Innovative approaches to methodological challenges facing ageing cohort studies: Policy briefing note

Tarani Chandola and Susan O'Shea

CCSR and Social Statistics, University of Manchester

Summary

Summary

Ageing cohort studies around the world face common methodological challenges of data collection, measurement and analysis, which become increasingly problematic as participants grow older. While these challenges are common to all longitudinal studies, ageing cohort studies in particular highlight complex methodological issues due to the nature of the population. The National Centre for Research Methods (NCRM) funded a series of workshops that brought together experts and researchers in longitudinal and ageing cohort studies to discuss some of these methodological challenges. The series was divided into workshops around the challenges of data collection, measurement and analysis in ageing cohort studies. This report summarises some of the work underlying the workshops and highlights some of the innovative solutions researchers have adopted to overcome these challenges. These include using mixed modes of data collection to deal with respondent burden, using the lifegrid history method to deal with recall bias for proxy respondents, using auxiliary variables to adjust for missing not at random mechanisms, and using a range of missing data analyses methods and simulation studies to assess the performance of a variety of different missing data mechanisms.

1. Introduction

Ageing cohort studies around the world face common methodological challenges of data collection, measurement and analysis, which become increasingly problematic as participants grow older. While these challenges are common to all longitudinal studies, ageing cohort studies in particular highlight complex methodological issues, for example, do we need different measures of health and functioning for studies of older people? Longitudinal studies use the same measures of health over and over again in order to be able to observe changes. But until recently we only did this between childhood and middle age. New measurement instruments may be needed to measure change in dimensions of participants' lives that become more important in older age, and that may not be captured well by instruments validated in younger populations. However, when a new measure is used in a longitudinal survey it breaks the continuity between past and future sweeps. What is needed is the right combination of continued repeat measures with newer ones that are more relevant in the later years.

Longitudinal cohort studies of ageing are a subset of longitudinal studies in which: (1) The sampling population an "older" person, which is usually someone above a specified age (usually 50 years and over); (2) the unit of analysis is typically the older person and their partner, and so has an emphasis on the household; (3) the study duration is the whole life course and so they are usually not restricted to just two waves of data; (4) there is an emphasis on collecting repeated measures and an accumulation of life history data focus on measures of relevance to older population.

A number of large longitudinal studies of ageing have been developed in recent years in many countries around the world. The U.S. Health and Retirement Study (HRS) was launched in 1992, and this study has provided a template for studies such as the English Longitudinal Study of Ageing (ELSA), the Survey of Health, Ageing and Retirement in Europe (SHARE), and many other more recent ageing cohort studies in Asia. Alongside these new ageing studies, which share a comparable template, some of the older birth cohorts such as the MRC 1946 Birth Cohort Study and the Whitehall II civil servants cohort are maturing with surviving respondents entering early old age. Furthermore, ongoing long running longitudinal studies such as the British Household Panel Survey and its successor, the Understanding Society Panel have always included older respondents who continue to be followed up and surveyed each year. All of these studies represent the different kinds of ageing cohort studies and although they have different origins, the share common methodological problems related to data collection and analysis.

Ageing cohort studies are inherently interdisciplinary, incorporating a range of biomedical and social science disciplines. As study members age and become increasingly frail, they may not be able to participate due to problems with mobility, problems with sight/hearing and dementia. This may require changes in the mode of data collection as well as the use of information from proxy respondents. While asking certain key questions from proxy informants is a standard method, asking relatives or carers for complex information on which health problems have occurred, when each one began and for how long it went on may be subject to considerable recall biases. Attrition and non-response are common problems for most longitudinal studies. Standard methodological approaches tend to assume there are no major differences between those who stay in the study and answer all the questions and those who do not (known as the "Missing At Random" assumption). But this assumption is hard to

justify in an ageing cohort. There are now more complex analysis strategies ("Not Missing At Random") which may become increasingly important as sample members leave ageing cohort studies.

A strategic review of panel and cohort studies (Martin et al. 2006) emphasised the importance of longitudinal studies to address key issues in the Economic and Social Research Council's (ESRC) Strategic Plan which includes research on ageing populations. Methodological challenges become more complex as participants in longitudinal studies become increasingly elderly. Studies need to address new research questions related to ageing while maintaining their existing instruments to measure change and adopt new modes of data collection that are appropriate for older populations. The methodological challenges raised by ageing cohort studies also reflect many of the key topics identified for UK research methods: these include longitudinal methods; interdisciplinary and mixed methods; data linkages to new digital data; survey methods; research ethics; and comparative research.

The National Centre for Research Methods (NCRM) funded a series of workshops that brought together experts and researchers in longitudinal and ageing cohort studies to discuss some of these methodological challenges. The key topics, innovations and recommendations are highlighted in this article. (http://www.methods.manchester.ac.uk/ageingcohort).

The article begins by highlighting trends in population ageing studies and taking a brief look at some key UK longitudinal studies. We then look at the following the key issues: non-response and attrition; selection and survivor bias; missing data; measuring cognitive change; mixed modes data collection; measuring and monitoring general health status; and psychometric considerations including questions around validity and ceiling and floor effects. Finally, recommendations are made for further developments in the area of ageing cohort studies resulting from the NCRM workshop series.

1.1 Trends in population ageing

In recent years there has been a rise in interest by policy makers and researchers in understanding the trends in population ageing worldwide due to increased longevity. The increase in people living longer has complex social and economic costs and benefits that require rigorous scientific understanding and these impacts, according to Howse (2012),

remain somewhat unclear. Howse suggests that the work of health care professionals in the future is likely to focus on strategies for the prevention and management of chronic disease in older age. Tucker et al. (2011) highlighted a trend towards emphasising prevention, enablement and personalisation of services for older people accessing community equipment, adaptations and care services. Their evaluation of self-assessment services suggested the potential of this type of referral system to address the preventative agenda.

Some other examples of the diversity of contemporary research in this area include: understanding the impacts of disability on an older population in order to assist policy-makers and health care managers plan appropriate services (Hebert et al., 2012); understanding the frequency and effects of dual impairments such as hearing and visual impairments in older people and their links with depression (Schneider et al., 2011); the influence of education on health and well-being in later life in different European contexts (Avendano et al., 2009); investigating how home delivered health care can improve well-being and reduce the overall need for care (Parsons et al., 2012); and the effects of social inequalities in health pertaining to physical functioning and mental health after retirement (Jokela et al., 2010) and the speed with which older people's physical health deteriorates in early old age dependent on prior occupational class (Chandola et al., 2007).

With the population of the USA set to double by 2025 (Reboussin et al., 2002) and by 2050 the number of people aged 60 or over worldwide will be set to triple (Jurges, 2009). The best way to improve our knowledge of individual and population ageing is, according to Jürges (2009), through multidisciplinary approaches and the use of longitudinal data. This process has already begun within the international research community, led in part by US survey research methodologists, with the development of integrated micro-data, with national and international data sources consisting of household-level, and individual-level data across a range of measures from economic well-being to health. Jürges suggests that it is the international dimension that will prove most useful in understanding ageing trends in different country, institutional and policy circumstances and as a means of learning about what policy interventions work best in particular ageing contexts. Jürges (2009) reports on many of the international data sources including the German *Socio-Economic Panel* (SOEP), the H*ousehold Income and Labour Dynamics in Australia* (HILDA), the *Cross-National Equivalent File* (CNEF) and the *Survey of Health, Ageing and Retirement in Europe*

(SHARE). This latter survey contains micro-data on social and family networks of individuals over 50 years old in fifteen European countries and includes indicators of their socio-economic status and health. One of the major advantages of this survey, according to Jürges (2009), is it is parlty harmonised with the *English Longitudinal Study on Ageing* (ELSA) and the U.S. *Health and Retirement Study* (HRS). Other surveys are currently in development that will build on many of the design elements of SHARE and HRS allowing further international comparability across a range of ageing related topics.

1.2 Longitudinal and ageing cohort studies in the UK

The UK is home to the largest and longest-running longitudinal studies in the world. Martin et al. (2006) in their review report advocate for continued investment and strategic planning for UK based longitudinal studies which continue to contribute to knowledge of individual and societal well-being trends and to act as aids to policy makers, particularly in light of ageing populations. The following list highlights some of the longitudinal studies available in the UK, many of which have national and internationally comparable elements. Full details of the studies, including access to data, are available via the UK data service (www.ukdataservice.ac.uk).

The *National Child Development Study* (NCDS) follows the lives of individuals born in a particular week in 1958 across Great Britain. The study's main aim is to understand human development across the whole lifespan and the factors that influence it. The 1970 *British Cohort Study* (BCS) covers England, Scotland and Wales and follows the lives of over 17,000 people born in a particular week in 1970. The first wave of the study was designed to be comparable with the 1958 cohort study and examined the social and biological characteristics of the mother in relation to neonatal morbidity. There have been several data collection sweeps which have developed over the years to include information not just on medical topics but also covering education, social development, health and the economic circumstances of individuals. Similarly, the *Millennium Cohort Study (*MCS) began in 2000 and focuses on the economic, health and social conditions facing children growing up from this period. All three cohort studies will help provide a better picture of the circumstances of ageing over time. They are housed at the *Centre for Longitudinal Studies* (CLS) where they are maintained and user support provided.

The *British Household Panel Survey* (BHPS) was conducted by the Institute for Social and Economic Research (ISER) at the University of Essex from 1991-2009 (Waves 1-18). This study formed the UK arm of the *European Community Household Panel* (ECHP) survey providing EU cross-national comparable micro-data on various living conditions such as work, health and housing. In 2009 the BHPS migrated to a new study called *Understanding Society*, also known as the *United Kingdom Household Longitudinal Study* (UKHLS). It is hoped the study will facilitate research on the general well-being of the UK over time with a particular focus on family life and social ties, health, work and finances. Lower level geographical datasets are made available under special licence from the Secure Data Service. The *English Longitudinal Study of Ageing* (ELSA) specifically focuses on older people's patterns of ageing, transitions to and beyond retirement and quality of life by examining the dynamics between social networks and participation, health and functioning, and economics. This study is modelled closely on the US *Health and Retirement Study*.

1.3 Measuring heterogeneity: the methodological and policy challenges

Observed heterogeneity amongst the older population; in functional status, health perceptions, satisfaction with health and emotional well-being; is one of the most consistent and remarkable findings about gerontological health. This heterogeneity has likely stimulated growing interest in using more multidimensional measures of health status in studies of elderly persons. This has resulted in research that not only looks at the difficulties associated with ageing but also at ways in which individuals age well. For example, Doyle et al. (2012) have recently used the ELSA and the *Health and Lifestyle Survey* (HALS) datasets to propose a model of successful healthy ageing that takes account of these widespread experiences combining both subjective and objective measures. Their model provides reproducible insights and confirms the role of active physical and social participation and personal resilience for ageing well. Utilising the SHARE survey, d'Uva et al. (2008) investigated the impact of educational level on the self-reported health status of older Europeans across six domains. They found that heterogeneity in reported emotional health, mobility, pain, breathing, sleep and cognition could mask health inequalities if patterns of reporting by educational status were not taken into account. Another study utilising SHARE, reported widespread heterogeneity in later-life mental health in Europe with higher educational attainment and being married associated with optimal mental health (Ploubidis and Grundy,

2009). Similarly, a study by Vonkova and Hullegie (2011) showed that in cases of self-assessment respondents may react differently to different types of research questions and that this heterogeneity in reporting behaviours may lead to biased results.

While methodological problems are common to all longitudinal studies, they are particularly relevant to ageing cohort studies due to the nature of the population. Frailty in the population, compounded by disability and health problems makes it difficult to measure many things in this population. Selective drop out from studies means that missing data analysis that relies on Missing At Random assumptions may not hold. It is more likely that the missing data mechanisms are Missing Not At Random, as a person's poor health status in the previous wave is likely to result in their being unobserved in the current wave. The following section looks at non-response and attrition in the context of ageing cohort studies.

2. Causes of bias in studies of ageing

All longitudinal studies are subject to problems with non-response and attrition, but these problems are magnified considerably in ageing studies where rates of attrition tend to be higher (Gardette et al., 2007). The challenges older study participants pose are not unique but their personal and medical circumstances present sufficient challenges for the design and analysis of clinical research studies to warrant attention. Van Ness et al. (2010) highlight the specific statistical challenges this group encompass design and analytical strategies for multiple outcomes, multipcompnent interventions, floor and ceiling effects, missing data, state transition models and mixed methods.

Sample loss arises in longitudinal surveys at every wave from a failure to locate sample members, or from a failure to contact them, or because contacted subjects choose not to cooperate with the survey. In ageing studies, the last reason becomes increasingly common as subjects may not be able to participate due to ill health and disability reasons and the external validity of a study is threatened (Mein et al., 2012). As the sample becomes progressively smaller it also becomes less representative of the population of research interest because the propensity for attrition varies in a systematic way. It has been argued that for a research findings to be interpreted correctly a detailed analysis of the different types and effects of attrition should be carried out to determine if attrition has occurred at random or if it is associated with particular participant characteristics (Gardette et al., 2007).

Non-response and attrition has the effect of reducing the sample size and as a result can have an impact on statistical efficiency. Researchers are becoming more aware of the need to account for non-response and attrition bias in their work and use a variety of techniques to do this. It can be dealt with by means of statistical procedures to reduce statistical bias and in relation to study design and follow up techniques at subsequent waves. A comparison of earnings measures from UK longitudinal and cross-sectional surveys (Francesconi et al., 2011) revealed notable and stable differences between the different data waves suggesting that non-response and attrition plays an important role between the first and the second comparison time points. David et al. (2012) tested the cost-effectiveness of different types of sample size maintenance programmes in a diabetes prospective cohort study. Their approach advocated the use of substitution sampling, finding it to be a justifiable economic means of maintaining sample sizes. This is one of the most common means of replenishing samples lost through attrition in ageing cohort studies. Matthews et al. (2004) suggest all longitudinal studies should investigate attrition or estimations could otherwise be biased. They suggest that these biases could be of particular import for estimates of movements from an older person's own home to a residential home and could affect the estimates of health and psychiatric diseases where risk factors are associated with attrition.

Conversely, in ageing studies, attrition due to ill health results in a healthy survivor bias which restricts inferences based on risk factors measured earlier on in life. A classic example of the healthy survivor bias is the association between cigarette smoking and dementia among older adults. While there is a strong association between smoking and cognitive performance in younger adults, this association generally attenuates in older samples (Whittington and Huppert, 1997) presumably because the lethality of smoking results in highly selected populations of very old smokers. Survivor bias in studies of ageing is thought to account for the mixed findings regarding the association between cigarette smoking and dementia among older adults(Liu et al., 2010). While this is a common problem for most longitudinal studies, standard methodological approaches (such as under a Missing At Random assumption) are hard to justify in an ageing cohort.

Another example is the reversal of the association between education and cognitive function in elderly populations; several studies deal with this issue (Basu, 2013; Ganguli et al. 2010;

d'Uva et al. 2008; Laursen, 1997). In general education promotes survival. However, the survival of people with no or little education into older ages suggests these selected elderly may have other unobserved, beneficial personal traits. This is something Doyle et al. (2012) highlight in their study on successful ageing in British populations, claiming personal resilience, along with participation in physical and social activities to be key. If unobserved traits influence health, they would bias any estimate of the association between education and health among the survivors. Because of the association between education and cognitive engagement, the same bias may apply to estimates of the association between adult cognitive activities and health. Common missing data methods can be used to analytically address selection bias, using missing data methods such as inverse probability weighting to account for competing risks if other factors are measured. In many instances, it has been shown that although this bias is a theoretical possibility, it is not severe unless survival is quite low (e.g., in very old samples). Studies of individuals above age 80 or 90, however, may face severe bias due to selective survival.

3. Dealing with missing data

All longitudinal studies are subject to problems with non-response and attrition, but these problems are magnified considerably in ageing studies. Selective drop out between waves of data collection is often on the basis of health problems. Some of these health problems may be observed in previous waves, but not all health problems and other factors related to drop out will be observed. Standard methodological approaches tend to assume there are no major differences between those who stay in the study and answer all the questions and those who do not (known as the "Missing At Random" assumption). But this assumption is hard to justify in an ageing cohort. As health and well-being measures are often the dependent variables in analyses of ageing cohort studies, such selective drop out results in missing data mechanisms that are missing not at random. Or in other words, a person's unobserved poor health status in the previous wave is likely to result in their being unobserved in the current wave.

In addition to unobserved poor health as a factor for missing data, there are other reasons why ageing cohort studies may particularly suffer from missing data problems, that sets them apart from standard longitudinal studies. There may be greater residential mobility by sample respondents at older ages due to "downsizing". Downsizing is a common lifecourse

occurrence with families after children become adults and leave their parental home. The parental generation reduces their need for housing consumption and require smaller and less expensive housing. This often results in downsizing and moving house, and so tracking older respondents in a longitudinal study becomes especially problematic when downsizing occurs. Differential 'respondent burden' is another oft-cited reason for non-response (Groves and Couper, 1998). There is evidence that surveys of longer duration have higher attrition rates (Zabel, 1998)

An important concern with all panel studies, and particularly those focused on an older population, is the potential for bias caused by individuals non-randomly dropping out of the survey over time. If attrition from a survey is systematically related to outcomes of interest or to variables correlated with these outcomes, then not only will the survey cease to be representative of the population of interest, but estimates of the relationships between different key outcomes, especially in a longitudinal context, may also be biased (Plewis, 2011).

There has been a substantial increase in missing data handling techniques that assume a missing at random (MAR) mechanism. MAR mechanisms assume that the propensity for missing data on an outcome is related to other analysis variables. Although MAR is often reasonable, there are situations where this assumption is unlikely to hold, leading to biased parameter estimates. One typical example is an ageing cohort study of functional decline where participants with the greatest declines also have the highest likelihood of attrition, even after controlling for other correlates of missingness. There is a large body of literature on missing not at random (MNAR) analysis models for longitudinal data, particularly in the field of biostatistics (Plewis 2011). Because these methods allow for a relationship between the outcome variable and the propensity for missing data, they require a weaker assumption about the missing data mechanism. There are 2 classic MNAR modeling approaches for longitudinal data: the selection model and the pattern mixture model (Plewis 2011). These models are now quite easy to estimate in popular statistical modelling packages such as STATA, R and Mplus.

More complex *Not Missing At Random* (NMAR) analysis strategies may become increasingly important as sample members leave ageing cohort studies due to ill health, lack of

engagement with the study, or death resulting in missing cases. While in many research circumstances employing data handling techniques that assume a *Missing At Random* (MAR) procedure may be suitable, in longitudinal ageing cohort studies this is not always appropriate.

4. Measuring cognitive change: methodological issues

Most studies on ageing measure aspects of cognitive function. This is because some of the basic cognitive functions most affected by age are attention, memory and perception. Although there is a general decline in cognitive function with ageing, enormous variability exists across individuals. Many older people out-perform young people, at least on some cognitive tasks, and others of the same age do at least as well as the young. However, there are a number of methodological issues that arise when measuring cognitive change in elderly populations.

The methodological challenges around cognitive ageing make it one of the best examples of the methodological challenges that ageing cohort studies face that are common and yet distinct from standard longitudinal studies. These include: the role of missing data (Kennison and Zelinski, 2005); the consequences of non-response (Bergman et al., 2010); learning effects in the use of standardised cognitive function measurement tools (Jenkins, 2012 and JacqminGadda et al., 1997); selection bias and the possibility of a test-examiner effect (Laursen, 1997); and intra-individual cognitive variability (short-term changes) that can highlight important predictive ageing related outcomes which can help explain the psychological process of adaptation (Martin and Hofer, 2004). Although cognitive function tends to decline with age, in particular memory, perception and attention, there is a wide range of variability across individuals. Liu et al. (2010) suggest that while some older people do as well as younger people on a range of cognitive function tests and may even out-perform them on some, there are still important methodological issues when measuring cognitive change in older people. The tension between continuing repeated measurements from earlier on in the lifecourse versus new and more relevant measures is typified in the challenges around cognitive decline. New measures of cognitive tests at older ages are often needed that distinguish between dimensions of cognitive functioning that are particularly important for older populations. Floor and ceiling effects may be marked in a number of cognitive measures with the majority of respondents in early old age obtaining maximum scores on

some cognitive tests, while a large proportion of those with dementia in later life obtaining the minimum scores. There is the problem of measuring change when participants are no longer able to participate in cognitive and other functioning tests, such as those with onset of dementia or other disabilities. Furthermore, cognitive tests in particular are subject to practice effects, which often result in improved cognitive scores after baseline tests. Finally, there are strong selection issues with missing data in cognitive tests as those with missing data are most likely to be the respondents with the poorest cognitive functioning.

4.1 Factors influencing performance on cognitive tests

Performance on cognitive tests can be affected by many factors including, cultural experiences, language usage, educational attainment, emotional and physical states, prior testing experience the testing environment, and measurement error (Ganguli et al., 2007 and Morris et al., 1999). Morris et al. (1999) claim it is for this reason that it becomes difficult to disentangle individual differences in cognitive scores that are due to the process or disease of interest from these other confounding factors. Their study focuses on measuring changes in cognition rather than cognition scores at a particular time only; and in doing so address various conceptual and methodological issues including the benefits of using longitudinal methods over cross-sectional methods, measurement variation, and the benefits of using multiple statistical models taking careful account of the impact of age and education. Morris et al. (1999) suggest that looking at cognitive trajectories rather than levels of cognitive functioning may be a more appropriate way of understanding the process of cognitive aging.

Ganguli et al., (2010) tested for the effects of age and education on cognitive tests from a population-based cohort and combined test scores into composite scores, something recommended by Morris et al. (1999). Participants were tested using the Mini-Mental State Examination (MMSE) and the Clinical Dementia Rating (CDR) scale. They found that older age and poorer education together was associated with worse neuropsychological test performance having a combined greater impact than other measured factors.

4.2 Ceiling and floor effects in measurement

The demands on cognitive tests used in epidemiological investigations are so severe that measurement errors frequently occur causing ceiling and floor effects. This is due to the need to take into account a wide range of cognition levels across different age groups in the older cohort, all with varying health status. Morris et al. (1999) note these issues and suggest that a broad range of cognitive functions should be covered in the study population at the beginning

of the study and during it. Ceiling and floor effects occur when levels of performance outside of the test range cannot be measured or when those participants with the highest or lowest scores can only change in one direct. According to Morris et al., random variation will not be evenly distributed around initial scores. Floor and ceiling effects are a particular problem for brief cognitive tests as they are unable to measure the full range of cognitive functioning in a satisfactory way that reflect the ways in which the dimensions of cognitive function change from early old age to the very old.

4.3 Challenges of collaboration

Ceiling and floor effects in testing are a challenge for epidemiologic studies, especially for the use of brief tests, but the problem is compounded when working in cross-disciplinary settings, as is usually the case in older age research. There can be additional issues in collaborative working due to the prevalence of particular tools covering different facets of cognition and the different epistemological and methodological approaches across disciplines increases the likelihood of experiencing ceiling and floor effects. Morris et al. (1999) acknowledge that compromising on these factors is challenging for collaborating statisticians, psychologists and epidemiologists. There are recommendations towards improved collaborative working between epidemiological and clinical researchers particularly those with a focus on working with older people suffering from depression and memory disorders (Steffens et al., 2006). Van Ness et al. (2010) also recognise the challenges and benefits of collaborative working in the complex area of geriatric health with participants with co-morbid conditions. However, they recommend the systematic training of biostatisticians intending to work with geriatricians, gerontologists and older participants as they will face particular challenges of design and analysis and need to deal with transition models, floor and ceiling effects, mixed methods, multiple outcomes and interventions.

4.4 Practice effects and measurement errors in repeated cognitive testing

Another issue for cognitive testing relates to measurement errors occurring due to practice effects. It is possible that practice effects (also known as learning or re-test effects) influence the scores of cognitive tests when they are repeated, as the earlier exposure to the tests may have interfered with the normal cognitive development that is being measured. Practice effects can occur during repeated cognitive testing, which is best practice for addressing

cognitive change, but can lead participants to experience a learning effect over time. This in turn can impact the estimates of change when the size of the effect is not appropriately accounted for. Morris et al. (1999) suggest that prior exposure to the same tests may lead to improved performance on function and cognitive tests issued at intervals. It is a difficult methodological task to differentiate practice effects from change scores. Separating out such practice effects from change scores is methodologically hard to do. Duff et al. (2010) tackle this issue by employing standardized regression-based (SRB) prediction formulas to determine if a change is meaningful and reliable. They were interested in the role of short-term practice effects in predicting test scores after one year. They concluded that baseline test performances were the best predictors of future test performances and including short-term practice effects added to the predictability of scores for nine different cognitive tests. Duff et al., also suggest that practice effects can have implications for longitudinal studies. However, they acknowledge that their models require validation and testing in clinical samples.

5. NCRM workshop recommendations

Innovations in data collection:

5.1 *Mixed modes data collection*: This is already done in many ageing studies. For example in ELSA, modes of data collection range from a self-completion questionnaire, nurse visits and CAPI. However, the actual measurement instruments tend to stay within the same mode of data collection between waves. For other ageing studies, there may be changes in the site of data collection. For example, the MRC 1946 BCS moved from the data collection mode of home visits by nurses to clinic visits. The Whitehall II study gives participants the choice of data collection from home visits and clinic visits with those with limiting health problems more likely to choose home visits.

The problems of mixing modes of data collection have been well documented in a number of studies. The Mode of questionnaire administration can have serious effects on data quality. One of the main primary data collection instruments in social, health and epidemiological research is the survey questionnaire. Even minor changes in question wording, question order or response format can result in differences in the type of response obtained, but can be difficult to separate out from other effects of different modes of administration.

The challenges around mixing modes of data collection are around data quality and reliability. However, if we do not mix modes of data collection, there is a possibility of more missing data from respondents unwilling to engage with a particular mode of data collection (such as preferring a short telephone interview to a face to face interview). Survey investigators often face the challenging question- "Is losing data from respondents a bigger disservice to the survey than collecting possibly compromised data using different methods?"

The benefits of mixing methods or modes includes less missing data and increased compliance with responses to questions, potentially resulting in a more representative sample of the study population. Multiple methods or modes within patient can reduce systematic missing data, especially data that is not missing at random (NMAR). Examples of missing data in ageing cohort studies include respondents who become too ill to complete surveys in person, or who develop impairments with hearing or sight. Mixing modes of data collection can take account of respondent's preferences for answering surveys.

There are well established statistical methods to measure equivalence of data collected using different modes. These include examining the correlation between methods, intra-class correlations, reliability tests, comparison of mean scores by methods, comparison of scores by sub-group or at particular range of scores, Differential Item Functioning (DIF) tests, Bland-Altman plots, and comparing the mixed-method/mode equivalence with test-retest equivalence of the instrument. The example of the Innovation Panel from Understanding Society (http://www.understandingsociety.org.uk/design/innovation/default.aspx) for carrying out experimental designs in a longitudinal context could be adopted in other ageing cohort studies.

5.2 *Using parallel forms of cognitive tests* in alternate waves of data collection (Houx et al. 2002). Although this reduces practice/learning effects associated with repeated cognitive tests, it does not get rid of them altogether. While direct learning is ruled out by administering parallel test versions, other forms of learning such as procedural learning (improving performance by merely doing something more than once) cannot be circumvented.

5.3 Using the lifegrid approach to retrospective data collection to minimise proxy recall bias. While asking certain key questions from proxy informants is a standard method, asking relatives or carers for complex information on which health problems have occurred, when each one began and for how long it went on may be subject to considerable recall biases. The lifegrid approach to retrospective data collection has been successful in reducing recall bias (Berney and Blane 2003) and has been adopted in a number of ageing cohort studies to collect life history data for the main respondents. Its use for obtaining life history data from proxy respondents could be extended. This could be especially useful when participants die or they are unable to answer questions. Information on dates and duration of life events could be obtained from proxy respondents using the prompts and memory aides from the lifegrid approach.

An event history calendar (EHC) is a data collection method used to gather retrospective data. Rather than a conventional list of survey questions, the EHC uses a calendar, with time going across the top in columns and key events of interest in the survey going down the side in rows. In contrast to a traditional linear questioning approach, the EHC method uses a series of semistructured questions and probes to encourage respondents to report from autobiographical memory using strategies such as sequential retrieval as well as cross-referencing one event with another. The method has been shown to improve the accuracy and quality of retrospective reports for some topic areas (Belli et al, 2004).

Innovations in analysis:

5.4 *Using short-term practice effects as an additional predictor variable in models of cognitive change*. Most studies collect repeated data within a wave of data collection to examine the validity and reliability of their measures. This enables us to examine the correlation between the practice effect (the difference in the mean of the cognitive measures from the wave and the repeat) and the baseline cognitive measure (Duff et al. 2010). If the practice effects are large for the subsample of participants with repeated measures, we can consider adjusting for the number of times the cognitive tests are taken to partially account for practice effects in analysing change in cognitive measures in the full sample.

5.5 Using *auxiliary variables to adjust for missingness in longitudinal analyses.*

One of the methods of reducing problems of bias due to missing data is to find those variables that predict whether a piece of data is missing, and which of those variables that predict missingness are also related to at least one out of possibly many outcomes of interest (Plewis 2011). Longitudinal researchers can draw on a wider range of potential predictors from earlier waves of data collection including:

- Variables of substantive interest that are measured on all the responding cases in the first wave of the study.
- Variables from the sampling frame for all sampled cases at the first wave.
- Variables that are related to aspects of data collection, either derived from administrative procedures used, for example, to track sample members over time or from data collected from respondents and interviewers during fieldwork.

These three groups of variables are sometimes referred to collectively as auxiliary variables although sampling statisticians tend to reserve this term for the variables derived from sampling frames and population registers (from the second group). Variables in last group are sometimes labelled paradata (Couper 1998), but are also referred to as instruments especially in the econometric literature on adjusting for non-response.

Multiple Imputation (MI) are now a standard way of dealing with missing data in cross sectional and longitudinal analyses. The strength of MI is its focus on a model of interest and thelink between this model and models for missingness. MI methods assume that data are missing at random (MAR), i.e. that missingness is ignorable conditional on the chosen set of predictors. This is not an assumption that can be supported by the data in many ageing cohort studies where missingness could depend on the wave t value of the variable, even after conditioning on measured wave t-k variables. In other words, the missingness mechanism ca not be ignored and then the data are ‘missing not at random’ (MNAR). One way of dealing with the problem of non-ignorable missingness is to jointly model the model of interest and the missingness mechanism (Plewis 2011).

Paradata could have a potentially important part to play in this joint modelling process because their association with response behaviour and their irrelevance to the model of interest means that they can be useful instruments that help to identify the joint model and thus improve the robustness of estimates from these models. This can be particularly

important when the model of interest includes more explanatory variables as controls. Candidate variables include the length of the interview at wave one, the proportion of questions answered at wave one (a variable found to be predictive of non-response by a number of researchers), the reluctance of the cohort child to attempt the educational tests, and variables like interviewer gender, ethnic group, age and experience. It is also possible to ask the interviewers to record observations about their contact with respondents, and to ask respondents to describe their experience of the interview and interviewer. On the other hand, there is a cost attached to collecting paradata. The question of whether it is cost effective to collect more variables of this kind must depend in part on an assessment of the possibilities and benefits of reducing the deleterious effects of non-response that they bring.

5.6 *Using a range of missing data analyses methods and simulation studies to assess the performance of a variety of missing data mechanisms.* While multiple imputation analyses are becoming increasingly standard, the Missing At Random assumptions may not be appropriate for many ageing cohort studies. Analytical methods such as Heckman selection models (Winship and Mare 1992) and the inclusion of auxiliary variables that predict missingness but not the substantive outcome (LaLonde 1986) could be useful tools in analysing longitudinal data from ageing cohort studies.

There are two main strategies used to investigate the effect of different methods of accounting for missing data (Landy, 2012). The first strategy involves analysing data from a dataset with missing data, using the missing data methods, and comparing the results. However, the true results remain unknown, making it impossible to compare the results achieved using missing data methods to the true values. Hence, it is not possible to assess which of the missing data methods produce the least biased results.

The second strategy is to use complete datasets and then delete values by simulating the missingness according to the postulated missing data mechanism. Simulation studies can be used to assess the performance of a variety of different missing data mechanisms in relation to a known truth. This entails generating simulated data from cohort studies and then deleting data according to potential missing data mechanisms. The performance of different missing data methods under various missing data mechanisms can be evaluated, using measures of bias, coverage and accuracy (Landy, 2012).

The analyses of interest are initially carried out on the complete dataset; then using missing data techniques on the dataset with the simulated missing data, and these results are compared with the results from the complete data analyses (Landy, 2012). One limitation of this method is that in practice the missing data mechanism is generally unknown. When both the true coefficients and the estimates from the various missing data methods are known (as in the second method above), the performance of different missing data methods under various missing data mechanisms can be evaluated, using measures of bias, coverage and accuracy (Schafer and Graham, 1999; Collins et al., 2001).

6. Conclusions

There has been a substantial increase in ageing cohort studies that are now available for researchers around the world. This has been driven by "new" ageing cohorts studies, as well as the maturation of sample members of ongoing longitudinal cohort studies and birth cohort studies. These ageing cohort studies share common methodological challenges which are also common to all longitudinal studies, but are particularly acute in older populations due to specific issues such as frailty, disability and ill health leading to data collection, measurement and analytical challenges.

The NCRM series of workshops have highlighted a number of methodological challenges that need to be taken into account for the design and analyses of ageing cohort studies. The rise in data from ageing cohort studies has also been paralleled by the development of methodological tools to cope with these challenges. These include using mixed modes of data collection to deal with respondent burden, using the lifegrid history method to deal with recall bias for proxy respondents, using auxiliary variables to adjust for missing not at random mechanisms, and using a range of missing data analyses methods and simulation studies to assess the performance of a variety of different missing data mechanisms. However, these are not standard methodological tools that most researchers involved in analysing longitudinal studies are familiar with. The training components associated with the NCRM workshop series were very popular and well received, suggesting there is a strong demand for such additional training tailored to non-statistical researchers.

References

Avendano M, Jurges H and Mackenbach JP. (2009) Educational level and changes in health across Europe: longitudinal results from SHARE. *Journal of European Social Policy* 19: 301-316.

Basu R. (2013) Education and Dementia Risk: Results From the Aging Demographics and Memory Study. *Research on Aging* 35: 7-31.

Belli, R.F., Lee, E.H., Stafford, F.P. and Chou, C. (2004) ACalendar and Question-List Survey Methods: Association between Interviewer Behaviours and Data Quality. *Journal of Official Statistics*, Vol 20, No. 2, pp. 185-218

Bergman P, Ahlberg G, Forsell Y and Lundberg I. (2010) Non-Participation in the Second Wave of the Part Study on Mental Disorder and Its Effects on Risk Estimates. *International Journal of Social Psychiatry* 56: 119-132.

Burns RA, Butterworth P, Kiely KM, Bielak AAM, Luszcz MA, Mitchell P, Christensen H, Von Sanden C and Anstey KJ. (2011) Multiple imputation was an efficient method for harmonizing the Mini-Mental State Examination with missing item-level data. *Journal of Clinical Epidemiology* 64: 787-793.

Chandola T, Ferrie J, Sacker A and Marmot M. (2007) Social inequalities in self reported health in early old age: follow-up of prospective cohort study. *British Medical Journal* 334: 990-993B.

Collins LM, Schafer JL, Kam CM. (2001) A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods* Dec;6(4):330-51.

d'Uva TB, O'Donnell O and van Doorslaer E. (2008) Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans. *International Journal of Epidemiology* 37: 1375-1383.

David MC, Bensink M, Higashi H, Donald M, Alati R and Ware RS. (2012) Monte Carlo simulation of the cost-effectiveness of sample size maintenance programs revealed the need to consider substitution sampling. *Journal of Clinical Epidemiology* 65: 1200-1211.

Doyle YG, McKee M and Sherriff M. (2012) A model of successful ageing in British populations. *European Journal of Public Health* 22: 71-76.

Duff K, Beglinger LJ, Moser DJ, Paulsen JS, Schultz SK and Arndt S. (2010) Predicting cognitive change in older adults: the relative contribution of practice effects. *Arch Clin Neuropsychol* 25: 81-88.

Enders CK. (2011) Missing Not at Random Models for Latent Growth Curve Analyses. *Psychol Methods* 16: 1-16.

Francesconi M, Sutherland H and Zantomio F. (2011) A comparison of earnings measures from longitudinal and cross-sectional surveys: evidence from the UK. *Journal of the Royal Statistical Society Series a-Statistics in Society* 174: 297-326.

Ganguli M, Snitz BE, Lee CW, Vanderbilt J, Saxton JA and Chang CCH. (2010) Age and education effects and norms on a cognitive test battery from a population-based cohort: The Monongahela-Youghiogheny Healthy Aging Team. *Aging & Mental Health* 14: 100-107.

Gardette V, Coley N, Toulza O and Andrieu S. (2007) Attrition in geriatric research: How important is it and how should it be dealt with? *Journal of Nutrition Health & Aging* 11: 265-271.

Groves, R., and M. Couper. (1998). Nonresponse in Household Interview Surveys. New York: Wiley.

Hebert R, Raiche M and Gueye NR. (2012) Survey disability questionnaire does not generate valid accurate data compared to clinical assessment on an older population. *Archives of Gerontology and Geriatrics* 54: E57-E62.

Houx PJ, Shepherd J, Blauw GJ, Murphy MB, Ford I, et al. (2002) Testing cognitive function in elderly populations: the PROSPER study. PROspective Study of Pravastatin in the Elderly at Risk. J Neurol Neurosurg Psychiatry. October; 73(4): 385–389.

Howse K. (2012) Healthy ageing: the role of health care services. *Perspectives in Public Health* 132: 171-177.

JacqminGadda H, Fabrigoule C, Commenges D and Dartigues JF. (1997) Longitudinal study of cognitive aging in non-demented elderly subjects. *Revue D Epidemiologie Et De Sante Publique* 45: 363-372.

Jenkins A. (2012) Participation in Learning and Depressive Symptoms. *Educational Gerontology* 38: 595-603.

Jokela M, Ferrie JE, Gimeno D, Chandola T, Shipley MJ, Head J, Vahtera J, Westerlund H, Marmot MG and Kivimaki M. (2010) From Midlife to Early Old Age Health Trajectories Associated With Retirement. *Epidemiology* 21: 284-290.

Jurges H. (2009) Provision for Old Age. National and International Survey Data to Support Research and Policy on Aging. *RatSWD Working Paper* 94: 1-15.

Kennison RF and Zelinski EM. (2005) Estimating age change in list recall in asset and health dynamics of the oldest-old: The effects of attrition bias and missing data treatment. *Psychology and Aging* 20: 460-475.

Lalonde RJ. (1986) Evaluating the Econometric Evaluations of Training-Programs with Experimental-Data. *American Economic Review* Sep;76(4):604-20.

Landy, R (2012) The influence of life course socioeconomic position on cognitive function and cognitive decline in older age: the impact of missing data. Doctoral thesis, UCL (University College London).

Laursen P. (1997) The impact of aging on cognitive functions - An 11 year follow-up study of four age cohorts - Preface. *Acta Neurologica Scandinavica* 96: 3-+.

Liu S, Jones RN and Glymour MN. (2010) Implications of lifecourse epidemiology for research on determinants of adult disease. *Public Health Reviews* 32: 489-511.

Martin J, Bynner, J., Kalton, G., Boyle, P., Goldstein, H., Parsons, S., Piesse, A. (2006) Strategic Review of Panel and Cohort Studies. *Report to the Research Resurces Board of hte Economic and Social Research Council* London: Longview, 1-85.

Martin M and Hofer SM. (2004) Intraindividual variability, change, and aging: Conceptual and analytical issues. *Gerontology* 50: 7-11.

Matthews FE, Chatfield M, Freeman C, McCracken C, Brayne C and Cfas M. (2004) Attrition and bias in the MRC cognitive function and ageing study: an epidemiological investigation. *Bmc Public Health* 4.

Mein G, Johal S, Grant RL, Seale C, Ashcroft R and Tinker A. (2012) Predictors of two forms of attrition in a longitudinal health study involving ageing participants: An analysis based on the Whitehall II study. *Bmc Medical Research Methodology* 12.

Morris MC, Evans DA, Hebert LE and Bienias JL. (1999) Methodological issues in the study of cognitive decline. *Am J Epidemiol* 149: 789-793.

Palmer RF and Royall DR. (2010) Missing Data? Plan on It! *Journal of the American Geriatrics Society* 58: S343-S348.

Parsons J, Rouse P, Robinson EM, Sheridan N and Connolly MJ. (2012) Goal setting as a feature of homecare services for older people: does it make a difference? *Age and Ageing* 41: 24-29.

Plewis I (2011) Auxiliary variables and adjustments for missingness in longitudinal studies. CCSR Working Paper Series. http://www.ccsr.ac.uk/documents/IanPlewisWorkingPaper2011-04.pdf

Ploubidis GB and Grundy E. (2009) Later-Life Mental Health in Europe: A Country-Level Comparison. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences* 64: 666-676.

Reboussin BA, Miller ME, Lohman KK and Ten Have TR. (2002) Latent class models for longitudinal studies of the elderly with data missing at random. *Journal of the Royal Statistical Society Series C-Applied Statistics* 51: 69-90.

Schafer JL. (1999) Multiple imputation: a primer. *Statistical Methods in Medical Research* Mar;8(1):3-15

Schneider JM, Gopinath B, McMahon CM, Leeder SR, Mitchell P and Wang JJ. (2011) Dual Sensory Impairment in Older Age. *Journal of Aging and Health* 23: 1309-1324.

Steffens DC, Otey E, Alexopoulos GS, Butters MA, Cuthbert B, Ganguli M, Geda YE, Hendrie HC, Krishnan RR, Kumar A, Lopez OL, Lyketsos CG, Mast BT, Morris JC, Norton MC, Peavy GM, Petersen RC, Reynolds CF, Salloway S, Welsh-Bohmer KA and Yesavage J. (2006)

Perspectives on depression, mild cognitive impairment, and cognitive decline. *Archives of General Psychiatry* 63: 130-138.

Tucker S, Brand C, O'Shea S, Abendstern M, Clarkson P, Hughes J, Wenborn J and Challis D. (2011) An evaluation of the use of self-assessment for the provision of community equipment and adaptations in English local authorities. *British Journal of Occupational Therapy*. 119-128.

Van Ness PH, Charpentier PA, Ip EH, Leng XY, Murphy TE, Tooze JA and Allore HG. (2010) Gerontologic Biostatistics: The Statistical Challenges of Clinical Research with Older Study Participants. *Journal of the American Geriatrics Society* 58: 1386-1392.

Vonkova H and Hullegie P. (2011) Is the anchoring vignette method sensitive to the domain and choice of the vignette? *Journal of the Royal Statistical Society Series a-Statistics in Society* 174: 597-620.

Whittington JE and Huppert FA. (1997) Smoking and cognitive decline. *Human Psychopharmacology-Clinical and Experimental* 12: 467-480.

Wiles JL, Allen RES, Palmer AJ, Hayman KJ, Keeling S and Kerse N. (2009) Older people and their social spaces: A study of well-being and attachment to place in Aotearoa New Zealand. *Social Science & Medicine* 68: 664-671.

Winship C, Mare RD. (1992) Models for Sample Selection Bias. *Annual Review of Sociology*;18:327-50.

Zabel J. (1998) An Analysis of Attrition in the PSID and the Survey of Income and Program Participation. *Journal of Human Resources* 33(2):479-506.