levels of evidence score, LI, grade A.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
		their music therapy referral. At week 14 data collection 2 participants were lost due to death (from the group allocated to music therapy first). 1 was lost due to hospitalisations, 1 through illness, 1 moved to another home and 1 was removed due to staff decisions.						
047a. Cooke et al. (2010) Intervention October 2008-March 2009	Australia Two mixed gender aged care facilities, site A and site B. Both were in Queensland and provided both low (assisted) and high (nursing) care.	47 in total. 33 (70%) were female and 14 (30%) were male. 3 were aged 65-74, 13 aged 75-84, 28 aged 85-94 and 3 older than 95. Intervention/Control: Following allocation- 1 deceased. At crossover- 3 refused to attend, 2 became ill, 1 did not participate as they experienced hearing problems during the intervention and 1 died. Control/Intervention: Following allocation- 2 refused to participate. At crossover- 3 withdrew from the study, 2 died and 1 left due to ill health. Another participant died before final data collection could take place.	Music- live music. Music and reading groups were carried out three times a week (Mon, Weds and Fri) for a period of 8 weeks. The music session comprised 30 minutes of live music and 10 minutes listening to pre-recorded music. The music activity was delivered by two musicians and the reading group by a research assistant.	To investigate the effects of a live music therapy programme on depression and quality of life for individuals with dementia. The study also looked at the long-term effects of the programme over six months.	Quality of life as measured using the Dementia Quality of Life Scale (DQOL) and changes in depression measured using the Geriatric Depression Scale (GDS). Administered at baseline, midway point (3 months) and immediately following the intervention. Sub-analysis was carried out with those who attended 50% or more of the sessions.	The only significant finding was that there was a significant difference in mid-point QOL belonging scores between the music and reading groups ($F(1, 45)= 6.672$, $p<0.05$. Participants who had experienced the reading control group first reported higher feelings of belonging (3.61) than those who experienced the music first (3.17). When the reading group crossed over to the music groups their scores decreased (3.61 to 3.46) but those in the music group first who crossed over to the reading group showed an increase (3.17 to 3.57). Two sub-analyses were carried out which showed significant differences: 1. In participants who attended greater than half of the music sessions (in either order, n=24) there were significant improvements in QoL self-esteem scored over time ($F(2, 46) = 4.471$, $p<0.05$) in scores from mid-point (3.36) to post-intervention (3.75). 2. In participants who had scores of less than 5 on the GDS (n=12) there was a significant difference in depression scores over time ($F (2, 22)= 8.129$, $p<0.01$. Depression scores decreased (8.25, 6.50, 4.42 respectively). This was more noticeable for those experiencing the music (9.00, 6.20, 4.40) compared with the reading group (7.71, 6.71, 4.43). This was analysed (GLM) and these findings occurred independent of carry-over effects as a non-significant order by treatment interaction was determined ($p=0.649$).	There was no significance difference between the reading group and the music group and there was no significant difference in the music group from baseline to follow up. Sub-analysis of those attending 50% or more music sessions (either first or second) showed significant improvements in Quality of life over time. There was also a significant difference in depression scores over time for music and reading those with low GDS scores initially, however this was more noticeable for the music condition.	GRADE score + + + + + + Levels of evidence score LI, grade A.
047b Cooke et al. (2010) Intervention October 2008-March 2009	Australia Two mixed gender aged care facilities, site A and site B. Both were in Queensland and provided both low (assisted) and high	47 in total. 33 (70%) were female and 14 (30%) were male. 3 were aged 65-74, 13 aged 75-84, 28 aged 85-94 and 3 were over the age of 95. Loss to follow-up: intervention/Control: Following allocation: 8.	Music- live music. Music and reading groups were carried out three times a week (Mon, Weds and Fri) for a period of 8 weeks. The music session comprised 30 minutes of live music and 10 minutes listening to pre-recorded music.	This study is tagged with 047b and features the same study population. The overall aim was to investigate the effects of a live music therapy programme on agitation and anxiety.	Changes in anxiety, as measured using the Rating Anxiety in Dementia (RAID) tool and anxiety, as measured using the Cohen-Mansfield Agitation Inventory Short Form (CMAI-SF). Sub-analysis was carried out with those who attended 50% or more of the sessions.	Analysis did not find an overall effect of the music programme in reducing agitation and anxiety over a six-month period. There was no evidence of any interaction effects of the first intervention arm (music or reading control group) at mid-point therefore suggesting there were no carry-over effects of either the music or reading group. Just one significant finding was shown for those who attending 50% or more of the music therapy sessions (n=24). Analysis showed a significant increase in frequency of verbal aggression over time regardless of group ($F(1, 46)= 3.534$, $p<0.05$). Mean scores increased from baseline (1.26, SD=0.590) to post-intervention (1.64, SD=0.767). A series of multiple regressions showed agitation levels at baseline and post-intervention were predicted by a number of factors:1. Participant MMSE scores ($\beta= -0.409$, $p<0.01$) significantly predicted overall agitation ($F(4,39)= 2.952$, $p<0.05$, adjusted $R2= 0.154$). Specifically, a higher level of agitation was associated with greater cognitive impairment. A shorter length of time living in the facility also appeared to be a predictor, although not at a significant level, of greater overall agitation (β 0.271, $p= 0.064$). 2. Although the models for baseline physical aggression	There was no significant effect of a live music therapy programme on behavioural agitation and anxiety in older people living in a residential care home. A sub-analysis of those who participated in 50% or more of the sessions did show there to be a significant increase in verbal aggression over time, however prevalence was still low and infrequent. A series of multiple regressions showed lower MMSE scores, length of stay and gender to be predictors of agitation. Anxiety was not a predictor of agitation.	GRADE score + + + + + + Levels of evidence score LI, grade A.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
	(nursing) care.	Control/Intervention: Following allocation: 9.	The music activity was delivered by two musicians and the reading group by a research assistant.			(adjusted R2= 0.124, p= 0.057) and baseline physical non-aggression (adjusted R2= 0.114, p= 0.069) were just outside of significance, inspection of values suggested that lower MMSE scores predicted higher levels of these two subtypes of agitation (β= -0.383, p= 0.013; β= -0.354, p= 0.021, respectively). 3. Non-significant results were found for baseline verbal aggression (adjusted R2= -0.026, p=0.579) and verbal non-aggression (adjusted R2= 0.055, p=0.187).		
048. Sung (2010) Intervention Dates not specified	Taiwan A long-term care facility with participants being selected from two units.	52 participants in total, 29 in the experimental group and 23 in the control group. Experimental group: 18 males (62%) and 22 females (38%). Mean age was 78.1 years (± 7.2). Control group: 11 males (48%) and 12 females (52%). Mean age was 82.7 years (± 7.4). Loss to follow-up: 1 hospitalised, unclear at which stage.	Music- preferred music. Those in the intervention group listened to preferred music for 30 minutes a day in mid-afternoon twice a week (total of 12 sessions over 6 weeks). Those in the control group continued with usual care. Music sessions were carried out by trained nurses and nursing assistants.	To explore the effects of a preferred music listening intervention on anxiety symptoms in individuals with moderate to severe dementia who lived within a nursing facility.	Changes in anxiety symptoms as measured using the Rating Anxiety in Dementia Tool (RAID). Administered at baseline and following the intervention period.	The experimental pre-test mean score was 10.93 (SD 5.46) and this decreased post-test to 8.93 (SD 4.86) which was a significant reduction (t=5.64, p <0.001). The mean control group was 9.52 (SD 4.09) and this decreased to 9.35 post-test, but was was not significant (t= 0.68, p=0.51). An ANCOVA was carried out which showed those in the intervention group had significantly lower anxiety scores following 6 weeks of a preferred music intervention in comparison to the control group who received usual care only (F=12.15, p=0.001).	This study found that residents with moderate to severe dementia had significantly lower anxiety scores (as measured using RAID) following participation in a 6-week preferred music intervention than before they started. They also had lower scores than those in the control group, who participated in usual care only.	GRADE score ? - - ? + - Levels of evidence score, LII, grade B.
049. Chen (2009) Intervention May 2004-May 2005	Taiwan A 350 bed nursing home in Changhua City, central Taiwan for older adults who are wheelchair users.	17 participants, 11 females (65%) and 6 males (35%). Mean age was 80.5, range 70-90. Loss to follow-up: None specified 4 subjects participated in 2 FGs and 13 in 1 FG only	Music Therapy Group music therapy carried out once a week for an hour for a total of three months, then once every 2 weeks for a year. Structured music therapy programme: warm up, dancing, moving, group play, listening to relaxing music, listening to a musical performance by a guest performer and a concluding phase. Sessions delivered by the authors (unclear if they are MTs or not)	To explore the perceptions of wheelchair-bound older adults within care homes when taking part in a music therapy intervention.	Focus Group Interviews carried out 6 months following the intervention. Guided on three questions: Can you please tell me your thoughts and feelings when you attended the group music activity? Which aspect of the group activity influenced you the most? What part of this music activity did you like best and why? These underwent thematic analysis.	Major themes and subthemes Strength derived from the group dynamic- sense of energy, distraction from suffering and confirmation as a person. Enhanced quality of life- variety added to lifestyle, motivation to exercise, learning positive behaviour and greater life satisfaction.	The participants in this study had positive views of the music therapy intervention. They discussed the benefits of the group dynamic and enhanced quality of life from the intervention.	Risk of bias score: Grade C, some flaws that may affect the credibility, transferability, dependability and/or confirmability of the study.
053. Millard, (1989) Intervention Dates not specified	USA A special unit for individuals with dementia within a nursing home.	10 participants in total were observed (however more attended the sessions). Mean age of the participants was 81.4 years (± 7.3), range 71-98 years. Seven were female (70%) and 3 were male (30%).	Music Therapy Sessions were held twice a week at 3.00pm for a total of five weeks. The experimenter facilitated these sessions which involved structured singing.	To explore the potential for singing to improve "quality of care" for individuals with Alzheimer's Disease.	Observations were carried out using 'behaviour mapping'. A checklist was used that measures frequencies of certain behaviours and this was modified to be completed by two observers with each subject being observed for 1 minute intervals, four times per session.	Within-subjects ANOVA's were calculated (A x B x S) where A served as the baseline condition (discussion) and B served as the treatment condition (singing) and S represented the subjects. Sitting (F=3.13, df 4, 36, p<0.026) and walking with others (F= 3.13, df=4, 36, p<0.026) were significantly different for the singing condition. A significantly higher number of vocal and verbal participation was observed for the singing sessions (F= 4.435, df 4, 36, p<0.005). Attendance was found to be significantly higher for the singing sessions compared with the discussion sessions (F= 2.61, df 4, 76, p<0.05).	There was a significant difference for two of the behaviours measured via the checklist 'sitting' and 'walking with others'. Verbal participation was significantly higher for the music condition compared with the discussion sessions, despite facilitation for verbal participation being greater for the discussion condition. However, it is also noted that there was more opportunity for a number of people to participate at once in the singing groups. The only significant interaction was observed for the 'walks with others' item. Both of the conditions positively impacted upon the scores	GRADE score ? ? - - Levels of evidence score, LII, grade B.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
		Loss to follow-up: None reported.			Observers also collected data for 30 minutes post-intervention.		of the checklist and verbal participation compared with not participating in any sort of activity.	
054 Guzman-Garcia (2013) Descriptive	UK Two private care homes located in Newcastle-Upon-Tyne	13 took part in the intervention: 3 were male (23%) and 10 were female (77%). Mean age was 80.5 (± 6.81). 7 took part in interviews: 5 females (71%(and 2 males (28%). Mean age was 82.4 (± 4.86). 9 staff members completed interviews. 7 females (71%) and 2 males (28%). Ages not reported.	Latin dancing- Danzon psychomotor intervention. Sessions were carried out bi-weekly for a total of six weeks and lasted for a total of 35 minutes. They comprised four parts: warm-up, Danzon practice, Danzon freestyle and cool down. Sessions were delivered by the author and facilitated by care staff.	To assess the implementation of a Latin ballroom dance intervention for people with dementia who were living in a residential facility.	Each participant giving interviews was interviewed twice and asked to identify the positive and negative aspects of the programme. Results were analysed using grounded theory.	The central category from interviews with staff was identified as 'benefits' which related to residents, staff and family members. This was further condensed into four subcategories: Category A- benefits on the residents who danced, with the following subcategories: behaviour, affective states, mental stimulation, socialising/communication, mobility, reminiscence and physical health. Category B- benefits on the spectator-residents, this related to those who observed the sessions but were not able to take part themselves. Category C- benefits on family members. Category D- benefits on care staff. The central category from interviews with residents was 'enjoyment'. Subcategories identified were: mood, behaviour, socialising, reminiscence, mental stimulation and mobility.	This study looked at the implementation of a Danzon music intervention for older people living within a care home facility. Interviews with staff members emphasised the benefits of taking part in the intervention for staff, family members and residents. Interviews with the residents taking part reflected their enjoyment of the intervention.	Risk of bias score: Grade C, some flaws that may affect the credibility, transferability, dependability and/or confirmability of the study.
055 Liu (2014) Intervention November 2012- February 2013	China A nursing home in the Shijingshan disctric of Beijing.	No demographic information reported, just that 25 were assigned to the control condition and 25 to the treatment condition. Does state no significant difference in age of groups at baseline ($p<0.05$). Loss to follow-up: none reported.	Music- Chinese 5 element. This involved both active and passive music therapy content, divided into 4 parts. Sessions lasted for 1-2 hours a week (Nov '12- Feb '13). Not clear who delivered sessions.	To explore the effect of music therapy consisting of five-element Chinese music, on older people with a diagnosis of seasonal affective disorder.	The Self-Rating Depression Scale (SDS) and Hamilton Depression Scale (HAMAD) were used to assess changes in depression levels. States qualitative elements but no methods for this reported.	T-tests were used to evaluate the difference between group means. Baseline mean SDS scores for control group 56.2 (± 19.3) for the treatment group 56.2 (± 19.1), ($p<0.05$). Baseline mean HAMAD scores for control group 14.7 (± 4.5) for the treatment group 15.1 (±5.9) ($p<0.05$). Following treatment mean SDS scores for control group 49.9 (± 18.8) for the treatment group 40.2 (± 18.1) ($p<0.05$). Mean HAMD scores 11.2 (±3.1) for the control group) and 8.8 (±4.9) for the treatment group. N.B. Reports there were qualitative elements, however simply states in the results section '*strength derived from the five-element group music therapy and emotional adjustment. The five-element group music therapy can reduce patients' psychological distress and let them feel inner peace and enhance their life satisfaction.* No actual analysis reported.	Those who took part in a Chinese five-element music programme with a pre-existing diagnosis of Seasonal Affective Disorder (SAD) showed lower depression scores following completion of the programme compared with those in the control group.	GRADE score ? ? - ? - - Levels of evidence score LI, grade A.
056. Lin (2010) Intervention August 2008- January 2009	Taiwan Three nursing facilities for older adults with dementia.	Experimental group: n=49. 26 female (53%) and 23 were male (47%). Mean age for the group was 81.5 years. Control group: n=51. 27 were female (53%) and 24 were male (47%). Mean age 82.2. Loss to follow-up: 3 from the intervention group withdrew and 1 from the control group (not known at what stage).	Music- listening to music. Information around music preferences was collected from participants and a music therapy protocol developed based on previous research. Sessions delivered by the researchers who had attended a music therapy course at university.	To ascertain the effects, if any, of a group music listening intervention on agitated behaviour exhibited by adults with dementia living within a nursing facility.	Chinese Cohen Mansfield Agitation Inventory (C-CMAI). Collected at baseline, following the sixth session and following the 12th session. Then follow-up one month post-intervention.	Mann-Whitney U-tests showed little difference between the average C-CMAI for the two groups for total score and each of the subscales at baseline. Longitudinal effects were analysed by using generalized estimating equations (GEE). There was no significant difference in C-CMAI scores for the control group (38.25, 38.55, and 37.75). ***Changes in overall score:*** There was a statistically significant decrease in agitated scores for the intervention group at each of the time points: 6th session scores versus pretest- the average score of the experimental group was 0.47 points lower ($p<0.001$). 12th session versus pretest- the average score of the intervention group was 0.44 points lower ($p<0.001$). One month follow up- the average score was 0.47 points lower ($p<0.001$). ***Changes in physically non-aggressive behaviours*:** There was a statistically significant decrease in physically non-aggressive behaviours at each of the time points for those participating in the intervention: 6th session versus pretest- the average score was 0.31 points lower ($p=0.004$). 12th session versus pretest- the average score was 0.26 points lower ($p=0.015$). 1 month follow-up- the average score was 0.34 points lower ($p=0.006$). ***Changes in physically aggressive behaviours***: There was a statistically significant decrease in physically aggressive behaviours in those taking part in the intervention at the different time points: 6th session versus pretest- the average score was 0.23 points lower ($p=0.28$). 12th session versus pretest- the average score was 0.20 points lower ($p=0.025$). At 1 month follow-up versus pretest the average score was 0.21 points lower ($p=0.018$). ***Changes in verbally non-aggressive behaviours***: There was a statistically significant difference in verbally non-aggressive behaviours at each of the time points for those who took part in the intervention: 6th session versus pretest- the average score of was 0.22 points lower ($p=0.042$). 12th session versus pretest- the average score was 0.28 points lower ($p=0.010$). 1 month follow-up- the average score was 0.26 points lower ($p=0.037$) ***Changes in verbally aggressive behaviours****:* There was a statistically significant decrease in	Agitation scores were collected at baseline, 6 weeks into the programme, at the end of the 12 weeks and at 1 month follow up. It was found that at each of these time points there was a significant reduction in CMAI scores overall, physical aggression, physical non-aggression and verbally non-aggressive behaviours. Verbally aggressive behaviours showed a significant decrease at 6 weeks but decreases at 12 weeks and follow up were not significant. The control group did not differ in terms of baseline agitation scores but did not experience the same reductions.	GRADE score + + ? + + + Levels of evidence score LI, grade A.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
						verbally aggressive behaviours in the intervention group at 6th session versus pretest with an average of 0.11 points (p=0.021). However the decrease at 12th session versus pretest was 0.09 points and this did not reach significance (p=0.104) and the difference at follow up (0.02 points) did also not reach significance (p=0.764).		
057. Ziv (2007) Intervention Dates not specified	Israel The Sophie and Abram Stuchynski Alzheimer Research and Treatment Centre (2 different residential wards).	28 participants. 25 were female (89%) and 3 male (11%). Mean age of the participants 82.6 (± 4.89). Loss to follow-up: not reported.	Music- listening to music A disk with songs totalling 16.48 minutes' duration comprised songs 'familiar' to participants. Took place on three separate Wednesdays between 13.15 and 14.25.	To explore the role of a music listening intervention on the behaviour of individuals with 'medium to advanced' Alzheimer's Disease, with focus on agitation and socialisation levels and affect.	Observation schedule devised using Cohen-Mansfield and Billing (1986- CMAI) conceptual framework observing three types of behaviours: positive, negative and neutral. Two observation sessions took place in each ward.	The mean number of positive behaviours for the group when listening to music was 5.18 (SD 3.62) and without music was 1.36 (SD 2.08), t=6.75, p=0.001. The mean number of negative behaviours when listening to music was 2.43 (SD 3.2) and without music was 5.96 (SD 4.13), t=-5.05, p=0.001. The mean number of neutral behaviours when listening to music was 4.39 (SD 3.61) and without music was 4.68 (SD 3.61), t=-0.37, (ns). Secondary analysis removed all behaviours that would only occur with music (i.e. swaying with music). 10 participants showed positive social behaviour only in the presence of music, 6 showed more positive behaviour in the presence of music than in absence and 9 participants did not show any difference in the amount of positive social behaviour. No participants showed more positive social behaviour without music than with it. Most negative behaviours were classified as repetitive. Aggressive and negative social behaviours were less frequent in general. Repetitive behaviours stopped during the music condition for 9 participants and were reduced with background music for 8 participants. 4 participants showed no difference and 2 showed repetitive behaviours only in the music condition. Aggressive behaviours appeared only without music and stopped during music in 4 participants and in a further 4 participants' music reduced aggressive behaviour. None of the participants showed aggressive behaviours when there was music playing.	Significantly more positive behaviours occurred when the participants were listening to music than when they were not. Furthermore, there was a significant reduction in negative behaviours when listening to music as opposed to when there was no music playing. In addition music was described as increasing positive social behaviours amongst participants as well as reducing negative repetitions and aggression.	Levels of evidence score LII-b, grade B.
058. Holmes (2006) Intervention Dates not specified	UK Residential homes within the south of England	32 in total. 28 were female (88%) and 4 (12%) were male. Mean age 84.9 (± 4.7). Range not reported. Loss to follow-up: None lost to follow up.	Music- both live interactive and passive listening. Participants took part in three sessions, each lasting for 30 minutes. Sessions were delivered by musicians.	To explore and compare the effects of a live music interactive session, passively listening to music and silence on behavioural symptoms of individuals with 'moderate to severe' dementia.	Dementia Care Mapping. Looks at 26 activities with a six point Likert-Scale. Category E of the scale correspond with engagement with a creative activity and was used for this study. Quality of engagement was rated every three minutes for each session from videotapes which were muted and had the live music activity obscured.	Live music showed significantly higher percentage of subjects showing positive engagement in comparison to silence X2 p<0.0001 for all participants and for those with moderate dementia only X2p<0.01 and severe dementia only X2p<0.01. There were no significant results when comparing the percentage of subjects engaged with pre-recorded music versus silence (all X2 >0.1). For those with severe dementia a greater level of participants showing engagement was found for the live music condition in comparison with the pre-recorded music (X2<0.01). In those with moderate dementia a greater percentage of engagement was found for live music compared with pre-recorded music (71% compared with 36%), but this was nonsignificant, X2 p=0.12. Wilcoxon signed rank tests for category E scores showed similar findings: ***All residents:*** Live music median +1 (range 0-3) vs. silence median 0 (range 0-1) p<0.001. Live music median vs. pre-recorded music median 0 (0-3) p<0.001. *Moderate dementia group:* live music median +1 (range 0-3) vs. silence median 0 (range 0-3), p=0.04. Live music vs. pre-recorded music median 0 (0-3), p=0.09. *Severe dementia group*= live music median +1 (0-3) vs.silence median 0 (range 0-1), p=0.04) and live music vs. pre-recorded music median 0 (0-1), p=0.04.	Live music in this study showed the greatest level on engagement for those with dementia. When accounting for severity of dementia, live music showed greatest engagement levels for both a group with severe and a group with moderate dementia. Pre-recorded music offered no additional benefit in terms of engagement when compared with silence. There were no ill effects reported from either of the music conditions.	GRADE score ? + + ? – Levels of evidence score LI, grade A.
059. Sung et al. (2006) Intervention Dates not specified	Taiwan A 160-bedded residential care home.	36 in total. Experimental group: n=18. 11 males (61%) and 7 females (39%). Mean age was 78.4 years (± 7.85). Control group: n=18. 15 males (83%) and 3 females (17%). Mean age was 78.4 years (± 7.9). Loss to follow-up: 4 left, 3 were hospitalised and 1 relocated to another care home. It is unknown at which stage they withdrew.	Music- music intervention with movement. Sessions were delivered twice a week for a total of four weeks and lasted for 30 minutes. Music familiar to the participants was okayed on a CD for 30 minutes and they were instructed on how to move. Sessions were delivered by were delivered in the afternoon by a nursing researcher and two research	To examine the effects of a group music with movement intervention on the frequency of agitated behaviours displayed by individuals with moderate to severe dementia.	Cohen-Mansfield Agitation Inventory (CMAI). This was modified to make observations for 60 minutes, counting behaviours that occurred using 10 minute periods. The CMAI was used at baseline, week 2 and week 4.	Baseline CMAI scores showed no statistical differences in agitated behaviours between the two groups (t=0.54, p=0.59). The mean number of agitated behaviours in the music intervention group significantly decreased by 1.17 at week 2 (m= 5.11 baseline, m=3.94 at week 2) and by a further 0.5 at week 4 (m=3.44). The mean number of occurrences of agitated behaviours in the music intervention group at week 4 decreased significantly in comparison to the control group (t=3.85, p=0.001). At week 2 the mean occurrence of agitated behaviours also decreased significantly in comparison with the control group (t=3.65, p=0.001), however, the control group did experience a slight decrease in mean occurrence of 0.16. The changes in occurrence of agitated behaviours were significantly different between the groups (F (2,33)=15.03, p<0.001. The music intervention group saw a mean decrease in occurrences of agitation of 1.67 and the control group saw a mean decrease of 0.22.	This study concluded that taking part in a music intervention successfully improved agitation in older people with dementia. Agitation scores decreased over time for those taking part in the music intervention such that at the end of the 4 weeks, a decrease in agitation occurrences was observed in a group of older adults with moderate to severe dementia. In comparison to the control group there was a significant difference in the occurrences of agitation in the group who took part in the music intervention at both 2 weeks and 4 weeks.	GRADE score + + - ? + - Levels of evidence score LI, grade A.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
			assistants who were trained to carry out the intervention					
061 Goddaer (1994) Intervention Dates not specified	Belgium Further information not specified.	29 completed the study. 23 were female (79.3%) and six male (20.7%). Mean age of the participants was 81.3 years (± 6.9). Loss to follow-up: None reported	Music- during dinner. Relaxing music" was chosen for the intervention. This was defined as music with a slow tempo; slow, irregular and unpredictable rhythm; no sound impulses; linear melody no evocation; a certain degree of homogenous monotony from the beginning to the end; no variation in intensity; and situated in the register of bass frequencies.	To evaluate the impact of music listening during dining on agitated behaviours in older adults who had severe cognitive impairment.	The CMAI, translated in to Dutch. The scale was reduced to a 0-1 scale, from its original format, to record presence (1) and absence (0) of behaviours. Scores were obtained at baseline (week 1), during music being played at mealtimes (week 2), during a week of no music (week 3) and when music was reintroduced (week 4).	A significant change was observed over the four week period on cumulative incidence of agitated behaviours ($F(3, 78)= 8.52$, $p<0.001$. Total agitated behaviours decreased by an average of 54% from week 1 to week 2. In week 3, the music was removed and the incidence of total agitated behaviours increased by 38.4%, giving a net loss of improvement of 15.6%. Reintroducing music at week 4 decreased scores by 43% from week 3 levels. The total reduction in agitation over the four weeks was 63.4%. Significant changes were observed on the dimension of physically nonaggressive behaviours ($F(3, 78)= 5.16$, $p<.003$). When music was introduced at mealtimes there was a 40.7% reduction in the level of this agitation measure. This level rose again by 26.3% (net loss of 14.4%) in week 3. During the final week there was a drop of 41.7%, indicating a total decrease of 56.3% for physically nonaggressive actions by those taking part. There were significant changes in verbally agitated behaviours ($F(3, 78)= 4.01$, $p<.01$. The decrease of these behaviours from week 1 to week 2 was 74.5%. Verbally agitated behaviours then rose again by 33.3% when the music was stopped in week 3 (net loss of 41.2%). By week four levels were the same as for week 2. Therefore from week 1 to week 4 there was a 74.5% reduction in these behaviours. There were no significant changes over time for physically aggressive behaviour ($F (3, 78) = 2.60$, ns) and hiding/hoarding ($F (3, 78) = .57$, ns). No percentage reductions were calculated for these domains.	Overall there were significant changes observed on the CMAI for the dimensions of: total agitated behaviours, physically nonaggressive behaviours and verbally agitated behaviours. No significant changes were observed for the other dimensions of the scale (physically aggressive behaviours and hiding/hoarding behaviours).	GRADE score ? ? ? – Levels of evidence score LII-2, grade B.
063. Ho (2011) Intervention May-December 2010	Taiwan A hospital based nursing home in Northeast Taiwan which had a 60-bed capacity.	22 in total. 12 were female (54.5%) and 10 male (45.5%). Mean age of the participants was 77.3 years (± 7.64), range 62-91 years. Loss to follow-up: None reported (9 withdrew prior to taking part).	Music Six different piano pieces were played on a CD player during mealtimes for a total of 4 weeks. The intervention took place twice a day between 11.30am and 12.30pm and 5.00pm and 6.00pm.	To evaluate a group music listening intervention consisting of music composed by the researcher and determine its effectiveness on agitation frequency in older adults with dementia.	The CMAI which was assessed at baseline, throughout the 4 week duration period and for a 2-week follow-up period. Staff nurses were trained by the researchers to use the CMAI prior to the intervention and kept a 24-hour record of agitated behaviour for the study duration.	Friedman tests revealed treatment effects across the different time points for the CMAI scale and subscale. The CMAI significantly changed from T1 (total mean 60.64) to T5 (total mean 42.99), showing a cumulative dose effect. At T5, CMAI scores had declined by 29.1% from baseline. The subscales had also decreased by 25.09%-35.91% by T5. At T6 the CMAI scores had increased slightly (total mean 46.14), but were still significantly lower than baseline data (all $p<0.008$). 15 participants (68.2%) said they liked the music played at mealtimes, 6 (27.3%) said they had no opinion and one participant said they did not like the music.	Music played at mealtimes significantly decreased agitation scores measured on the CMAI (total, non-aggressive verbal, non-aggressive physical, aggressive physical and aggressive verbal scores) in a group of participants with dementia. Significantly reduced CMAI scores were also shown at one-week follow up (T5) and two-week follow-up, although at two week follow-up scores were higher, suggesting the effects may decrease with time.	GRADE score ? ? – Levels of evidence score LII-2, grade B.
064. Chang (2010) Intervention Dates not specified	Taiwan Further details not reported.	41 in total. 15 were male (36.6%) and 26 were female (63.4%). Mean age 81.7 (± 6.4). Age range: 69-94 years. Loss to follow-up: Not reported.	Music. Nature music was played during the time immediately following lunch, where agitated behaviours were identified as being most prevalent.	To implement and evaluate a lunch time music programme for individuals with dementia living within a care home. The effects of the programme on agitation levels in those taking part were also explored.	The Chinese CMAI (C-CMAI). This was administered by six students (either nurses or social workers) at baseline and during Week 2, Week 4, Week 6 and Week 8 (music weeks) and Week 3, Week 5 and Week 7 (non-music weeks).	Line plots and GEE analysis are provided. Following adjustment for time trend and covariates the weeks CMAI scores during music player were significantly lower (B= 1.06, p=0.04) and physical aggression (B=0.39, p=0.04) and verbal aggression (B=0.49, p=0.02). This also carried over to the following week for physical aggression (B=-0.55, p=0.01) and verbal aggression (B-0.49, p0.03).	Music played during the period of time following lunch significantly reduced the incidences of physical and verbally disruptive behaviours for weeks when music was playing. Furthermore, a carry-over effect was observed with the following weeks' CMAI scores also showing an effect of the music.	GRADE score ? ? ? – Levels of evidence score LII-b, grade B.
065 Simard (2010) Intervention Dates not specified	USA EPOCH Senior Living Healthcare Centres located in Massachusetts	86 in total, 69 (80%) were female and 17 (20%) were male. Mean age was 84.5 (± 7.1), with a range of 68-103 years. Loss to follow-up: Not reported.	Namaste Care Program The program takes place 7 days a week for around 5 hours a day and is staffed by nursing assistants called Namaste carers.	No aims explicitly stated, however the paper discusses a specially created Namaste Care Program and evaluates its implementation with reference to challenging behaviour, agitation, delirium and anti-	Minimum Data Set (MDS) forms were completed prior to the resident enrolling in the programme and then more than 30 days after their enrolment in the programme. MDS forms did not contain any personal information on the residents.	No difference in depression scores before and after enrolment (p values and/or significance not reported). Only 8 residents had a score larger than 2 (considered depressed) and 3 of these were not treated for depression. No significant differences between behavioural symptoms before and after enrolment in the whole study population. When only residents who had withdrawal or reduced social interaction were included in the analysis, the Interest subscale of the MDS Challenging Behaviour Profile was significantly decreased following enrolment (3.27, SD 0.30 vs. 2.00, SD 0.47, n= 11, p=0.046) indicating less impairment in social interaction. Scores from the Agitation subscale of the MDS Challenging Behaviour Profile for residents with a CPS score between 1 and 3 decreased following Namaste enrolment, (1.04, SD 1.25 vs. 0.81, SD 1.27, n= 26, p=0.18). Lower levels of agitation were shown for residents with CPS scores between 4 and 6 and these remained unchanged (actually they increased slightly) (0.40, SD 0.69 vs. 0.50,	The discussion section says that the results from this study indicate involvement in Namaste Care improved interest in the environment, however this does not appear to have been measured by any of the outcome measures. The Minimum Data Set used within the study did not report significant difference across any of the measures for the whole group when their baseline scores were compared with those following enrolment in the programme. Deletion of some questions from the delirium	GRADE score ? ? ? – Levels of evidence score LII-b, grade B.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
				psychotic medication use.		SD 0.89, p values not reported). Total indicators of delirium also decreased non-significantly following enrolment (2.52 SD 1.94 vs. 2.35 SD 1.88, p=.079. When the individual items of this subscale were compared it was found that ratings for 'periods of restlessness' and 'periods of lethargy' were unchanged. When these 2 items were deleted there was a significant decrease of ratings of the remaining delirium indicators following enrolment (2.00 SD 1.53 vs 1.81 SD 1.51, p=.02). Comparison of psychoactive medication administration days did not find any difference in the number of anti-psychotics, antidepressants or hypnotics., Significant decrease in days when anti-anxiety medications were administered (0.80 SD 2.18 vs. 0.49 SD 1.79, p=0.35). The differences were caused by 11 residents who were receiving anti-anxiety mediations prior to the Namaste enrolment and only 6 who were receiving such medications after enrolment. The number of residents receiving antidepressants reduced by two (54 vs. 52) and there was a decrease in the number of days when hypnotic medication was administered to one of the residents (4 vs. 1).	subscale resulted in a significant decrease in delirium, suggesting some benefits, along with a decrease in the number of days that antianxiety medications were required.	
068. Svansdottir (2006) Intervention Dates not specified	Iceland Two nursing homes and two psycho-geriatric wards.	47 participants in total. Age range 71-87 years. No further demographic information provided. Loss to follow-up: 11-8 moved to a nursing home, 2 had their dementia deteriorate and 1 participant died.	Music therapy sessions. Carried out in groups of 3-4 participants by a registered music therapist, 3 times a week for a total of 6 weeks (sessions lasted 30 mins). Involved the music therapist playing instruments and singing songs, with residents encouraged to join in.	To assess the effects of a group music therapy programme on the behavioural and psychological symptoms of those with 'moderate to severe' Alzheimer's Disease.	Behavioural Pathology in Alzheimer's Disease Rating Scale (BEHAVE-AD). Participants were rated at baseline, following the intervention (6 weeks) and follow-up (four weeks after completion of the programme).	Most reported BEHAVE-AD symptoms presented at baseline were activity disturbances and paranoid and delusional ideation. Wilcoxon signed ranks showed a significant decrease for activity disturbance scores in the therapy group (mean 1.6 pre vs.0.7 immediately following, p<0.02) after six weeks but not for the control group (mean 1.4 pre vs. 1.0 immediately following, p>0.5). This effect decreased over the next four weeks and was not significantly lower than baseline. Total BEHAVE-AD scores were lower at six week follow up in the therapy group, however this was not significant (mean 5.5 pre vs. 4.4 at six weeks, p=0.3) and a small decrease in the control group which was also not significant (mean 5.4 pre vs. 4.7 at six weeks, p=<0.5). No other subscales showed a significant decrease in symptoms for either group. Grouping together activity disturbances, aggressiveness and anxiety scores showed a significant decrease in symptoms for the therapy group (p<0.01) but not the control group (p=0.5). At 4 week follow-up all of the scales showed no lasting benefits.	This study explored the impact of music in improving behavioural symptoms associated with dementia. Activity disturbance scores were shown to significantly decrease for those taking part in the intervention, however this was not the case at 4 week follow-up, showing time limited effects. Additionally no other subscales showed a significant decrease. When combined, there was a significant improvement in activity disturbances, aggressiveness and anxiety scores was shown for the intervention group, but not those in the control group. Again, there were no lasting effects at four week follow-up.	GRADE score ? ? + - ? - Levels of evidence score, L-I grade A.
069. Ashida (2000) Intervention Dates not specified	USA Two care homes located in Florida.	20 in total, mean age of the group was 86.2 years with a range of 73-94 years. 17 females (85%) and 3 males (15%). Loss to follow-up: Not reported.	Music- drumming, singing and reminiscence with discussions. Participants were divided in four small groups based upon the homes they lived in (the average number of participants in a session was 4.8). Participants took part in the music therapy every day for a week. The average length of the sessions was 42.95 minutes, and they ranged from 38 to 45 minutes in duration.	To explore whether there was any effect on depressive symptoms when a group of older people with dementia took part in music therapy sessions which involved playing instruments (drums), singing and using reminiscence to shape the content of the sessions.	Cornell Scale for Depression completed by activity staff who interviewed the staff responsible for the residents. Data were collected at the end of the first week (pre-test), the end of the second week (no treatment) and the end of the third week (post-test).	***Cornell Scale for Depression in Dementia***: Data were only included for participants who had attended at least four of the music therapy sessions, the total number included was 20. Ten of the participants had attended four sessions and the others had attended five. One-way ANOVA, F test was used to test significance of differences. Significant difference between groups was assumed given the obtained F value (3.77) was larger than the F critical value (3.23) for df 2, 57, p<.05. A Newman-Keuls Multiple Comparison Procedure was then used to determine which weeks produced significant differences in scores. Values for this test were df=57, α=.05 and critical values 3.032 (two-step) and 3.646 (three-step). Obtained values were .3 between O1 (pretest) and O2 (no treatment phase); 3.4 between O1 and O3 (follow up); and 3.7 between O2 and O3. Significant differences were found between baseline and at the end of the intervention week (O1 and O3) and between the control week (with no music, O2) and treatment week (O3). No significant differences were observed between baseline and the control week (O1 and O2). Data collection sheets were used to ascertain the levels and characteristics of participants in each group and to see if any changes occurred in these as the weeks progressed. Each of the participants were observed for 30 second periods at intervals of 2.5 minutes. Observed behaviours included positive affect, on-task, active participation and passive participation	This study looked at the effects of a drumming intervention for residents of a care home who had dementia.. There were no significant differences in depression scores at baseline and following the control week. Significant improvements in depression scores were recorded between baseline and following the intervention and at the end of the control week and following the intervention. Observations were also carried out to measure affect, on-task behaviour and both active and non-active participation levels in the residents. The collected data hints at their being more evidence of high participation in one of the groups compared with the others and that there was an increase in active participation and a decrease in passive participation for all of the groups, aside from the high functioning group, however no statistical analysis is carried out on any of the observational data (nor any visual presentation of results, mean scores etc.). Videotapes of drumming at the start and end of the sessions were analysed, these failed to show a significant difference in 'perceived mood' (no validated scale is mentioned for measuring this, nor how the scale was explored) of the residents (i.e. they did not show an increase in 'happy' drumming).	GRADE score ? ? - - Levels of evidence score LII-b, grade B.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
070 Ragneskog (1996a) Dates not specified.	Sweden A nursing home located near Gothenburg, Sweden, with a special care Dementia ward.	10 participants from a potential 25 were selected (those with the highest Multi-Dimensional Dementia Assessment Scale Scores- therefore the highest level of behavioural disturbances). 5 were analysed (5 had their back to the camera). Ages were 69, 80, 84, 85 and 85 (mean age 80.6, ± 2.4). 4 female (80%) and one male (20%). Loss to follow-up: none reported.	Music- during dining. The music intervention lasted for around 30-45 minutes and was played in a group setting during lunch time (1pm). Three types of music were played (soothing, music from the 20s and 30s and rock and pop) and the sequence was decided randomly. During period 1 (baseline) no music was played.	To assess the effects of different genres of dinner music on symptoms of irritability, restlessness and uninhibited behaviours in nursing home residents with dementia.	Recorded observations were coded such that the researchers analysing them were blind as to treatment. For each segment the following behaviours were noted as present or absent: eating the food, sitting at the table without touching the food, pushing away the plate, taking or trying to take a fellow patients food, seeking attention, rising from the table, eating food, smearing food, picking, being spoon fed successfully/unsuccessfully. Time spent with dinner was recorded also.	Four of the five participants spent more time with their food during the music periods than the control period. Longest time spent with dinner was during soothing music, second longest during the time when music from the 20s and 30s were played, third longest during pop music being played. Total mean time spent with a meal increased by 22% during music being played compared with the control period. Dinner time decreases from playing of the soothing music towards the control period ($p<.05$). Nursing staff fed the participants significantly more often during the soothing music periods ($p<0.05$). Individual case studies are presented for each of the five participants.	This study assessed the impact of music on agitated behaviours which may affect consumption of food at mealtimes. It was found that participants spent more time with their food during music playing periods of time, than during the control period. Soothing music was observed to have the greatest impact on both participants and nursing staff members.	GRADE score + - - - Levels of evidence score LII-b, grade B.
071 Ragneskog (1996b) Intervention Spring 1991	Sweden A nursing home located in Bohslan	20 completed, from an initial 25 recruited. 10 (50%) were male and 10 (50%) female. Mean age of participants was 80 years. Range 69-94 years. Loss to follow-up: 25 patients located within the unit, one was excluded on the basis that they ate their dinner alone. A further four died during the study period and therefore did not complete the programme, it is unclear at which part of the study they died.	Music- during dining. No music played during period 1 (5 days), soothing music during period 2 (8 working days), no activity for a week, period 3 where tunes for the 20s and 30s were played for 10 work days, no activity for a week, period 4 (pop & rock music from the 80s) for 8 days followed by a control period of 9 days.	To evaluate the effects of music being played at dinner time on symptoms commonly seen in dementia, such as depressed mood, irritability and restlessness and determine whether a certain type of genre were more beneficial than others. Also to evaluate the effects of the music intervention on food intake.	Mood changes assessed via the GBS (scored by a blinded psychologist) Amount of food consumed, measured by weighing the plate of food before dinner and then following. Weight of participants was recorded at the end of each mealtime session. Questionnaires on music preferences were administered to staff members pre-study. Measured at baseline, during soothing music (period 2), during 20s and 30s music (period 3), during pop and rock music (period 4).	Friedman rank sum tests showed significant improvements for dimensions of irritability, fear-panic and depressed mood compared with the control period ($p<0.05$), with results most pronounced for when soothing music was played. All the other dimensions were reported as not significant. A two-way ANOVA showed food intake performance was significantly worse when turns from the 20s and 30s were played ($p<0.05$) and emotional blunting was significantly poorer when soothing music was 20s and 30s were played ($p<0.05$). No significant differences were observed for impaired motor performance, intellectual impairment and emotional impairment. Fischers test showed staff members served larger portions to the participants when music was playing, compared with the control period ($p<0.001$) for both main course and dessert. Participants consumed more food during the music period than the control period, however this was mainly found for dessert ($p<0.001$). There was a significant correlation between served and consumed food ($r=0.85$, $p<0.001$). Heavier participants were administered larger helpings ($r=0.63$, $p<0.01$) and consumed more food ($r=0.58$, $p<0.01$).	This study explored the role of music being played during mealtime on changes on food intake and mood. Participants within the study consumed more food during the time periods when music was played, this effect was most noticeable for the time when pop music was played, but was also found for other music genres. Staff members also served more food to the participants when music was being played and more food was consumed when portion sizes were bigger. Ratings on the GBS scale showed improvements in anxiousness, depression and irritability during the times when music was played. Depressed mood, however, only showed a significant improvement from baseline to when soothing music was played.	GRADE score + ? ? ? Levels of evidence score LII-b, grade B.
074 Palo-Bengtsson (1998) Descriptive February-May 1995.	Sweden A nursing home located in Stockholm, Sweden	6 participants (4 resided in the care home and 2 were attendees at the day unit). Age range= 76-88. Mean age= 83.3 (± not reported) 4 were female (67%) and 2 were male (33%). Loss to follow-up: 4 lost from 10 participants who were selected to take	Dancing- social dance sessions. Already carried out within the home, once a month and took place between 10.30am and 11.15am (lasting 45 minutes in total). Number of people attending the sessions ranged from 20 to 50. Residents attended with their caregivers.	To assess how participants with dementia would respond when taking part in social dance sessions.	Content analysis was performed on qualitative data obtained from video recordings of the sessions using three steps: 1. Define the content, 2. Identify the characteristics, 3. Perform content analysis.	Group category: motor functions. Variables: physical activity, spontaneous activity. Group category: intellectual functions. Variables: orientate themselves in space, recent memory, timing, personal orientation, distant memory, wakefulness, concentration, increased tempo, collected, long-winded, distractibility. Group category: emotional functions. Variables: showing emotions, control of emotional reactions, motivation. Group category: different symptoms common in dementia. Variables: confusion, irritability, anxiety, agony (mental discomfort), reduced mood, restlessness.	This study determined that social dancing was a feasible activity for older people in residential care.	Levels of evidence score LIII, grade B.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
		part (prior to the start of the intervention).	Dance music was performed by a local dance band which consisted of three male musicians and a female singer.					
075 Ledger (2007) Intervention Dates not specified	Australia 12 Nursing Homes in Queensland and 1 in Victoria. Homes were made up on community, church and privately owned facilities and offered a mixture of high and low level nursing care.	45 in total. Experimental group: n=26. 22 were female (84.6%) and 4 male (15.4%). 3 were aged 71-75, 4 aged 76-80, 6 aged 81-85, 9 aged 86-90, 3 aged 91-95 and one aged between 96-100. Control group: n=19. 18 were female (94.7%) and one was male (5.3%). Three were between the ages of 76 and 80, 7 aged 81-85, six aged 86-90, 2 between Loss to follow-up: 15.	Music therapy- group sessions. Sessions were carried out weekly, lasting between 30 and 45 minutes and for a minimum of 42 weeks of the year. Each site varied in the time sessions were carried out and group sizes varied from 2-10. Four of the music therapy groups were conducted by the first author (a qualified music therapist) and one group was conducted by a research assistant (also a qualified music therapist).	To assess whether participants who took part in a music therapy programme would have a reduction in the frequency and severity of agitated behaviours compared with a control group not taking part in the intervention.	The CMAI was used to assess the agitation levels of the participants. Levels were recorded at baseline, at three months, six months, nine months and a year. The CMAI was completed with staff members who had spent time with the participants to provide information as to agitation levels during the previous two weeks.	Baseline CMAI scores were significantly higher for the control group ($t=2.17$, $p<0.05$), with both groups showing high standard deviations (control group mean= 39.05, ± 22.15, experimental group mean= 25.95, ± 15.42). Repeated measures multivariate analysis showed significant effects within-participants over time ($F=2.61$, $p<0.05$) but not within-participants over time by group (experimental or control $F=1.61$, $p=0.432$), showing there was no significant difference between the groups. There was no significant difference between the groups for any of the subscales verbal non-aggressive $F=0.33$, $p=0.57$; verbal aggressive $F=0.59$; $p=0.45$; physical non-aggressive $F=0.62$, $p=0.44$; physical aggressive $F=0.78$, $p=0.38$. Univariate analysis: significant effect for time and group for the verbal aggressive scale ($F=2.70$, $p<0.05$) Anecdotally, the log kept by the music therapist reported that those taking part in the intervention showed less wandering, fidgeting, grabbing, insults, complaints and anxious statements after sessions.	This study followed two groups, those taking part in music therapy and a control group, for a year and collected their agitation scores prior to taking part in the intervention and at 3 month intervals. CMAI scores fluctuated throughout the different time points for both the group taking part in the intervention and those who were not. There was no significant differences for CMAI scores in those taking part in music therapy interventions compared with normal controls. Univariate analysis did show there to be a reduction for the verbally aggressive subscale, suggesting this study may have suffered from a lack of statistical power.	GRADE score - - + - Levels of evidence score LII-a, grade B.
076 Chang (2013) Intervention November 2006-October 2007.	Taiwan Nursing Home of St Martin de Porres Hospital in Chiayi.	33 completed the intervention. Ages not reported. 27 (81.8%) were female and 6 (18.2%) were male. Loss to follow-up: 15 participants were lost to follow up, 10 had died, 2 refused to complete at follow up and 3 had been discharged from the nursing home.	Environmental stimulation. Wall murals were created on each floor using acrylic paint. 4 themes were chosen to reflect the life experiences of the residents living there. Floor 2- decorated to look like Alishan, Floor 3- decorated to look like old streets and shops, Floor 5- decorated to look like an old veterans' village, Floor 6- decorated with rural images and the sunny rooms were decorated to look like gardens. Researchers and artists painted the wall murals.	The focus of the study was to enhance the environment of a nursing home by using visual arts aimed at promoting reminiscence. The study also aimed to determine reliability and validity for a scale to assess residents' satisfaction with their living environment (Satisfaction with Living Environment Nursing Home Scale-SLE-NHS) and use this to determine effects of the intervention with regards to satisfaction with the living home environment.	The Satisfaction with Living Environment Nursing Home Scale (SLE-NHS) was developed and validated for use in this study. This 18-item questionnaire measures responses on a Likert-type scale and has three factors: memory recall, pretty and pleasurable and convenience. Scores were obtained at baseline, prior to the installation of the mural and following installation.	T-tests were used to compare pre and post scores on the SLE-NHS subscales. The recalling old memories scale showed significant improvement (mean 1.41 pre vs. -3.47 post, $t=13.32$, $p<0.001$). All of the floors showed a significant improvement for this subscale (floor 2, $t=5.99$, $p<0.05$; floor 3, $t=6.97$, $p<0.01$, floor 5, $t=7.29$, $p<0.001$, floor 6, $t=10.88$, $p<0.001$). A slight improvement (3.47 pre and 3.60 post, ns) was observed for the pretty and pleasurable subscale. A significant increase was observed for those participants on the second floor at pre-test ($t=13.86$, $p<0.05$). Convenience scores were not significantly different ($p>0.05$).	This study used visual imagery on each floor of a nursing home to depict scenes that would have been familiar to the residents. A questionnaire was formulated called the Satisfaction With Living Environment at Nursing Home Scale (SLE-NHS). Overall the imagery was found to increase scores on the SLE-NHS, however only memory recall in the residents, however only memory recall showed a significant increase, scores for 'convenience' and 'pretty and pleasurable' subscales did not reach significance. When residents were analysed based upon the floor they lived on, significant differences were found for all floors in relation to the recalling old memory subscale and participants on the second floor showed significantly increased scoring on the pretty and pleasurable subscale.	
082 Remington (2002) Intervention	USA Four different	68 in total. 59 female (87%) and 9 (13%) male. Mean age 82.4	Music- calming music, hand massage.	To assess the effectiveness of two interventions;	The CMAI recorded frequency of agitated behaviours on four occasions: immediately	No difference in agitation level at baseline (F test for group differences $F=1.1$, $p=.36$), showing that the participants had been adequately randomly allocated.	Both hand massage and calming music were seen to reduce agitation, as measured by the CMAI. This effect was observed for both	GRADE score + + + ? - ?

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
Dates not specified	long term care facilities.	years. Range 62-99 years. Loss to follow-up: not reported.	CM condition: music was played during this condition for ten minutes. HM condition: participants received ten minutes of hand massage based on a protocol from previous research. CM and HM: those in this condition received both	exposure to calming music and hand massage in reducing agitated behaviours of individuals with Alzheimer's Disease. The effectiveness of combining both interventions was also assessed.	before the intervention, during the intervention, immediately following the intervention and one hour following the intervention. The CMAI was administered by a research assistant who was trained in its use.	Significant difference over time in agitation levels for the four groups (Fcons=6.47, p<.01). Tukey's HSD test found the control group was had significantly higher agitation scores than any of the experimental groups during the intervention period, no difference was found between the three intervention groups at any of the intervention time-points (2, 3 or 4). No significant difference over time for physically aggressive agitated behaviours (Fcons=1.93, p=.09) for the groups. Physically nonaggressive behaviours showed a significant difference (Fcons=3.78, p<.01). At time point 2, 3 and 4 the control group exhibited significantly greater scores for this subscale (Tukey's HSD post-hoc tests with a significance level of .05). Verbally agitated behaviours were also not effected by time (Fcons=1.92, p=.10). Tukey's post-hoc analysis (with a significance level of 0.5) showed there was a significant difference between the control group and the hand massage group at time points 2 and 3. A significant difference in scores was also found for all of the treatment groups and the control group.	interventions individually and when the two were carried out together. These results were sustained for up to an hour following the intervention. A wide range of baseline agitation existed within the participants, with scores from 4 to 54. None of the interventions showed a significant effect for physically aggressive behaviours for any of the time points. The authors state that this may be as the initial levels were too low to show a decrease, but these are not presented in the results section and therefore it is difficult to comment on this accurately. Physically non-aggressive behaviours were significantly lower in all three intervention groups in comparison to the control group. Verbally agitated behaviours were reduced at all of the time points for those receiving hand massage and for all of the interventions compared with the control group at time point 4.	Levels of evidence score LI, grade A.
084. Vink (2013) Intervention Dates not specified.	Holland 6 nursing homes, 4 described as 'typical Dutch nursing homes' and two others based on alternative medicine (anthroposophic care.	Intervention group: n=43. Mean age 82.4 (± 7.6). 29 (67%) female and 14 (33%) male. Control group: n=34. Mean age 81.8 (± 5.7). 25 (74%) were female and 9 were male (26%). Loss to follow-up: Five residents (four from the control condition and one from the music) died during the study; seven residents (all from the control group) were excluded as they had missing baseline data; five had no agitated behaviours evidenced at baseline. This left 77 residents, 43 for the intervention condition and 34 for the control group. Three (one from the music condition and two from the control) were missing data at the 4th hour time point, but were included in sensitivity analysis.	Music- group music therapy. All of the homes had a music therapist employed. Sessions took part over a four month period in a small group setting (no more than 5), twice a week for 40 minutes. Frequency, content and duration of sessions was based on a previous study. OTs provided content for the control condition (board games, puzzles etc.).	To determine the immediate short-term effects of a music therapy intervention with regards to reduction of agitated behaviours in individuals with Dementia.	Cohen-Mansfield Agitation Inventory (CMAI). Recorded by nurse carers who were trained in its use. Data were obtained for each session 1 hour prior to the start, 2 hour following the session, 2 hours following and 4 hours following.	Results were calculated excluding those where agitated behaviours were not present prior to the intervention (n=) and then again with these participants included. Following adjustments for agitated behaviours before the intervention and session numbers the reduction in mean CMAI scores (in residents who demonstrated agitated behaviours prior to taking part in the intervention) was greater for the music therapy group rather than the general activities group, however this effect was not significant (F=2.885; p=0.090) This effect was lessened when GDS was adjusted for (F=1.500, p=0.222). The median CMAI score in those taking part in music therapy was 2 (IQR: 1-3, range 1-9) 1 hour before the session commenced. This was 1 (IQR 0-2, range 0-7) at 4 hours following the session. For those in the control group the median score was 2 (IQR 1-3, range 1-13) 1 hour before the start of the sessions and was 1 (IQR 0-2, range 0-7) 4 hours post-intervention. No interaction was shown between the type of intervention and the session number for CMAI scores 4 hours following the intervention (F=0.275, p=0.603). There was also no interaction effect between the type of intervention and CMAI score before the session for the CMAI scores 4 hours after it took place (F=0.225, p=0.635).	Music therapy showed no greater a reduction in agitation scores (in the short term) than 'general activities' provided by occupational therapists. Overall CMAI average scores did show a decrease from one hour before the session to four hours after the session, however this was not significantly different among the two groups.	GRADE score ? + - + + - Levels of evidence score LI, grade A.
085. Lord (1993)	USA	60 recruited. 42 females (70%) and 18	Music group: big band music from 20s	To assess the impact of music therapy on	Questionnaire developed for the study specifically (and un-	ANOVA showed a significant effect between groups was observed for recall, social interactions and general mood (F= 4.11, p<0.01).	Those taking part in a music therapy interventions showed significantly improved	GRADE score ? ? ? –

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
Intervention Dates not specified	A privately funded care home for older adults.	males (30%). Age range 72-103. Participants were non-systematically separated into three groups of 20 participants each with 14 females and 6 males. One group took part in the intervention, one carried out puzzle activities and one participated in usual care. Loss to follow-up: Not reported.	and 30s played and participants given 'children's' instruments to play. Carried out in a group of 20. Each session lasted 30 minutes and took place for a total of six sessions. Sessions were delivered by the authors. Nursing assistants (two) and an activities coordinator accompanied the sessions.	the mood, memory and social interactions of persons with Alzheimer's Disease.	validated). The beginning of the questionnaire asked demographic information, the second section contained questions to aid memory and the third section focused on social interactions. 30 second observations were carried out for each participant for the first two weeks of the study.	T-ratios for scores on recall, social interaction and mood questionnaire, showed a significant difference for the music groups scores when compared with both the placebo (M 2.0, t 1.93, p<0.05) and control group (M 2.3, t 2.7, p<0.05). There was no significant difference between scores for the control and placebo group. Correlation calculations showed there to be a significant correlation between pre and post questionnaire scores for the music group (*r*= 0.468, p<0.01), but no effect for either the control (*r*= 0.712, p=reported ns) or placebo group (*r*= 0.728, p= reported ns). Within the music group significant mean gain scores were observed for pre and post observations for patient mood (M=1.51, t= 6.71, p<0.01) and social interaction (m=1.6, t= 6.71, p<0.01).	scores for recall, social interaction and mood changes, compared with a placebo and control group. Those in the music therapy group also showed significant improvements from pre-test to follow-up for measures of mood and social interaction. This effect was not found for either the control or placebo group.	Levels of evidence score LII-b, grade B.
090. Bennett (1988) Intervention Dates not specified	Australia Participants resided in either a hostel or care home environment	26 in total. Mean age of the treatment group was 82.2 (± 7.8) and for the control group was 81.2 (± 7.8). Gender differences not reported. Loss to follow-up: not reported.	Music- music based life review. Carried out within a group setting once a week for six weeks. 45 minutes in duration. The sessions involved listening to a piece of music and then having a discussion about their feelings in relation to the song. The control group took part in a verbal life review process.	To explore the effectiveness of a music-based life review carried out with individuals who resided either within a care home or hostel. The effectiveness of the music-based life review was also compared with that of a verbal life review therapy (the most common form of life review therapy).	Life Satisfaction Index A (LISA) and Ego Integrity Subscale of the Ego Adjustment Scale. Administered pre-treatment and post-treatment. Participants were also given a questionnaire asking about their enjoyment of the intervention.	***LISA:*** **2x2** repeated measures ANOVA with two within-subjects factors; treatment (music based life review) and time (pre and post treatment), showed a significant F statistic for the interaction between treatment and time for the LISA (f (1, 12)= 26.674, p<0.001). The pre-treatment scores for the life review group were higher, but this trend was reversed at post-treatment. ***Ego Integrity:*** Participants in the music group showed significantly higher scores compared with the verbal life review group (F (1, 12)= 9.07, p<0.05), however they also showed higher pre-intervention scores. There was no effect of treatment type and time for ego integrity scores. ***Enjoyment Questionnaire****:* those in the music-based condition showed higher enjoyment scores (m= 3.23) than those in the verbal review group (m= 2.31), t (112)= 3.21, p<0.01. Those in the music condition also reported they found the treatment significantly more helpful (m= 2.07) than the verbal life review group (m= 2.00), t (12)= 3/27, p<0.01. A highly significant moderate correlation was shown between life satisfaction scores and enjoyment and perceived helpfulness of the intervention.	This study compared a music-based life review process and a verbal life review process. Those in the music condition showed higher life satisfaction scores (LISA) and higher ego integrity scores following the intervention compared with those from the verbal life review group. Life satisfaction scores were shown to be significantly impacted upon by time (strong increase for the music group and slight decline for the verbal group), however this treatment effect was not observed for the ego scores. Those in the music group reported higher enjoyment of the intervention and greater perceived helpfulness compared with the verbal life group.	GRADE score ? - - ? ? – Levels of evidence score LII-a, grade B.
091. Smith (1986) Intervention Dates not specified	USA Two homes located in Philadelphia	12 participants, four of whom were located from one of the homes and eight from the other. No further demographic information was provided. Loss to follow-up: 1 participant became ill for a session for the musically cued reminiscence.	Music- musically cued reminiscence. Three groups met for a total of six sessions, each lasting for half an hour. They received musically cued reminiscence, verbally cued reminiscence and a musical activity on its own.	To determine the effects on cognitive functioning during three separate interventions: musically cued reminiscence, verbally cued reminiscence and musical activity (without the reminiscence aspect). To compare the interventions to assess which had the greatest effect on cognitive functioning.	Mini Mental State Questionnaire (MMSQ- now the MMSE). Administered at baseline and then to 2 members of the group (from a potential four)	Total MMSQ scores were increased following musically cued reminiscence sessions (mean 10.9 pre VS. 12.1 post), however this was not significant (t (11)= 1.07, p>0.05). No significant differences for orientation or attention subscales (t (11)= .37, t (11)= .80, p<0.05), but a significant improvement in the language subscale (mean 5.5 pre vs. 6.1 post, t (11)= 1.83, p>0.05). **Verbally cued reminiscence**:(10.9 pre compared with 12.0 post), however not significantly so (t (10)= 1.42, p >0.05). There were no significant increases for either the orientation or attention subscales (t (10= .44, t (10)= .56, p >0.05). A significant improvement was shown for the language subscale (t (10)= 1.84, p<0.05). **Musical activity alone**: (t (11)= 1.83, p<0.05), but no significant increase for any of the subscales (orientation t (11)=0.92; attention t (11)=0.18 and language t (11)=1.68, p>0.05.) No intervention improved MMSQ scores significantly more so than the other (total F (2, 32)=0.003; orientation F (2, 32)=0.028, attention F (2, 32)-0.022 and language F (2, 32)=0.033, p <0.05).	Musical activity on its own improved total MMSQ scores, but neither the musically cued or verbally cued reminiscence tasks showed this improvement. Language was improved via the two reminiscence interventions and this improvement was not observed for the musical activity alone. Orientation subscale scores were low at baseline and showed little improvement throughout the study period. Pre-treatment scores did not predict which treatments would be the most effective,	GRADE score - ? ? – Levels of evidence score LII-b, grade B.
094. Vanderark (1983) Intervention Dates not specified	USA Location not specified	43 in total, 20 from the experimental group and 23 from the control group. Mean age of 78 for the intervention	Music- group music programme Two music sessions per week for a total of five weeks. Treatment was	To examine whether there would be an effect on 'concerns and attitudes' (measured by self-concept, life	The authors devised the measure used within this study. The measures included Items that would indicate a change in attitude about life satisfaction, socialization, self-	T-tests on pre-test mean scores showed the groups to be similar at baseline, however significant differences were found for measures of self-concept (t=1.75, p<0.05) and music self-concept (t= 2.31 p<0.05). The music intervention group showed significant gains for measures of life satisfaction (t= 3.74, p<0.05), music attitude (t= 2.64, p<0.05) and self-concept in music (t= 2.54, p<0.05).	Taking part in a music therapy programme improved measures of life satisfaction, music attitude and self-concept in music for older people living within a care home.	Levels of evidence score LII-b, grade B.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
		group and 82 for the control group. Loss to follow-up: not reported.	planned by the experimenter and modified accordingly. Sessions were delivered by the researchers.	satisfaction, socialisation, music attitudes and/or music self-concept) in those taking part in a music participation programme.	concept, music or music self-concept.	It was found that the experimental group was significantly improved over the control group when all criterion variables were considered simultaneously (p = .031).	Those taking part in the music intervention also showed a significant improvement when compared with a control group.	
097. Doric-Henry (1997) Intervention February 1995-April 1995.	USA Evangelical Home of Saline (church based residential nursing home in Michigan)	40 in total; 20 in the intervention and 20 in the control. Intervention (n=20)- 19 female (95%) and 1 male (5%). Mean age 83.5 (range 50-95). Control (n=20)- 16 female (80%) and 4 male (20%). Mean age 85.9 (range 76-99). Loss to follow-up: not reported.	Pottery Sessions were held once a week for eight weeks for an hour. A number of different ceramic processes including wedging, throwing, drying, trimming, bisque-firing, glazing and glaze firing which enabled them to complete a piece of pottery by the end of the course. Sessions were delivered by the researchers.	To explore the role of a pottery intervention on levels of anxiety, self-esteem and depression levels in older adults.	Coopersmith Self-Esteem Inventory. Beck Depression Inventory. State-Trait Anxiety Inventory. Descriptive observational records were also kept throughout the intervention period and a subjective questionnaire about their self-esteem, depression and anxiety was also completed. Administered pre & post-test.	The pottery group showed significantly lower depression scores following the intervention than before (mean 4.6 (SD 5.5) pre and 2.1 (SD 2.5) post), alpha level .05. The pottery group also showed significantly improved self-esteem scores following the intervention (72 (SD 13.9) pre and 81.6 (SD 8.4) post), alpha level .05. Trait and State anxiety scores also reduced post-test (28.7 (SD 7.7) vs. 22.6 (SD 5.8)). No changes on any of the outcome measures were observed for the comparison group. Matched pair t-tests showed that those who began the intervention with higher self-esteem scores (mean of 75.2) did not show a significant improvement in their self-esteem scores following the intervention, however those with lower mean scores (60.8) pre-test did (alpha level .05). This was also found for depression scores those with low (mean 1.0) did not show improvements but those with high pre-test scores (mean 8.2) did (alpha level .05). Those with high state (mean 34.3) and trait (mean 34.5) anxiety also showed significant improvements, whereas those with low scores (state anxiety mean 23.0 and trait anxiety 23.5) did not (alpha level .05).	Individuals who took part in a pottery programme showed significant improvements in measure of anxiety, self-esteem and depression. The comparison group did not show any improvements. Those who had lower self-esteem scores prior to the intervention showed the greatest levels of improvement in this measure. Furthermore those with high levels of anxiety and depression at baseline also showed highest levels of improvement.	GRADE score ? - ? – Levels of evidence score LII-a, grade B.
098. Houston (1998) Intervention Dates not specified	UK Seven care homes within the UK (one carried out the pilot and the remaining six took part in the actual study).	61 residents from six homes. 44 were female (72%) and 17 male (28%). Mean age 83.7 (± 6.79) range 63-97. 31 in the intervention condition and 30 in the control. Loss to follow-up: not reported.	Humour and sing-a-long The researchers sang along with a backing track and danced in a comical fashion and participants were given props and asked to join in.	To ascertain whether there would be an impact on the psychological wellbeing of older people taking part in a 'laughter inducing humorous activity' consisting of a sing-a-long and dance activity. Focused on anxiety and depression scores in those taking part in the intervention.	The general health questionnaire (GHQ-28) the Hospital Anxiety and Depression Scale (HADS). Carried out at baseline and immediately following the intervention.	Baseline t-tests showed no significant differences (p>0.05) for scores on the GHQ and HADS and subscales. ***GHQ:*** For the GHQ anxiety subscale ANCOVA showed significant correlation between Time 1 and Time 2 F (1, 58)= 2.73, p=<0.001 and significant effect of condition when Time 1 scores were adjusted for F(1, 58)= 15.48, p<0.001. GHQ anxiety mean score intervention 2.11 and control 5.02, F (1, 58)= 15.48, p<0.001. Not significant for somatic symptoms (6.29 intervention vs 6.33 control), p=0.5; severe depression (8.24 vs. 9.38), p=0.43 or social dysfunction (intervention 5.78 vs 3.26 control), p=0.43. ***HADS****:* for anxiety significant effects of the covariate F (1, 58)= 45.37, p<0.001 and condition once covariate was adjusted for F(1, 58)= 19.99, p<0.001 (mean intervention 3.15 vs. 6.04 control). For depression effect of covariate was not significant F(1, 58), p=0.17 but effect of condition was, F (1, 58)= 8.89, p=0.004 (mean intervention 3.27 vs. 5.26 control).	This study shows a significant effect of the intervention for anxiety and depression, whilst covariate scores did not meet significance the effect of the intervention did. Somatic symptoms, severe depression and social dysfunction subscales of the GHQ did not show any changes.	GRADE score ? ? ? ? ? ? Levels of evidence score LI, grade A.
101. Sole (2014) Intervention Dates not specified.	Spain Setting not specified.	16 in total 15 female (94%) and 1 male (6%). Mean age of participants 87.5 (± 5.6) and range 76-91. Group 1- GDS score 3-4, n=9 Group 2- GDS score 5, n=5 Group 3- GDS score 6-7, n=2 Loss to follow-up: not reported.	Music- group music therapy. Activities included: singing, listening to music, playing instruments, composition and movement. Took place once a week for 12 weeks, lasting 45-60 minutes. All sessions were carried out by a qualified music therapist.	To analyse whether there were changes in affect and participation levels in individuals with dementia (mild, moderate and severe) who were taking part in group music therapy sessions.	GENCAT-Quality of Life, administered at baseline and then follow-up three months later (end of programme). Observation schedule focusing on verbalisations, physical contact, visual contact, active participation and emotions.	Several GENCAT measures were reportedly 'not addressed' by the intervention so only dimensions of emotional wellbeing, interpersonal relations and personal development were analysed. *Emotional wellbeing:* entire group pre-test median 12 (IQ 4) post-test median 23 (IQ 4), Z= -2.176, p=0.03. Median for 'mild' Group 1 at pre-test 22 (IQ 4) and post-test 23(4): z=-2.047, p=0.041. No other significant effects for different groups. *Interpersonal relations:* whole group pre-test median= 27 (IQ 6) post-test 25.5 (IQ 6), Z= -2.074, p=0.03). **All groups:** Behaviours observed four or more times per session were all positive, aside from agitation which was present for two of the participants with GDS scores of 5 and 7. Positive behaviours occurred when the participants were interacting with the music therapist (V1, V5 and L1) and active participation in activities when playing musical instruments (P1, P3, P4). Negative verbalisations (V2, V6) had low values whereas expression of negative emotions did not appear (aside from agitation). Positive emotions with smiles (E1) were low level throughout. Initiation of spontaneous physical contact did not occur. **Group 1:** 10 behaviours.. Behaviours appearing on four or more occasions: positive verbal responses to the music therapist, playing musical instruments and improvisation. There were no negative verbalisations. There was a slight trend in the appearance of positive emotions, with smiles in particular. There were no signs of negative emotions aside from one session where	This study looked at a music therapy programme for three groups of participants, those with 'mild', 'moderate' and 'severe' dementia based on their Global Deterioration Scale Scores. Although not significant, quality of life scores were lower for the total group and also for each individual group before taking part in the programme, showing quality of life was not improved by the intervention, rather it was worse following. Emotional wellbeing was shown to be significantly improved following the intervention for the group as a whole and those with mild dementia. Descriptive accounts of behaviours are also presented, with positive behaviours being reported more common than negative ones.	GRADE score ? ? - - Levels of evidence score LII-b, grade B.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
						crying was observed. No positive spontaneous physical contact occurred with the therapist or other members of the group. **Group 2:** 11 behaviours were observed. Behaviours recorded four or more times were interaction with the music therapist, in particular verbalisations and looks and participation in playing instruments. There was also participation in singing activities but to a lesser extent than those in the mild group. No negative verbalisations were seen. One case of agitation was observed. Signs of restfulness were high, however there was a decrease in these behaviours. No sign of spontaneous physical contact with the therapist or others. **Group 3:** these were observed on a lower level to those in the other. Nonverbal responses to demands of the music therapist were observed. Non-verbal communication with the therapist reduced from the start of the program to the end where there was almost no eye contact between participants. Instrument playing either spontaneously or directed remained high throughout. Once case of agitation was observed, but no other emotional expressions.		
102 de Guzman (2011) Descriptive Dates not specified.	The Philipines 'Group home'.	3 older adults, aged 61, 73 and 86 (mean age 72.2). Loss to follow-up: none, only three interviewed.	Art and Craft-Traditional Filipino Arts (TFA) Puni-making (no detail on what this entails, there are some pictures in the paper).	To explore the potential for traditional Filipino arts as a recreational therapy for older adults.	Phenomenological analysis was used to analyse individual videotaped interviews carried out with the participants.	Themes and Subthemes Me, myself and melancholy- wearing out, walking away and wanting more. Will not let my worth wither- making it through, making it happen and making a difference.	This study presents the experiences of three older people living within a care home who took part in traditional arts activities. They reflected on their lives prior to the care home, their experiences within the home and their experiences of the craft activity. Two major themes were identified, the first 'me, myself and melancholy' outlines the struggles related with becoming institutionalised and the second 'I will not let my worth wither' highlights the desire to feel worthwhile and enjoying creativity.	Risk of bias score Grade B, some flaws unlikely to affect the credibility, transferability, dependability and/or confirmability of the study Levels of evidence score LIII.
106. Hammar (2011) Intervention Dates not specified.	Sweden Two care homes for people with dementia.	Ten participants took part in each stage of the study. Mean age was 81.3 years with a range of 66-92 years. Six were female (60%) and four males (40%). Loss to follow-up: 12 participants were originally to participate in the study, two died during data collection and so 10 took part in the whole study, data is presented for these 10 only.	Music- music therapeutic caregiving Two Caregivers attended an MTC course) and the remainder were taught how to undertake MTC by the study author. Morning care sessions (washing, dressing etc.) were then carried out along with caregiver singing of 'familiar' songs.	To incorporate music therapeutic caregiving (MTC) into the morning routines of people with dementia and assess the impact on expression of emotions and restiveness to care.	The Observed Emotion Rating Scale (OERS) and the Restiveness to Care Scale (RTCS). A total of 8- sessions were recorded, eight for each participant (four at baseline and four during the intervention).	**Restiveness to Care Scale:** At baseline, there was a longer duration of pulling away in comparison to when the MTC was being carried out (148.8 seconds compared with 49.3 seconds) and this result was significant (p=0.031). Grabbing objects: The mean length of time for grabbing objects was significantly greater in the baseline phase (81.3), compared with the intervention phase (32.7), p=0.020. Behaviour adduction: Mean length of time was significantly lower for those in the MTC phase (30.8) compared with the baseline (78.5).Screaming: The only behaviour which lasted for longer than 2 minutes, observed in both phases. 55% (n=22) of observations within the baseline phase involved a participant screaming, however only 32.5% (n=13) of observations in the MTC did. Length of time was 205.8 for the no singing condition and decreased to 104.5 for the singing condition, however this result was not statistically significant. **Observed Emotion Rating Scale:** The mean length of time that pleasure was observed for was 281.8 for the no singing condition and 1387.5 seconds for the singing condition and this was a statistically significant difference (p=0.016). 20% of the participants expressed pleasure for more than one minute at baseline, compared with 50% during the intervention. General alertness occurred for 2010.2 seconds on average, compared with 2703.3 seconds during the intervention (p=0.042).. For negative expressions of emotion showed that anger lasting longer than 16 seconds decreased from 25 to 40 observations within the baseline condition to 14 out of 40 during the intervention. None of the participants showed anger for longer than five minutes during the MTC. Anxiety/fear lasting longer than 16 seconds occurred for half of the observations at baseline, versus only 37.5% during the intervention. Sadness was the least common expression for both conditions, 75% expressed no sadness during the usual care condition and 85% expressed no sadness during the MTC condition.	There was a significant decrease in pulling away, grabbing objects and adduct when participants were exposed to carer singing during their morning care routine, furthermore there was a significant decrease in the length of time these behaviours occurred for when the MTC was carried out. There was also a significant increase in the level of positive emotions expressed when the MTC was carried out. There was also a decreased during the intervention, however this result was not statistically significant.	GRADE score - ? + + Levels of evidence score LII-b, grade B.
108 Gotell (2007) Descriptive Dates not specified.	Sweden A special care unit within an Urban area of Sweden.	9 residents, with a mean age of 84 years and a range of 80-90 years. Seven were female (78%) and two (22%) were male. Five female caregivers also participated within the study.	Music- carer singing Three conditions were observed: the usual morning care routine, the morning care routine with background music and the morning care routine with care-giver singing.	To explore whether there were any effects for people with dementia who were exposed to carer singing during their morning care routines.	Observations- each of the participants were observed for each of the conditions, giving 27 observation periods in total. Video-recordings were transcribed into text and then analysed using content analysis.	A major theme was developed for each of the three conditions which represented the emotions and moods that occurred during the interactions between the carer and the resident. *The usual morning care session*: disjoint vitality *Background music:* mutual vitality infused with playfulness *Caregiver singing:* mutual vitality infused with sincerity.	This study observed three different conditions: morning care routines as they would usually be carried out, morning care routines carried out with background music and morning care routines carried out with caregiver singing. They study found that during usual care with no music the participants were disjointed with the caregivers, with the caregivers showing energy and enthusiasm that was not met by the participants with dementia who responded in a	Risk of bias score Grade B, some flaws unlikely to affect the credibility, transferability, dependability and/or confirmability of the study

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
		Loss to follow-up: not reported.					listless and disinterested manner. During the music condition the participants with dementia displayed more positive emotions and a reduction in aggression. Music was shown to aid cooperation in carrying out the morning care activities.	
112 Myskja (2008) Intervention October 2003-January 2004	Norway A nursing home in Oslo (Valerengen bo-og service-senter) with three wards.	72 participants regularly (85.7% of all at the home). 51 (71%) were female and 21 (29%) were male. Mean age was 87.5 and range was 63.2-95.8 for females and 80.9 (57.5-93.2) for males. Loss to follow-up: of the 72 who took part, 9 withdrew. Therefore 63 residents could be involved in the analysis of the music sessions. At the second round of measures 9 withdrew.	Music- group music therapy. 45 minutes in duration carried out twice a week. Delivered by a music therapy aide who played the piano and led singing of familiar and preferred songs.	The aim of this study was to restart a music therapy programme that had been previously discontinued due to staff changes and assess changes in depressive symptoms of residents following the programme starting again.	Montgomery Asberg Rating Scale (MADRS) administered pre-test when the usual music therapist had been absent for 11 weeks and 2 months after the MT had been reinstated. Scored by the nurse leading the programme and the named nurse of the participant. Level of participation: 3- always/nearly always; 2- usually; 3- sometimes; 1- never.	Baseline depression scores varied for each ward; Ward 2 (mixed somatic and dementia ward) = 27.5 (± 3.8); Ward 1 (somatic)= 18.6 (± 5.8) and Ward 3 (dementia)= 21.7 (±5.8). Frequency analysis showed significant reduction in depression scores (all only reported as $p<0.05$) for the music condition for all results combined (mean 20.4 vs 12.2 following) and for each individual ward (ward 1 18.6 vs 11.1; ward 2 27.5 vs. 15.5 and ward 3 21.7 vs. 12.7). The relationship between participation levels and MADRS values, linear model showed participation levels predicted a change in depression scores (t -2.18, $p<0.05$). N.B participation measures are not reported in the outcome measure section of the paper but are reported in the results section. The authors also refer to semi-structured interviews within the discussion section, but again these are not presented in the results section.	This study examined depression levels in participants who had a regular music therapy programme which had not taken place for eleven weeks. Depression levels were significantly lower following the programme restarting than at baseline levels. Additionally, greater levels of participation were predictors of greater reductions in depression. Lower levels of participation, however, were also linked to advanced disease.	GRADE score ? + - - Levels of evidence score LII-b, grade B.
115. Mohammadi (2011) Intervention	Iran 'Long term care facility'.	19 in total. 11 males (58%) and 8 females (42%). Intervention group 11, six males 55% and five females (45%) mean age 69.9. Control group 5 males (63%) and 3 females (37%). Mean age 68.9	Music- Iranian music therapy. A total of ten sessions, once a week lasting for 1 and a half hours. Those taking part were encouraged to be involved with the intervention, as well as taking part in group discussions prompted by the music. They were also encouraged to sing the words to popular songs and play the instruments with the music.	To assess the effects of an Iranian group music therapy programme on levels of stress, anxiety and depression in older adults living within a care home.	Depression Anxiety Stress Scale (DASS-21) Measures were determined at baseline and then immediately following the intervention.	No significant differences between groups at baseline for depression Z-0.91 (p=0.36), anxiety Z-0.96 (p=0.96) or stress Z- 0.48(p=0.48). Post-test significant differences between groups for depression Z=-3.60, $p <0.001$; anxiety Z=-2.91, p=0.004 and stress Z=-3.20, p=0.001.	Prior to the intervention there were no differences in groups in terms of depression, anxiety or stress. Following the music therapy sessions the group who took part in the intervention had significantly improved scores on each subscale for depression, anxiety and stress.	GRADE score ? ? ? ? - ? Levels of evidence score LI, grade A.
116. Hicks-Moore (2007) Intervention Dates not specified.	Canada Three large care homes located within New Brunswick, Canada. Each had special care units, two of which had 24 beds and	40 in total. 32 were female (78%) and 9 male (22%). Mean age of participants was 84.5 years (± 6), range 67-92 years. Consent was obtained for 56 residents. Of these, 41 completed the entire study. 9 did not show any	Music and hand massage (individually then both at once). Favourite music-preferences were collected from family members and a 10 minute disk of favourite music created. This was then played on a CD player to the residents.	To explore the effects of hand massage or favourite music (carried out individually) on agitated behaviours in individuals with dementia. Then, to assess the effects of both interventions carried out at the same time on agitation levels in	A modified version of the CMAI which records the frequency of agitation as it occurs rather than retrospectively. The CMAI was measured at baseline (10 minutes prior to treatment), immediately following the intervention and then one hour post-treatment.	A 3x3 repeated measures ANOVA was used to determine whether any change in aggressive, physically non-aggressive and verbally agitated behaviours. **Aggression:** (Fcorr =1.91, p=.17) or across the three treatment types (Fcorr=2.97, p=0.8). There was also no significant interaction between treatment type and time. Therefore neither of the treatments was shown to significantly reduce physically aggressive behaviours. **Verbally agitated behaviours:** (Fcorr =24.74, p=0.001). Scores decreased following treatment and then levelled off between post-treatment and follow up. Bonferroni-corrected pairwise comparisons also indicated significant differences between pre-treatment and the follow-up ($p<0.001$). There were no significant differences or interactions between the treatment types. **Non-aggressive behaviours:** (F= 123.38, p=.001) but not over the three treatment conditions (F=.47, p=.63). Pairwise comparisons confirmed this significant difference between pre-treatment and post-treatment/follow up ($p <0.001$). Scores for both this measure and that of non-aggressive	Residents who took part in hand massage, listening to favourite music or both conditions combined showed a significant reduction in levels of agitation following the intervention and one hour after the intervention had finished compared with a control group. No significant effect was shown for aggressive behaviours; however both non-aggressive and verbally aggressive behaviours were significantly reduced. None of the interventions was shown to be more effective than the other and combining	GRADE score ? ? - ? + - Levels of evidence score LI, grade A.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
	one which had 34 beds.	agitated behaviours, three transferred to another unit and 3 passed away.	Hand massage- five minutes for each hand based on a protocol developed by a previous study.	those with dementia. Finally, to compare whether both treatments carried out simultaneously were more or less effective in reducing agitation levels than either of the interventions carried out on their own.		behaviours decreased quickly following treatment and then remained stable between post-treatment and follow up.	the two was not significantly more effective than when they were carried out individually.	
118. Kincaid (2003) Intervention Dates not specified.	USA A special care unit located within Courtland Terrace Nursing Home, North Carolina.	12 in total (who exhibited exit seeking behaviours from 17 who consented to the study). Of these ten were female (83.3%) and 2 were male (16.7%). Mean age of the participants was 78. Loss to follow-up: not reported.	A mural produced by a local arts class which covered from the floor to the ceiling and the whole door and walls. The home administrator had requested that the colours and themes be in keeping with the existing décor of the home. Residents had an input in the final design of the mural.	To explore the effects of a wall mural on exit-seeking behaviours exhibited by individuals with dementia who were living in a nursing home.	Data were collected in 2 hour blocks for six weeks prior to the installation of a wall mural and then at the same frequency for six weeks following installation of the wall mural. Four different door testing behaviours were observed.	The overall mean score for door testing behaviours prior to the intervention was 55.67 (*df*=11; SD=67.57) and following the intervention was13.42 (*df*=11; SD= 28.41). This represents a significant mean decrease of 42.25 (*t*=2.6; *p*=0.24) All 12 residents showed exit-seeking behaviours prior to the intervention and only 3 remained active following. **Type 1 behaviours** pre-installation mean average was 35.67 and following 6.17, a mean significant decrease of 29.5 behaviours (t=2.622, *p*=0.024) **Type 2 behaviours** pre-installation mean of .67 and .17 follow. No significant difference (p=0.5). **Type 3 behaviours** pre-installation mean average was 3.42 and .75 afterwards, a significant mean decrease of 2.67 (t=2.432, p=0.03). **Type 4 behaviours** pre-installation mean average was 15.92 and 6.33 following, a mean decrease, that was not significant (p>0.05).	Observation of four types of door-testing behaviours: 1. Walking to the door and calmly pushing/pulling it. 2. Waiting patiently for someone to walk out the door and then trying to exit 3. Teaming with another resident to try and open the door 4. Actively testing the door with exerted force and showing agitation or hostility. Qualitative data was also obtained during the study to enhance findings.	GRADE score ? - - Levels of evidence score LII-b.
120 Suzuki (1998) Intervention Dates not specified	USA Residential care facility in the North-West.	8 took part in the study from an initial 11 recruited. Five were female (63%) and three male (37%). Mean age 83.5 (± not reported). Range 75-93. Loss to follow-up: 3.	Music- group singing and lyric composition. Sessions delivered 3 time a week for total of 9 times over 3 weeks. Sessions made up of three parts- sing-along; music making and relaxation.	The overall aim of the study was to assess the role of music therapy on mood and congruent memory recall in older adults who had depression.	Positive and Negative Affect Schedule (PANAS). Memory Retrieval Test. Pre-test took place before the intervention and post-test no more than 90 minutes following.	There was no significant difference in scores for the positive affect aspect of the scale, however the change in negative affect were significant (p=0.26). This show there was a significant decrease in negative mood following the music therapy. There was also no significant difference in the percentage of pleasant memories pre and post-test. There was a significant change in the percentage on unpleasant memories (p=0.034), this shows that following the music therapy there was a significant decrease in negatively recalled memories (table 2). No significant correlation between Positive affect and pleasant memories (r=.71, p>0.5), Positive Affect and unpleasant memories (r=-.66, p>0.5), Negative Affect and pleasant memories (r=-.60, p>0.5) and Negative Affect and unpleasant memories (r=.31, p>.05) The discussion section reported analysis by gender but this is not included within the results section and values are not given. N.B used GDS to select participants but not as an outcome measure.	Older adults who initially presented with symptoms of depression showed a significant decrease in negative mood and a decrease in recall on unpleasant memories that was also significant. No significant effect was shown for congruent mood.	GRADE score - - + ? Levels of evidence score LII-b, grade B.
121 Carruth (1997) Intention Dates not specified.	USA Florida nursing home- an Alzheimer's specific unit and skilled nursing unit.	Seven participants in total, all were female. Mean age was 87 (range 79-90). Loss to follow-up: not reported.	Music therapy 14 sessions in total, sessions 7-10 involved music. Participants were given a photo of the staff member who then gave their name and sang a song, participant was encouraged to join in. if they did not the carer sang the song again.	To determine whether there was any improvement in face-name recognition when older people with dementia took part in a music therapy intervention.	Target-stimulus recall. This tested whether the participant could successfully recall the name of the care giver. If the participant failed to correctly recall the name on the photograph then further timed intervals were carried out (10 seconds, 20 seconds, 40 seconds, 1 minute, 1.5 minutes, 2 minutes, 3 minutes, 4.5 minutes, 7 minutes, 10 minutes, 1 day). Participants were given three chances for each criterion, if they were still unable to recall the name then the data collection was stopped.	Only four participants showed greater mean response times for the music condition (S1 88% music vs. 83% none; S2 96% music vs. 86% none; S3 100% music vs. 67% none and S7 88% music vs. 75% none). No statistical comparisons for any means.	Findings from this study are limited and there was no comparison of findings, perhaps due to the small sample size. Face-name recognition was improved for four of the participants but not for the other three.	GRADE score ? ? ? – Levels of evidence score LII-b, grade B.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
					An observation form was also completed by a graduate student.			
122. Brontons (1996) Intervention Dates not specified	USA Four North-Western care facilities.	47 initially recruited, 20 completed ad data only given for these. 17 (85%) female and 3 (15%) male. Mean age of participant was 82 (± 6.6). Range was 70-96 years. Loss to follow-up: Of the 47 who began the study, only 20 completed all aspects of the study. The rest were reported to have either not attended all of the sessions or failing to meet inclusion criteria.	Music therapy-group. Took place in groups of 3-4 participants. 5 sessions in total lasting for 30 minutes. Sessions started with a hello song, then dance or movement based activity, then composition, followed by percussion and ending with singing songs. Sessions were carried out by two certified music therapists.	To carry out a live music therapy intervention and assess its impact on agitated behaviours in older people with Alzheimer's Disease.	Agitation behaviour scale section of the Disruptive Behavior Rating Scale (DBRS) administered at baseline, during the intervention and follow-up. Medical charts to obtain P.R.N medication use. Musical backgrounds of participants were collected via questionnaire.	9 of the participants= musical background, 5= no musical background, no information for the other 6. 40% of the sessions were analysed by two independent observers and interrater reliability was established to be .99. There was no effects with regard to music background (F1=1.79, p=.20) but there was a significant main effect of agitation behaviour (F3=16.33, p=.0001). Fisher's PLSD tests showed participants to be significantly more agitated before music therapy (m=11.46) than either of the two time points during the intervention (m=7.68 and m=7.52) and following the music therapy sessions (m=8.37). There was no two-way interaction between music background and agitation behaviour (F3=1.34, p=.28). 40% of the total observations were independently observed. A t-test for independent samples was calculated and showed there was no significant differences in agitation scores between music therapists and caregivers [t(78)=.67, p=.50]. The scores means and standard deviations for music therapists and caregivers were m=7.90, SD=2.01 and m=7.62, SD=1.63 respectively. Not enough information on P.R.N medications for an analysis.	Study showed that agitation scores were significantly lower following the classes than at baseline, but tended to fluctuate during the session and from session to session. Furthermore, there was an increase in agitation scores at baseline from session 1 to session 5, showing that overall agitation levels did not decrease over time, suggesting a limited effect of the intervention (for the period immediately following the music therapy). The musical background of the participants had no influence on agitation scores, suggesting that music therapy can be beneficial irrespective of musical knowledge. There was not enough information on PRN medications for an effect to be determined.	GRADE score - - + - Levels of evidence score LII-b, grade B.
123 Clair (1996) Intervention Dates not specified	USA A 'Midwest residential facility'.	26 in total. Age range was 62-83 years. 22 were female (85%) and 4 were male (15%). Loss to follow-up: not reported.	Music- live singing. Sessions involved the participants being exposed to all of the conditions, each lasting for two minutes each, one after the other: 4 x 2 minutes of singing, 4 x 2 minutes of reading, 4 x 2 minutes of silence, preceded by 2 minutes of baseline. These sessions were carried out by the researcher.	To assess alert-responses in individuals with dementia who were non-ambulatory and non-verbal during a live singing intervention and compare this with silence and listening to a reading.	Graduate students recorded presence or absence of behaviours at 15 minute intervals. The following items were identified: rotating the head to localise sound, rotating the head to localise the experimenter, opening eyes, turning eyes in the direction of the experimenter, moving arms, hands, legs and/or feet. Making vocal sounds and changing facial expression.	Participants were separated into three groups to conduct within-group analysis, Group 1 (n=3) were described as being 'agitated'. Group 2 (n= 16, including all males) were subdued/closed and Group 3 (n=7) were classified as subdued/open. T-tests were calculated in order to establish whether there was any difference between each of the three conditions (all participants in all groups). There were significant differences between singing and silence and silence and reading (p values not reported). No statistically significant effect was shown for the singing and reading comparison. One-way ANOVA showed there to be no statistically significant difference between the three groups at baseline. ***Second Session:***. The singing mean scores were significantly different between the groups (df=2 & 23, F=3.743, p=.039). Reading mean scores were also significantly different for the second session (df=2 & 23, F=3.45, p=.048). According to the multiple range test, Group 1 (agitated, M=8.67) and Group 3 (subdued/open, M=29.43) were significantly different (at .05 confidence level). Group 2 (subdued /closed, M=17.31) and Group 3 (subdued/open, M=29.43) were also significantly different (at .05 confidence level). Therefore those who were open were more likely to respond to the reading condition than those who were agitated or showed closed behaviours. ***Third session***:. Singing differences (df=2 & 23, F=5.06, p=.015) among the groups occurred between Group 2 (subdued/closed, M=11.56) and Group 3 (subdued/open, M=29.14), but no significant difference was observed for the agitated Group 1 (M=12.67) and any of the other groups. An ANOVA also showed significant differences among the groups (df= 2 & 23, F=5.78, p=.009) for responses during the silence with follow-up tests confirming differences between Groups 1 (agitated, M=5.33) and 3 (open, M=24.00) as well as between Groups 2 (subdued/closed, M=8.13) and Group 3 (subdued/open, M=24.00), at the .05 confidence level. Those in Group 3 who had their eyes open were not very active responders during the silence condition. ***Fourth Session****:* At session 4 there were no significant differences among the behavioural groups for any condition and the mean scores showed that Group 1 (agitated) and Group 2 (subdued/closed) had increased in their responses under all the conditions. ***Total:*** When the total group scores for all of the sessions were compared for each of the three experimental conditions statistically significant differences were found for singing (df= 2 & 23, F=4.4682, p=.022) and reading (df=2 & 23, F=3.44376, p=.049) but no statistically significant differences for silence. Multiple range follow up tests for singing showed Groups 1 (agitated, M=64.33) and Group 2 (subdued/closed, M=68.75) were not significantly different from one another, however they were both significantly different from Group 3 (subdued/open, M=116.43). This was also shown for the reading condition where it was found that Group 1 subjects who showed agitation	This study involved a small group of participants who were separated into three different groups based on their presentation. Group 1 was participants who presented with agitation, Group 2 was those who presented with subdued and closed off symptoms and Group 3 was those who presented with subdued but open symptoms. Overall there were significantly more responses in the music condition compared with the silence and the reading condition compared with the silence. However, there was no difference between reading and singing. Responses changed over time and the participants became increasingly responsive. Those in Group 3 were shown to respond the greatest to the interventions and this indicates a greater response to the carers singing and reading in those who exhibit subdued but open behaviours rather than those with agitation or those who show closed symptoms.	GRADE score ? ? ? ? Levels of evidence score LII-b, grade B.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
						(M= 70.33) and Group 2 subjects (subdued/closed, M=65.81) showed significantly fewer responses than those in Group 3 (subdued/open, M=106.86).		
125 Scott (2014) Intervention Dates not specified	Australia 'Aged care facilities' within the South East Queensland area.	Biophilia group (n= 15). Mean age was 82.2 years (± 8.9). 10 were female (66.7%) and 5 were male (33.3%). 10 staff members also took part. Reminiscence group (n=10). Mean age 83.3 years (± 8.6). 6 of the participants were female (60%) and 4 (40%) were male. 6 staff members also took part. Control group (n=8). Mean age was 83.3 (± 5.7). 6 were female (75%) and 2 were male (25.5%). Loss to follow-up: Attrition rate was reported as 13% for the biophilia group, 13% for the reminiscence group and 10% for the control group.	Sensory enrichment. Comparison of two different wall murals, one inspired by nature and a second inspired by reminiscence.	To implement a 'biophilia' environmental stimulation involving placing plants and natural elements within care homes. To assess the benefits of the intervention and compare these with a similar installation which comprised of non-natural factors (a reminiscence intervention). To compare the benefits of both interventions with a control condition which involved no stimulation.	A specifically designed resident questionnaire collected information on the satisfaction residents felt with their living environment. The Geriatric Depression Scale (GDS) and Geriatric Anxiety Inventory (GAI) scores were collected at Time 1, Time 2 and Time 3. A question was used to assess environmental satisfaction and two items were used to assess satisfaction with opportunities to keep occupied. Wellbeing and engagement were also measured using Likert-type scale of six questions and a staff questionnaire were administered to staff.	No significant differences were shown for baseline measures of depression, anxiety or QoL-AD scores. There were also no significant differences for scores on the satisfaction with environment or satisfaction for opportunities scales. Both those from the biolphilia and reminiscence conditions stated that they were happy with their environments. 90% (n=9) of the staff reported they liked the biophilia installation and 7 (87.5%) from the reminiscence installation. Significant interaction effect where responses differed as a function of group condition and time (T1, T2 and T3) ($F=(2, 22)= 2.85$, $p<0.05$, $n^2 p= 0.21$). Post-Hoc analysis showed there to be a significant difference between the groups at time 2 for social engagement for the biophiilia group (m=2.17) and reminiscence group (m=2.42) compared with the control group (m=2.27). A similar was shown at T3 where social engagement was significantly greater for the biolhilia group (m=2.27) and the reminiscence group (m=2.27) compared with the control group (m=3.20). There was no difference across timepoints for levels of social engagement in the control group ($p>.05$). Levels of social engagement were significantly different across time for the biophilia condition ($p<0.001$) and the reminiscence condition ($p<.05$). No significant changes for satisfaction or mood measures (all $p>.05$), however improvements were observed at T2 and T3 for the biophilia condition. Staff social engagement measures also showed no increase ($p>.05$). 60% of staff reported they felt the biophilia condition improved memories in participants and all staff in the reminiscence condition reported improvements.	The study looked at a biophilia installation, consisting of benches and plants and a multi-sensory reminiscence based installation which involved articles such as old telephones and paintings. There was also a control condition. There were significant improvements in measures of social engagement (as measured using a specific questionnaire) for both the biophilia and reminiscence groups when taking part in the intervention compared with measures obtained prior to the installation. Similarly improvements were shown at the time following the intervention compared with baseline. Both types of sensory enrichment were therefore beneficial to improving the living environments within care homes.	GRADE score + ? ? ? + + Levels of evidence score LI, grade A.
126 Vankova (2014) Intervention.	Czech Republic. 7 different nursing homes.	Total number participants= 162. Mean age 83 (± 7.98). 149 female (92%) 13 male (8%). Experimental group= 79 participants. Mean age 83 ((± 8.23). 76 female (92%) 3 male (8%). Control group= 83 participants. Mean age 83 ((± 7.87). 73 (92%) female, 10 (8%) male. Loss to follow-up: 216 participants agreed to take part 54 withdrew.	Exercise dance EXDASE classes held once a week for a total of three months. Sessions lasted for an hour- 10 minutes warm up, 40 minutes main intervention and 10 minutes cool down. Does not report who delivered the sessions.	To implement an exercise dance intervention for seniors (EXDASE) and evaluate its role on symptoms of depression in older people residing in a nursing home.	A Czech language version of the Geriatric Depression Scale (GDS). Scores were obtained at baseline (one week prior to the intervention) and within one week following the intervention (post-test).	GDS scores were significantly improved following the dance therapy intervention (5.71 mean (± 3.8) versus 5.00 (± 3.3)). In contrast, the control group showed worsening of symptoms (4.86 (± 3.2) pre-test versus 5.27 (± 3.3) post-test), however not significantly so ($p=0.81$). The group was separated into those taking antidepressants and those not, both groups showed improvements, however only the subgroup not taking antidepressants showed a significant improvement ($p=0.008$). Generalised linear model: significant for dance therapy (Wilks lambda $F=10.58$, $p=0.001$). Controlling for antidepressant usage showed similar results, suggesting changes occurred as a result of the intervention (Wilk lambda $F=10.58$, $p=0.001$) rather than antidepressant intake (Wilk lambda $F=0.107$, $p=0.744$). Further GLM analysis confirmed that the results occurred as a result of the intervention (Wilks lambda $F=10.75$, $p=0.001$) rather than the nursing home (Wilk lambda $F=0.26$, $p=0.621$).	Taking part in an exercise dance class improved depression scores in older people who resided within different nursing homes. Analysis showed that antidepressant use and the specific nursing facility the participant resided in did not have any effect on depression scores, showing changes in scores were likely due to the effect of the intervention.	GRADE score ? ? + + - + Levels of evidence score LI, grade A.
130 Hammar (2010)- this study is tagged with 106. Intervention Dates not specified.	Sweden Two care homes in urban Sweden.	Older people- 10 in total, demographic information not provided. Carers at the home- six in total. Age range 31-54.	Music- therapeutic caregiving (MTC). Two Caregivers attended an MTC course) and the remainder were taught how to undertake MTC by	This study was part of a larger research project aimed at understanding the influences of music therapeutic caregiving on people	Group interviews were carried out with the caregivers, interviews lasted around an hour each (four in total). Observations- each person with dementia was observed for a total of four times (once	Main theme arising from the interviews related to the normal morning care situation- 'Struggling for Care in Communion', with four subthemes: hampered communication, physical and mental struggle with aggression, struggling with ethical demands and the reward- consolation and love. The second main theme which arose from the interviews related to the MTC morning care situation was 'consolidating care in communion' with two subthemes: awakening cooperation and feelings of wellbeing.	This study explored the role of MTC in improving morning care routines with older people with dementia who resided within a care home. Caregivers provided qualitative interviews presented within this paper. Interactions were referred to by the care givers as being difficult during the normal care routine, with	

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
		Loss to follow-up: not reported.	the study author. Morning care sessions (washing, dressing etc.) were then carried out along with caregiver singing of 'familiar' songs.	with dementia and their caregivers.	a week) during usual care and four times (once a week) during MTC taking place- these are described not analysed in this paper.		participants resisting care and needing to be restrained and interacting poorly with caregivers. Interactions during the MTC sessions were said to be more enjoyable, with less restraint needed and improved communication between the caregiver and the person with dementia. This improved the mood of the caregivers which, in turn, improved the mood of the residents.	
132 Guetin (2009) Intervention September 2007-April 2008.	France Les Violettes Nursing Home.	30 in total. 22 females (73%) and 8 males (27%). 15 assigned to the music therapy group. 13 females (87%) and 2 males (13%) with a mean age of 85.2 (± 6) range 75-93. 15 assigned to the control group 9 females (60%) and 6 males (40%). Intervention group: 3 withdrew but all 15 in ITT analysis. Control group: 4 withdrew, all 15 included in ITT analysis.	Music therapy Sessions carried out weekly, for 16 weeks in total. A control group were offered sessions such as reading at the same time. A "U sequence" was created using preferred music which was then listened to via headphones in the participants' bedrooms or preferred areas.	To assess the impact of a short and medium term music therapy programme on anxiety in those with mild to moderate Alzheimer's disease who lived within a nursing home. The study also aimed to assess depression scores at 2-months post intervention to assess long-term impact.	Hamilton Scale for Anxiety Geriatric Depression Scale (GDS) Took place at baseline (D0), week 4 (W4), week 8 (W8), week 16 (W16) and week 24 (W24- follow up). Carried out by a neurologist and psychologist.	An ANOVA with repeated measures showed a significant difference ($p<0.001$), showing that the two groups progressed differently during the follow up period. At baseline there were comparable levels of anxiety between the two groups (22, SD 5.3 for the music therapy group and 21.1, SD 5.6 for the control group). However, at week 16 the anxiety scores for the music therapy had decreased to 8.4 (SD, 3.7) whereas the control group had only decreased to 20.8 (SD, 6.2) ($p>0.001$). ***Anxiety:*** 2 months following cessation of the sessions there was a significant difference between the two groups with respect to their anxiety scores ($p<0.0001$). Week 24 mean music therapy 10.6 (± 6.3) and control mean 20.5 (± 5.4) ($p<0.001$). Week 16-24 music therapy mean change in points 2.1 (± 3.7) and control mean change points -.08 (± 2.8) and baseline-week 24 change in mean score music therapy group -11.5 (7.2) and control group -1.5 (± 6.8). ***Depression:*** Baseline music therapy mean score 16.7 (± 6.2) control mean 11.8 (SD 7.4) repeated ANOVA was significant ($p=0.001$). Overall changes were not significant over time, however each group progressed in a different way during follow-up (significant time/group interaction $p= 0.0095$). W16 scores for music therapy group 8.9 (± 3.3) and control mean 11.2 (± 6.1). ANOVA significant ($p=0.002$). After 16 weeks the improvement was approximately 7.7 (± 4.6) points (47%) in the music therapy group, with a mean depression score of 16.7 (± 6.2) versus an improvement of around 0.2 (± 4.4) points (1.7%) on the control group with a mean depression score of 11.8 (± 7.4). Scores at W24 were compared between the two groups. ANOVA with repeated measures showed a significant difference ($p=0.006$). Scores progressed differently during follow-up up to the six-month point. Depression score at W24 was 12.5 (± 6.4) for music group and 12.1 (± 7.6) in the control group ($p=0.03$).	Assessed anxiety and depression at a number of intervals prior to, during and following a music therapy programme. Anxiety scores fell between baseline and week 4 for the music therapy group but remained unchanged for the control group. Those in the music therapy group showed lower anxiety scores at 2 month follow up (W24) in comparison with the control group. Significant reductions across time were observed for anxiety scores over time for the music therapy group. Depression scores increased by more points for those in the music therapy group and there was a significant difference observed for intergroup differences between baseline and follow-up (W24) scores, suggesting reduction in depression may continue up to 2 months.	GRADE score + ? + + + + Levels of evidence score LI, grade A.
135 Noice (2006) Intervention Dates not specified.	USA Plymouth Place Continuing Care Facility (La Grange, Illinois), which offers varying support from independent living with meals provided to	18 participants. Mean age was 82.3 years (± 5.6), with a range of 72-95 years. 5 were male (27.2%) and 13 female (72.2%). Loss to follow-up: 21 participants expressed a willingness to be involved with the study, 3 withdrew prior to starting.	Drama A course was delivered on Tuesdays and Thursdays for a total of one hour. Emphasis was placed on genuine involvement and communicating thoughts and feelings. These sessions were delivered by a trained drama instructor.	To determine whether a theatrical intervention had bearing on cognition in older people residing within a continuing care facility.	The course took place twice a week (Tues and Thurs) and the sessions lasted for a total of an hour. The drama instructor was a professional actor who stressed the necessity for 'genuine involvement' in the scenes and corrected mistakes. Each participant took turns in taking part in the activity, which typically involved improvised scenes around a phrase, and the other members of the group became audience members.	**Cognitive battery**: A significant MANOVA was computed for the variables of time (within-subjects variable) and cognitive measures (dependent) ($p<0.001$). Therefore univariate ANOVAs were carried out and showed significant improvements for word recall ($F(2, 34)= 4.41$, $p=0.02$), problem solving, ($F (2, 34)= 15.46$, $p<0.001$) and set span ($F(2, 34)= 6.43$, $p<0.01$). Word recall was significantly increased from pre-test 2 to post-test ($t(17)= 2.70$, $p=0.015$) but not between pre-test 1 and pre-test 2 ($p=0.59$). Problem solving was significantly improved between pre-test 2 and post-test ($t (17)= 2.64$, $p=0.002$) and also between pre-test 1 and pre-test 2 ($t(17)= 3.01$, $p=0.01$). Marginal increases were observed for working memory from pre-test 1 to pre-test 2 ($p=0.70$) and pre-test 2 to post-test ($p=0.13$) but not significantly so. **Psychological Functioning**: MANOVA, significant effect of time ($F (2, 14= 4.24$, $p=0.02$) and univariate analyses did not reach significance ($p>0.05$). T-tests showed no significant self-esteem changes from pre-test 1 to pre-test 2 ($p=0.62$) and a non-significant improvement from pre-test 2 to post-test ($t(17)= 1.63$, $p=0.12$). No significant increases were found for quality of life. Scores for the Present Ability subscale showed a moderate positive correlation with Effort utility scores ($r 16=0.48 p<.0.05$) and a negative correlations with Inevitable Decrement ($r(16)=-0.54$, $p=0.02$). Those who believed they were able to improve their memory showed the greatest increase in recall performance $r(16=0.52$, $p=0.028$). However, positive correlation was also observed for Inevitable Decrement and problem solving scores $r(16)=0.56$, $p=0.016$.	Cognitive battery 1. Word Recall Task using 30 common nouns Listening Span 2. Means End Problem Solving Procedure (MEPS) The Psychological Wellbeing Battery 1. Self-Esteem Scale using a 10 item scale which explores dimensions of self-worth 2. Self-Reported Psychological Health- utilises 3 scales measuring personal growth, self-acceptance and positive relations with others. 3. Memory Controllability Index- self reported measure of memory. Also included was the Current Activity Index, which looks at ongoing participation in mentally stimulating activities.	GRADE score ? ? - ? Levels of evidence score LII-b, grade B.
139 Burrack (2004) Intervention Dates not specified	USA 'Large urban nursing home'	16 agreed to take part, 13 completed. Of these 9 female (69%) and 4 males (31%). Mean age 82 (± 8.92) range 64-93 years.	Music- therapy, individualised individually listened to 'individualised' music within their rooms through a personal CD player for a one-off session	To explore the effects of individualised music therapy on immediate satisfaction and overall quality of life.	Global quality of life assessed via the Quality of Life Alzheimer's Disease (QoL-AD). Music background questionnaire and music intervention questionnaire.	***Access and Frequency of Listening to Music***: participants reported that they listened to music significantly less often following entering the nursing home ($m=3.58$, $SD=1.16$) than they had before ($m=4.50$, $SD-0.52$), $t(11) = 2.73$, $p<.05$. Twelve participants reported that they used to attend live music performances. Eight participants used to play a musical instrument and five stated that they used to sing. In the nursing home the most frequently reported mode of listening to music was by TV (8 participants). Seven participants said they listened to music by radio and three through a tape recorder. Seven reported they had the opportunity within the nursing home to listen to music they enjoyed, one reported they sometimes had the opportunity	There were no significant differences before and after the music intervention with respect of global quality of life measures. However, all of the residents reported that they would like to engage in the activity again. Residents were mixed as to whether or not they felt that others would enjoy the intervention, five said they thought others would enjoy it, seven were	GRADE score - ? - - Levels of evidence score LII-b, grade B.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
		Loss to follow-up: Three participants withdrew from the study.	which lasted for 30 minutes		Administered at pre-test and then immediately following the arts activity.	and five reported that they did not have the opportunity to listen to music they liked. When asked if they attended music activities, four responded that they did, seven said no and one said that they liked listening to music in their own room. There were no significant differences between the pre and post-test on any of the global quality of life items or combined scales. Descriptive data based on participants' enjoyment was also reported.	unsure and one reported they did not feel others would enjoy the activity	
140 Orsulic-Jeras (2000) Intervention 9-month time period, dates not specified.	USA Menorah Park Centre for Senior Living, an orthodox Jewish facility. Participants were recruited from the special care unit (SCU).	25 took part in the whole study (from 44 recruited). 23 females (92%) and two males (8%). 13 residents were assigned to the treatment condition and 12 to the control, matched on their MMSE scores and performance on other cognitive measures. Loss to follow-up: 44 participants from the living facility were initially recruited to the study, the duration of which was 9 months. 19 of the participants dropped out of the study due to death (n=3), transfer to another unit within the facility (n=12), excessive absence from the sessions (n=4).	Montessori group activities Participants took part in the programme twice a week and involved group, individual and QAR sessions. Group programming sessions lasted 25-45 minutes; individual activities 10-30 minutes and QAR (structured reading and discussion) lasted 30-60 minutes. Those in the control group took part in other activities including hand massage, aromatherapy and tai chi. It is not reported who carried out the sessions.	To explore the role of Montessori-based activities in improving engagement levels and reducing the frequency of problem behaviours in residents of a special care unit located within a senior living facility.	Outcome measures were carried out at pre-test and then at follow-up nine months later at the end of the programme. The Multidimensional Observation Scale for Elderly Subjects (MOSES) was used to assess functional ability, the CMAI was used for agitation and the Cornell Scale for Depression (CSD) for depressive symptoms. MMSE scores were also obtained at pre- and post-test.	**MMSE and MOSES**: both showed significant effects over time for those taking part in the intervention (F[1, 23]=5.3, p<.03) and (F [1, 23]= 6.0, p<.02) respectively. Initial MMSE scores showed participants to be in the moderate to advanced stages of dementia and MOSES showed moderate ability levels, both showed a steady decline across the study for the sample. No other effects reached significance for any other measure. Pre-test scores for both agitation and depression were low and these remained low for the study. **Engagement***:* Non-engagement and self-engagement were rare and therefore not statistically analysed. ***Constructive engagement****:* Significant effects were found for group (F (1, 23)= 133, p<0.001) and time (F (2, 46)= 20.9, p<0.001). Group-time interaction was also significant (F[2, 46]= 23.4, p<.001). One-tailed independent t-tests were found groups did not significantly differ at baseline (p>.05) but they did differ significantly at post-test 1 (p<.001) and post-test 2 (p<.001*). **Passive engagement* **significant** main effects were found for group (F [1, 23]= 29.1, p<.001) and time (F [2, 46]=3.6, p<.04. group-time interaction was not significant (F [2, 46]= 1.4, p<.27). One-tailed independent t-test carried out found the two groups were significantly different from each other at baseline (p<.02) and at post-test 1 (p<.001) and post-test 2 (p<.001). For the control group there was no significant change in passive engagement levels across time (F [1, 12], p<.1) whereas the treatment group did show a significant change across time (F [1, 12]= 8.1, p<.002). A 2 x 2 repeated measures ANOVA was used to analyse passive and engagement measures where the within-subjects factors were "type of programming" (treatment vs. control) and "time" (post-test 1 vs. post-test 2). Significant types of programming effects were found for both constructive engagement (F [1, 12]= 126, p<.001) and passive engagement (F [1, 12]= 18.8, p<.001). **Affect:** anger, sadness and anxiety were observed infrequently and therefore not statistically analysed**.** ***Pleasure****:* significant main effects were found for group (F[1, 23]= 26.7, p<.001) and time (F [2, 46],= 7.5, p<.002). Group-time interaction was significant (F [2, 46]= 5.7, p<.006). One-tailed independent t-tests showed the two groups were not significantly different to each other at baseline (p<.05) but were significantly different at both post-test 1 (p<.001) and post-test 2 (p<.001). ***Passive engagement****:* significant main effects were found for group (F [1, 23]= 29.1, p<.001) and time (F [2, 46]=3.6, p<.04. group-time interaction was not significant (F [2, 46]= 1.4, p<.27). one-tailed independent t-test found the two groups were significantly different from each other at baseline (p<.02) and at post-test 1 (p<.001) and post-test 2 (p<.001). For the control group there was no significant change in passive engagement levels across time (F [1, 12], p<.1) whereas the treatment group did show a significant change across time (F [1, 12]= 8.1, p<.002).	This study showed significant improvements in constructive engagement and decreased incidences of passive engagement in those taking part in Montessori-based activities. There was also a significant effect over time, indicating improvements along the duration of the programme. Comparison at post-follow up also found that those who were taking part in both usual activities and Montessori-activities showed increased engagement in the Montessori activities only. For affect, only pleasure was analysed and this was shown to be greater for the Montessori-group than the usual activities group. However, pleasure scores significantly decreased with time for those in the Montessori group from immediately following the programme to follow-up 2.	GRADE score - ? - - Levels of evidence score LII-a, grade B.
144 Koike (2012) Intervention Dates not specified	Japan Details not specified	15, 10 females (67%) and 5 males (33%). Loss to follow-up: not reported.	Vibroacoustic therapy Delivered in the afternoon between 15.30 and 16.00 using classical music.	To determine whether vibroacoustic therapy would improve the levels of psychological symptoms in individuals with depressive symptoms.	Dementia Mood Assessment Scale (DMAS) for affect changes. Dementia Behavioural Disturbance measure. MMSE scale for cognition changes Physiological changes: temperature, Sp02, blood pressure and pulse rate. Sleeping hours using a wristwatch type actigraphy.	***DMAS:*** significant changes from pre to post intervention for total scores obtained, mean pre 49.66 (± 16.17) and mean post 43.8(± 17.93), p< 0.05. Significant changes for depression and sadness mean pre 36.66 (± 11.13) and post mean 32.93 (± 14.05). No significant MMSE, DBD or dementia severity changes (all p>0.05). ***Physiological changes****:* significant change in temperature mean, pre 36.46 vs. post 36.31, p<0.001 and pulse rate, mean pre 72.28 (± 9.88) and post 70.66 (± 9.44), p<0.001. Sp02 and blood pressure not significant. ***Sleep time****:* total sleep in the second week was significantly decrease, compared with the first (626.54 vs. 559.61 mins, p<0.05). Wake time was significantly increased in the second week compared with the first (813.45 mins vs. 838.06 mins, p<0.05). Night time sleep and night time sleep efficiency were not significantly different.	Following taking part in Vibroacoustic therapy, participant with pre-existing depression showed improved overall mood and depression and sadness scores. They also showed significantly lessened sleeping time and an increase in wakefulness, slightly increased temperature a lower pulse rate. There were no changes in cognition, behavioural disturbances, dementia severity or any other physiological measures.	GRADE score ? ? + ? Levels of evidence score LII-a, grade B.
146 Hicks-Moore (2005) Intervention Dates not specified	Canada A Specialised Care Unit (SCU) within a 120 bed nursing	30 in total. Mean age 82.4, range 70-101. 21 female (70%) and 9 male (30%).	Music- relaxing music during meal times. Baseline Week 1 (days 1-7) the music was played during week 2 (8-14), a	To explore the impact of a relaxing music intervention on agitation levels in older people who had dementia and	The CMAI- recorded the presence or absence of behaviours during the meal period. Baseline data were obtained during baseline (week 1) and	Total agitated behaviours had a potential range of 0-203 (29 possible behaviours x 7 days). The actual range of behaviours observed for week 1 (no music) was 0-69 and there was a mean incidence of 9.85. For week 2 (music) the range of behaviours was 0-32 and the mean incidence was 4.57. During week 3 (no music again) the mean incidence of total agitated behaviours had increased to 7.29 with a range of 0-51 and in week 4 (music) had decreased again to a mean of 3.34 with a range of 0-24. No statistical analysis carried out using these means.	This study used a modified version of the CMAI inventory to assess the impact of music during mealtimes on agitated behaviours in a group of care home residents with dementia. The results section is largely descriptive, but does show that there were more agitated	GRADE score - ? + - Levels of evidence score LII-a, grade B.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
	home in urban South Eastern Canada. The unit has 34 beds for residents with severe cognitive impairment.	Loss to follow-up: not reported.	control period during week 3 (15-21) and week 4 (22-28) music was played again.	resided within a nursing home.	at throughout each of the weeks, with a mean per week (i.e. week 2 score, week 3 score etc).		behaviours shown for the participants during the weeks when music was playing than during the weeks when music was not playing.	
153 La Cour (2005) Intervention Qualitative	Sweden A hospital and nursing home which specialises in rehabilitation, palliative and geriatric care	8 residents from the hospital and nursing home and 7 occupational therapy staff members. Demographic data not provided. Loss to follow-up: not reported.	Creative activity The creative activities that participants engaged in included woodwork, pottery, silk painting, soap making, knitting and gardening. Participants either took part individually or within a group. Sessions were delivered by occupational therapy staff members.	To explore the components of engagement with creative activity in the form of occupational therapy for older people who were dealing with a life-threatening illness rom perspectives of participants taking part and those delivering sessions.	The interviews with the older people were carried out either in their own rooms or the occupational therapy department (e.g. in the kitchen area). All of the therapists' interviews were carried out in the occupational therapy department. Each of the interviews were between 30 and 45 minutes in length and were audio tape-recorded and transcribed verbatim.	The core category identified from the analysis was "creating connections to life". Three subcategories were also identified with separate characteristics of engaging in creative activity. The first was "generous receptive environment" which showed the sociocultural setting as necessary for engaging in creative activity. The second was "unfolding creations- an evolving process" which relates to the practical making and process involved with the end products of the creative activity. The third category is "reaching beyond for possible meaning horizons" which related to meaning beyond the current conditions of the body and engagement in the activity.	Taking part in a creative activity allowed participants to become connected to their environment. The sessions allowed those taking part to relive memories and experiences and reaffirmed their sense of self, rather than being seen as 'clients' or 'patients'. Furthermore, participants expressed their joy at taking part in the activities. However, also noted was the potential for the creative activities to remind participants of their limitations.	
162 Klages (2011) Intervention Dates not specified.	Canada An urban not-for-profit residential home located in Ontario Canada, with over 390 residents.	24 in total. 9 in the intervention group and 10 in the control group. Intervention group: mean age was 84 years (± 6.6). 7 (78%) were female and two (22%) were male. Control group: mean age was 89 years (± 3.2). 6 females (60%) and four males (40%). Loss to follow-up: Total of 5 participants. Two of the participants from the intervention group withdrew from the study due to lack of interest in the Snoezelen room. Two participants in the control group and another one in the intervention group had incomplete data and were excluded from the analysis.	Snozelean Relaxed Snozelen sessions to place twice a week for a total of six weeks, with a two day period separating weeks. Preferences of the residents were taken into account and activities included those to stimulate tactile, visual and proprioceptive sensations. Sessions were delivered by the lead author.	To assess the role of Snozelean rooms on the incidence of falls and improvements in balance in older people with dementia living within a care home.	Detailed journals were completed by the facilitators and these recorded balance-enhancing effects such as body posture changes. Positive and negative emotional reactions were also recorded. Four tests were administered to record pre and post intervention balance- the Functional Reach Test, Sharpened Romberg and the Timed Up and Go test. Patient notes were also examined for incidences of falls across the study period.	The split=plot MANOVA showed no significant effects of the Snozelen room on balance of participants. Multivariate effect of time was negative from pre to post test (F (4,14)= 1.13, p=0.38) and group was also negative (F (4, 14)= 0.92, p=0.48). Both groups showed small balance improvements over time, however not significantly so. Secondary analysis (incidence of falls) recorded 16 falls in the intervention group compared to 44 in the control group over an 18 week period, however an outlier in the control group was responsible for the majority of these falls (21 out of 44) and these were removed. The intervention group had 5 prior, 7 during and 4 at follow-up and the control had 8 prior, 8 during and 7 at follow up, these changes were not significant at pre, during or post (p=0.29, p=0.47, p=0.47 respectively). The different aspects of the snozelen room and their effects on balance are also discussed.	This study looked at incidence of falls in older people who took part in Snoezelen activities and those in a control group. The Snoezelen comprised of activities to improve eye and head movement, vibration and creative activities such as listening to music. Functional balance tests showed that there were no significant changes in balance and falls rated in older people who took part in the Snoezelen activities. Both groups actually showed an improvement in balance over time (none were significant, though). There were also no significant differences between the groups. Furthermore there was no reduction in falls for those taking part in the intervention either during or at follow up. The groups were also not statistically different with respect to falls incidences at any of the three time points (before, during or after the intervention). These were unstructured sessions and as such each of the participants experienced the Snoezelen in a different way and took part in different activities dependent upon their interests.	GRADE score + + - - + - Levels of evidence score LI, grade A.
164 Hong (2011) Intervention	South Korea Not specified,	30 in total. Mean age of the group was 78.3 years (± 6.3).	Creative song writing (3 stages)	To explore the effects of a song-writing intervention	Korean Mini Mental State Examination (MMSE-K) scores.	**Changes over time**: Overall mean MMSE-K scores rose for the intervention group by a percentage of 22.16%, compared with a decline of 0.06% recorded for the control group. MMSE-K percentage changes for the intervention group were 22.16% for orientation, 15.78% for	A significant increase in MMSE-K scores was shown following participation in a song-writing programme for older people with dementia. 3	GRADE score + + ? ? –

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
Dates not specified.	just referred to as 'nursing home'.	28 (93.3%) of the participants were female and 2 (6.7%) were male. Loss to follow-up: not reported.	Preparing songwriting, songwriting and then reinforcing the songwriting. Sessions were carried out once a week for a 16 week period and lasted for an hour. Those in the control group were given free time during the sessions. Not reported who carried out the arts activity.	on cognitive functioning of older people with dementia living within a nursing home.	Baseline scores were obtained two weeks prior to the programme and the final assessment was carried out one week following the programme. Each participant was rated by a researcher and a social worker that was employed by the care home, with average scores being calculated.	memory, 206.69% for attention and calculation, 40.35% for language functions and 37.93% for comprehension and judgement. While large relative changes were observed for the categories of 'memory' and 'comprehension and judgement' their magnitude meant that their overall MMSE-K score contribution was low. **MMSE-K scores** were significantly increased from 14.6 to 18.40 for the intervention group following the programme ($p<0.001$). The control group showed a mean total MMSE-K score of 15.00 prior to the intervention period and this decreased significantly to 14.13 at the end ($p<0.001$). Significant changes were observed for the 'orientation' scale (from 5.73 to 7.00, $p<0.001$), 'memory' scale (from 5.07 to 5.87, $p<0.02$) and 'language' scale (from 2.80 to 3.93, $p<0.01$). Significance was set at the $p<0.01$ level meaning improvements in 'comprehension and judgement' ($p=0.02$) and 'attention and calculation ($p=0.04$) were not significant. None of the subscales for the control group reported significant improvements at follow-up. Between groups: Mean-MMSE scores were not different at baseline. Following the intervention period, there were significant differences between the groups for total MMSE-K scores ($p=0.001$), 'memory' ($p<0.001$) and 'language function' ($p=0.003$). Orientation scores were close to significance ($p=0.013$). 'Attention and calculation and 'comprehension' subscales did not show significant improvements ($p=0.499$ and $p=0.164$ respectively).	sub-items were identified as contributing most to the increase: language functions, orientation and memory. In contrast, a control group who did not take part in the intervention showed a significant decrease in overall MMSE-K scores over time and no significant changes for any of the subscales.	Levels of evidence score LI, grade A.
168 Clarkson (2007) Intervention Dates not specified.	Canada 'Urban nursing home'	17 in total, 8 in the moderate dementia subgroup and 9 in the severe dementia subgroup. Moderate dementia group- mean age 86 (± 8.9), 7 (88%) female and 1 (22%) male. Severe dementia group- mean age 86 (± 7.6), all female (100%). Loss to follow-up: Completion was defined as attending at least two of the five concerts that took place during the intervention period. Two of the participants did not meet this criteria and three participants died before the music phase. Therefore 17 completed the study.	Music- live performance The live music concerts were held 5 days a week for a total of three weeks, taking place in the hour after lunch. Music for the intervention was selected based on 'appropriateness for the older adults' generation and all participants were encouraged to join in with the singing. A volunteer group carried out the music concerts.	The overall aim was to determine the effectiveness of a live music intervention carried out by volunteers, on agitation and depression levels in older people with dementia living within a nursing home. Secondary to this to determine whether there were any differences in presence of agitation depending on the stage of dementia.	Cohen-Mansfield Agitation Inventory (CMAI) for agitation, scored by two nurses. P.R.N sedative use to assess any changes in need of anti-anxiety medications, obtained from information contained in patient charts. Cornell Scale for Depression, scored by the nurse who delivered the CMAI. Actigraphy (rest/movement cycles) measures physical agitation in a random sample (n=4) participants using an actiwatch. These were recorded weekly for the study period duration.	***Agitation:*** total number of concerts attended ranged from 8-15, low levels of agitation were observed. For the first 2, only 2 and 3 participants showed agitation (from an average attendance of 16.6). only 2 participants showed agitation during the third weeks' session. Most common behaviours were nonaggressive vocalisations and mannerisms. ***Physical Aggression:*** significant effect of phase ($F[2, 30]=5.77$, $p<.001$) and a marginally significant effect of phase ($F[2, 30]= 3.20$, $p<.055$). One-way ANOVAs comparing the three study phases were significant for the group with moderate dementia only ($F[2, 14]=5.02$, $p<0.23$). T-tests showed a near-significant decrease in physical aggression in the music phase compared with baseline (baseline vs. music mean [SD]=14.0 [4.3] vs. 11.7 [1.5], $t[7]=2.23$, $p<.061$) for the group with moderate dementia and similar results were shown for the post-intervention phase (post-intervention mean [SD]=11.4 [1.1[, $t[7]=2.27$, $p<0.58$. ***Verbal aggression*:** The two-by three ANOVA revealed a significant interaction effect of phase-by-group ($F[2, 30]=3.56$, $p<0.4$). Despite this, one-way ANOVAs did not show any significant changes across the study period for either of the groups. ***Verbal non-aggression*:** two-by-three ANOVA showed a significant effect of study phase ($F[2,30]=4.28$, $p<.023$) and a non-significant interaction of phase-by-group ($F[2, 30]=1.06$, $p>.05$). Follow up t-tests (collapsing the severity groups, n=17) examining the phase effect showed a significant rise in verbal nonaggression during the music phase (baseline vs. music mean [SD]=10.7 [5.4] vs.13.0 [5.5], $t[17]=2.74$, $p<0.015$) which lasted during the post intervention phase (baseline vs. post intervention mean [SD]= 10.7 [5.4] vs. 12.4 [6.7], $t[17]=2.93$, $p<.01$.	This study examined the effects of a live music intervention on individuals with dementia who live in a nursing home. The study separated participants for analysis based on the severity of their dementia and assessed differences in agitation scores in the two groups. Verbal aggression scores did not differ significantly for either of the groups of participants during the intervention or following. Both groups showed a rise in verbally nonaggressive agitation during the music phase, but not following the intervention. Anecdotally it was noted by staff that this rise in apparent verbal agitation may have been a reflection of positive, rather than negative verbalisations such as participants singing to themselves. For the group with moderate dementia, there was a reduction in physically nonaggressive behaviours both during and following the intervention, however these only reached statistical significance during follow up. There were no changes in depression scores throughout the intervention.	GRADE score - ? - + Levels of evidence score LII-b.
169 Cohen-Mansfield (1997) Intervention Dates not specified.	Israel 'Nursing home'.	32 in total. 6 (18.8% male) and 26 (81.3%) female. Mean age of the participants was 86.8 (± 1.16), range not reported. Loss to follow-up: Initially 60 participants, seven of whom died before data collection. Of the 53 participants	Music (with three other interventions). Medical examination- to ascertain physical causes of behaviour. Exposure the music- half hour audio tapes based on preferred music. Use of family-generated videotape. One-to-one social interaction.	To assess and compare the effectiveness of four different interventions on incidences of verbally-disruptive behaviours in older people with dementia. The four interventions were: medical examination in order to remove	Tape recordings: carried out for an hour a day, 15 minutes prior to the intervention, 30 minutes during and 15 minutes following. Coded by trained research assistants. Standardised observations: using the Screaming Behavioural Mapping Instrument (SBMI). Informant ratings: staff members completed the CMAI for participants by their	***Medical Examination:*** for 23 participants (78%) there was no physiological reasons identified for their VDB following assessment by a physician. 7 were found to be in some form of pain (from mild (n= 2) to discomfort (n= 4) to distressing (n= 1). Physicians also noted that despite the pain in these participants this was not the source of their disruptive behaviours. ***Audio-tape data:*** the effect of time (before and during the intervention) was significant ($F (1, 23)= 137.1$, $p <0.01$) as well as the interaction between time and type of intervention ($F (3, 124)= 7.7$, $p <0.01$). The effect of intervention was significant ($F (3)= 45.5$, $p <0.001$) and all the interventions were significantly better than no treatment. Behaviours decreased by 56% during the social interaction, 46% during the videotape, 31% during the music and 16% during the no-intervention condition. ***Observational data:*** A repeated measures ANOVA was carried out to assess changes for different types of VDB. Shouting, complaining and/or inappropriate verbal and other VDB all showed significant differences between nothing and social interaction ($p<0.01$), nothing and videotapes	This study looked at the difference in benefit for three different interventions: social interaction, simulated family presence and music for individuals with dementia and associated Verbally Disruptive Behaviours (VDB). All of the interventions showed a significant reduction in verbally disruptive behaviours compared with no intervention, the two interventions which promoted social interaction showed the greatest effect (face to face and simulated family presence).	GRADE score - - - + Levels of evidence score LII-b, grade B.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
		that were left, five died before completing the interventions, 11 'became quiet', 3 refused to continue their participation and 2 were physically restrained and excluded (based on evidence that VDB increases when restraint is used).		any physiological explanations for disruptive behaviours, music, stimulated family presence using a videotape and use of one-to-one social interaction.	care givers following the intervention.	(p<0.01) and nothing and music (p<0.01) (shouting: F (3, 318)= 11.8, p<0.05, complaining: F (3, 345)= 11.5, p<0.05 and social interaction F (3, 183= 6.6, p<0.0). the repeating words measure showed significant decreases between nothing and social interaction (p<0.01), nothing and social interaction and videotape (p<0.01) and nothing and music (p<0.01) (F (3, 294)= 18.8, p<0.01. requests for attention showed a significant difference between social interaction, videotape and music (p<0.01) (F (3, 90)= 29.3, p<0.05. Hallucinations showed a significant difference between nothing and videotape (p<0.01) and nothing and music (p<0.01) (F (3, 39)= 3.6, p<0.01). ***Nurses assessments**:* duration of VDB decreased 'considerably' during the first three minutes of each intervention (apart from the no treatment) then remained stable for the half hour intervention before declining post intervention. Appropriate verbal behaviours increased significantly during the one-to-one social interaction and videotapes (F (6, 248)= 25.57, p<0.01).	The effect of all three interventions was short lived. Reductions in VDB were only observed during the intervention period and the nurses reported no reductions in VDB long-term.	
173 Philips (2010) Intervention December 2008- July 2009.	USA Four nursing homes and two assisted living facilities within central Missouri.	Intervention group (n= 28). Mean age 83.4 years (± 8.0). 23 were female (82%) and 5 were male (18%). Control group (n= 28). Mean age 85.8 (± 6.8). 26 (93%) were female and 2 (7%) were male.	Creative storytelling TimeSlips- an established group story telling programme. This was delivered to the intervention group for a total of six weeks, each session lasted for one hour and group size varied between 6 and 12 participants. Sessions were delivered by the principle investigator and a research nurse, both of whom were trained in delivering the TimeSlips programme.	To assess the role of TimeSlips in relation to communication skills, quality of life, neuropsychiatric symptoms and observed emotions in participants with dementia living within a long-term care facility. It was suggested that TimeSlips would improve cognition and psychological wellbeing.	Observation to record emotion at week 0 (baseline), 1 week post-intervention (Week 7) and 4-weeks post intervention (week 10) for both groups and weeks 3 and 6 for the intervention group. This was measured using the Observed Emotion Rating Scale (OERS). Modified Cumulative Illness Rating Scale. Mini Mental State Examination (MMSE). Cornell Scale for Depression in Dementia (CSDD). Neuropsychiatric Inventory-Nursing Home Version (NPI-NH) Quality of Life in Alzheimer's Disease (QOL-AD) Functional Assessment of Communication Skills (FACS)	A significant group X time interaction was observed for the measures of NPI-NH, social communication and OERS (p<0.05, p<0.001 and p<0.001 respectively). After adjusting for baseline and MMSE scores, OERS pleasure scores were significantly higher at week 3 (p<0.001), week 6 (p<0.001) and weeks 7 (p<0.05) for the intervention group than the control group. Social communication scores were higher for the intervention group at week 7, however not significantly so (p not reported). Week 10 scores were better for the control group compared with the intervention group, but again not significantly so. There was no significant effect for the CSDD scores. Following adjustment for baseline and CIRS scores, however these were significantly lower for the control group at week 10.		GRADE score ? ? + + Levels of evidence score LII-a, grade B.
180 Martin (2004) Descriptive	Canada A not-for-profit long term care facility which consisted of four 15 bed houses and a adult day programme with 54 clients.	63 in the nursing home, with most taking part at least once. Mean age 84, range 60-102. 46 (73%) were female and 17 (27%) were male. 54 took part from the day programme. Mean age 78, range 54-92. Loss to follow-up: Not reported.	Music- drum circles The sessions were carried out on average twice a week and the observation time was a period of 5 months. Participants were free to join or leave the drum circle at any time. Any of the instruments chosen by the participants would be acceptable and residents were able to switch from one instrument to another.	To assess the feasibility of a drum circle intervention for older people who lived within a care home and had functional cognitive abilities. Questions focused on the observed experiences of the participants and facilitators.	Ethnographic field notes; both participant and non-participant observations, real time interviews focus groups and interviews with facilitators. 2 facilitators recorded notes in 3 of the 20 sessions.	Group comparison showed nursing home participants were more likely to be female, older and severely regressed (if they had dementia)- statistical values not given. Length of sessions ranged from 21-52 minutes (mean 37) for the nursing home group and 25-46 minutes (32 minutes mean) for the day programme. Day programme group numbers ranged from 12-17 (mean= 15) and nursing home groups from 8-17 participants (mean= 10). Emergent thematic behaviours: empowerment, bringing out and bringing together. These behaviours magnified to create energy and community resonance.	This study examined the feasibility of carrying out a drum circle programme with residents from a nursing home and older adults who attended a day-programme. The study showed that taking part promoted inclusion, encourage empowerment, and togetherness. Also found was the way in which the programme adapted with input from residents and evolved across the duration of the programme.	Risk of bias score: Grade C, some flaws that may affect the credibility, transferability, dependability and/or confirmability of the study.
185 Hilliard (2004) Descriptive	USA 15 different nursing homes, 7 in	Total of 80; 40 females (50%) and 40 males (50%).	Music- therapy Cognitive-Behavioural music therapy carried out	To explore the use of music therapy for people receiving hospice care in a	Patient notes were utilised from individuals who had died. Information on length of life and amount of time spent	Mean number of days within the hospice (length of life) was significantly greater for the music therapy group (141.28 (± 129.5), than the group not receiving music therapy (24.28, ± 44.44), t-5.867, df=78, p<0.001.	This study retrospectively examined the care notes of individuals living within a nursing home for terminally ill residents who either did or did not take part in a music therapy	GRADE score ? ? ? ? - + Levels of evidence score LI, grade A.

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
Dates not specified.	Tallahassee and 8 in rural areas of nearby counties. Each offered services for the terminally ill.	Intervention group- 50% male (n=20) and 50% female (n=20), mean age male 78 (range 49-95) and mean age female 77 (range 48-98). Control group 50% male (n=20) mean age 70 (range 22-89) and female mean age 82 (range 48-98). No loss to follow-up as retrospective.	within nursing homes by a certified music therapist with all-live music and sessions planned by the therapist.	nursing home environment. Questions identified were: whether patients receiving music therapy lived longer than those who didn't, whether there were gender differences in length of life on the hospice programme and did patients die closer to the time of a music therapists' visit than other hospice professionals' visits.	with the therapist were extracted.	Music therapists participated in more visits with participants than social workers (mean 10.88 vs. 6.53 respectively) t=5.447, df=39, p<0.001. Music therapists spent more minutes with the residents than the social workers (mean 399.98 vs 163.88 respectively), t=5.889, df=39, p<0.001. Care plans showed cognitive issues were most addressed by nurse (n=47), then music therapists (n=3) then social workers (n=1). Physiological needs were also most often addressed byy the nurse (n=273), then music therapist (n=8) then social workers (n=1). Emotional needs were addressed by both the social workers (n=38) and music therapists (n=39). Social needs were addressed by the social worker (n=33) then the music therapist (n=1) but not the nurses (n=0). Music therapists and social workers addressed spiritual need (n=7, n=2 respectively) but not nurses.	programme. Results showed that those who took part in the music therapy programme, lived longer than those who did not. Care plans showed different needs addressed by different professionals, with cognitive and physiological needs being addressed by the nurses and emotional needs and spiritual needs addressed by the social worker and music therapist.	
200 Roe (2015) Descriptive June-December 2012	UK Manchester Museum and Whitworth Gallery Manchester	8 participants from a supported living facility and nine from a care home (n=17 total). 4 members of staff from the museum and gallery and a supported living project worker. Comments were also provided from ab artist and activities coordinator from the care home. No demographic information provided as those taking part were seen as members of the public attending a museum.	Museum and gallery intervention. Programme schedule consisted of monthly sessions: 1. Tour of Manchester gallery, 2. West African arts, 3. Living worlds gallery, 4. West African arts, 5. Tour of vivarium and 6. Meeting artists and printmakers. Sessions alternated between the two venues (Manchester Museum weeks 1, 3, 5, Whitworth weeks 2, 4, 6). Exhibits were facilitated by gallery and museum staff members and the researchers.	To evaluate the feasibility of an arts for health program carried out within museums and galleries for older people who reside within a care home. Particularly to identify the benefits of the program in relation to the wellbeing of those taking part.	Non-participant observations were carried out along with field notes recorded by the research team. Spradley's framework of nine dimensions was used for the observation data and field notes consisting of: space, objects, acts, activity, event, time, actors, goals and feelings. Independent content analysis was carried out for each session using the Spradley's framework by two research team members. A final group staff interview was conducted with an interview schedule of 10 questions.	*Session and participants:* discusses the structure of the sessions, five of which were carried out as planned and another session which had to be rearranged. *Impact of arts for health programme and sessions:* acts and actors which discussed the structure of the sessions (acts) and those taking part (actors); space and time which evaluates the logistical aspects of carrying out the programme; goals, events, object and activity which presents how engaging with creative activity was beneficial for those taking part; feelings which discusses the positive feelings the participants felt when taking part in the study. *Benefits, impact on wellbeing, feasibility of the sessions and programme:* this category outlines the need for museum staff to be flexible in their approach to the programme. *Feasibility of the programme, sessions and the future:* the gallery and museum staff members expressed their interest in taking part in a similar programme in the future and how they felt this would be achieved.	This study allowed individuals from a care home to access programmes at a museum and gallery and discusses the challenges and benefits of doing so. Sessions were carried out mainly how they were intended to be conducted, however the need for flexibility was outlined, particularly if sessions needed rearranging due to commitments from the care staff. Also discussed was challenges faced with carrying out such a programme, including transport to and from the venue, issues related to disabilities such as toilet and lift facilities within the museums and level of staff support needed. Benefits to the participants in terms of wellbeing were also reported.	Levels of evidence score, LIII, Grade C.
201. Gerdner (2000) Intervention Dates not specified.	USA Six participating long term care facilities located within Iowa.	39 in total. 30 females (67%) and 9 males (33%). Mean age 82.6 years. Loss to follow-up: not reported.		To compare immediate and 30 minute residual effects of those exposed to 'classical relaxation music' with those exposed to individualised music	A modified version of the Cohen Mansfield Agitation Inventory (CMAI). This instrument had been modified using 10 minute increments to allow for the 60-minute observation period in order to provide definitive assessment of immediate and residual effects	14 of the 16 demographic variables were not statistically different between the two groups. Significant differences were shown for the variables of urinary incontinence and assistance needed with ambulation but these variables were not associated variables of gait or toileting and therefore the groups were regarded demographically similar. There were occasions of data missing from the MCMAI and when this was the case least square means were calculated using individual values to predict missing values. ANOVA revealed a significant two-way interaction between phase (df (2, 74), F=32.93, p=0.001) and minute (df 5, 2763, F=53.28, p=0.001) and the main effects of phase and minute was also significant (df(10, 2763), F= 4.77, p=0.001). No significant effect was found for phase and week (F= 4.77, p=.07) or phase, week and minute (F= 0.64, p=0.99). Frequency of agitated behaviours was also significantly less during each of the three 10 minute increments during which individualised music was playing. Frequency of agitated behaviours was significantly less during the 30 minutes immediately following presentation of individualised music compared with the corresponding three 10-minute intervals of baseline assessment. Finally, no significant difference occurred in frequency of agitation between baseline and first 20 minutes of classical music. A significant decrease in agitated behaviours did occur during the final	Both classical and individualised music showed lower incidences of agitation during the time when music was played than baseline. When these mean decreases were compared, there was a significant reduction in agitation scores during the individualised music intervention compared with the classical music intervention. Further to this individualised music showed a significant reduction in agitated behaviours during the study period and for the 30 minutes' follow-up period compared with baseline. For the classical music condition there was only a reduction in agitated behaviours for the final 10 minutes of the class and for the first ten minutes of the follow up period.	

Li Reference of Paper & Study ID	Location	Population	Arts Type	Aims of the study	Outcome measures	Findings	Conclusions	Quality
						10 minutes of classical music compared to baseline. This continued only during the first 10 minutes following presentation of classical music.		
202 Cox (2004) Intervention Dates not specified.	Australia Rice Village Residential Facility.	24 residents participated, 23 were female (96%) and one male (4%). Ages not reported. Six caregivers and six visitors consented to be interviewed for the qualitative study element. Loss to follow-up: not reported.	Each participant took part in each of the activities, the living room (usual care), Snozelen and Garden installations for 16 minute sessions. A total of nine 16-minute observation periods were carried out for each participant (3 of each condition?).	To explore the impact of two different types of multisensory environments: Snozelen rooms and a landscape garden on improving the wellbeing of older people and compare these interventions with the usual living room environment within the residential home.	The Affect Rating Scale (ARS) was used to carry out observations with the oparticioants. This contains six different and discrete categories of affect: pleasure, anger, anxiety or fear, sadness, interest and contentment. Immediately prior to entering the installation and at 4 minute intervals during the sessions an observed used to ARS to record participant responses	Wilcoxon signed rank tests were used to examine individual differences prior to and during the intervention. When looking at the interventions individually, we can see that pleasure was increased during all three of the interventions (living room mean 1 before, 44 following, $p < 0.001$, garden mean 0 before, 47 following, $p<0.001$ and snozelen 0 before, 38 following, $p<0.001$). interest scores increased only for the living room, mean beforehand 12, following, 31, $p<0.01$. Contentment significantly reduced for each condition (mean before living room 67, following was 18, $p<0.001$; mean garden prior 65 and following, 24, $p<0.001$ and snozelen before 56 and following, 26, $p<0.01$). Anxiety or fear decreased only for the living room condition (mean prior 14 and following, 4, $p<0.05$). Qualitative- states nine general themes, however only four are reported for each group. Caregiver group, 4 themes were identified: first impressions, something special, changing patterns, what gets in the way. Relative/visitor group identified 4 themes: changing the patterns, feeling sad about the medication, privacy and meditation, fascination.	This study implement two different installations and compared them with the usual living area within a residential facility. There were no significant differences observed for agitation levels between the different conditions for any of the measures, other than sadness which was reportedly lower for the snozelen and garden conditions during the intervention when compared with the living room condition. Individually pre and post-test There were reductions in contentment during the intervention for each condition, but improvements in pleasure. Anxiety reduced only for the usual care condition.	

Appendix 4

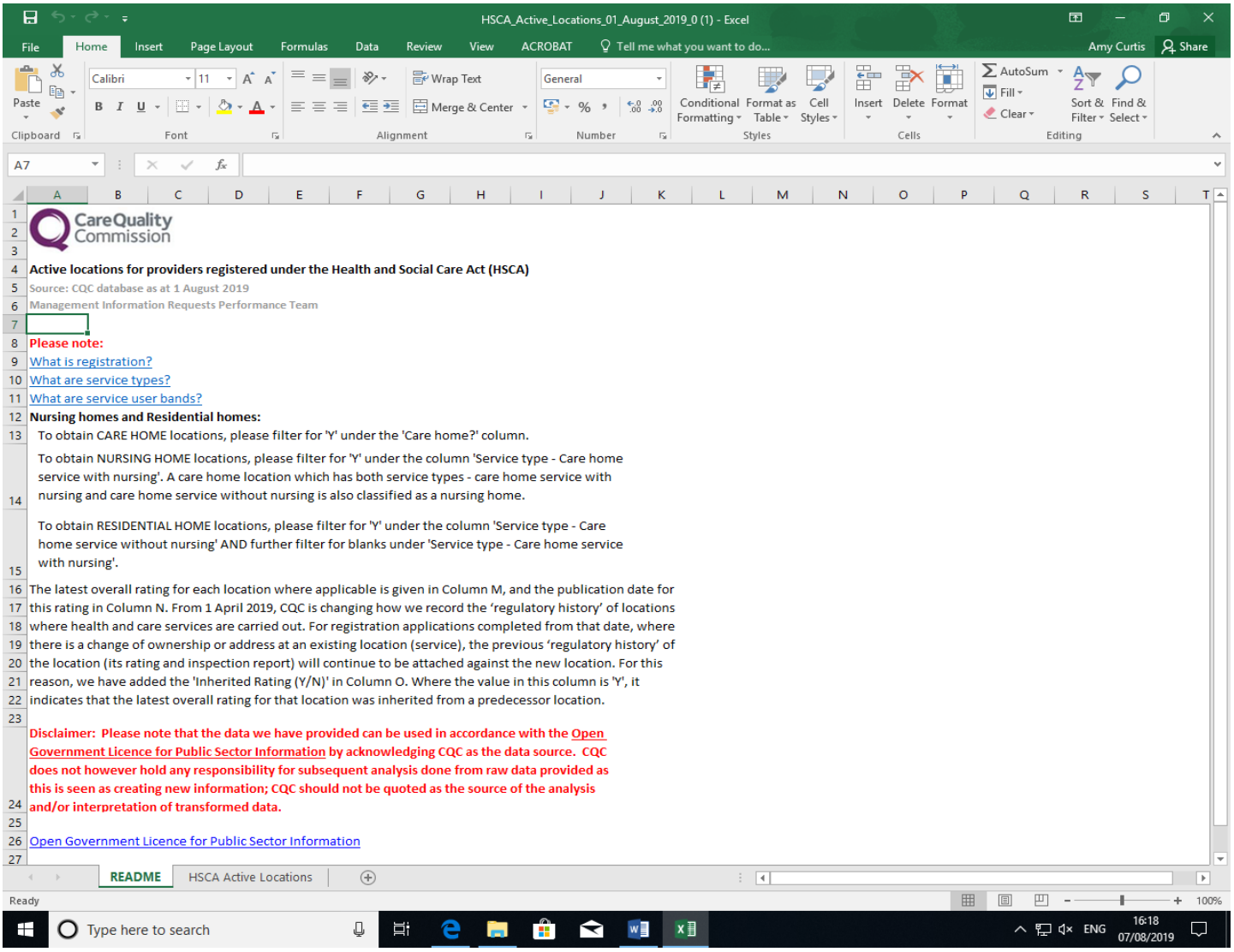
Database	Hits	Not in English	Not Relevant	Potentially relevant	Duplicate
Medline	252	34	134	84	
PsycInfo	245	10	137	44	54
CINAHL	184	7	82	27	67
AMED	70	0	50	3	17
EMBASE	340	1	166	13	160
Total	**1091**	**52**	**569**	**169**	**301**

Reasons for exclusion at title and abstract search ('not relevant')						
Reasons for non-relevance	Medline (n=134)	PsycInfo (n= 137)	CINAHL (n=82)	AMED (n=50)	EMBASE (n=166)	**Total**
Not arts for health	52	40	25	42	100	**259**
Not correct study population	10	12	3	5	10	**40**
Not empirical (systematic review, literature review or meta-analysis)	12	7	8	1	25	**53**
Not empirical (interview or article)	60	78	46	2	31	**217**
Total						**569**

Reason for exclusion			
001	Reminiscence- no life story work	080	Reminiscence- no life story work
002	Reminiscence- no life story work	081	No relevant outcome measures
003	No relevant outcome measures	083	Reminiscence- no life story work
004	Reminiscence- no creativity	087	No relevant outcome measures
005	Reminiscence- no creativity	088	Reminiscence- no life story work
006	Reminiscence- no creativity	089	No relevant outcome measures
007	No relevant outcome measures	092	No relevant outcome measures
008	No relevant outcome measures	093	Not empirical
009	No relevant outcome measures	095	Not empirical
011	Reminiscence- no life story work	097	No relevant outcome measures
012	Not empirical- case studies	100	Reminiscence- no life story work
013	No relevant outcome measures	103	No relevant outcome measures
014	No relevant outcome measures	104	No relevant outcome measures
015	Reminiscence- no life story work	105	Reminiscence- no life story work
016	No relevant outcome measures	107	No relevant outcome measures
019	No relevant outcome measures	109	No relevant outcome measures
022	No relevant outcome measures	111	Letter- no full paper
023	Reminiscence- no life story work	119	No relevant outcome measures
024	Reminiscence- no life story work	124	No relevant outcome measures
025	Not arts for health	127	Reminiscence- no life story work
026	Reminiscence- no life story work	131	Not care home residents
027	Reminiscence- no life story work	133	Full paper not available
030	Reminiscence- no life story work	134	Not care home residents
036	No relevant outcome measures	137	No relevant outcome measures
038	Not arts for health	141	No relevant outcome measures
042	Not care home residents	142	No relevant outcome measures
No relevant outcome measures		145	No relevant outcome measures
045	No relevant outcome measures	149	No relevant outcome measures
050	Not arts for health	150	Reminiscence- no life story work

051	No relevant outcome measures	154	Reminiscence- no life story work
052	Reminiscence- no life story work	155	No relevant outcome measures
060	No relevant outcome measures	158	Not care home residents
062	Not empirical- letter with no reported outcome results	159	Reminiscence- no life story work
		161	Reminiscence- no life story work
067	Not arts for health	166	Reminiscence- no life story work
072	Not arts for health	167	No relevant outcome measures
073	No study data available- protocol only	169	Reminiscence- no life story work
077	Reminiscence- no life story work	170	Not arts for health
078	Reminiscence- no life story work	171	Not arts for health
079	Reminiscence- no life story work	172	Reminiscence- no life story work

Appendix 5


HSCA_Active_Locations_01_August_2019_0 (1) - Excel
File
Home
Insert
Page Layout
Formulas
Data
Review
View
ACROBAT
Tell me what you want to do...
Amy Curtis
Share
Paste
Calibri
11
Wrap Text
Merge & Center
General
Conditional Formatting
Format as Table
Cell Styles
Insert
Delete
Format
AutoSum
Fill
Clear
Sort & Filter
Find & Select
Clipboard
Font
Alignment
Number
Styles
Cells
Editing
A7
CareQuality Commission
Active locations for providers registered under the Health and Social Care Act (HSCA)
Source: CQC database as at 1 August 2019
Management Information Requests Performance Team
Please note:
What is registration?
What are service types?
What are service user bands?
Nursing homes and Residential homes:
To obtain CARE HOME locations, please filter for 'Y' under the 'Care home?' column.
To obtain NURSING HOME locations, please filter for 'Y' under the column 'Service type - Care home service with nursing'. A care home location which has both service types - care home service with nursing and care home service without nursing is also classified as a nursing home.
To obtain RESIDENTIAL HOME locations, please filter for 'Y' under the column 'Service type - Care home service without nursing' AND further filter for blanks under 'Service type - Care home service with nursing'.
The latest overall rating for each location where applicable is given in Column M, and the publication date for this rating in Column N. From 1 April 2019, CQC is changing how we record the 'regulatory history' of locations where health and care services are carried out. For registration applications completed from that date, where there is a change of ownership or address at an existing location (service), the previous 'regulatory history' of the location (its rating and inspection report) will continue to be attached against the new location. For this reason, we have added the 'Inherited Rating (Y/N)' in Column O. Where the value in this column is 'Y', it indicates that the latest overall rating for that location was inherited from a predecessor location.
Disclaimer: Please note that the data we have provided can be used in accordance with the Open Government Licence for Public Sector Information by acknowledging CQC as the data source. CQC does not however hold any responsibility for subsequent analysis done from raw data provided as this is seen as creating new information; CQC should not be quoted as the source of the analysis and/or interpretation of transformed data.
Open Government Licence for Public Sector Information
README
HSCA Active Locations
Ready
100%
Type here to search
ENG
16:18
07/08/2019

Appendix 6

Edge Hill University

Invitation Email

Dear____

My name is Amy Curtis and I am a PhD student as part of a team from Edge Hill University exploring the role of arts for health activities for older people living in care homes. Your care home has been randomly selected from a list of all homes within England, provided by the CQC and I would be grateful if you would consider participating in this research by completing a short online survey, which it is anticipated will take around 10 minutes to complete. There will also be the option to take part in a telephone interview, however this is in addition to the online questionnaire. Further information is provided on this at the start of the survey.

If you would be willing to complete the survey please click on to the link below where you will be provided with information on the project and will be able to complete the questionnaire.

[LINK]

Thank You

Amy Curtis (PhD student, Edge Hill University- curtisa@edgehill.ac.uk)

Appendix 7

Edge Hill University

Provision of Arts for Health activities in care homes: telephone interview participant information sheet

Research team: Amy Curtis (curtisa@edgehill.ac.uk) Professor Brenda Roe, Professor Mary O'Brien and Dr Lucy Gibson.

Information

Thank you for answering the online survey and for providing your details to be contacted about participating in the telephone interview. This information sheet will explain the process of the interview and why they are being carried out. Please take the time to read this information and decide if you still wish to take part.

Why are the interviews being carried out?

The interview are being carried out as part of a PhD project exploring the potential benefits of arts activities for older people who reside in care homes. You have already participated in the online survey which was designed to provide information as to what arts activities are currently taking place in your care home and any benefits that you might have identified. The telephone interview are designed to provide further information about these activities and benefits.

Do I have to take part?

No, you do not have to take part in the interview, your participation is entirely voluntary. If you no longer wish to be contacted to take part in the interview then you can contact Amy Curtis and let her know (curtisa@edgehill.ac.uk).

What will happen if I decide to take part?

Amy will contact you at the date and time that you provided at the end of the online survey and the interview will be carried out over the phone. If this time is no longer convenient then the interview can be rearranged. If you agree, then the interview will be audio-recorded in order to make sure all of the information is remembered. You will be asked a few questions related to arts activities that are undertaken within your home. The interview will last around 20 minutes, however they may take slightly less time or slightly longer depending on the level of information that you give.

What will happen to the information I provide in the interview?

Your answers will be used along with the answers from the online survey to better understand what arts activities are currently taking place in care homes within England. With your permission, the interview will be recorded so that all the information you provide is remembered. These audio recordings will be deleted at the

end of the project. Your answers will be stored confidentially and anonymised upon write up such that no individual or individual care home could be recognised from the data. It is not anticipated that any information of a sensitive nature will be disclosed within the interview, however if any disclosures are made these will need to be discussed with the CQC and this would be fully discussed with you.

What happens if I decide that I no longer wish to take part?

If you do not wish to take part in the interview anymore then please email Amy Curtis at curtisa@edgehill.ac.uk to let the research team know. If you decide you wish to withdraw your data then you have seven days following the interview to do so, please email the address mentioned above to do this.

Will there be any benefits to taking part in the study?

The end aim of the research project is to gain better information as to what arts activities may be beneficial to older people who reside within care homes and you will be provided with a copy of the final write up.

Who has reviewed the study?

This study has been reviewed by the Faculty of Health and Social Care Research and Ethics Committee within Edge Hill University.

Who can I contact if I wish to discuss the project further?

You can email Amy Curtis at curtisa@edgehill.ac.uk if you wish to discuss any aspect of the research project.

Concerns

If you have any concerns about the research project and wish to speak to a member of the supervisory team, you can contact Brenda Roe at roeb@edgehill.ac.uk. If you wish to speak to someone external from the research team you can contact Clare Austin who is the Associate Dean for Research and Innovation in the Faculty of Health and Social Care at Edge Hill University on 01695 650 772 or austincl@edgehill.ac.uk

Thank you for reading this information sheet.
Please keep hold of it in case you wish to refer to it again.

Appendix 8

Edge Hill University

Provision of Arts for Health activities in care homes: survey

Information

The aim of the survey is to collect information on arts for health activities provided within or from care homes in England. The survey comprises two sections, the first of which will ask for some background information on your care home and the second will ask for information regarding arts activities. It is estimated that the study will take around 10 minutes to complete.

All of the responses given will be anonymous, and no care home will be identifiable. Upon completion of the project a summary report of the findings will be made available to you. The email address that you provide will be stored confidentially and a copy of the results will be made available upon completion of the project.

The survey is not asking any questions which are sensitive in nature and is not asking for personal information about residents. It is therefore not anticipated that any safeguarding issues should arise, however in the unlikely event that anything is disclosed then this will be passed on to the CQC, and you will be advised of this.

If you are happy to proceed with the survey then please click on the below to begin. If not, then you may close this page and do not have to take part. If you have any questions then I can be contacted on the email address below.

Thank you

Amy Curtis (PhD student, Edge Hill University) – curtisa@edgehill.ac.uk

Research team; Professor Mary O'Brien, Dr Lucy Gibson, Professor Brenda Roe, Faculty of Health & Social Care, Edge Hill University. See weblink: www.edgehill.ac.uk/eprc

[CONTINUE]

Background

1. What category does your care home fit into:
    - Residential care older people
    - Residential care older people with dementia care specialty
    - Nursing care older people
    - Nursing care older people with dementia care specialty
    - Dual registered (residential & nursing) care older people
    - Dual registered (residential & nursing) with Dementia Care speciality
    - Other, please state……

2. Is the care home:
    - Private
    - Public
    - Mixed private and public
    - Not-for-profit

3. How many residents in total does the care home have places for?

4. How many residents do you currently have?

5. Are any of the places for respite care?
   - Yes
   - No

6. What proportions of residents have a confirmed diagnosis of dementia?
   - Less than 25%
   - 26-50%
   - 51-75%
   - 76% and above

7. What proportions of residents have memory problems but do not have confirmed dementia?
   - Less than 25%
   - 26-50%
   - 51-75%
   - 76% and above

8. What proportion of residents would you see as having challenging behaviour?
   - Less than 25%
   - 26-50%
   - 51-75%
   - Above 76%

9. What proportion of residents would you see as having depression/low mood?
   - Less than 25%
   - 26-50%
   - 51-75%
   - Above 76%

10. What proportion of residents have the following:
    - Incontinence- Less than 25% / 26-50% / 51-75% / above 76%
    - Pressure sores- Less than 25% / 26-50% / 51-75% / above 76%
    - Risk of falls - Less than 25% / 26-50% / 51-75% / above 76%
    - Multiple medications (more than 3) - Less than 25% / 26-50% / 51-75% / above 76%

11. Please could you tell us:
    - How many staff are employed in your care home?
    - How many of these are registered nurses?
    - How many volunteers help out in your care home?

12. How often do residents have contact with the following professional groups?
    - GP- weekly / monthly / rarely / never
    - Consultant (e.g. Geriatrician or Psychiatrist)- weekly / monthly / rarely / never
    - Pharmacist- weekly / monthly / rarely / never
    - Social Worker- weekly / monthly / rarely / never

Arts activities and interventions

13. Please select the arts activities which currently take place within or from your care home and the frequency upon which they take place.

Activity	More than once a day	Once a day	Every other day	2 times a week	Weekly	Fortnightly	Monthly	Less often
Music								
Dance								
Singing								
Drama/Theatre								
Poetry								
Reading								
Life story work								
Storytelling								
Creative writing								
Painting								
Collage								
Pottery								
Sewing								
Knitting								
Gardening								
Engaging with museums or galleries (in-reach or out-reach)								
Other (please describe)								

14. Which member/s of staff facilitates or are involved with these activities? (Select as many as apply).

Activity	Care staff	Nurse	Activities co-ordinator	Volunteers	Staff not directly employed by the care home (please describe below)	Other (please describe below)
Music						
Dance						
Drama/Theatre						
Poetry						
Reading						
Life story work						
Storytelling						
Creative writing						
Painting						
Collage						
Pottery						
Sewing						
Knitting						
Gardening						
Engaging with museums or galleries (in-reach or out-reach)						
Other (please describe)						

15. Do you receive any additional funding to carry out these activities? If so, please could you provide information on this below:

16. What do you estimate the costs of running these activities?

17. Are there any benefits to residents in terms of health, wellbeing or quality of life that you would identify with the arts activity, or activities? If so what are they?

18. Are there any benefits to staff that you would identify with the arts activity or activities? If so what are they?

19. Are there any benefits to your care home in general that you would identify with the arts activity or activities? If so what are they?

Thank you for completing this online survey. Is there anything we have missed that you would also like to add?

[FREE TEXT BOX]

We would also like to conduct telephone interviews with a small sample of care home managers to provide more contextual information. If you are willing to participate in a telephone interview, please could you provide your name, contact number and a time when it would be convenient to contact you?

Name_____

Name of Care home_____

Telephone number_____

Time when it would be most convenient to call_____

Thank You for your help with this survey.

[SUBMIT]

Appendix 9

Interview Topic Guide

Thank you for agreeing to take part in the survey, my name is Amy Curtis and I am a PhD student at Edge Hill University.

The questions within this interview will follow on from the responses that you gave within the survey to help provide some more information. If you would still like to take part in the interviews I would like to audio record them so that I collect all the relevant information, these audio recordings will be destroyed once I have typed them up into transcripts.

How long the interviews take will depend upon the amount of detail provided within your responses, but are likely to take around twenty minutes. All the responses will be anonymised when typed up and no care home will be identifiable from the report. As we are not discussing sensitive information it is unlikely any disclosure would be made that would need reporting, however if this occurred then the CQC will be informed. If there are any questions that you don't wish to answer you can skip and move on to the next one. If once you have completed the interview you decide you want to withdraw your data, you can do so for up to two weeks.

Can I just check that you would still like to take part in the interview and you are happy for me to audio record?

1. Could you provide some information on what arts activities your home currently carries out? Have these activities changed at all since the survey was completed
    1.1 How often are these carried out? (Daily, weekly etc.)
    1.2 Where are they carried out? Do they take place within the home or elsewhere?

2. If no arts activities are offered/ undertaken why is this? Would you like to offer arts activities in the future? What would help with this?

3. Who decides what arts activities take place within your care home?
    3.1 Are the arts activities carried out by the same person that chooses the activities and their location?
    3.2 How are the activities chosen and where they are offered?
    3.3 Do residents have any input into choosing activities?

4. How are residents able to take part in activities?
    3.2 can they volunteer to take part?
    3.3 are any residents excluded from taking part? Or is there criteria for being able to take part?

5. How do you think the arts activities impact on the residents?
    5.1 In terms of benefits: for residents, staff members, and the care home in general.

6. Do you receive any additional funding for carrying out these activities?
    6.1 How often do you receive the funding? Where from? Does it cover the cost of organising and running the activities?

7. Are there any challenges to carrying out these activities?
    7.1 Behaviour/financial/logistical

8. How important to do you think it is to carry out these activities?

9. Is there anything else that you would like to add that you feel I may have missed?

Thank you very much for taking part in the survey and the interview.

Appendix 10

Edge Hill University

Amy Curtis

12th July 2016

Dear Amy,

Thank you for submitting your revised ethics application '*The Effectiveness of Arts for Health on the health, wellbeing and Quality of Life of Older People in Care Homes*' (Project Ref: FOSH120) to the Faculty of Health & Social Care Research Ethics Committee.

I have pleasure in informing you that I have reviewed your resubmission, and recommended that your study is granted Faculty of Health & Social Care research ethics approval, subject to the following conditions:

1. Ethical approval covers only the original study for which it is sought. If the study is extended, changed, and / or further use of samples or data is needed the Committee Administrator, Daniel Brown, must be contacted for advice as to whether additional ethical approval is required.

2. (NHS studies only) NHS Research governance processes must be adhered to. An application must be made to the HRA for approval for the research to be conducted in the NHS. All NHS R&D departments (in Trusts where data is being collected) will also need to be approached for Trust permission to proceed.

3. If the project requires HRA approval and/or NHS ethical approval, please forward evidence of the approval(s) to Daniel Brown (browdan@edgehill.ac.uk) before commencing the study

4. The Principle Investigator is responsible for ensuring that all data are stored and ultimately disposed of securely in accordance with the Data Protection Act (1998) and as detailed within the approved proposal.

5. The Principle Investigator is responsible for ensuring that an annual monitoring form and an end of study form, where appropriate, is sent to the Committee Administrator (browdan@edgehill.ac.uk). The form will be sent to you at the appropriate time by the Committee Administrator.

The study documentation that has been reviewed and approved is detailed below:

<doc title>	**<version no & date>**
Arts for Health invitation email	**V2, May 2016**
Proposal	**V2, March 2016**
Survey – provision of arts for health in care homes	**V2, May 2016**
Telephone interview information sheet	**V2, June 2016**

p.p.

On Behalf of Dr Lucy Bray
Acting Chair of Faculty of Health & Social Care Research Ethics Committee Edge Hill University
St Helens Road
Ormskirk
Lancashire
L39 4QP
brayl@edgehill.ac.uk

Appendix 11

Edge Hill University

Amy Curtis

13th April 2017

Dear Amy,

Thank you for submitting your revised ethics documentation for '*The Effectiveness of Arts for Health on the health, wellbeing and Quality of Life of Older People in Care Homes*' (FOSH120) to the Faculty of Health & Social Care Research Ethics Committee.

I have pleasure in informing you that I have reviewed your amended documents and approved the changes made to your research in the submitted documents.

The amended documentation that has been reviewed and approved is detailed below:

<doc title>	**<version no & date>**
Arts for Health invitation email	**V2, May 2016**
Proposal	**V2, March 2016**
Survey – provision of arts for health in care homes	**V2, May 2016**
Telephone interview information sheet	**V2, June 2016**
Interview Topic Guide	**V1, March 2017**

Yours sincerely

Dr Lucy Bray

Acting Chair of Faculty of Health & Social Care Research Ethics Committee Edge Hill University
St Helens
Road Ormskirk
Lancashire
L39 4QP
brayl@edgehill.ac.uk