Modelling and Predicting Change in Population Psychological Distress: An Analysis of the General Health Questionnaire in a Representative UK Sample

Christopher Smyth

Degree of Doctor of Philosophy (PhD)

Faculty of Life and Health Sciences of Ulster University

September 2020

I confirm that the word count of this thesis is less than 100,000 excluding the title page, contents acknowledgements, summary or abstract, abbreviations, footnotes, diagrams, maps, illustrations, tables, appendices, and references or bibliography.

Contents Page

Acknowledgements

The list of people I should acknowledge may be longer than the thesis if it were to be exhaustive; however, I feel that several people deserve special mention as, without them, this thesis would never have been completed.

My two supervisors, Professors Murphy and Adamson, have displayed boundless patience putting up with my incessant questions and tendency to land at their office in a tizz at 4.30 pm on a Friday when they wanted nothing more than to put their feet up for the weekend.

I have made a friend for life during my PhD when I moved into the beautiful Portrush, '*Ulster's Gold Coast*' with my friend Ben. After a long day arguing with MPLUS, he was always on hand with a vegetarian burger and 'a cold one' on the balcony. He also introduced me to the Atlantic Bar, which served as a bar, office, and social hub of my university life

My officemates have supported me through my time in Coleraine. They were always there to support me; either by bringing me copious and downright unhealthy volumes of Redbull or being there to have some deep and important conversations with. I'm truly grateful to them all for making my PhD experience the truly life-changing event that it has been, but my good friends Izzy and Dean deserve special mention and I wish them well in their own PhDs.

Finally, my wonderful family have supported me during this time providing reassurance, logistical support and most importantly, a steady supply of good food to keep the engine running over what has felt like an eternity. To them and so many more, I cannot say thankyou enough.

Declaration

"I hereby declare that with the effect from the date on which this thesis is deposited in Research Student Administration of Ulster University, I permit the Librarian of the University to allow the thesis to be copied in whole or in part without reference to me on the understanding that such authority applies to the provision of single copies made for study purposes or for inclusion within the stock of another library. I permit this thesis to be made available through Ulster Institutional Repository and / or EThOS under the terms of Ulster e-Theses deposit arrangement which I have signed. IT IS A CONDITION OF USE OF THIS THESIS THAT ANYONE WHO CONSULTS IT MUST RECOGNSE THAT THE COPYRIGHT RESTS WITH THE AUTHOR AND THAT NO QUOTATION FROM THE THESIS AND NO INFORMATION DERIVED FROM IT MAY BE PUBLISHED UNESS THE SOURCE IS PROPERLY ACKNOWLEDGED."

Abbreviations

AIC- Akaike Information Criterion

ALMR- Adjusted Lo Mendel Reuben

BHPS- British Household Panel Study

BIC- Bayesian Information Criteria

CFA- Confirmatory Factor Analysis

CFI- Confirmatory Fit Index

CI- Confidence Intervals

CMI- Cornel Medical Index

CTCM- Correlated Traits Correlated Methods

CTCU- Correlated Traits Correlated Uniqueness

DF- Degrees of Freedom

EFA- Exploratory Factor Analysis

EMB- Ethnic Minority Boost Sample of Understanding Society

FML- Full Maximum Likelihood

GHQ- General Health Questionnaire

GMM- Growth Mixture Modelling

GP- General Practitioner

GPS- General Population Sample of Understanding Society

GRO- Government Records Office

HADS- Health and Depression Scale

IP- Innovation Panel of Understanding Society

LCA- Latent Class Analysis

LPA- Latent Profile Analysis

LPS- Land and Property Services

MHI-5 – Mental Health Index 5 Item Version

MLR- Maximum Likelihood with Robust Standard Errors

ONS- Office for National Statistics

OR- Odds Ratios

PCA- Principal Component Analysis

PSU- Primary Sampling Unit

RMSEA- The Root Mean Square Error of Approximation

RSE- Rosenberg Self-esteem Scale

SCL-90- Symptoms Checklist 90 Item Version

SF-12- Short Form 12

SRMR- Standardised Root Mean Square Residual

SSABIC- Sample Size Adjusted Bayesian Information Criteria

STD- Standardised

STDX- Standardised along the X-axis

STDY- Standardised along the Y-axis

SWEMBS- Short Warwick Edinburgh Mental Wellbeing Scale

TLI- Tucker Lewis Index

TMHQ- Thai Mental Health Questionnaire

UK- United Kingdom

USA – United States of America

UKHLS- Understanding Society: United Kingdom Household Longitudinal Study

WHO- World Health Organisation

List of Tables

List of Figures

Thesis Abstract

Evidence has shown that mental health varies within and across individuals and, importantly, over time. The current study aimed to identify trajectories of mental health over time and to analyse what predisposed individuals to exhibit these trajectories. Data from first five waves of the Understanding Society database, a sample of over 100,000 participants which was representative of the wider UK population, was used in order to conduct this analysis.

Ye's (2009) model was deemed an appropriate dimensional representation and was found to be stable over time and to display concurrent validity with a range of associated covariates. Four trajectories, characterised by slope and intercept were extracted from the data and these represented stable periods of poor and good psychological distress, with steadily increasing or decreasing levels respectively. The stable group of low levels of psychological distress was labelled as the reference group. These trajectories had a wide range of biological social and psychological covariates regressed upon them. The analysis showed that a wide range of biological, social and psychological variables affected individual's trajectories over time, with social variables such as income and job satisfaction having the largest affect on class membership. Personality characteristics such as neuroticism was also to have a strong association with individuals exhibiting persistently elevated psychological distress. Generally, biological characteristics had a smaller affect on class membership with the majority of ethnicities displaying no statistically significant relationship with class membership.

This research has utility as it demonstrates how individuals may exhibit similar levels of psychological distress at a given time period but may have vastly different trajectories in how they arrived at these points. This research demonstrated how through analysing longitudinal trajectories, mental health practitioners can develop a wider perspective of

how psychological distress can predispose individuals to poorer outcomes and could be used to inform treatment options accordingly.

Chapter 1- Mental Health and its Measurement

1.1.- Introduction

The purpose of this chapter was to give an overview of the concept of mental health generally with specific reference to the General Health Questionnaire (GHQ) and the concept of psychological distress. A detailed overview of this instrument was provided including the various refinements from the original 140-item GHQ test to the more recent 12-item test, its historical context and its properties including how it performed relative to other measures of mental health. The GHQ-12 was then evaluated using existing literature on established criteria such as its validity, reliability and dimensionality. It was also discussed how the GHQ-12 performs in numerous settings, including cultural variations and clinical and general population applications and over time. Finally, this chapter detailed the general rationale and hypotheses that underpinned the thesis and further chapters.

1.2.- Historical Context of Mental Health and its Measurement

Mental health is a concept, which does not have a single agreed definition. When searching relevant literature and documentation, it was clear that while numerous definitions had broadly similar thrusts, several variations existed, all of which emphasised slightly different aspects of the concept. Within the literature, many studies have attributed mental health to the absence of mental health disorders (Fuller, Edwards, Procter, & Moss 2000) and have operationalised their research into the search for mental health problems. Others have viewed it as a continuum (Keyes, 2002) with participants being placed along a continuous scale from good to poor mental health.

NHS England defined mental health differently, focusing on the positive aspect of mental health by defining it as "*a positive state of mind and body, feeling safe and* *able to cope, with a sense of connection with people, communities and the wider environment*" (DoH, 2011, para. 7). In a similar vein to 'NHS England', the World Health Organisation (WHO) described mental health as, "*a state of well-being in which every individual realises his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community*" (WHO, 2004, para. 2).

The Health Education Authority, a now-defunct branch of the UK government, described mental health in terms of resilience with the following quote "*the emotional and spiritual resilience which enables us to survive pain, disappointment and sadness. It is a fundamental belief in our own and others' dignity and worth*" (Creek and Lougher, 2006, p. 12).

From the different sources, there appeared to be a clear rationale for each of the organisations to adopt their respective definitions. Medical organisations subject to public scrutiny tend to focus on the positive aspect of mental health. In contrast, researchers tended to operationalise the concept differently, placing greater emphasis on measurable elements in their research papers. The ambiguity in the definitions was noticeable; however, the general thrust of all definitions referred to a state of mind which enables functionality.

When one considers how to measure mental health, parallels may be drawn between the measurements of mental health with that of physical health. In the 1900s, physical health research was focused on the area of infectious and communicable diseases, with a particular focus on illnesses such as malaria and tuberculosis (Harpman et al., 2003). This focus slowly changed over time however a significant paradigm shift occurred in 1993 with a World Bank and World Health Organisation report which accounted for not only mortality, i.e. an individual's risk of death, but also morbidity, which attempted to measure an individual's well-being.

This significant change in focus shone a light on mental health as it suggested that an individual's state of being, later referred to as their mental health, was a major predictor of morbidity even though it was a relatively minor predictor of mortality. This research looked at a concept known as 'burden of disease' which is defined by the world health organisation (Murray, 1996) as

- The extent to which an individual's life expectancy is shortened
- The extent to which an individual is affected by a disability caused by a disease

While primarily an economic document, the World Bank and World Health Organisation report had a profound effect on how society understood health as a general concept (Harpman et al., 2003). It was able to demonstrate the economic benefits to nations of addressing their mental health needs in order to have a healthy and productive workforce.

Due to the debilitating effects on a nation's economy and a general shift in focus from mortality to morbidity, it became increasingly necessary to develop methods which could measure mental health in a reliable, practical and cost-effective manner. It also became necessary to be able to differentiate the severity of the effects of mental health disorders in a way that merely looking at incidence rates was not capable of doing.

In an effort to improve mental health services globally, the WHO produced a report designed to encourage countries, especially developing countries, to put measures in place to enhance both understanding and treatment of poor mental health. The WHO found that only 48% of countries had strategies in place to address poor mental health

(World Health Organization, 2001). The report found that mental health issues are primarily dealt with by nations GP's rather than specialist mental health professionals (Manderscheid et al. 1993), and therefore any strategies had to operate within the confines of that country's existing medical structures. This, in turn, led to researchers attempting to design simple, easily interpreted tests to measure mental health. These tests were designed to be used by medical practitioners, especially GP's who may not have had a specific specialisation in mental disorders and several of them are detailed throughout this chapter in sections 1.3 and 1.10. Poor mental health was found to be particularly prevalent amongst developing nations and areas of deprivation. Both of these, at that time at least, had limited literacy and numeracy skills in their population and limited access to technology such as telephones and computers. From a practical point of view, the methods of data collection were limited to interviews and questionnaires. It was also vital for the tests to be widely used, and that they were worded in a way that a layperson could understand (Harpman et al., 2003). Due to limitations around accessing treatments and other geographically distinct circumstances, it is estimated that between 76% and 85% of people suffering from mental health issues globally receive minimal if any treatment (Demyttenaere, 2004).

Within a global context, as previously mentioned, mental health is increasingly being seen and appreciated as a significant component of living a healthy life. As per the 2013 burden of disease analysis (Vos et al., 2013), the predominant mental health condition experienced worldwide is depression, with anxiety, schizophrenia and bipolar disorder representing other major contributors to mental health issues respectively. It is important to note that there is a two-way relationship between mental health issues and some physical health conditions (Ferrari et al., 2013) with the prevalence of depression being positively correlated with an increased incidence of heart attacks and strokes. It is vital to consider this relationship as the aforementioned studies cite this relationship as one of the ways that mental health contributes to the global burden of disease and reduction in quality of life.

In an effort to quantify the concept of mental health, techniques were developed, some of which were created long before the aforementioned WHO report. These techniques experienced a rise in their profile and use in line with increased public and policymakers understanding following the publication of the report. These early measures were useful to medical personnel and to policymakers, however as time progressed, these tests came under increased scrutiny as to their cost-effectiveness, usefulness and validity.

The established method of testing the effectiveness and validity of tests is conducted by investigating a test's psychometric properties. (Werneke, Goldberg, Yalcin & Ustun, 2000). This involves investigating their reliability and validity. Reliability refers to the extent to which a measure remains stable and returns consistent results, whereas validity refers to the extent to which a measure actually measures what it claims to. These can then be supplemented by investigating the sensitivity and specificity of the measurements. Sensitivity refers to the extent to which a test can identify the presence of a condition, and specificity refers to a test's ability to detect negative results (Altman and Bland, 1994).

The validity of these measurement techniques became the subject of much research with studies discovering methodological shortcomings that could distort data and led researchers and clinicians to come to false conclusions about the population they were examining.

One such example was that of social desirability which refers to the extent to which participants modify their answers to questions in order to avoid being portrayed as displaying what others may perceive as bad behaviour (Krumpal, 2013). Social desirability biases were found to be particularly prevalent when investigating issues that the participant considered sensitive (Van de Mortel, 2008). Due to the sensitivity and stigma associated with mental health problems, it was clear that techniques measuring mental health would be susceptible to distortion due to social desirability. It also became apparent that techniques and measurements of mental health performed differently when used in different geographical areas (Goldberg, 1998). This was a significant deviation from conventional medical practice as medical techniques were generally ubiquitous in the detection of conditions of a physical nature, regardless of the population being examined.

Within the medical field, mental health measures such as the Cornel Medical Inventory (CMI) and the General Health Questionnaire (GHQ) began to be used by medical practitioners worldwide in the 1950s. The CMI was published in 1949 (Brodman et al., 1949) and was a four-page questionnaire with 195 items. It was designed to collect information about a patient's medical history and allow medical practitioners to make inferences about an individual's health without necessarily seeing the patient face-to-face (Brodman, Erdmann and Wolf, 1949). The large size of the questionnaire inevitably led to its supersession with the GHQ overtaking the CMI as the most commonly used mental health questionnaire used in a medical setting in the 1970s (Goldberg et al., 1997).

The GHQ was created in response to widely divergent practices by General Practitioners of Medicine when analysing psychiatric conditions. Shepard et al., (1966) found such significant variations in the methods employed, diagnosis rate and surveys used that it was claimed that no useful conclusions could be drawn from them. Shepard et al. (1966) claimed that as much as 51% of the variations in his findings could be attributed to ecological and observational factors. In response to this Goldberg and Blackwell identified the need for a measurement tool that would 'eliminate observer variation so that comparisons could be made about the amount of psychiatric illness found in different areas' (Goldberg and Blackwell, 1970). Such a measurement tool would also need to be useful and applicable for use by a general practitioner of medicine as there was a growing appreciation that more mental health cases are dealt with by GPs than by the mental health sector (Manderscheid et al. 1993).

The GHQ was effectively the successor to the Cornell Medical Inventory (CMI), which was the most frequently used scale at the time (Brown and Fry, 1962). Goldberg and Blackwell (1970) conducted a similar test to the one that was conducted on the GHQ which is explained in the relevant section, where 2245 patients at a GP's surgery were asked to complete the CMI and subsequently given a psychiatric evaluation. CMI scores were evaluated, and participants were labelled as being at high risk of developing a mental health disorder or not. Following this, the correlation between the participants identified by the CMI and the clinical assessment of a psychiatrist was calculated. Under these conditions, the correlation between the psychiatrist's assessment and the Cornell Medical Inventory was only +0.19. Furthermore, of the 1484 patients who were mental health outpatients, 30% were misdiagnosed as they fell within the range of '*normal*' respondents (Goldberg and Blackwell, 1970).

In contrast to other earlier tests, the GHQ was specifically created for use in a General Practitioner and community setting to identify psychological distress (see section 1.3) as opposed to a specialised mental health facility (Goldberg and Blackwell, 1969). It aimed to provide information about an individual's mental state and did not attempt to investigate personality traits which leave an individual vulnerable to developing a mental disorder or any other predictive statements (Goldberg and Blackwell, 1970). Goldberg and Blackwell (1970) attributed its reliability in part to the GHQ's avoidance of personality traits. When the GHQ and Cornell Medical Inventory were subjected to validity testing the +0.19 correlation between test score and clinical assessment achieved by the Cornell Medical Inventory, did not compare favourably with the GHQ which scored a correlation of +0.80 under similar conditions (Shepard et al., 1966).

With the rise in prominence of statistical techniques such as Confirmatory Factor Analyses and Principal Component Analysis, which will be detailed more thoroughly in Chapter 3, it also became apparent that some measures of mental health, including the GHQ, were not performing as expected when subjected to such analysis. In the case of the GHQ-12, which claimed to measure one concept- an individual's predisposition to developing psychological conditions (Goldberg & Blackwell, 1969), numerous researchers claimed that they had found evidence of multiple factors being present in their analyses (Gratez, 1991; Worsely and Gribbin, 1977; Politi, 1994) This may have been attributable to a number of reasons including geographical variations, methodological issues or the presence of a wording effect, however in the absence of clear understanding of the factor structure of a measure, it could not be definitively proposed that it measured what it claimed to measure (Hankins, 2008). For this reason, later analysis will investigate the factor structure of the mental health measure in the context of a representative UK population.

1.3.- The GHQ

In this section, the GHQ and its purpose are discussed. The term GHQ refers to a family of tests ranging from the original 140 item pilot through to the 12 item version used throughout this thesis. Initially, the GHQ family of tests will be discussed; generally, however, specific reference will be made to the various versions of the test later in this chapter.

The General Health Questionnaire (GHQ) was originally created in 1970 by Sir David Goldberg (Goldberg and Blackwell, 1970) and was intended to be a way of detecting "psychiatric disorders...in community settings and non-psychiatric settings" (Goldberg, 1988, p. 191). It has been widely used by clinicians as an appropriate and reliable way of identifying numerous psychological conditions, including anxiety and depressive symptoms (Gelaye et al., 2015), (Patel et al., 2008), (Araya et al., 1992), (Abubakar & Fischer, 2012), (Padron et al., 2012). The GHQ is a method used to quantify the risk of developing psychiatric disorders which is referred to henceforth as psychological distress. This instrument targets two areas – the inability to carry out normal functions and the appearance of distress to assess psychological morbidity. It primarily focuses on the concepts of normal functioning in everyday tasks such as 'concentration' (GHQ-12 item 1), 'decision making' (GHQ-12 item 4) and 'sleep (GHQ-12 item 2)' and the emergence of symptoms which are likely to be associated with increased mental distress such as 'low confidence' (GHQ-12 item 10), 'heightened stress levels' (GHQ-12 item 5) and 'aversion behaviours' (GHQ-12 item 8) (Goldberg and Williams, 1988). It is not, however, recommended to be used as a diagnostic tool for specific conditions (Tait, Hulse and Robertson, 2002).

The GHQ is not a specific test but instead refers to a number of versions which are derived from the original 140 items piloted which are listed later in this chapter. All

GHQ variations comprise a number of standardised questions, a full list of GHQ-140 question headings is provided in table 1.1 alongside the shorter versions of the GHQ in order to display the items in each version. All items are responded to via Likert Scales and are equally weighted. These responses are then converted into scores which are summed. While numerous scoring procedures exist, invariably high scores indicate that the individual is likely to be at high risk of developing a psychological illness, whereas low scores indicate low risk. In larger versions, a mixture of positively and negatively scored items were included; however, the GHQ-12 only contains positively scored items to aid interpretation.

Table 1. 1

The Items of the Various Versions of the GHQ

Question topic	GHQ-140	GHQ-60	GHQ-30	GHQ-28	GHQ-12
Feeling well and in good health	✓	✓		✓	
Feeling in need of a good tonic	✓	✓		✓	
Run down and out of sorts	✓	✓		✓	
Feel that you are ill	✓	✓		✓	
Worried about losing weight	✓				
Putting on too much weight	✓				
Getting pains in your head	✓	✓		✓	
Really bad headaches	✓				
Noise in your ears	✓				
Tightness or pressure in your head	✓	✓		✓	
Couldn't give mind	✓				
Pins and needles in hands and feet	✓				
Hands shaking and trembling	✓				
Bothered by noise	✓				
Blushing easily	✓				
Might have a terrible disease	✓				
Concentrate on what you are doing	✓	✓	✓		✓
Worried about your heart	✓				
Collapse in a public place	✓	✓			
Aware of heart-thumping	✓				
Had palpitations	✓				
Frightened heart might suddenly stop	✓				
Hot or cold spells	✓	✓		✓	
Perspiring a lot	✓	✓			
Getting short of breath	✓				
Suffer from backache	✓				
Aches and pains	✓				
Been off food	✓				
Feeling nauseated	✓				
Getting indigestion	✓				
Food doing you no good	✓				
Griping pains in the belly	✓				

Troubled by wind	✓				
Any diarrhoea	✓				
Suffered from constipation	✓				
Waking early, unable to sleep	✓	✓	✓	✓	
Sleep hasn't refreshed	✓	✓			
Long time to get going	✓	✓			
Too tired to eat	✓				
Lacking in energy	✓	✓			
Lost sleep over worry	✓	✓	✓	✓	✓
Mentally alert or wide awake	✓				
Full of energy	✓	✓			
Restless and unable to relax	✓				
Easily fatigued	✓				
Spells of exhaustion	✓				
Too tired in evenings	✓	✓			
Tired and ready for bed	✓	✓			
Difficulty getting to sleep	✓	✓			
Sleeping pills	✓				
Staying asleep	✓	✓			
Unpleasant dreams	✓	✓			
Getting up to pass water	✓				
Restless and disturbed nights	✓	✓			
Giving vent to feelings	✓				
On the brink of tears	✓				
Given way to tears	✓				
Keep busy and occupied	✓	✓	✓	✓	
Having to do things repeatedly	✓				
Putting things off	✓				
Taking longer over things	✓	✓		✓	
Losing interest	✓	✓			
Sit down doing nothing	✓				
Losing interest in appearance	✓	✓			
Less trouble with clothes	✓				
Biting nails	✓				
Getting out of the house	✓	✓	✓		
Enjoy going out in the evenings	✓				
Able to go to the shops	✓				
Afraid to go out alone	✓				
Managing well as most people	✓	✓	✓		
Doing things well	✓	✓	✓	✓	
Time off work	✓				
Late to work	✓				
Difficulty keeping up	✓				
Satisfied with tasks	✓	✓	✓	✓	
Worried about the effect on	✓				
Getting on well with those close	✓				
Feel warmth and affection	✓	✓	✓		
What is going to happen to…	✓				
Losing temper	✓				
Feeling that you are a burden	✓				
On guard even with friends	✓				
All right with neighbours	✓				
Easy to get on with others	✓	✓	✓		
Spent time chatting	✓	✓	✓		
Others regard you as touchy	✓				
Need to talk to others	✓				
Afraid to say anything in case….	✓	✓			
Others seem to have misunderstood	✓				
Brooding over things and people	✓				

Others getting on your nerves	✓				
Difficulty speaking to strangers	✓				
Properly valued by others	✓				
No good to anybody	✓				
Playing a useful part	✓	✓	✓	✓	
Contented with your lot	✓				
Capable of making decisions	✓	✓	✓	✓	✓
Not able to make a start	✓				
Dreading things	✓	✓			
Constantly under strain	✓	✓	✓	✓	✓
Couldn't overcome a difficulty	✓	✓	✓		✓
Afraid to express foolish mistakes	✓				
Frightened to be on own	✓				
Confident about public places	✓				
Afraid of papers, TV	✓				
As though you were not really there	✓				
Finding life a struggle	✓	✓	✓		
Blaming yourself when things go wrong	✓				
Enjoy normal activities	✓		✓	✓	✓
Taking things hard	✓	✓	✓		
Edgy and bad-tempered	✓	✓		✓	
Scared and panicky for no reason	✓	✓	✓	✓	
Able to face problems	✓	✓	✓		✓
Worry over money	✓				
Everything on top of you	✓	✓	✓	✓	
Easily upset over things	✓				
Little annoyances, upset and angry	✓				
People looking at you	✓	✓			
Feeling easily hurt	✓				
Feeling unhappy and depressed	✓	✓	✓		✓
Losing confidence in yourself	✓	✓	✓		✓
Thinking of yourself as worthless	✓	✓	✓	✓	✓
Life entirely hopeless	✓	✓	✓	✓	
Hopeful about the future	✓	✓	✓		
Reasonably happy	✓	✓	✓		✓
Worrying unduly	✓				
Afraid you might lose control	✓				
Afraid something awful is about to happen	✓				
Going to have a nervous breakdown	✓				
Feeling nervous and strung up all the time	✓	✓	✓	✓	
Anxious someone may have been harmed	✓				
Felt that life isn't worth living	✓	✓	✓	✓	
Possibility of Suicide	✓	✓		✓	
Can't do anything due to nerves	✓	✓	✓	✓	
Thoughts going round and round	✓				
Unwelcome thoughts	✓				
Wishing you were dead	✓	✓		✓	
Suicidal thoughts	✓	✓		✓	

The GHQ has been described as being one of the most extensively used self-report questionnaires used to examine mental health (El-Metwally et al., 2018) and has been translated into different languages in order to facilitate its use worldwide. This is detailed more explicitly later in this chapter under 'Cultural and Demographic Variation'. It is important to note, however that while psychological distress and mental health refer to similar concepts, psychological distress refers specifically to an individual’s risk of developing a psychological condition.

When using the GHQ, it is important to note that while the concept of psychological distress could be considered an aspect of mental health and one could reasonably expect the two concepts to be highly correlated, that there is a distinction between the two concepts. Its relative popularity stands as a testament to its utility as a screening test for mental health problems and an indicator of psychological distress. That said, the GHQ has developed over time, and as it was refined, items in larger versions that were found to have very high correlations with other items and could be claimed to measure the same concept were removed. This is known as multicollinearity. This process inevitably resulted in redundant items being removed, and accordingly, several shorter versions have been developed.

1.3.1- GHQ- 140

The pilot version of the GHQ contained 140 items. These items were equally divided between 4 constructs that Goldberg (1972) described as 'depression', 'felt psychological disturbance', 'observable behaviours' and 'hypochondriasis'. These constructs or 'factors' represented latent variables with which specific items would exhibit strong correlational relationships. The 140 item pilot version of the test was originally given to 553 consecutive patients in a general practitioner's surgery, with a sample of 200 of these being given a subsequent assessment of their mental state by an independent psychiatrist who utilised a standardised psychiatric review (Goldberg and Blackwell, 1970). Following this pilot, the 140 items were subjected to 'item analysis' as it was felt that large numbers of items would undermine the aim of the questionnaire, specifically "*to be acceptable to a large range of respondents*" (Goldberg and Williams, 1988, p.35). Item Analysis refers to a range of statistical procedures which evaluate the characteristics of items within a questionnaire, usually for the purpose of informing a decision as to their inclusion in a scale. Item analysis has a long history of use within research with documented uses in the mid 1930s (Guilford, 1936). While numerous methods exist, item analysis generally consists of the discarding of items which display any of the three criteria listed below (Guildford, 1936)

- Demonstrating little variation in the participants' responses
- Exhibiting strong correlations with other items as to suggest multicollinearity
- Weakly correlate with all other items and resulting in a reduction of internal consistency

During item analysis of the GHQ-140, items were subject to Principal Component Analysis (PCA), a data reduction technique with aims to reduce large numbers of items into a smaller number of latent variables. These latent variables are referred to as Principal Components. A detailed discussion of PCA is given in Chapter 3. This analysis resulted in 60 items being selected, and this, in turn, led to the creation of the GHQ-60. A more detailed examination of this process is given when examining the GHQ-60, below. Over time a number of versions such as the GHQ -12, GHQ-20, GHQ-28 and GHQ-30 were developed in order to fulfil a number of different roles and be used in different settings.

1.3.2.- The GHQ-60

The GHQ-60 is described in the user guide as the ideal instrument for detecting participants displaying psychological distress (GL Education Group Ltd, 2018) however the claim that this item contains redundant items is acknowledged by the GHQ's own website (GL Education Group Ltd, 2018). The GHQ-60 contains 60 items and, as was earlier mentioned, was created after item analysis of the 140 item pilot study (Layton & Rust, 1986). This analysis allowed researchers to determine the usefulness of specific items and was used to inform researchers as to what items should be included in the GHQ-60. Forty items were removed in a screening exercise which focused on an individual's understanding and endorsement of the items, while the rest were subject to PCA. A description of PCA and other statistical techniques mentioned is provided in Chapter 3.

PCA showed that there were five principal components which were labelled 'The General Factor', 'Psychic Depression Versus Somatic Depression', 'Agitation Versus Apathy', 'Anxiety at Night versus Anxiety During Daytime' and 'Personal Neglect Versus Irritability'. Items deemed worthy of inclusion in the GHQ-60 included the 21 items which were closely associated with 'The General Factor' and 36 items which had close associations with the other 4 Principal Components. Finally, three items which were positively worded were included to ensure a balance of positively and negatively worded items (Goldberg, 1988). This was the first version actually published by Goldberg, and while undeniably containing some redundancy (GL Education Group Ltd, 2018), it remains an effective way of identifying those at risk of developing a mental disorder (GL Education Group Ltd, 2018). The GHQ-60 was designed to be unidimensional and therefore generated a single cumulative score, which was then compared against a pre-determined cut-off score. Those who fail to meet the cut-off score were, according to the questionnaire, at risk of developing psychological conditions.

Due to the large number of items in this test, there was a greater likelihood of numerous factors being present in the data than more refined versions of the GHQ. Numerous studies have been conducted examining the dimensionality or factor structure of the GHQ-60 with as many as eight different factors being extracted (Worsely, Walters & Wood, 1978). It is important to note that much of the research conducted into the psychometric structure of the GHQ-60 was conducted before more advanced techniques of factor analysis were developed, primarily using principal component analysis (PCA), a precursor to factor analysis. While similar to factor analysis, PCA does not assume an underlying correlation between variables. Using this technique, researchers commonly found that that there were six groups of variables which accounted for maximum variance in the data, known as Principal Components when tested in community settings (Vazquez-Barquero, 1988).

1.3.3.- GHQ-30

This version of the GHQ was developed from the larger and earlier developed GHQ-60 and involved removing all questions relating to physical illness in order to focus on psychological distress. This version of the GHQ is according to the user guide (Goldberg, 1988), the most widely validated version of the GHQ, with 29 different validation studies listed in that guide. This claim, however, may not be accurate at present as considerable research into the GHQ-12 has been conducted since the publishing of the user guide. Similar to the GHQ-60, this version produces a single cumulative score which is measured against a pre-determined cut-off score, where those who score over a specific score are considered to be at risk of developing a psychological condition.

1.3.4.- GHQ-28

This version was unique amongst GHQ variations as instead of a single cumulative score, this version contained four distinct subscales. These are listed below

- Somatic symptoms (questions 1-7)
- Anxiety and insomnia (questions 8-14)
- Social dysfunction (questions 15-21)
- Severe depression (questions 22-28)

The GHQ 28 was described in the user manual as the most well-known and popular version of the GHQ. While no individual thresholds for the subscales are set, and assessments are arrived at by using the sum of the subscales. The subscales provide a useful insight into providing individual diagnostic information and building up a profile on individuals. This has inevitably lead to difficulties as inherent in this method was an assumption that all subscales were equal in their ability to determine an individual's mental health. As the scores of each of the subscales are summed, there was a questionable rationale of treating the subscales separately.

1.3.5.- GHQ-12

The GHQ 12 is a questionnaire consisting of 12 questions and was created following in-depth item analysis of the GHQ-60 in order to determine the items which would provide the greatest utility in the smallest number of items. The top 12 items were derived from the GHQ-60 and selected by item analysis using PCA (Goldberg, 1988). It is considered to be reliable and 'remarkably robust' in comparison to some of the longer and more complex instruments such as the GHQ 28 (Goldberg et al., 1997). The GHQ-12 is the most extensively used version of the GHQ due to numerous factors such as its speed of completion, perceived ease of completion and simplicity of interpretation (Abubabkar & Fischer, 2013). This method was relatively free from redundancy as only the most useful items from the larger studies have been included (Goldberg et al., 1997). Like both the GHQ-60 and GHQ-30, this test outputs a single cumulative score from all items, and these are compared against a pre-determined cut-off point. Its obvious strength is that due to the small number of items, it is easy to administer and in a more clinical setting, is less likely to overwhelm respondents. Research by Smeeth and Fletcher (2002) clearly showed that an added benefit of short questionnaires was that they had improved response rates and therefore the data integrity lost through only using a short questionnaire can be compensated for by larger response rates and decreased likelihood of extremely damaging practices such as random answering.

This version of the GHQ is particularly appropriate when dealing with time constraints or participants with attention or reading difficulties, and research shows that the GHQ-12 remains a good indicator of psychological distress, despite being considerably shorter than other variations (Holi, Martunnen & Aalberg, 2010).

1.4.- Scoring Methods

The GHQ 12 utilises a scoring metric where items are responded to using a Likert Scale response with four different options. The items are unweighted and are summed to generate a single score which is measured against a pre-determined cut off point. Individuals who exceed the cut off are deemed to be at risk of developing a mental disorder. Implicit in this scoring method is the assumption that all items in the measure are equally important and that they all measure a single construct. Andrich and van Schoubroeck (1989) suggested that this approach may be too simplistic and conducted analysis into the different scoring methods used in the GHQ-12.

The questionnaire can be scored in numerous ways. Within the literature, Likert scoring metrics are frequently used. This metric consists of scoring the items using the four response options as a sliding scale and can be expressed as a 0,1,2,3 scoring system. Graetz (1991) stated that Likert style scoring provided "a more acceptable distribution of scores". Andrich and Shouebreck also stated that use of a Likert scale would allow a measure of the intensity of symptoms to be expressed by respondents, however, Andritch and Shoubreck (1989) also stated that it was standard practice to collapse the positive and negative scores into what they describe as a dichotomous scale. Schmitz (1999) described 'collapsed scoring' as the most common scoring method used in General Practice. Andritch and Shouebreck (1989) investigated four different scoring methods which were described as follows. (0,1,2,3), (0,0,1,1), (0,0,1,2) and a more complex weighted measure derived from analysis of three different groups with different levels of mental health. Further scoring methods were proposed by Goodchild and Duncan-Jones (1985) albeit, on the GHQ 30, that different scoring methods could be employed for the positive and negative items as they believed that the more traditional scoring methods were ill-suited to overcome the 'floor effect'. This effect occurs when a measure has a lower score limit, and a large number of participants score close to this lower limit. It usually indicates that a measure is ill-suited to deal with poor scoring participants. They also believed that the response of 'no more than usual' in relation to a negatively worded item, should be treated as an indicator of ill health, rather than wellness. Goodchild and Duncan Jones also suggested that trichotomous scoring warranted further investigation. Their proposal suggested a scoring procedure on positively worded items of (1,0,3,3) and (0,2,3,3) on negative items. Regardless of the merits of the numerous scoring procedures put forward by Duncan- Jones, in personal communications (Goodchild & Duncan-Jones, 1985) it was stated that "treating the GHQ response categories as providing even ordinal information is somewhat dubious, at least for the positive items". It must also be noted that while some researchers prefer a Likert scale, others state that this scoring method can yield spurious results as individuals can prefer moderate or extreme responses depending on their individual character traits (Goldberg & Williams, 1988).

1.5.- Psychometric Properties of the GHQ-12

The GHQ user guide mentions 43 validation studies (Goldberg & Williams, 1988) and since publication, the GHQ-12 has been further tested in terms of its psychometric properties. This section will give an overview of research into the psychometric properties, of the GHQ-12 in a number of settings, however more detailed discussions around the difference in the performance of the GHQ-12 in various cultures and settings are detailed in the relevant sections later in this thesis.

The accepted process for testing a questionnaire involves the investigation of its dimensionality, validity and reliability (Guilford, 1954), which collectively can be described as its psychometric properties. While the concepts of reliability and validity have already been explained, dimensionality refers to the presence of latent or unmeasured variables which only express themselves through relationships with items in a measure. These dimensions or factors as they are commonly known, allow researchers to ascertain if a scale is measuring a single concept or many. This concept is further explained in Chapter 3.

Winefield, Goldney and Winefield, (1989) conducted research into the reliability and internal consistency of the GHQ 30, 28 and 12 and found all versions to be reliable and internally consistent. The GHQ-12 was shown to be reliable in a German primary care setting; however, raised concerns over specific items in this setting. Specific items demonstrated poor positive response rate and low levels of specificity when investigated in this setting (Schmitz, 1999), however, the author concluded that in its entirety the GHQ-12 demonstrated high levels of reliability and internal consistency. High levels of internal consistency and reliability were also observed when investigating specific populations such as urological patients (Quek et al., 2001) and very elderly patients (Boey & Chiu, 2008), those with facial disfigurement (Martin & Newel, 2005), cancer patients (Rueter & Harter, 2001) and patients with skin conditions (Picardi, Abendi & Pasquini, 2001). Research into internal consistency was supplemented with studies investigating the sensitivity and specificity of the measurements of the GHQ (Gureje & Obikoya, 1990). This research found acceptable levels of sensitivity and specificity within the GHQ-12. Boey & Chiu (2008) found acceptable levels of sensitivity and specificity but demonstrated lower levels than other tests from which it was compared in this analysis.

As the GHQ is a self-report measure, it is subject to the myriad of limitations that these types of study commonly encounter. These limitations have led some clinicians to question its use as a diagnostic tool (Gilbody, Touse & Sheldon, 2001). As with all self-report measures, social desirability represents a significant bias that must be considered. Specifically to the GHQ, Parkes (1980) found evidence of social desirability in an occupational setting and Pevalin (2000) was able to find similar evidence in a general population sample.

While the debate about the factor structure of the GHQ-12 will be discussed in detail in Chapter 3, it must be noted that there is no clear consensus about the appropriate factor structure of the GHQ-12 and while this debate continues, the overall utility of the GHQ-12 remains questionable.

Hankins (2008) argued that any measure that is intended as a unidimensional measure could not be considered valid if it actually measures numerous factors. He also argues that most of the reliability analysis has assumed the unidimensionality of the GHQ, and that should this be found to be incorrect the reliability research would cease to be appropriate. He argued that if the measure was indeed multidimensional, then each of the factors must have their reliability tested individually. His conclusion was that while there was considerable research that supported the soundness of the GHQ-12 through reliability, validity and internal consistency investigations, they did not guarantee it without clarity around factor structure.

Finally, it has been shown that GHQ-12 responses are subject to the Hawthorne Effect. This phenomenon refers to the extent to which individuals modify their behaviour as a consequence of their awareness that their responses are being observed (Frank & Kaul, 1978). De Amici et al. (2000) reported clear evidence of the Hawthorne Effect by demonstrating that participants who were aware that they were part of a larger study reported higher levels of mental health than those who did not. This undermines the validity of a number of studies where participants were aware of their responses being analysed and as a result, may have significantly over-reported their mental health.

1.6.- Cultural and Demographic Variation

The GHQ has been extensively translated into at least 38 different languages including Arabic (Daradkheh, Ghubash & ElRufaie, 2001), Spanish (Graetz, 1991), Portuguese (Laranjeira, 2008) and Swedish (Winzer et al., 2014). Goldberg et al. (1997) stated that *'the GHQ works as well in the developed as in the developing world and only loses a small amount in translation'* (p.197). This was in response to a study which investigated the reliability and validity of the GHQ 28 and GHQ 12 in 11 different population centres, namely, Ankara, Athens, Bangalore, Berlin, Groningen, Ibadan, Mainz, Manchester, Nagasaki, Paris, Rio de Janeiro, Santiago, Seattle, Shanghai and Verona. In this study, it was found that sensitivity of the GHQ 12 ranged from 70.6% in Ankara to 86.7% in Bangalore, whereas the specificity ranged from 89.3% in Manchester to Verona 65.3%. Goldberg (1997) also states that while language effects were not significant, a very small difference was present in the performance of the GHQ-12 across various locations while this was not present for the GHQ-28.

Anomalous results were recorded when investigating Australian participants. Donath (2001) showed that the GHQ-12 was less effective in an Australian setting with the author writing, "in Australia the GHQ-12 appears to be a less useful instrument for detecting mental illness than in many other countries." His work was subsequently supported in 2009 when Whinefield et al. (2009) found GHQ-12 misclassification rates to be "unacceptably high" in this area. This unique finding stands in contrast to earlier work in Australian populations by Tennant (1977), which found the GHQ-12 to be effective and reliable. The research conducted by Whinefield and Donath (2009) and Tennant (1977) were conducted in different populations which may account for divergent results. Tennant's (1977) research was conducted in Sydney, a large city whereas Donath used data gained from the Australian National Survey of Health and

Wellbeing, and Winefield and Donath (2009) used a sample from South Australia, which included rural and urban dwellers. Secondly, it must be noted that Winefield and Donath's (2009) research specifically limit their samples to young adolescents, whereas Tennant (1977) does not. This divergence would suggest that the GHQ-12 is limited in its effectiveness either in Australian adolescents or that it is limited in Australian contexts outside of Sydney.

Research has shown that the pre-determined cut-offs on which the GHQ-12 depends may vary depending on the population investigated. Holi, Marttunen and Aalberg (2010) proposed unique cut-offs which they proposed were more appropriate to Finish people, whereas Makowska et al. (2002) identified optimised cut-offs for Polish people. Goldberg (1988) attempted to address why thresholds varied from place to place and suggested that these variations were due in part to the prevalence of multiple diagnoses in areas with higher thresholds being advised in areas with high rates of multiple diagnoses. He also suggested that areas where the discriminatory power of the GHQ-12 was lower than average, that a comparatively low threshold was advisable to protect the measure's sensitivity. These explanations were however subsequently described as unconvincing (Willmot et al., 2004), and researchers suggested that using pre-determined thresholds may not be appropriate in some circumstances, favouring a continuum-based approach (Furukawa, 2001).

While the above research demonstrated that the GHQ-12, when viewed in its entirety, did vary according to population, research also investigated how individual items performed in various cultural contexts. Sriram et al. (1989) conducted novel research into comparisons between the Indian and English versions of the GHQ. Firstly, using a technique known as 'translation-back-translation', they identified where the items of the GHQ might be altered in translation. They then administered English and

Indian versions of the GHQ over the course of a week to bilingual students and analysed the results. They found that all versions of the GHQ exhibited high coefficients for both internal consistency and reliability and that these figures were almost identical between Indian and English version. Specific item analysis showed considerable differences in items with endorsement ratings for the 60-item version only being similar in 4 items. This was attributable to semantic differences between the two languages. It must also be noted that, due to the fact that there was only a week between retests, it is possible that results were subject to retest biases. While the GHQ has been shown to be resilient to retest effects, most research into the retestability of the GHQ were conducted over longer periods such as three months (Gibbons, de Arévalo, & Mónico, 2004) or a year (Pevalin, 2000).

In Thailand, the translated version of the 28 item version of the GHQ was compared against a measure specifically created in Thai called the Thai Mental Health Questionnaire (TMHQ) (Phattharayuttawat, Kongsakorn & Ngamthipwattana, 2018). The test investigated the correlation between scores and mental illness diagnoses. It was found that GHQ-28 was found to identify similar proportions of participants as cases and similar correlations to clinical assessments as the TMHQ. While this research which used the GHQ-28 was deemed relevant as in this scenario, individual items were found to retain their meaning when translated into Thai.

In a Spanish population, Molina and Andrade (2002) wished to discover if the GHQ-28 was constructed in a Spanish population, would it have been constructed using different items. They concluded that if the process followed by Goldberg in constructing the GHQ-28, i.e. a PCA of the GHQ-60, with the items performing best forming the GHQ-28, that different items would have been selected. This would suggest that Spanish populations may respond to questions differently than English speaking populations and that it may be beneficial to introduce a bespoke version based on his analysis. It was noted however that the sample used in this sample did differ from the one used in the original validation study of Goldberg in that this study was based in a clinical setting, whereas Molina and Andrade (2002) used a general population setting.

Finally, research has identified how the factor structure of the GHQ-12 altered depending on the population tested. Molina and Rodrigo (2013) conducted a meta-analysis, an investigation of a number of independent studies all on the same topic, in order to ascertain if there were underlying trends and patterns present. Their research clearly illustrated one of the issues which has prevented researchers from coming to an agreed conclusion as to the most appropriate factor structure of the GHQ-12, was that of cultural variation. Molina and Rodrigo (2013) detailed how previous researchers had conducted research in different populations and compared how these different populations had responded when investigated. This meta-analysis was important because it was done in cognisance of more modern and advanced dimensional representations of the data such as Hankins (2008) which used a correlated errors approach and Ye (2009) which introduced a method factor. They found that within African populations, three-factor models tended to exhibit better fit (Abubakar & Fischer, 2012), and this was similar within Chinese participants (Ye, 2009). However, Smith et al. (2013) which investigated English participants suggested that a correlated errors approach was more appropriate in UK populations and this conclusion was corroborated by Aguardo et al. (2012) and Molina and Rodrigo (2013) in Spanish populations. Generally, it would appear that African populations are less affected by wording effects (Tomas, Guiterez & Sancho, 2015) and tend to favour three distinct factors, whereas western especially UK samples are more likely to be affected by wording effects and favour one factor with correlated errors to model wording effects

(Hankins 2008). More information on these dimensional representations is given in Chapter 3.

1.7.- Populations Tested

Research has been conducted on the use of the GHQ-12 on adolescents, adults and older adults (Costa, Barreto, Uchoa, Firma, Lima-Costa, & Prince, 2006). While none of the GHQ versions are recommended for use with children, the user guide of the GHQ-12 notes that some researchers have used the GHQ on children with apparent accuracy and reliability (Meltzer, 2003). For the most part, however, children should be assessed using a bespoke measure known as the Child Health Questionnaire. It must also be noted however that the term 'child' is not defined and research has struggled to identify a fixed age or developmental stage where the GHQ-12 becomes valid (Tait, Hulse & Robertson, 2002). Of the studies into adolescence, only one was found which used a sample that covered the entire age range of 11-20 (Shek, 1989) and this study did not report GHQ-12 scores by age. Within adolescents especially, there have been large reported gender differences in GHQ-12 response rates (Steptoe & Butler, 1996). These differences show that adolescent females report higher levels of psychological distress than their male counterparts. Furthermore, in adolescents, evidence of an age effect was found (Fichter et al., 1988; D'Arcy & Siddique, 1984) which has been attributed to life events such as examination stress and financial responsibility. This research demonstrates the difficulty of comparing GHQ responses between different age groups and different populations. This concept was further investigated in 2019 when Furham and Cheng (2019) were able to show that GHQ-12 scores exhibited statistically significant increases between the same people when aged 16 and then when retested at 30 in a UK population; furthermore, this effect was demonstrated in a number of settings with Kaur and Kaur (2018) demonstrating an age effect in Punjab.

In relation to the GHQ's performance as a screening tool amongst young people, generally, a specificity score of 0.8 is considered acceptable (Seva et al., 1992). With the exception of the aforementioned Australian anomaly, (Winefield, 1989), few studies failed to meet this threshold. One that didn't was an Italian study (Politi, Piccineli & Wilkinson, 1994) which failed to demonstrate sufficient specificity and another was Banks (1983) which found that while the GHQ-28 met the threshold the 12 and 30 item version did not. It must be noted that these two studies, in particular, tended to use younger participants than other studies using adolescent participants such as Seva et al., (1992) in Spanish populations and Radovanovic (1983) in former Yugoslavian participants which found the GHQ to have acceptable specificity.

The GHQ has also been extensively tested in regards to different occupational groups, including civil servants (Laaksonen et al., 2007) academic staff, (Campbell & Knowles 2007) and nurses (Elovinio et al., 2010). While the GHQ-12 has been shown to be acceptable in these populations, when investigating nurses, the GHQ-12 was compared against the PsychNurse Methods of Coping Scale, which was a measure specifically designed for nurses. The above test outperformed the GHQ-12 in terms of inter factor correlations, internal consistency and predictive validity (McElfatrick et al., 2000) suggesting that while the GHQ-12 remains an effective tool in this context, it can be outperformed by custom-built scales for specific circumstances. While this was not unexpected, it showed that research into very specialised populations might be better suited to less generalised measures.

Martikanen et al. (2003) conducted research in the specific area of white-collar workers at Whitehall. This research suggested GHQ scores were predicted by a number of variables, especially those relating to health and social position, rather than wealth which the researchers expected to be a significant predictor. It was suggested that these findings were due to the psychosocial benefits of being employed and having a strong social group rather than wealth. The fact that the non-financial benefits of employment were so emphasised in this study would suggest that the context of one's social and professional life play a significant role in their psychological distress. It is also interesting to note that of those who are employed, those with poor job security performed almost as poorly as those who were not employed at all highlighting the complexity of the relationship between employment and mental health.

1.8.- Mental Health Over Time

Longitudinal research has suggested that generally, self-reported mental health increases over time as a function of increasing age (Aldwin et al., 1989), however, this idea has been disputed and there continues to be disagreement about the 'normal' progression of mental health through life. Within the literature, some research into mental health generally suggested that mental health trajectories through life are U shaped (Schwandt, 2013). This phenomenon, commonly known as mid-life crises, however, was undermined by Bell's (2014) work where he proposed that this relationship was spurious. While initial assessment would suggest that GHQ-12 responses do exhibit a U-shaped design over time, his analysis, which was conducted on the period 1991-2008 and used the BHLS as a sample, argued that once confounding variables had been removed, that the trajectory of mental distress was linear over time in a representative UK population. Further research is given in both Chapter 4 and Chapter 7 when the longitudinal aspects of the GHQ-12 are discussed in detail.

1.9- Clinical/ General Population

The GHQ-12 was originally tested in a General Practitioner setting (Goldberg, 1970); however, its creator envisaged a test that detected psychological morbidity in both a clinical and non-clinical setting. Fernandes and Vasconcelos-Raposo (2012) found through multi-group comparison analysis that clinical populations had a considerably higher mean than the general population suggesting a higher level of distress in the clinical population which would be expected. The GHQ-12 was shown to be an effective screener for psychological conditions for individuals with a cancer diagnosis, however, was shown to be less specific than for example the Hospital Anxiety and Depression Scale (HADS) which was specially designed for detecting anxiety and depression in a hospital setting (Reuter & Harter, 2006). Aydin and Ulusahin (2001) showed the GHQ-12 to be effective in the clinical settings of COPD and Tuberculous patients. In line with earlier research, this would suggest that the GHQ-12 remains an effective screener, however pre-determined cut-offs may be malleable depending on the population tested. Picardi, Abendi and Pasquini (2008) found that while the GHQ-12 was a reliable and valid measure of psychological distress, the mean scores of different clinical groups varied. For example, those with more widespread skin conditions had higher mean GHQ-12 scores than those with more limited skin conditions. This may of course been attributable to the higher levels of distress experienced by these participants. Shelton and Herrick (2009) suggested that the lack of an item relating specifically to guilt may have been responsible for relatively poor performance in identifying individuals who were at risk of postnatal depression and suggested that the inclusion of such an item may make the instrument more effective in this scenario.

In 2017, a test of model stability over time between clinical and non-clinical populations was conducted, and invariance was found between the two populations (Ruiz, Garcia-Beltran & Suarez-Falcon, 2017). A detailed explanation of measurement invariance testing is given in Chapter 5. The level of invariance demonstrated between the two groups was sufficient to facilitate comparing means and suggested that important characteristics of the two groups stable over time.

In conclusion, the literature suggested that the GHQ-12 remains an appropriate scale to use in clinical and non-clinical populations, however, a number of important issues must be accounted for, especially cut-offs within specific clinical populations.

1.10.- Comparison with Other Measures

While the GHQ-12 has been described as the most popular technique for measuring mental health through self-report (El-Metwally et al., 2018), several alternative methods for measuring mental health exist. In this section, the literature which compared the GHQ-12 with other measures has been detailed below to inform why the GHQ-12 would be a more appropriate measure to accomplish the research aims of this thesis. It is also important to note that while the GHQ-12 is commonly used as a measure of mental health, it actually measures psychological distress. The differences between these two concepts are given in section 1.3.

Cornel Medical Inventory

As previously stated, the GHQ was primarily the successor to the Cornel Medical Inventory (CMI). The CMI exhibited significant divergence in its assessments of patient mental health when compared against practitioner assessments. When tested, the CMI exhibited a correlational relationship of 0.19 between the assessments derived from participant completion of the CMI and a Medical Professional's assessment (Shepard, 1966) compared to a 0.8 correlation achieved by the GHQ. The GHQ effectively superseded the CMI because it was able to provide more reliable results in a shorter and easier to interpret format.

Short Form-12

The mental health component of the Short Form (Ware, Kosinki & Keller, 1998) (SF-12), known as Mental Health Inventory (MHI-5), consisted of a five-item questionnaire which was part of a much larger general health questionnaire. It can be utilised as a measure of mental health and has been extensively validated and yields consistent, reliable results (Kelly et al., 2008). When compared against the GHQ-12 the MHI-5 compares favourably to the GHQ-12 in terms of being shorter, demonstrating similar internal consistency and it being part of a comprehensive questionnaire which can give a wider picture of health in the individual participant (McCabe et al., 1996). A major drawback, however, is that the MHI-5 does not have clinically validated cut-offs as of yet (Hoeymans, 2004), while the GHQ-12 does. It has been proposed to use the GHQ-12 as a benchmark to derive MHI-5 cut-offs points, however, this in itself does not match the clinical validation of GHQ-12 which has seen extensive validation in numerous populations which are detailed in the '*Cultural and Demographic*' section of this chapter.

Symptom Check List 90

The Symptom Check List 90 (SCL-90) is also a well utilised self-report measure in the field of mental health. It consists of 90 items which evaluate symptoms. This measure is much less practical for evaluating the general population as it is considerably longer than its competitors. It also does not converge on a single factor but instead yields nine scores relating to specific symptoms and three relating to global distress. Schmitz et al. (1999) found little difference in the performance of the SCL-90 and the GHQ-12 in relation to identifying cases with psychological issues. The GHQ-12 is superior to the SCL-90 as it is shorter, easier to score without sacrificing reliability and validity.

Short Warwick Edinburgh Test of Mental Well-being

The Short Warwick Edinburgh (SWE) test is a measure of well-being rather than mental health or associated concepts. While designed for use in the general population the scale's user guide states explicitly that the SWE was not intended to detect exceptionally poor mental health and therefore no strict cut-offs were developed or adopted (Stewart-Brown & Janmohamed, 2008). When compared against the GHQ-12 in a representative UK population, the SWE was found to have a correlation of -0.53 (McFall & Garrinton, 2011). The researchers described this as 'relatively low'. The SWE measures well-being in contrast to the GHQ-12, which measures psychological distress. The researchers suggested that these two concepts may not be congruent and that this may account for the poor correlational relationship. They also claimed that participants might report high levels of anxiety which would impact heavily on well-being scores but may not necessarily equate this with poor mental health (McFall & Garrinton, 2011). The GHQ-12 is a more appropriate mental health test for this analysis than the SWE for three identified reasons.

-The absence of pre-determined cut-offs

- The fact that the SWE is a measure of well-being and this appears to be a distinct and separate concept from mental health

-The SWE may struggle to accurately measure participants with extremely poor mental health.

Mental Health Prescriptions

Tseliou, Donnelly and O'Reilly (2018) evaluated if using self-report data was an appropriate and valid way of collating data to determine an individual's risk of psychological morbidity. GHQ-12 responses were compared against individual's medicinal prescriptions for 7489 patients in Northern Ireland**.** This research investigated whether questionnaires such as the GHQ could compete with techniques which identified participants as suffering from mental health issues based on their medical history. The GHQ was scored using a collapsed scoring technique which provided a smaller range of available scores than Likert scoring but is commonly used in clinical practice. Participants who scored over a predetermined score were identified as ‘cases’, as were participants who were currently receiving medication associated with mental health conditions.

Despite the two measures identifying different individuals as 'cases', both measures identified similar population distributions, however, of those participants who were identified as "cases" by the GHQ-12, only 53.6% of them received medication for mental health conditions. It is important to note that the GHQ-12 claims to measure psychological morbidity, which is not the same as medication usage, however, the authors argued that one could reasonably expect a strong correlation between the two. The researchers claimed that certain trends in the data which were observed by examining medication uptake, were not present in GHQ-12 analyses and therefore while acknowledging the differences between the two measured variables were not confident of the predictive properties of the GHQ-12.

Conclusion

In conclusion, the GHQ-12 represents an effective tool for use in general population settings when investigating poor mental health through psychological distress, both cross-sectionally and longitudinally. This is due to its tested reliability, validity, sensitivity and specificity in both clinical and general population settings. Its extensive use as a screener for mental health stands as a testament to its use in this field and ensures that in comparison to other measures of mental health it remains an excellent measure of mental health 40 years after it was published. That said, it would not be appropriate to conclude this comparison without acknowledging minor drawbacks, which have been alluded to throughout this chapter. While many of these such as social desirability and the 'Hawthorne Effect' are common to all self-report measures, researchers who use this measure should be cognisant of the continued disagreement over factor structure and variations over its performance according to geographical location.

1.11.- Rationale

The end goal of this thesis was to provide a unique perspective of psychological distress and how it changes relative to covariates over time in a representative UK population. Previous longitudinal research has focused on simple regressions (e.g. Sacker et al., 2011) or comparisons at two points in time (e.g. Gutherie et al., 1998) using a relatively limited range of covariates. The growth mixture modelling (GMM) techniques used in the latter parts of this thesis expanded upon the literature both in terms of the range of covariates selected and the complexity of the techniques used. GMM techniques can be used to separate participants displaying certain behaviour trajectories over time into specific groups, called classes. It also allows the effect of change in various covariates on class membership to be calculated.

In order to be in a position to conduct GMM based analyses, a number of sequential preparatory steps were completed to determine the properties of the GHQ-12, both initially at a cross-section in time and later in a longitudinal context. While each of the steps could be described as preparatory to the final chapter's analysis, each provided a unique contribution in their own right.

Before any longitudinal analysis was conducted, an appropriate dimensional representation of the data was determined in a cross-sectional context. This was particularly pertinent due to historic disagreements of the most appropriate dimensional in the literature. These disagreements have been particularly pertinent in the literature given recent additions using more complex statistical techniques which attempt to simulate method effects (Ye, 2009; Hankins, 2008). These are detailed more thoroughly in Chapter 3. In an effort to further inform the most appropriate dimensional representation of the data in the UKHLS, covariate analysis examined the relationships between factors contained within the various factor structures that demonstrated sufficient fit in Chapter 3. This was conducted on the basis that numerous studies (Shevlin & Adamson, 2004; Gao et al., 2004) which identified that dimensional representations which demonstrated good fit did not always represent a valid or meaningful representation of the data. These two chapters represented necessary preparations in order to conduct later analysis but also had merit in their own right. These Chapter's analyses determined the most appropriate dimensional representation of the data in light of the most recent additions to the literature and provided a rationale as to why some of the most popular dimensional representations of the data (Graetz, 1991) may not have been appropriate.

Once an appropriate dimensional representation of the data was determined, an analysis was conducted to determine whether important characteristics of this dimensional representation remained stable over time. This analysis was vital to the research aims of this thesis as if important characteristics of the factor structure changed over time, then the extent to which longitudinal analysis could be conducted would have been curtailed. While studies into longitudinal stability have previously been conducted, and stability was demonstrated, these were conducted using more conventional multidimensional representations of the data (Mäkikangas et al., 2006; Smith et al., 2012).

Once an appropriate dimensional representation was identified, and measurement invariance was established, the findings of the previous chapters facilitated longitudinal analysis which took the form of growth mixture modelling. This analysis identified if subpopulations of participants existed and separated the participants into groups characterised by the slope and intercept of their GHQ-12 responses. This analysis was necessary as the identification of an appropriate class structure was fundamental for the final chapter's analysis. It also represented a meaningful finding as research into longitudinal psychological distress trajectories had traditionally used more conventional dimensional representations of the data than was done in this analysis and were usually conducted in more niche populations than a general population sample (Høyer Holgersen et al., 2011).

The final stage of preparation consisted of identifying appropriate variables which were likely to influence psychological distress trajectories over time. Variables were chosen due to their correlation with the biopsychosocial model of mental health (Gathchel et al., 1996). Some of these variables, i.e. those that were time-varying required extra preparation to ensure that they were in a format that could be analysed in a GMM context. Covariates that were transformed were checked to ensure that linear interpretations of each covariate were appropriate in their specific context. This stage was fundamental to ensuring that the final chapter's analysis utilised appropriate covariates and that these covariates were in an appropriate format to facilitate this analysis.

Once previous chapter's analyses had identified an appropriate dimensional representation which remained stable over time, and appropriate trajectories displayed by the participants were extracted, covariates which were appropriately formatted to facilitate analyses would be investigated. This analysis would attempt to explain why participants exhibited the various trajectories extracted previously. This type of analysis would provide an insight into how a change in various covariates related to the trajectories displayed by participants. This form of analysis can therefore provide a perspective of the effect of various covariates in a way that more simplistic analytic frameworks could not.

1.12.- Research Objectives

Early research by numerous researchers had suggested that the GHQ-12 was not unidimensional and that it actually measured a number of sub aspects of mental health (Graetz, 1991; Schmitz, 1999; Worsely & Gribbin, 1997). More recent research had suggested that these multidimensional representations of the GHQ-12 were spurious and were attributable to method effects that were present in the wording of the items (Andrich & van Schoubroeck, 1989; Hankins, 2008; Ye, 2009). The first chapter's hypothesis was that the GHQ-12 could be treated as unidimensional once method effects were accounted for and that factor structures which accounted for these method effects would demonstrate comparable or better fit than those which suggested that the GHQ-12 was a multidimensional construct.

Once several factor structures were identified which demonstrated good fit, literature guidelines suggest that other considerations such as parsimony, interpretability and validity (Vanheule & Bogaerts, 2005) should be taken into consideration when determining an appropriate dimensional representation. This was particularly pertinent as research had previously identified that multidimensional representations of the GHQ-12 had demonstrated good fit but that the various component factors offered no unique predictive validity over a unidimensional model (Shevlin & Adamson, 2005). As a result, the hypothesis of this chapter would be that the factors of multidimensional representations of the GHQ-12 would offer no unique utility over treating the GHQ-12 as unidimensional.

Once an appropriate dimensional representation of the data was identified, preparatory analyses were required in order to determine if characteristics of the model changed over time and if so, to what extent. Previous research had suggested that both unidimensional (Smith et al., 2012) and multidimensional representations (Mäkikangas et al., 2006) of the GHQ-12 had demonstrated stability over time using various techniques which are detailed more thoroughly in Chapter 5. As research had implied model stability over time, this Chapter operated under the hypothesis that whatever dimensional representation of the GHQ-12 was selected would also demonstrate stability over time.

Longitudinal analyses were then conducted, and the extent of these analyses were dependent on the degree to which model stability was demonstrated previously. As strong measurement invariance was demonstrated, longitudinal analysis was not curtailed by the lack of stability over time (Geiser, 2010).

Longitudinal analysis was conducted using growth mixture modelling techniques which divided the participants in the sample into latent classes based on the trajectories of the GHQ-12 scores. This step was necessary as it provided a longitudinal framework from which to conduct the analysis of the final chapter. Research had previously suggested that in various contexts, latent subpopulations did exist in the populations that they tested. Høyer Holgersen et al. (2011), for example, extracted multiple classes in the context of those who had experienced a natural disaster. In light of the research which is more thoroughly detailed in Chapter 6, it was hypothesised that numerous latent subpopulations were present in the data and that these would become apparent when growth mixture modelling (GMM) techniques were utilised.

Finally, the covariates which would form the final chapter's analysis were identified and prepared in such a way as to be compatible with GMM techniques. While this chapter was primarily descriptive in nature and therefore would not require specific hypotheses, time-varying covariates were tested to determine if a linear model of their change over time was appropriate. For this analysis only, it was hypothesised that a linear interpretation of the time-varying covariates was appropriate.

The final chapter, as previously mentioned, used growth mixture modelling techniques to analyse how change in both time-varying and time-invariant covariates explained psychological distress trajectory over time. As it was a relative step into the unknown, research using this type of analysis was limited. More simplistic research had, however, suggested that individually all aspects of the biopsychosocial model (Gathchel et al., 1996) had a material impact on one's mental health. This research is discussed in detail in Chapter 7. Based on this research, it would be reasonable to hypothesise that when analysed together these relationships would manifest and that all aspects of the biopsychosocial model (Gathchel et al., 1996) would display a statistically significant relationship with class membership.

References

Abubakar, A., & Fischer, R. (2012). The factor structure of the 12-item General Health Questionnaire in a literate Kenyan population. *Stress and Health*, *28*(3), 248-254.

Aguado, J., Campbell, A., Ascaso, C., Navarro, P., Garcia-Esteve, L., & Luciano, J. V. (2012). Examining the factor structure and discriminant validity of the 12-item General Health Questionnaire (GHQ-12) among Spanish postpartum women. *Assessment*, *19*(4), 517-525.

Aldwin, C. M., Spiro, A., Levenson, M. R., & Bossé, R. (1989). Longitudinal findings from the normative aging study: I. Does mental health change with age?. *Psychology and Aging*, *4*(3), 295.

Altman, D. G., & Bland, J. M. (1994). Diagnostic tests. 1: Sensitivity and specificity. *BMJ: British Medical Journal*, *308*(6943), 1552.

Andrich, D., & Van Schoubroeck, L. (1989). The General Health Questionnaire: a psychometric analysis using latent trait theory. *Psychological Medicine*, *19*(2), 469-485.

Araya, R., Wynn, R., & Lewis, G. (1992). Comparison of two self administered psychiatric questionnaires (GHQ-12 and SRQ-20) in primary care in Chile. *Social Psychiatry and Psychiatric Epidemiology*, *27*(4), 168-173.

Aydin, I. O., & Uluşahin, A. (2001). Depression, anxiety comorbidity, and disability in tuberculosis and chronic obstructive pulmonary disease patients: applicability of GHQ-12. *General hospital psychiatry*, *23*(2), 77-83.

Banks, M. H. (1983). Validation of the General Health Questionnaire in a young community sample. *Psychological medicine*, *13*(2), 349-353.

Boey, K. W., & Chiu, H. F. K. (1998). Assessing psychological well-being of the old-old: A comparative study of GDS-15 and GHQ-12. *Clinical Gerontologist*, *19*(1), 65-75.

Brodman, K., Erdmann, A. J., & Wolff, H. G. (1949). *Cornell medical index health questionnaire: Manual*. Cornell University Medical College.

Brodman, K., Erdmann, A. J., Lorge, I., Wolff, H. G., & BROADBENT, T. H. (1949). The Cornell Medical Index: an adjunct to medical interview. *Journal of the American Medical Association*, *140*(6), 530-534.

Brown, A. C., & Fry, J. (1962). The Cornell Medical Index Health Questionnaire in the identification of neurotic patients in general practice. *Journal of Psychosomatic Research*, *6*(3), 185-190.

Campbell, A., & Knowles, S. (2007). A confirmatory factor analysis of the GHQ12 using a large Australian sample. *European Journal of Psychological Assessment*, *23*(1), 2-8.

Centofanti, S., Lushington, K., Wicking, A., Wicking, P., Fuller, A., Janz, P., & Dorrian, J. (2019). Establishing norms for mental well-being in young people (7–19 years) using the General Health Questionnaire-12. *Australian Journal of Psychology*, *71*(2), 117-126.

Costa, E., Barreto, S. M., Uchoa, E., Firmo, J. O., Lima-Costa, M. F., & Prince, M. (2006). Is the GDS-30 better than the GHQ-12 for screening depression in elderly people in the community? The Bambui Health Aging Study (BHAS). *International psychogeriatrics*, *18*(3), 493.

Creek, J., & Lougher, L. (2011). *Occupational therapy and mental health*. Elsevier Health Sciences.

Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and psychological measurement*, *10*(1), 3-31.

Daradkeh, T. K., Ghubash, R., & El-Rufaie, O. E. (2001). Reliability, validity, and factor structure of the Arabic version of the 12-item General Health Questionnaire. *Psychological Reports*, *89*(1), 85-94.

D'Arcy, C., & Siddique, C. M. (1984). Psychological distress among Canadian adolescents. *Psychological Medicine*, *14*(3), 615-628.

De Amici, D., Klersy, C., Ramajoli, F., Brustia, L., & Politi, P. (2000). Impact of the Hawthorne effect in a longitudinal clinical study: the case of anesthesia. *Controlled clinical trials*, *21*(2), 103-114.

Demyttenaere, K., Bruffaerts, R., Posada-Villa, J., Gasquet, I., Kovess, V., Lepine, J.P., … & Chatterji, S. (2004). Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. JAMA, 292(21), 2581–2590.

Department of Health, (2011). *No health without mental health*. UK: Department of Health,. http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_123766

Donath, S. (2001). The validity of the 12-item General Health Questionnaire in Australia: a comparison between three scoring methods. *Australian & New Zealand Journal of Psychiatry*, *35*(2), 231-235.

El-Metwally, A., Javed, S., Razzak, H. A., Aldossari, K. K., Aldiab, A., Al-Ghamdi, S. H., ... & Al-Zahrani, J. M. (2018). The factor structure of the general health questionnaire (GHQ12) in Saudi Arabia. *BMC health services research*, *18*(1), 595.

Elovainio, M., Kuusio, H., Aalto, A. M., Sinervo, T., & Heponiemi, T. (2010). Insecurity and shiftwork as characteristics of negative work environment: psychosocial and behavioural mediators. *Journal of Advanced Nursing*, *66*(5), 1080-1091.

Enis, B. M., Cox, K. K., & Stafford, J. E. (1972). Students as subjects in consumer behavior experiments. *Journal of Marketing Research*, *9*(1), 72-74.

Fay, M.T., Morrissey, M., Morrissey, M. and Smyth, M., 1999. *Northern Ireland's troubles: The human costs*. Pluto Press.

Fernandes, H. M., & Vasconcelos-Raposo, J. (2012). Factorial Validity and Invariance of the GHQ-12 Among Clinical and Nonclinical Samples. *Assessment*, *20*(2), 219-229.

Ferrari, A.J., Charlson, F.J., Norman, R.E., Patten, S.B., Freedman, G., Murray, C.J.L., … & Whiteford, H.A. (2013). Burden of Depressive Disorders by Country, Sex, Age, and Year: Findings from the Global Burden of Disease study 2010. PLOS Medicine, 10(11).

Fichter, M. M., Elton, M., Diallina, M., Koptagel-Ilal, G., Fthenakis, W. E., & Weyerer, S. (1988). Mental illness in Greek and Turkish adolescents. *European archives of psychiatry and neurological sciences*, *237*(3), 125-134.

Franke, R. H., & Kaul, J. D. (1978). The Hawthorne experiments: First statistical interpretation. *American sociological review*, 623-643.

Fuller, J., Edwards, J., Procter, N., & Moss, J. (2000). How definition of mental health problems can influence help seeking in rural and remote communities. *Australian Journal of Rural Health*, *8*(3), 148-153.

Furnham, A., & Cheng, H. (2019). GHQ score changes from teenage to young adulthood. *Journal of psychiatric research*, *113*, 46-50.

Furukawa TA,Goldberg DP (1999) Cultural invariance of likelihood ratios for the General Health Questionnaire. Lancet 353: 561–562

Furukawa, T. A., Goldberg, D. P., Rabe-Hesketh, S., & Üstün, T. B. (2001). Stratum-specific likelihood ratios of two versions of the General Health Questionnaire. *Psychological medicine*, *31*(3), 519.

Gao, F., Luo, N., Thumboo, J., Fones, C., Li, S. C., & Cheung, Y. B. (2004). Does the 12-item General Health Questionnaire contain multiple factors and do we need them?. *Health and Quality of Life Outcomes*, *2*(1), 63.

Gatchel, R. J., Peng, Y. B., Peters, M. L., Fuchs, P. N., & Turk, D. C. (2007). The biopsychosocial approach to chronic pain: scientific advances and future directions. *Psychological bulletin*, *133*(4), 581.

Geiser, C. (2012). *Data analysis with Mplus*. Guilford press.

Gelaye, B., Tadesse, M. G., Lohsoonthorn, V., Lertmeharit, S., Pensuksan, W. C., Sanchez, S. E., ... & Anderade, A. (2015). Psychometric properties and factor structure of the General Health Questionnaire as a screening tool for anxiety and depressive symptoms in a multi-national study of young adults. *Journal of affective disorders*, *187*, 197-202.

Gibbons, P., de Arévalo, H. F., & Mónico, M. (2004). Assessment of the factor structure and reliability of the 28 item version of the General Health Questionnaire (GHQ-28) in El Salvador. *International Journal of Clinical and Health Psychology*, *4*(2), 389-398.

Gilbody, S. M., House, A. O., & Sheldon, T. A. (2001). Routinely administered questionnaires for depression and anxiety: systematic review. *BMJ*, *322*(7283), 406-409.

Gl-assessment.co.uk. (2019). *General Health Questionnaire (GHQ)*. [online] Available at: https://www.gl-assessment.co.uk/products/general-health-questionnaire-ghq/ [Accessed 7 Jun. 2019].

Goldberg DP, Gater R, Sartorius N et al. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. Psychological Medicine; 27:191–197

Goldberg DP, Williams P: A User's Guide to the General Health Questionnaire. 1988, Windsor: nferNelson

Goldberg, D. (1972). The detection of psychiatric illness by questionnaires. London: Oxford University Press.

Goldberg, D. P., & Blackwell, B. (1970). Psychiatric illness in general practice: a detailed study using a new method of case identification. *Br med J*, *2*(5707), 439-443.

Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O., & Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological medicine*, *27*(1), 191-197.

Goodchild, M. E., & Duncan-Jones, P. (1985). Chronicity and the general health questionnaire. *The British Journal of Psychiatry, 146*(1), 55-61.

Graetz, B. (1991). Multidimensional properties of the general health questionnaire. *Social psychiatry and psychiatric epidemiology*, *26*(3), 132-138.

Griffith, G., & Jones, K. (2019). Understanding the population structure of the GHQ-12: evidence for multidimensionality using Bayesian and Exploratory Structural Equation Modelling from a large-scale UK population survey. *bioRxiv*, 584169.

Guilford, J. P. (1936). Psychometric methods. New York: McGraw-Hill.

Gureje, O., & Obikoya, B. (1990). The GHQ-12 as a screening tool in a primary care setting. *Social psychiatry and psychiatric epidemiology, 25*(5), 276-280.

Guthrie, E., Black, D., Bagalkote, H., Shaw, C., Campbell, M., & Creed, F. (1998). Psychological stress and burnout in medical students: a five-year prospective longitudinal study. *Journal of the Royal Society of Medicine*, *91*(5), 237-243.

Hankins, M. (2008). The factor structure of the twelve item General Health Questionnaire (GHQ-12): the result of negative phrasing?. *Clinical Practice and Epidemiology in Mental Health*, *4*(1), 10.

Hankins, M. (2008). The reliability of the twelve-item general health questionnaire (GHQ-12) under realistic assumptions. *BMC public health*, *8*(1), 355.

Harpman, T., Reichenheim, M., Oser, R., Thomas, E., Hamid, N., Jawsal, S., Ludermir, A. and Aidoo, M. (2003). How to do (or not to do). (20).. *Health Policy and Planning*, *18*(3), pp.344-349.

Hoeymans, N., Garssen, A. A., Westert, G. P., & Verhaak, P. F. (2004). Measuring mental health of the Dutch population: a comparison of the GHQ-12 and the MHI-5. *Health and quality of life outcomes*, *2*(1), 23.

Holgersen, K. H., Klöckner, C. A., Jakob Boe, H., Weisæth, L., & Holen, A. (2011). Disaster survivors in their third decade: Trajectories of initial stress responses and long-term course of mental health. *Journal of traumatic stress*, *24*(3), 334-341.

Holi, M. M., Marttunen, M., & Aalberg, V. (2003). Comparison of the GHQ-36, the GHQ-12 and the SCL-90 as psychiatric screening instruments in the Finnish population. *Nordic journal of psychiatry*, *57*(3), 233-238.

Kaur, M., & Kaur, R. (2018). Using the 12-item General Health Questionnaire (GHQ-12) to assess the Mental Health of Farmers of Punjab. *Int. J. Pure App. Biosci*, *6*(6), 905-912.

Kelly, M. J., Dunstan, F. D., Lloyd, K., & Fone, D. L. (2008). Evaluating cutpoints for the MHI-5 and MCS using the GHQ-12: a comparison of five different methods. *BMC psychiatry, 8*(1), 10.

Keyes, C. L. (2002). The mental health continuum: From languishing to flourishing in life. *Journal of health and social behavior*, 207-222.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, *47*(4), 2025-2047.

Laaksonen, E., Martikainen, P., Lahelma, E., Lallukka, T., Rahkonen, O., Head, J., & Marmot, M. (2007). Socioeconomic circumstances and common mental disorders among Finnish and British public sector employees: evidence from the Helsinki Health Study and the Whitehall II Study. *International Journal of Epidemiology, 36*(4), 776-786.

Laranjeira, C. A. (2008). General health questionnaire-12 items: adaptation study to the Portuguese population. *Epidemiology and Psychiatric Sciences, 17*(2), 148-151.

Layton, C., & Rust, J. (1986). The factor structure of the 60 item General Health Questionnaire. *Social Behavior and Personality: an international journal*, *14*(2), 123-131.

Mäkikangas, A., Feldt, T., Kinnunen, U., Tolvanen, A., Kinnunen, M. L., & Pulkkinen, L. (2006). The factor structure and factorial invariance of the 12-item General Health Questionnaire (GHQ-12) across time: evidence from two community-based samples. *Psychological assessment*, *18*(4), 444.

Makowska, Z., Merecz, D., Moscicka, A., & Kolasa, W. (2002). The validity of general health questionnaires, GHQ-12 and GHQ-28, in mental health studies of working people. *International journal of occupational medicine and environmental health*, *15*(4), 353-362.

Manderscheid, R. W., Rae, D. S., Narrow, W. E., Locke, B. Z., & Regier, D. A. (1993). Congruence of service utilization estimates from the Epidemiologic Catchment Area Project and other sources. *Archives of General Psychiatry*, *50*(2), 108-114.

Martikainen, P., Adda, J., Ferrie, J. E., Smith, G. D., & Marmot, M. (2003). Effects of income and wealth on GHQ depression and poor self rated health in white collar women and men in the Whitehall II study. *Journal of Epidemiology & Community Health*, *57*(9), 718-723.

Martin CR, Newell RJ: Is the 12-item General Health Questionnaire (GHQ-12) confounded by scoring method in individuals with facial disfigurement?. Psychology and Health. 2005, 20: 651-659. 10.1080/14768320500060061.

McCabe, C. J., Thomas, K. J., Brazier, J. E., & Coleman, P. (1996). Measuring the mental health status of a population: A comparison of the GHQ–12 and the SF–36 (MHI–5). *The British Journal of Psychiatry*, *169*(4), 517-521.

McCambridge, J., Witton, J. and Elbourne, D.R., 2014. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *Journal of clinical epidemiology*, *67*(3), pp.267-277.

McElfatrick, S., Carson, J., Annett, J., Cooper, C., Holloway, F., & Kuipers, E. (2000). Assessing coping skills in mental health nurses: is an occupation specific measure better than a generic coping skills scale?. *Personality and Individual Differences*, *28*(5), 965-976.

McFall, S. L., & Garrington, C. (2011). Understanding society: early findings from the first wave of the UK's Household Longitudinal Study.

McNemar, Q., 1946. Opinion-attitude methodology. *Psychological bulletin*, *43*(4), p.289.

Meltzer, H., Gatward, R., Goodman, R., & Ford, T. (2003). Mental health of children and adolescents in Great Britain. *International review of Psychiatry*, *15*(1-2), 185-187.

Mitchell AA, Werler MM, Shapiro S: Analyses and reanalyses of epidemiologic data: Learning lessons and maintaining perspective. Teratology. 1994, 167-168. 10.1002/tera.1420490304.

Mitchell, A. J., Rao, S., & Vaze, A. (2011). Can general practitioners identify people with distress and mild depression? A meta-analysis of clinical accuracy. *Journal of affective disorders*, *130*(1-2), 26-36.

Molina, J. D., & Andrade, C. (2002). The factor structure of the GHQ-60 in a community sample: a scaled versión for the spanish population. *Avances en salud mental relacional*, *1*(2).

Molina, J. G., Rodrigo, M. F., Losilla, J. M., & Vives, J. (2014). Wording effects and the factor structure of the 12-item General Health Questionnaire (GHQ-12). *Psychological assessment*, *26*(3), 1031.

Murray, C.J., Lopez, A.D. and World Health Organization, (1996). The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020: summary.

Padrón, A., Galán, I., Durbán, M., Gandarillas, A., & Rodríguez-Artalejo, F. (2012). Confirmatory factor analysis of the General Health Questionnaire (GHQ-12) in Spanish adolescents. *Quality of Life Research*, *21*(7), 1291-1298.

Parkes, K. R. (1980). Social desirability, defensiveness and self-report psychiatric inventory scores. *Psychological medicine*, *10*(4), 735-742.

Patel, V., Araya, R., Chowdhary, N., King, M., Kirkwood, B., Nayak, S., ... & Weiss, H. A. (2008). Detecting common mental disorders in primary care in India: a comparison of five screening questionnaires. *Psychological medicine*, *38*(2), 221-228.

Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of consumer research*, *28*(3), 450-461.

Pevalin, D. J. (2000). Multiple applications of the GHQ-12 in a general population sample: an investigation of long-term retest effects. *Social psychiatry and psychiatric epidemiology*, *35*(11), 508-512.

Phattharayuttawat, S., Kongsakorn, R., & Ngamthipwattana, T. (2018). General health questionnaire (GHQ) Thai version and the Thai mental health questionnaire (TMHQ). *Siriraj Medical Journal*, *52*(9), 599-607.

Picardi, A., Abeni, D., & Pasquini, P. (2001). Assessing psychological distress in patients with skin diseases: reliability, validity and factor structure of the GHQ-12. *Journal of the European Academy of Dermatology and Venereology*, *15*(5), 410-417.

Politi, P. L., Piccinelli, M., & Wilkinson, G. (1994). Reliability, validity and factor structure of the 12-item General Health Questionnaire among young males in Italy. *Acta Psychiatrica Scandinavica*, *90*(6), 432-437.

Quek, K. F., Low, W. Y., Razack, A. H., & Loh, C. S. (2001). Reliability and validity of the General Health Questionnaire (GHQ-12) among urological patients: A Malaysian study. *Psychiatry and Clinical Neurosciences*, *55*(5), 509-513.

Radovanović, Z., & Erić, L. J. (1983). Validity of the General Health Questionnaire in a Yugoslav student population. *Psychological Medicine, 13*(1), 205-207.

Reuter, K., & Härter, M. (2001). Screening for mental disorders in cancer patients–discriminant validity of HADS and GHQ-12 assessed by standardized clinical interview. *International Journal of Methods in Psychiatric Research, 10*(2), 86-96.

Roering, K. J., Schooler, R. D., & Morgan, F. W. (1976). An evaluation of marketing practices: Businessmen, housewives and students. *Journal of Business Research*, *4*(2), 131-144.

Ruiz, F. J., García-Beltrán, D. M., & Suárez-Falcón, J. C. (2017). General Health Questionnaire-12 validity in Colombia and factorial equivalence between clinical and nonclinical participants. *Psychiatry research, 256,* 53-58.

Schmitz, N., Kruse, J., & Tress, W. (1999). Psychometric properties of the General Health Questionnaire (GHQ–12) in a German primary care sample. Acta Psychiatrica Scandinavica, 100, 462– 468.

Schmitz, N., Kruse, J., Heckrath, C., Alberti, L., & Tress, W. (1999). Diagnosing mental disorders in primary care: the General Health Questionnaire (GHQ) and the Symptom Check List (SCL-90-R) as screening instruments. *Social psychiatry and psychiatric epidemiology*, *34*(7), 360-366.

Schooler, R. (1971). Bias phenomena attendant to the marketing of foreign goods in the US. *Journal of international business studies*, 71-80.

Seva, A., Sarasola, A., Merino, J. A., & Magallon, R. (1992). Validity test of the GHQ-28 items in a subsample of young people. *The European journal of psychiatry*.

Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual review of psychology*, *46*(1), 561-584.

Shek, D. T. (1989). Validity of the Chinese version of the General Health Questionnaire. *Journal of Clinical Psychology*, *45*(6), 890-897.

Shelton, NJ; Herrick, KG; (2009) Comparison of scoring methods and thresholds of the General Health Questionnaire-12 with the Edinburgh Postnatal Depression Scale in English women. **PUBLIC HEALTH** , 123 (12) 789 - 793.10.1016/j.puhe.2009.09.012.

Shepard, R. N. (1966). Metric structures in ordinal data. *Journal of Mathematical Psychology*, *3*(2), 287-315.

Shepherd, M., Irving, D., & Davies, G. (1966). Psychiatric illness in general practice London: Oxford University Press.

Sheth, J. N. (1970). Are there differences in dissonance reduction behavior between students and housewives?. *Journal of Marketing Research*, *7*(2), 243-245.

Shevlin, M., & Adamson, G. (2005). Alternative factor models and factorial invariance of the GHQ-12: a large sample analysis using confirmatory factor analysis. *Psychological assessment*, *17*(2), 231.

Smeeth, L., & Fletcher, A. E. (2002). Improving the response rates to questionnaires: Several common sense strategies are effective. *BMJ: British Medical Journal*, *324*(7347), 1168.

Smith, A. B., Oluboyede, Y., West, R., Hewison, J., & House, A. O. (2013). The factor structure of the GHQ-12: the interaction between item phrasing, variance and levels of distress. *Quality of Life Research*, *22*(1), 145-152.

Sriram, T. G., Chandrashekar, C. R., Isaac, M. K., & Shanmugham, V. (1989). The general health questionnaire (GHQ). *Social Psychiatry and Psychiatric Epidemiology, 24*(6), 317-320.

Steptoe, A. S., & Butler, N. (1996). Sports participation and emotional wellbeing in adolescents. *The Lancet, 347*(9018), 1789-1792.

Stewart-Brown, S., & Janmohamed, K. (2008). Warwick-Edinburgh mental well-being scale. *User Guide. Version*, *1*.

Tait, R. J., Hulse, G. K., & Robertson, S. I. (2002). A review of the validity of the General Health Questionnaire in adolescent populations. *Australian and New Zealand Journal of Psychiatry, 36*(4), 550-557.

Tennant, C. (1977). The General Health Questionnaire: a valid index of psychological impairment in Australian populations. *Medical journal of Australia*, *2*(12), 392-394.

Tomás, J. M., Gutiérrez, M., & Sancho, P. (2015). Factorial validity of the general health questionnaire 12 in an Angolan sample. *European Journal of Psychological Assessment*.

Tseliou, F., Donnelly, M., & O'Reilly, D. (2018). Screening for psychiatric morbidity in the population-a comparison of the GHQ-12 and self-reported medication use. *International Journal of Population Data Science*, *3*(1).

Van de Mortel, T.F., 2008. Faking it: social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, The*, *25*(4), p.40.

Vazquez-Barquero, J. L., Williams, P., Diez-Manrique, J. F., Lequerica, J., & Arenal, A. (1988). The factor structure of the GHQ-60 in a community sample. *Psychological Medicine*, *18*(1), 211-218.

Vos, T., Barber, RM., Bell, B., Bertozzi-Villa, A., Biryukov, S., Bolliger, I., …Murray, CJ.. (2013). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: A systematic analysis for the Global Burden of Disease study. The Lancet, 386(9995), 743–800.

Werneke, U., Goldberg, D. P., Yalcin, I., & Üstün, B. T. (2000). The stability of the factor structure of the General Health Questionnaire. *Psychological medicine*, *30*(4), 823-829.

WHO. (2001) What are social determinants of health? Retrieved from who.int/social_determinants/sdh_definition/en/ [Accessed 26/08/16].

WHO. (2014) What are social determinants of health? Retrieved from who.int/social_determinants/sdh_definition/en/ [Accessed 26/08/16].

Willmott, S. A., Boardman, J. A., Henshaw, C. A., & Jones, P. W. (2004). Understanding general health questionnaire (GHQ–28) score and its threshold. *Social Psychiatry and Psychiatric Epidemiology*, *39*(8), 613-617.

Winefield, H. R., Goldney, R. D., Winefield, A. H., & Tiggemann, M. (1989). The General Health Questionnaire: reliability and validity for Australian youth. *Australian and New Zealand Journal of Psychiatry*, *23*(1), 53-58.

Winzer, R., Lindblad, F., Sorjonen, K., & Lindberg, L. (2014). Positive versus negative mental health in emerging adulthood: a national cross-sectional survey. *BMC Public Health*, *14*(1), 1238.

World Bank. (1993). World Development Report 1993 : Investing in Health. New York: Oxford University Press. © World Bank. https://openknowledge.worldbank.org/handle/10986/5976 License: CC BY 3.0 IGO."

Worsley, A. & Gribbin, C. C. (1977). A factor analytic study of the twelve-item General Health Questionnaire. Australia and New Zealand Journal of Psychiatry 11, 269–272.

Worsley, A., Walters, W. A. W., & Wood, E. C. (1978). Responses of Australian patients with gynaecological disorders to the General Health Questionnaire: a factor analytic study. *Psychological medicine*, *8*(1), 131-138.

Worsley, A., Walters, W. A. W., & Wood, E. C. (1978). Responses of Australian patients with gynaecological disorders to the General Health Questionnaire: a factor analytic study. *Psychological medicine*, *8*(1), 131-138.

Ye, S. (2009). Factor structure of the General Health Questionnaire (GHQ-12): The role of wording effects. *Personality and I*

Chapter 2 - Overview of the Suitability of the Understanding Society Database for Proposed Research

2.1.- Introduction

The overall aim of this thesis was to model and explain psychological distress trajectories over time. In order to do this, Understanding Society was selected as an appropriate data source to facilitate the aims of this thesis as it contained several mental health measures collected over a number of time points, had a multitude of relevant covariates, was representative of the UK population and contained a large enough sample to generate sufficient statistical power to maximise the likelihood that findings would be statistically significant. The purpose of this chapter was to outline the Understanding Society database. Firstly, a historical overview of the database was provided for contextual purposes. Secondly, the structure of the data was discussed alongside an evaluation of the properties that made this resource suitable for the proposed analyses. Finally, an overview of the content that would be relevant to the thesis was provided. Weighting techniques which ensured that the sample remained representative of the UK population were also explained. Finally, missing data was investigated for the GHQ-12 data that will be used in subsequent analyses.

2.2.- Understanding Society Historical Overview

The Understanding Society; The UK Household Longitudinal Study (UKHLS) database was described as "*the largest panel study in the world intended to inform social and economic research*" (GL Education Group Ltd, 2019). It was designed as the successor to the British Household Panel Survey (BHPS) and included markers that allow UKHLS data to be linked to the BHPS participants. It was described as a much more ambitious survey than its predecessor and covered a much larger range of topics than the BHPS (Buck & McFall, 2012). The Survey examined approximately 40,000 UK households and approximately 100,000 individuals across a representative sample of the UK population. On its website, its stated aim was to provide high-quality longitudinal data to inform research across multiple disciplines (GL Education Group Ltd, 2019).

The UKHLS formed part of an international network of longitudinal surveys such as 'The Household, Income, and Labour Dynamics in Australia Survey' (Wilkin et al., 2015), 'The Panel Study of Income Dynamics' (Hill, 2001), 'The German Socio-Economic Panel Study' (Wagner et al., 1993), 'The Swiss Household Panel' (Budowski et al., 2001), 'The Survey of Labour and Income Dynamics in Canada' (Webber, 1994) amongst others. All these studies, while providing data at the individual level, incorporated a household element in their participant selection.

A household design differs from a more traditional cohort-based research as it takes participants from households and attempts to ensure that these households are representative of the general population, rather than basing representativeness on an individual's characteristics. This method of data collection has been shown to be particularly effective since it was piloted in the USA in the 1960s (Ferber, 1962) (Sirken, 1970). It is important to note, however, that household membership may not remain constant over time and if one simply investigated households that remain constant over time, then the sample would not be representative of the general population (Duncan and Hill, 1985). While participants were selected according to household membership data was recorded at the individual level. Household membership is not the recommended level at which data should be analysed and is suggested to be used as a contextual characteristic, rather than the unit of data collection (Buck & McFall, 2012). As participants are added and removed from households due to births and deaths, there is also more scope for the sample to remain representative over time than simple cohort design surveys (Buck & McFall, 2012).

2.3.- Structure of the Understanding Society Database

The UKHLS had a complex structure, the annual data collection technique and the response rates are detailed below, as are the various components of the sample.

2.3.1- Waves

In order to facilitate longitudinal analysis, data was collected annually via interviews and through an online survey in the Innovation Panel Component (see section 2.3.3.4 for more information on the Innovation Panel Component). The first wave of data was collected between January 2009 and December 2010, with subsequent waves following a similar pattern each subsequent year. Participants over 16 years old completed the full survey, however, those under 16 completed a youth questionnaire. Participants were added and removed at each wave as a result of births, deaths etc. It is also the case that some participants may not have responded to the interview or may have become unreachable, through address changes or migration etc. The UKHLS allows designated proxies to complete responses in the place of other householders if they are temporally unreachable however some questions, especially those which relate to an individual's mood are deemed inappropriate to be answered by proxy and are therefore not recorded.

2.3.2- Response Rates

In relation to Wave 1, interviewers were issued 93,712 household addresses. After initial screening, only 45,431 'General Population' household and 10,253 'Ethnic Minority Boost' households were found to be eligible. In order to retain desirable qualities within the data, a further 797 eligible households were added 'in the field'. Of the eligible participants, 30,169 households were successfully interviewed, with 50,994 individuals completing full or proxy interviews. Below, household response rates are detailed.

Table 2. 1

Household Response rates in Wave 1 of the Understanding Society Database.

	General Population (N)	Ethnic Minority Boost Sample (N)
Issued	49,915	43,797
Initial Eligible	44,916	9,971
Added Households	515	282
Total eligible	45431	10,253
Productive	26,075	4,060
Refusal	16,479	3,104
Non-Contact	1,777	989
Other Unproductive	1,118	2,100

While large numbers of addresses were issued to interviewers, large numbers of addresses were found to be ineligible. As can be seen from Table 2.1, a large number of the ethnic minority boost sample were found to be ineligible with comparatively small numbers in the general population sample. Reasons for ineligibility varied and the breakdown of reasons for ineligibility is shown below in Table 2.3.

Table 2. 2

Reasons for Ineligibility in Wave 1 of the Understanding Society Database

Reason for Refusal	General Population (N)	General Population (%)	Ethnic Minority Boost (N)	Ethnic Minority Boost (%)
Not eligible	4,999	100	33,828	100
Not yet built/under construction	64	1.3	66	0.2
Demolished/derelict	293	5.9	312	0.9
Vacant/empty housing unit	3,000	60	2,264	6.7
Non-residential address	501	10	697	2.1
Address occupied, no resident household	694	13	178	0.5

Communal establishment/institution	96	1.9	100	0.3
Resident household, not eligible for survey	4	0.1	29,993	88.7
Other ineligible	347	6.9	214	0.6

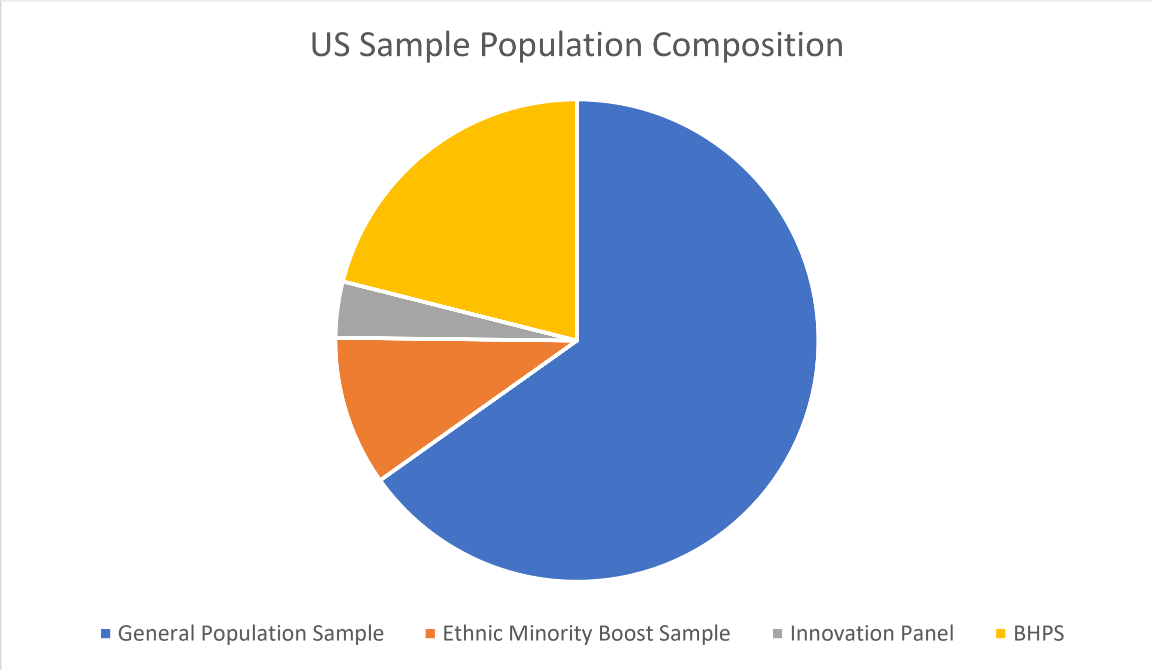
The large number of EMB households found to be ineligible was a result of the researchers wishing to retain a balance of ethnic minorities in line with the population of the UK. The largest reason for ineligibility for the GP was the address being unoccupied, which proportionally, was much higher than the EMB sample.

2.3.3- Database Components

The UKHLS survey population consisted of four components, a general population sample (GPS), the ethnic minority boost sample (EMB), the legacy British Household Panel Study (BHPS) and the Innovation Panel (IP). The boost samples were included to purposefully oversample specific demographics to capture variation in the data that may have been obscured if the sample was representative. Each of these components will be discussed in detail below, as each has unique features (Lynn, 2009). Of the total 40,000 households selected, 26,000 were from the general population sample, 4,000 from the Ethnic Minority Boost Sample, 1,500 from the Innovation Panel and 8,400 were chosen from the British Household Panel Survey, see Figure 2.1

Figure 2. 1

Sample Proportions for the Various Components of Understanding Society Database



2.3.3.1- General Population (GP) Sample

The Purpose of the GP was to provide a representative sample of the UK population, and a rigorous participant selection regime was adopted to ensure this representativeness. Within the GP sample, there were discrepancies between how households were selected between participants living in Great Britain, i.e. England, Scotland and Wales and participants living in Northern Ireland. Within Great Britain, the sample was clustered, stratified and subjected to weighting, however, in Northern Ireland, the sample was not clustered due to administrative constraints. Addresses were selected using what the database designers referred to as 'primary sampling units' (PSU's). These were postcodes or groups of postcodes from which the sample was drawn.

PSU's were selected using the following steps. PSU's with less than 500 addresses were merged with neighbouring PSU's, so long as their centres were within 15km of each other. Initially, PSU's were stratified by Government Regional Office (GRO), excluding Northern Ireland. PSU's were further stratified by the proportion of householders who held non-manual jobs as identified in the 2001 census. This resulted in 36 strata with three strata for each of the 12 GRO's. PSU's were further stratified into three bands by population density resulting in the 36 strata mentioned above being transformed into 108. Finally, PSU's were sorted according to the proportion of the population that was from an ethnic minority background as per the 2001 census returns.

PSU's were selected using systematic random sampling techniques with the likelihood of selection being adjusted for the number of addresses in each PSU. Overall, 2,640 PSU's were selected in Great Britain. In each of these PSU's, 18 addresses were selected, resulting in 47,520 participant households. In Northern Ireland, 2,400 participants were selected from The Land and Property Services (LPS) database of domestic properties and was an unclustered sample.

Due to methodological constraints, which resulted in a two-year cycle of fieldwork, the sample was divided into 24 sub-samples relating to their month of data collection. Each of these sub-samples was independently representative of the UK population to guard against response distortion that may have resulted from data collection timing e.g. such as the phenomena known as 'seasonal affective disorder' which is more prominent in winter months (Partonen & Lönnqvist, 1998).

2.3.3.2- Ethnic Minority Boost Sample (EMB)

The UK has been described as a society which has large numbers of minority groups which individually account for small proportions of the general population

(Buck & McFall, 2011). Participants were chosen to ensure that the five most common ethnic minorities, i.e. Indian, Pakistani, Bangladeshi, Caribbean and African, had samples exceeding 1000 participants. Within the dataset, there was deliberate oversampling of ethnic minorities, which was done in order to generate sufficient statistical power to capture variance within these minorities and to allow these minorities to be analysed in detail. This, however, was offset by the inclusion of clustering, weighting and stratification variables which were designed to adjust the influence these oversampled cases have on the overall dataset and to ensure that the data remains representative of the wider UK population. These variables were designed to allow for oversampling in certain cohorts of the study but to retain the overall integrity of the sample by manipulating the influence that individual cases have on a dataset. In order to ensure that oversampled minorities do not unduly influence an entire sample, minorities responses are reduced in influence to ensure that cumulatively they would affect the dataset as if they were sampled representatively. A more in-depth description of weighting, clustering and stratification variables are detailed in section 2.5.

Using census data, supplemented with more recent survey data, geographic areas of ethnic minority populations density exceeding 5% were selected. Further sub-sampling, based on the likelihood of residents belonging to an ethnic minority, was conducted as a way of increasing the efficiency of the fieldwork. In order to ensure that as many ethnic minorities as possible were represented in the data in sufficient quantities to generate sufficient statistical power, households which belonged to the less commonly encountered ethnic minorities were selected with greater priority, whereas some of the more commonly encountered ethnic minorities such as ‘Indian’ were occasionally deselected.

2.3.3.3- BHLS Component

From Wave 2, UKHLS incorporated participants who participated in the BHLS. They included unique identification variables which allowed researchers to link data from the BHLS and UKHLS and to conduct longitudinal analysis over a time period which otherwise would not be possible, albeit with a reduced sample than the UKHLS. A good example of a study which did this was Jenkins and Taylor (2011) who used linked UKHLS and BHLS data to demonstrate changes in employment in the period between 1991- 2009, which encompassed two major economic recessions on different demographics of the population. Jenkins and Taylor’s (2011) research does, however, detail the difficulties in comparing the BHLS and UKHLS due to methodological differences and sample inconsistencies.

The BHLS consisted of 5500 participants, however, it did not draw participants from the Scottish Highlands and Islands. While a relatively small part of the population, especially in such a small sample size, it does somewhat distort the representativeness of the sample population. In 1999 and 2001, booster samples in both Scotland and Northern Ireland respectively were introduced. This was primarily to investigate the effect of devolution on these regions.

2.3.3.4- Innovation Panel (IP)

This component of the sample was designed to test different methods of data collection. In total, the IP comprises of 2,760 participants selected from 120 postal sectors. It did not draw its participants from Northern Ireland nor from north of the Caledonian Canal. This resulted in unrepresentative findings when compared to the general UK population. Researchers attempted as far as was possible within the overall purpose of IP to mirror the techniques and sample selection of this component in order to maximise the ability to compare the data gained in the IP with general population data.

2.4.- Content of the Understanding Society Database

In line with the project aims, a number of criteria were selected that would ensure that the data was appropriate for the proposed analyses. These criteria are listed below.

- One or more established measures of mental health contained within the dataset
- A longitudinal aspect to investigate change over time
- A number of supplementary variables that have an established relationship with mental health.
- That data is collected at the individual level rather than aggregated across cohorts
- A rigorously selected sample to ensure that any findings can reasonably be defended as representative of the target population
- A large enough sample to ensure that all findings are likely to be statistically significant

The Understanding Society (UKHLS) database fulfilled the criteria listed above. The data contained within the UKHLS dataset is stored by the UK Data Service and is available to those who obtain the appropriate licences and permission from the Service. Data was collected longitudinally on an annual basis (referred to hereon as waves). The frequency of data collection allows researchers to track changes in participants over relatively short periods of time and has facilitated research into the impact of specific events such as the economic depression (Gush & Taylor, 2012).

As previously stated, the UKHLS investigated approximately 100,000 participants across 40,000 households. The size of this sample provided researchers with considerable statistical power, and it also afforded researchers the opportunity to study subpopulations. It has also allowed for more advanced statistical techniques, such as latent class analysis, to be used effectively (Davillas & Jones, 2019).

The designers of the UKHLS grouped questions on similar topics into modules, with specific modules being rotated over time while others were retained at every wave. Furthermore, some modules were specific to certain components of the sample population. The design overview (Burton, Laurie & Lynn, 2011) provided a table of content rotation and population-specific modules, which was provided below.

Table 2. 3

Module’s data collection schedules

Annual repeated measures	Rotating Modules	Ethnic minority boost	Youth self-completion
Basic demographic characteristics for all household members	Health-related behaviour, diet, exercise and sleep	Language and functional English literacy	Relationships with family and friends
Housing characteristics	Leisure and cultural participation	Migration history networks	Social networks Illicit/risky behaviour
Housing expenditure	Wealth, assets and debts	Remittances	Experience of education and aspirations
Household facilities and car ownership	Psychological/personality traits (‘Big 5’) Illicit/risky behaviour	Employment discrimination	Bullying at school and between siblings
Consumption and expenditure	Family and social networks	Harassment	Use of leisure time
Changes between waves– employment, fertility, partnering, geographic mobility, education and training, diagnosis health condition	Family relationships	British identity Additional items on family and social	Health, diet and obesity, exercise
Health status (e.g. SF12) , disability, GHQ-12	Local neighbourhood	Additional items on political engagement	Self-esteem and satisfaction with life
Education aspirations and expectations	Social support	Use of smokeless tobacco	Strength and Difficulties Questionnaire (SDQ)
Labour market activity and employment status, job search	Environmental attitudes and behaviour	Ethnic identity	Future aspirations for job, family, independence
Current job characteristics, basic employment conditions, hours of paid work, second jobs	Political engagement	Financial literacy and financial inclusion	Social and political attitudes and values
Childcare, other caring within and outside household	Employment conditions	Religious practice	Financial behaviour and paid work
Income and earnings	Travel behaviour	Civic capital/use of services	Caring responsibilities
Life satisfaction	Time use preference and risk		Ethnic and religious identity
Political affiliation	Trust		
Transport and communication access	Mental health and well-being		
Child development and parenting (from Wave 3 for children aged 3, 5 and 8 years)			

Expectations of retirement, Initial conditions, place of birth, education, family background, relationship and fertility information (at first interview only)	

The wide array of variables contained within the dataset facilitated the proposed research in later chapters. Of particular note was the mental health measures, i.e. the GHQ-12 the SF-12 which were conducted annually and measures of wellbeing which were conducted biannually. A number of social variables with established relationships with mental health such as income, life satisfaction and demographic data were also collected annually which would be utilised when conducting analysis utilising covariates as proposed in Chapters 4 and 8. Personality traits were classed as a rotating module, however, were only collected once every three years and therefore were only collected once, within the confines of the data used in this thesis, i.e. waves one through 5.

2.4.1- Participants

The characteristics of the population of the sample were described below. All information referred to below is taken from wave 1 of the UKHLS unless stated otherwise.

2.4.2- Geographic

US drew participants from all over the UK, and participants were selected from PSU's as described in section 2.2.3.1. Table 2.4 details the representation of certain geographical areas in the UKHLS database, compared against the Office for National Statistics population estimates for the same time period (ONS, 2018).

UKHLS provided a stratification variable based on the PSU, which adjusts the influence of participants from oversampled areas in order to retain the representativeness of the sample to the UK general population. That said, descriptive statistics for the graphical location were presented to provide context for the data collected.

Table 2. 4

A table showing the population breakdown of the UK alongside the sample breakdown of Understanding Society

Region	Actual population as per 2001 census (%)	Sample in UKHLS (%)	Variance
London	13.3	28.2	-14.9
SE England	13.7	11.6	2.1
East England	9.3	7.7	1.6
South West	8.4	5.3	3.1
North East	11	10.1	0.9
East Midlands	7.2	5.8	1.4
West Midlands	8.8	10	-1.2
Yorkshire	8.3	6.9	1.4
North East	4	2.9	1.1
ENGLAND	84	88.5	-4.5
WALES	4.8	3.2	1.6
SCOTLAND	8.3	5.8	2.5
NI	2.8	2.5	0.3
Total	99.9	100	

As can be seen from table 2.4, the sample population of the UKHLS database demonstrated significant under-representation for England (4.5%) but significant over-representation of participants in London (14.9%) relative to ONS figures (2018). It must also be noted that London based participants represent an exclusively urban cohort, whereas areas of northern England and the Regions are much more rural. This may be pertinent as research has suggested that London based residents are more susceptible to mental health issues as early as 1973 (Rutter, 1973).

Scotland was over-represented (variance between UKHLS and ONS proportions= 2.5%), with Wales being slightly over-represented (variance between UKHLS and ONS proportions= 1.6%). Northern Ireland was close to being represented appropriately (variance between UKHLS and ONS proportions=0.3%). There are large discrepancies within the UK regions in relation to how their participants' mental health was both measured and mental health prevalence in the regions (Mental Health Foundation, 2016)

2.4.3- Age

Participants of the UKHLS database were drawn from a variety of ages, consequently, a histogram detailing the age profile of the sample investigated was provided for context.

Figure 2. 2

A histogram of the Age Distribution of Participants at Wave 1

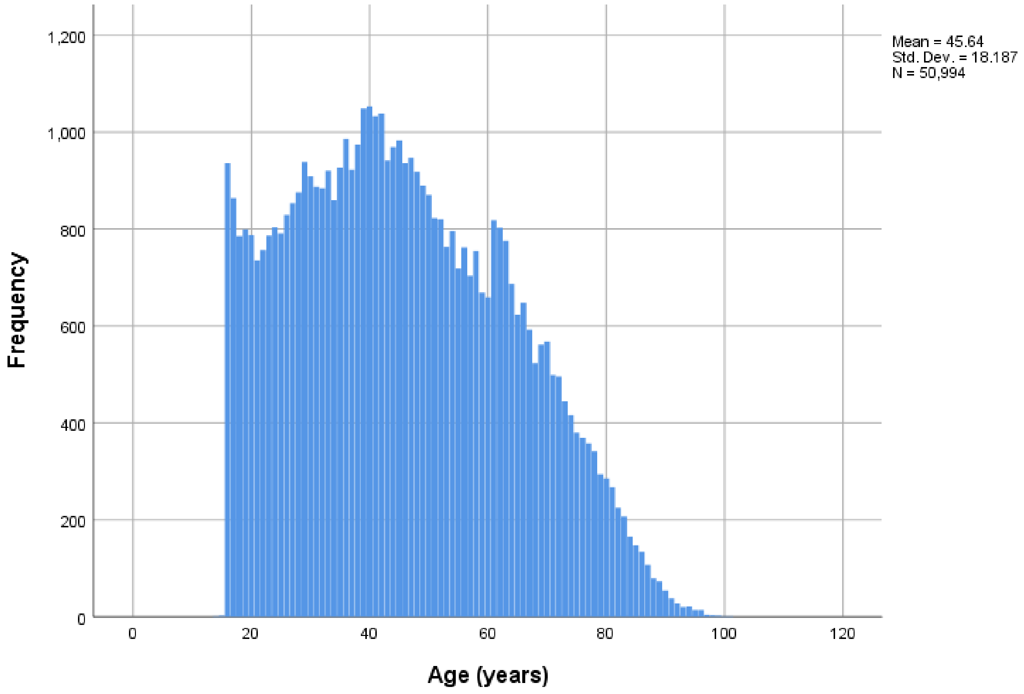


Figure 2.2 demonstrates that the age profile of participants was fairly normally distributed, albeit with positive skewed and with localised spikes at 18 and 61 years. No participant was aged under 18 years, however, subsequent participants may be added when they reached 18 years old at subsequent waves. It is important to note that techniques which handle missing data may extrapolate findings from participants who only turned 18 at later waves and therefore some responses from participants who were not yet 18 may be estimated.

2.4.4- Sex

According to census figures, as of 2001, 51% of the population was female and 49% male. While the statistics presented at wave 1 (see below) display a slightly larger female cohort than male, females were slightly oversampled.

Table 2. 5

Sex of Participants of Wave 1 of Understanding Society

Sex	Frequency in UKHLS	Proportions in UKHLS %	Proportion in 2001 census %
Male	23208	45.5	49
Female	27786	54.5	51
	50994	100	100

The UKHLS sample included 3.5% more females than would be representative of the UK population. This variation may have happened as participants were selected on households, not individual characteristics.

2.4.5- Ethnicity

As can be seen from the table, there was an element of oversampling of ethnic minorities. Ethnicities of the UK population were derived from census returns, however as they use different categorisations, comparisons between UKHLS and Census returns were difficult. It was also difficult to harmonise different census as the Scottish Census used different categorisation from other areas of the UK. In the table below, it was noted that the African categories used in Scotland could potentially capture White/Asian/Other African in addition to Black identities" (Office for National Statistics, 2013), which could not occur in non-Scottish census returns.

Table 2.6

Ethnicity Figures from the Understanding Society Database and 2011 Census

Response	Frequency	Sample %	Population %
Missing	28	0.1	N/A
proxy respondent	3262	6.4	N/A
Refused	19	0.0	N/A
don't know	7	0.0	N/A
British /English /Scottish / Welsh/	35881	70.4	90.1

Northern Irish			
Irish	717	1.4	N/A
Gypsy or Irish traveller	16	0.0	0.1
any other white background	1378	2.7	N/A
white and black Caribbean	332	0.7	N/A
white and black African	135	0.3	N/A
white and Asian	163	0.3	N/A
any other mixed background	192	0.4	2
Indian	1897	3.7	2.3
Pakistani	1435	2.8	1.9
Bangladeshi	1126	2.2	0.7
Chinese	318	0.6	0.7
any other Asian background	567	1.1	1.4
Caribbean	1119	2.2	N/A
African	1405	2.8	N/A
any other black background	85	0.2	N/A
Arabic	172	0.3	N/A
any other ethnic group	740	1.5	N/A
Total	50994	100.0	

Note. N/A figures are recorded due to the different criteria that different sources have adopted.

While administrative constraints rendered comparisons between the UK population and UKHLS sample difficult, it was clear that British, English, Welsh and Northern Irish participants were under-sampled relative to their proportion of the UK population. Census forms provided fewer options for participants to respond to, as an example, apparent under-sampling of '*other Asian*' participants may have been attributable to UKHLS participants having the opportunity to designate as a specific Asian subpopulation such as Arabic.

The inclusion of weighting variables ensured that oversampled ethnic minorities relative to their proportion of the UK population had their influence on the dataset reduced to protect the representativeness of UKHLS, however, these figures were provided to contextualise any research that used this database.

2.4.6- Mental Health Measures

In the UKHLS, three self-reported measures of mental health were collected, the GHQ-12, the SF-12 and the SWEMB. All of these reports have been explained in detail in Chapter 1. The GHQ-12 has been used extensively to investigate the effect of life events and certain sociodemographic factors on an individual's mental health, both in UKHLS and its predecessor, the BHPS (Hughes & Kumari, 2017) (Sacker et al., 2017) (McManus & Lord, 2012) (Bartley, 1994).

Within the BHPS, the GHQ-12 was found to be a reliable and stable measure across time (Pevalin, 2000) showing itself to be resistant to retest effects when the survey was administered annually. Even with the comparatively limited data afforded by the BHPS, Jenkins and Taylor (2011) were able to find statistically significant links between an individual's financial standing and their mental wellbeing. They also proposed that the effect was more pronounced in males. Due to the reduced sample of the BHLS and the absence of specific cohorts of the population, it was possible that the results were not representative of the entire UK population and that the results are not as sound as later investigations using the UKHLS database. Within the UKHLS database's first wave, it was found that GHQ-12 scores exhibited a negative correlation of -0.53 with the Short Warwick Edinburgh Test of Mental Wellbeing (SWEMWB) (Booker & Sacker, 2011). This correlation was described as 'relatively low', and the report suggests that positive wellbeing as claimed to be measured by the SWEMWB was a distinct concept to that of psychological distress, which the GHQ-12 claimed to measure. There was also a claim that participants may report high levels of anxiety but may not necessarily equate that to poor mental health. The same report analysed relationships between another measure of mental health, the SF-12 and GHQ-12 which were described as strong, albeit with high variation in the results obtained.

Furthermore, it has been suggested that UKHLS GHQ-12 scores may have been subject to bias. Brown et al. (2018) found that males especially were susceptible to under-reporting 7 out of the 12 items and that this is more pronounced in older people. Brown argues that given the applications of the GHQ-12, under-reporting in such a vulnerable population was a cause for concern. This research was conducted on BHPS waves 1-18 and UKHLS waves 1-7. The fact that the waves account for 25 years allows for extensive longitudinal research, and subsequent researcher should be aware that specific populations may under-report psychological morbidity.

During wave 6 of the UKHLS, the Innovation Panel was added. As previously stated, this component of the sample was primarily utilised to test the methodological effects of different data collection methods. GHQ-12 responses were found to be stable, whether they were administered using conventional or computer-based systems (Allum et al., 2014).

Table 2. 7

Mental Health Measures at Wave 1 of Understanding Society

	GHQ-12	SF-12 PCS	SF-12 MCS	SWE
Valid (N)	39700	47400	47400	38395
Valid (%)	77.85	92.95	92.95	75.29
Missing (N)	11294	3594	3594	12599
Missing (%)	22.15	7.05	7.05	24.71
Mean	11.05	49.49.	50.49	25.18
Median	10.00	53.4700	53.0400	26
SD	5.359	11.48891	10.11810	4.544
Range	36	70.57	77.11	7

N= Number of participants

SD= Standard Deviation

The figures shown in table 2.7 demonstrated that higher numbers of participants completed the SF-12 components than completed other self-report measures of mental health. The differentials in mean and median between the various tests were attributable to the different scoring mechanisms that these tests use with SF components being scored out of 100 and GHQ-12 scores scored out of 36. Further descriptions of these various measures are given in both Chapter 1 and Chapter 4.

2.5.- Missing Data

Within any database, there is the potential for data to be incomplete. How this data is treated will undoubtedly have a major impact on the analysis. As a result, this section will explore the properties of the missing data in the UKHLS GHQ-12 responses. As UKHLS participants have a number of responses which were categorised as non-valid, a table of possible responses was provided to contextualise further discussions.

Table 2. 8

Non-Valid Data Response Options in Understanding Society

Coded	Meaning
-9	Missing
-8	Inapplicable
-7	Proxy respondent
-2	Refused
-1	Don’t Know

Missing, inapplicable and proxy responses were categorised as missing as it would not have been appropriate to derive any further meaning from these responses, however, the interpretation of ‘don’t know’ and ‘refusal’ responses was less clear cut. As Dunrar (1998) explained in his research, analysts must consider what exactly a response of ‘don’t know’ or ‘refusal’ indicates. These responses could either denote missing or midpoint responses and should be treated accordingly. It was also important to note that GHQ-12 responses did not have a neutral response, which means that participants who wished to state a neutral response may have wished to utilise ‘don’t know’ responses as a way of indicating a response which was neither positive nor negative. Research by Durand and Lambert (1988) has shown that the percentage of a sample that gives a non-committal response, such as ‘refusal’ or ‘don’t know’ was affected by the lack of a specific neutral response. While it was possible that ‘don’t know’ may be used as an indication of a neutral response, it may also be used as a way of expressing that the participant does not understand the question (Durand and Butler, 1988).

Durand and Butler (1988) suggested three criteria which must be fulfilled in order to derive meaning from ‘don’t know’ or ‘refused’ responses, one of which was that there must be uniformity in what constitutes a non-committal response. It could not be said with any confidence that all participants who responded ‘don’t know’ were expressing a midpoint response, and therefore, it was difficult to justify interpreting these responses as anything other than missing.

Table 2. 9

Participants who completed the GHQ-12 at Each Wave

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Valid	39700	43437	40612	38852	37196
Valid (%)	77.85	79.56	81.65	82.39	82.84
Mean	11.05	11.20	11.08	11.01	11.18
Median	10.00	10.00	10.00	10.00	10.00
Std. Deviation	5.359	5.514	5.520	5.597	5.646
Range	36	36	36	36	36

The table shows that across the waves, a relatively constant number of participants completed the GHQ-12 task. The mean scores across time fluctuated between 11.01 and 11.20 however remained relatively constant throughout. At all waves, 10 was the most common response by participants, and standard deviations ranged from 5.359 to 5.646, and the standard deviation displayed a trend towards small increases over time, albeit negligibly.

2.5.1- Weighting Variables

The purpose of weighting variables were to ensure that despite oversampling in certain minorities, samples remain representative of the larger population. This will ensure that any findings are as representative as possible and can be generalised to, in the case of the UKHLS, a wider UK population. UKHLS provides weighting variables for both households and individuals but also allows for weighting of individuals who only completed certain parts of the survey, such as those who only completed the Adult Self Completion Instrument.

Knies (2015) provided a summary of the weighting strategy used in this database. During this paper, he suggests assumptions which researchers implicitly make if they chose to not use weighting variables. These are listed verbatim below (p. 67)

- *That all estimates of interest are the same in Northern Ireland as in the rest of the UK;*
- *That people of ethnic minority origin are the same as British;*
- *That people who live at an address with more than three dwellings or more than three households are the same as those who don't*
- *That people who responded at Wave 1 are the same with respect to your estimates as those who did not*
- *That people who continued to respond at later waves are the same as those who did not*
- *That people who responded to each particular instrument used in the analysis (individual interview, self-completion questionnaire etc.) are the same as those who did not, see Lynn, Burton et al. (2012).*

In order to ensure that all analyses could be as accurate as possible, it was deemed vital to include weighting, clustering and stratification in any analyses conducted within the course of this thesis. It was felt that the assumption around ethnic minorities was particularly appropriate as Modood et al. (1997) suggested that ethnic minorities experience differences in health and social scenarios ethnic majorities. Furthermore, Vega and Rumbaut (1991) suggested that ethnic minorities may respond differently to standardised tests, especially in a mental health setting than ethnic majorities. Furthermore, it was found that in longitudinal studies, attrition is not constant amongst the general population and failure to account for this will inevitably lead to samples becoming unrepresentative and subject to bias (Young, Powers & Bell, 2006).

2.5.2- Missing Data Patterns

While table 2.10 detailed the participants, who completed at least one item of the GHQ-12, a table which detailed the frequency of responses which was able to identify partial missingness in responses was provided below.

Table 2. 10

The Top 50 Patterns of Responses by Frequency of GHQ responses.

Frequency	Wave A	Wave B	Wave C	Wave D	Wave E
15473	✗	✗	✗	✗	✗
8311	✓	✗	✓	✓	✗
2234	✗	✗	✓	✓	✓
2162	✗	✗	✗	✗	✓
1837	✓	✗	✓	✓	✓
1576	✓	✓	✓	✗	✗
1244	✓	✓	✗	✓	✓

858	✓	✓	✓	✗	✓
687	✗	✓	✓	✓	✗
673	✓	✓	✓	✓	✗
578	✓	✓	✗	✗	✗
499	✗	✓	✓	✗	✓
480	✗	✓	✗	✗	✗
449	✗	✓	✓	✓	✗
412	✓	✗	✗	✓	✓
404	✓	✓	✓	✗	✗
399	✓	✗	✗	✗	✗
321	✓	✓	✗	✗	✓
307	✗	✗	✗	✓	✗
304	✗	✓	✓	✗	✗
275	✗	✗	✓	✗	✓
254	✗	✓	✗	✗	✓
241	✗	✗	✓	✗	✗
239	✗	✓	✓	✗	✗
213	✓	✗	✓	✗	✓
206	✓	✓	✗	✓	✗
187	✓	✗	✗	✗	✓
167	✗	✓	✗	✓	✗
129	✓	✗	✓	✓	✗
118	✓	✗	✗	✓	✗
83	✓	✗	✗	✓	✗
43	✓/✗	✓	✓	✓	✓
24	✓/✗	✓	✓	✓	✓
22	✓	✓/✗	✓	✓	✓
22	✓	✓/✗	✓	✓	✓

Note.

✓= fully completed

✘= fully missing

✘/✓= partially completed

From the table shown, the vast majority of missing data can be attributed to participants not completing entire waves. Partially completed waves are likely those which contain ‘refused’ and ‘don’t know’ responses which have been recoded as missing in order to make the data readable for MPLUS. Overall the patterns show little to no discernible configuration of participants who completed certain waves. As per the design of the study, participants may only have had the opportunity to complete the GHQ-12 at certain time points.

A total of 892 responses exhibited partially completed waves, however, the majority of these cases only expressed missing data on less than three variables, representing a maximum of 5% of their data being recorded as ‘don’t know’ or ‘refused’. Considering that 892 respondents accounted for 2% of the sample population and as previously mentioned, the majority of these participants exhibit less than 5% of their data being under these conditions, the effect was considered as negligible.

2.5.3- Listwise deletion

When confronted with missing data, decisions must be made as to whether to delete all incomplete data, known as listwise deletion or to conduct analysis on partially complete data. Listwise deletion was described as statistically problematic (Jeličić, Phelps & Lerner, 2009) and was even described as ‘evil’ (King et al., 1998) by researchers in the past. While this was considered hyperbolic, it was recognised that listwise deletion would result in the sample being used, reducing in size from 42,000 to closer to 10,000 across waves 1 to 5 using GHQ-12 data. The reduction in sample size was deemed to be unacceptably large, therefore it was decided to avoid listwise deletion unless absolutely necessary.

2.6.- Summary

The Understanding Society Database represents a suitable data source from which to investigate mental health trajectories over time. The longitudinal aspect of data collection alongside the wide array of covariates which supplemented the mental health measures facilitated the proposed analysis in future chapters. Further descriptive statistics on relevant data for each of the chapters are detailed in their respective '*methods*' sections.

References

Alcock, I., White, M. P., Wheeler, B. W., Fleming, L. E., & Depledge, M. H. (2014). Longitudinal effects on mental health of moving to greener and less green urban areas. *Environmental science & technology*, *48*(2), 1247-1255.

Allum, N., Auspurg, K., Blake, M., Booker, C. L., Crossley, T. F., d'Ardenne, J., ... & Lynn, P. (2014). Understanding Society Innovation Panel wave 6: results from methodological experiments.

Bartley, M. (1994). Unemployment and ill health: understanding the relationship. *Journal of Epidemiology & Community Health*, *48*(4), 333-337.

Bayliss, D., Olsen, W., & Walthery, P. (2017). Well-being during recession in the UK. *Applied research in quality of life*, *12*(2), 369-387.

Bell, A. Life-course and cohort trajectories of mental health in the UK, 1991-2008–a multilevel age-period-cohort analysis.

Binder, M. (2016). "… Do it with joy!"–Subjective well-being outcomes of working in non-profit organizations. *Journal of Economic Psychology*, *54*, 64-84.

Blackman, T., Harvey, J., Lawrence, M., & Simon, A. (2001). Neighbourhood renewal and health: evidence from a local case study. *Health & place*, *7*(2), 93-103.

Booker, C. L., & Sacker, A. (2012). Psychological well-being and reactions to multiple unemployment events: adaptation or sensitisation?. *J Epidemiol Community Health*, *66*(9), 832-838.

Booker, C. L., Rieger, G., & Unger, J. B. (2017). Sexual orientation health inequality: evidence from understanding society, the UK longitudinal household study. *Preventive medicine*, *101*, 126-132.

Booker, C. L., Rieger, G., & Unger, J. B. (2017). Sexual orientation health inequality: evidence from understanding society, the UK longitudinal household study. *Preventive medicine*, *101*, 126-132.

Brown, S., Gray, D., & Roberts, J. (2015). The relative income hypothesis: A comparison of methods. *Economics Letters*, *130*, 47-50.

Brown, S., Harris, M. N., Srivastava, P., & Taylor, K. (2018). Mental Health and Reporting Bias: Analysis of the GHQ-12.

Buck, N., & McFall, S. (2011). Understanding Society: design overview. *Longitudinal and Life Course Studies*, *3*(1), 5-17.

Budowski, Monica, Robin Tillmann, Erwin Zimmermann, Boris Wernli, Annette Scherpenzeel & Alexis Gabadinho. (2001) The Swiss Household Panel 1999-2003: Data for research on micro-social change, ZUMANachrichten 49, Jg. 25, November 2001.

Burton J, Lauire H and Lynn P. (2011) Appendix: Understanding Society Design Overview. In S. L. McFall and C. Garrington. eds. Understanding Society: Early findings from the first wave of the UK's household

Chan, T. W. (2018). Social mobility and the well-being of individuals. *The British journal of sociology*, *69*(1), 183-206.

Clark, A. E., & Oswald, A. J. (1994). Unhappiness and unemployment. *The Economic Journal*, *104*(424), 648-659.

Clark, A. E., Westergård-Nielsen, N., & Kristensen, N. (2009). Economic satisfaction and income rank in small neighbourhoods. *Journal of the European Economic Association*, *7*(2-3), 519-527.

Davillas, A., & Jones, A. M. (2019). *A Latent Class Approach to Inequity in Health Using Biomarker Data* (No. 2019-09). Institute for Social and Economic Research.

Demey, D., Berrington, A., Evandrou, M., & Falkingham, J. (2013). Living alone and psychological health in mid-life: the role of partnership history and parenthood status.

Demey, D., Berrington, A., Evandrou, M., & Falkingham, J. (2014). Living alone and psychological well-being in mid-life: does partnership history matter?. *J Epidemiol Community Health*, *68*(5), 403-410.

Dorsett, R., Rienzo, C., & Weale, M. (2015). Intergenerational and inter-ethnic well-being: an analysis for the UK. *London: National Institute of Economic and Social Research (NIESR Discussion paper, no 451)*.

Duncan G and Hill M. (1985) Conceptions of Longitudinal Households: Fertile or Futile?, Journal of Economic and Social Measurement, 13, 361-375.

Durand, R. M., & Lambert, Z. V. (1988). Don't know responses in surveys: Analyses and interpretational consequences. *Journal of Business Research*, *16*(2), 169-188.

Ferber, R. (1962). Research on household behavior. *The American Economic Review*, *52*(1), 19-63.

Ferrer-i-Carbonell, A. (2005). Income and well-being: an empirical analysis of the comparison income effect. *Journal of public economics*, *89*(5-6), 997-1019.

Giles P. (2001) An overview of the Survey of Labour and Income Dynamics (SLID). Canadian Studies in Population, 28, 363-375

Gl-assessment.co.uk. (2019). *General Health Questionnaire (GHQ)*. [online] Available at: https://www.gl-assessment.co.uk/products/general-health-questionnaire-ghq/ [Accessed 7 Jun. 2019

Glozah, F. N., & Pevalin, D. J. (2015). Factor structure and psychometric properties of the General Health Questionnaire (GHQ-12) among Ghanaian adolescents. *Journal of Child & Adolescent Mental Health*, *27*(1), 53-57.

Gush, K., & Taylor, M. (2012). Employment transitions and the recession. *UNDERSTANDING SOCIETY: FINDINGS 2012*.

Hill MS. (1992) The Panel Study of Income Dynamics: a User’s Guide. Sage Publications, Newbury Park CA

Hughes, A., & Kumari, M. (2017). Unemployment, underweight, and obesity: Findings from Understanding Society (UKHLS). *Preventive medicine, 97*, 19-25.

Jeličić, H., Phelps, E., & Lerner, R. M. (2009). Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. *Developmental psychology, 45*(4), 1195.

Jenkins, S., & Taylor, M. (2011). Non-employment, age, and the economic cycle. *Longitudinal and Life Course Studies*, *3*(1), 18-40.

Jokela, M., Batty, G. D., Vahtera, J., Elovainio, M., & Kivimäki, M. Socioeconomic inequalities in common mental disorders and.

King, G., Honaker, J., Joseph, A., & Scheve, K. (1998, July). List-wise deletion is evil: what to do about missing data in political science. In *Annual Meeting of the American Political Science Association*, *Boston*.

Luttmer, E. F. (2005). Neighbors as negatives: Relative earnings and well-being. *The Quarterly journal of economics*, *120*(3), 963-1002.

Lynn P. (2009) Sample design for Understanding Society, Understanding Society Working Paper 2009-01, No. 2009-01, Colchester: ISER, University of Essex.

Lynn, P., & Kaminska, O. (2010). *Weighting strategy for Understanding Society* (No. 2010-05). Understanding Society at the Institute for Social and Economic Research.

Lynn, P., Burton, J., Kaminska, O., Knies, G., & Nandi, A. (2012). An initial look at non-response and attrition in Understanding Society.

McBride, M. (2001). Relative-income effects on subjective well-being in the cross-section. *Journal of Economic Behavior & Organization*, *45*(3), 251-278.

McDool, E., Popli, G., & Ratcliffe, A. (2017). ETHNIC IDENTITY AND YOUNG PEOPLE’S OUTCOMES.

Mguni, N., Bacon, N., & Brown, J. F. (2012). The wellbeing and resilience paradox. *The Young Foundation, London*.

Modood, T., Berthoud, R., Lakey, J., Nazroo, J., Smith, P., Virdee, S., & Beishon, S. (1997). *Ethnic minorities in Britain: diversity and disadvantage* (No. 843). Policy Studies Institute.

Mohan, G., Longo, A., & Kee, F. (2017). Evaluation of the health impact of an urban regeneration policy: Neighbourhood Renewal in Northern Ireland. *J Epidemiol Community Health*, *71*(9), 919-927.

Office for National Statistics (2013) 2011 Census: Key Statistics and Quick Statistics for Local Authorities in the United Kingdom. Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/keystatisticsandquickstatisticsforlocalauthoritiesintheunitedkingdom/2013-10-11#ethnicity-and-country-of-birth

Oskrochi, G., Bani-Mustafa, A., & Oskrochi, Y. (2018). Factors affecting psychological well-being: Evidence from two nationally representative surveys. *PloS one*, *13*(6), e0198638.

Partonen, T., & Lönnqvist, J. (1998). Seasonal affective disorder. *CNS Drugs*, *9*(3), 203-212.

Paykel, E. S., Abbott, R., Jenkins, R., Brugha, T. S., & Meltzer, H. (2000). Urban-rural mental health differences in Great Britain: findings from the National Morbidity Survey. *Psychological medicine*, *30*(2), 269-280.

Pevalin, D. J. (2000). Multiple applications of the GHQ-12 in a general population sample: an investigation of long-term retest effects. *Social psychiatry and psychiatric epidemiology*, *35*(11), 508-512.

Rutter, M. (1973). Why are London children so disturbed?.

Sacker, A., Ross, A., MacLeod, C. A., Netuveli, G., & Windle, G. (2017). Health and social exclusion in older age: evidence from Understanding Society, the UK household longitudinal study. *J Epidemiol Community Health*, *71*(7), 681-690.

Sacker, A., Ross, A., MacLeod, C. A., Netuveli, G., & Windle, G. (2017). Health and social exclusion in older age: evidence from Understanding Society, the UK household longitudinal study. *J Epidemiol Community Health*, *71*(7), 681-690.

Schwandt, H. (2013). Unmet aspirations as an explanation for the age U-shape in human wellbeing. Centre for Economic Performance discussion paper.

Senik, C. (2004). When information dominates comparison: Learning from Russian subjective panel data. *Journal of Public Economics*, *88*(9-10), 2099-2123.

Senik, C. (2008). Ambition and jealousy: Income interactions in the 'Old'Europe versus the 'New'Europe and the United States. *Economica*, *75*(299), 495-513.

Sirken, M. G. (1970). Household surveys with multiplicity. *Journal of the American statistical Association*, *65*(329), 257-266.

Sorokin, P. A. (1959). Social and cultural mobility. *New York*, *4*, 99-145.

Stillwell, J. (2010). Ethnic population concentration and net migration in London. *Environment and Planning A*, *42*(6), 1439-1456.

Vega, W. A., & Rumbaut, R. G. (1991). Ethnic minorities and mental health. *Annual Review of Sociology*, *17*(1), 351-383.

Wagner G, Burkhauser R, and Bheringer F. (1993) The English language public use file of the German Socio- Economic Panel. Journal of Human Resources 28, 429-433.

Webb, E., Panico, L., Bécares, L., McMunn, A., Kelly, Y., & Sacker, A. (2017). The inter-relationship of adolescent unhappiness and parental mental distress. *Journal of Adolescent Health*, *60*(2), 196-203.

Webber M. (1994) Survey of labour and income dynamics: an overview. In Report1994, Cat. 75-201E. Statistics Canada, Ottawa.

Wilkins, R. (2015). *The household, income and labour dynamics in Australia survey: Selected findings from waves 1 to 12*. Melbourne: Melbourne Institute of Applied Economic and Social Research, The University of Melbourne.

Young, A. F., Powers, J. R., & Bell, S. L. (2006). Attrition in longitudinal studies: who do you lose?. *Australian and New Zealand journal of public health*, *30*(4), 353-361.

Zwysen, W. (2015). The effects of father's worklessness on young adults in the UK. *IZA Journal of European Labor Studies*, *4*(1), 2.

Chapter 3- The Dimensionality of the GHQ-12

3.1.- Abstract

Introduction

The General health questionnaire (12 item version) is one of the most extensively used self-report questionnaires used to assess mental health. Since its creation, there has been debate about its psychometric properties, especially its dimensionality. This chapter used confirmatory factor analysis techniques to test competing dimensional representations that have been identified within the literature and to identify the most appropriate dimensional representation of the GHQ for the participants of the Understanding Society database.

Methods

Confirmatory Factor Analysis was conducted to test numerous competing models identified in the literature using data obtained from wave 1 of the UKHLS. These models ranged from the unidimensional model envisaged by its author (Goldberg, 1997) to multidimensional models (e.g., Graetz, 1991; Martin, 1999 & Worsely & Gribbin, 1997), to those that model method effects (Hankins, 2008 & Ye, 2009). Fit statistics, factor loadings, and factor correlations were investigated to determine the most appropriate representation of the UKHLS participants to inform suitable models which would be taken forward into further analysis in this thesis.

Results

The results showed that several models had acceptable fit to warrant further analysis in Chapter 4. Politi's two-factor model, (1994) Graetz's three-factor model, (1991), Hankins' correlated errors model (2008) and Ye's method factor model (2009), performed strongly and will be investigated further in Chapter 4. Generally, those

models which attempted to model method effects had highly correlated factors, whereas multidimensional models demonstrated comparatively weak factor loading.

Conclusions

The results suggested that some of the dimensional representations were inappropriate for this data and exhibited poor fit and/or poor factor loadings. Some of the models that performed well, modelled wording effects (e.g. Ye, 2009), while others proposed a number of distinct constructs within the data (e.g., Graetz, 1991). The validity of these factors will be investigated in a subsequent analysis in Chapter 4 of this thesis.

3.2.- Introduction

The purpose of this chapter was to identify the most appropriate dimensional representation of the GHQ-12 for the population of the Understanding Society: United Kingdom Household Longitudinal Study (UKHLS) database. Within the literature, there has been much disagreement surrounding the most appropriate dimensional representation of the GHQ-12, which is detailed below.

Initially, the GHQ-12 was designed as a unidimensional measure of psychological distress as evident by the scoring matrix (see chapter 1), however early research into the dimensionality of the measure identified a number of multidimensional dimensional representations to demonstrate superior fit to the unidimensional model originally envisaged (Politi, 1994; Andrich & Schoubroeck, 1989; Schmitz, 1999; Martin, 1999; Graetz, 1991; Worsely & Gribbin, 1997). The studies suggested that the GHQ-12 measured several distinct concepts, rather than as previously suggested, the single concept of psychological distress. Until 2008, Graetz’ three factor (1991) dimensional representation was the most commonly accepted in the literature, however, both Hankins (2008) and Ye (2009) demonstrated subsequently that techniques that simulated method effects caused by the wording of the items of the GHQ-12 demonstrated superior fit than the previously mentioned multidimensional representations. Finally, a meta-analysis by Molina and Rodriguez (2013) investigated previous studies which tested the factor structure of the GHQ-12 and found that either Graetz (1991), Hankins (2008) or Ye (2009) demonstrated superior fit over competing models depending on the population tested.

This chapter investigated the properties of all established dimensional representations within the literature utilising a battery of fit statistics to investigate fit and subsequently investigated the factor loadings and correlations to inform the identification of the most appropriate dimensional representation further.

3.2.1- Factor Structure of the GHQ

Researchers have come to very different conclusions as to the factor structure of the GHQ. A Unidimensional model was envisaged by the GHQ's author (described in chapter 1), however, multidimensional models (e.g., Graetz, 1991 Martin, 1999 & Schmitz, 1999) were found to demonstrate superior fit. As factor structures will feature prominently in this chapter and others, an outline of the debate regarding factor structures is outlined below.

The GHQ-12 claims to measure the predisposition of an individual towards a psychiatric disorder, a singular concept, which is important for a number of reasons. The scoring metric used, which is described in chapter 1, depends on the measure's unidimensionality. The scoring method used by the GHQ-12, a Likert Scale ranging from 1-4, with participants scores being compared against predetermined cut-off scores (see chapter one) implies a number of things:

- That all items are equally valid measures of the factor in question
- That all items have an equal gradation in their scores
- That all items contribute to an aggregate score of one factor
- That the population will behave relatively uniformly in relation to predetermined cut-offs

Rey et al. (2014) hypothesized that the various findings regarding the factor structure of the GHQ-12 (see table 3.1) were explained by three methodological factors.

1- That negative responses had an element of ambiguity which may appear as separate factors under certain conditions
2- Multiple scoring schemes may yield different results (see chapter 1)
3- Inappropriate estimation methods (see chapter 1)

Rey's research suggests that when the above factors are accounted for, the GHQ remains a unidimensional measure. Research conducted by Shevlin and Adamson (2006) suggested that while multidimensional models in the literature demonstrated good fit of the data, the factors generated are highly correlated and do not demonstrate unique predictive power when regressed on relevant covariates. As a result, Shevlin and Adamson (2006) suggested that from a practical point of view, there may be no benefit in treating the GHQ-12 as multidimensional.

Furthermore, much of the research which underpinned the GHQ-12 was conducted on the basis of assumed unidimensionality of the scale. While this supporting research has been described as '*extensive*' (Navarro et al., 2007), Hankins (2008) argued that this supporting literature would be inapplicable if the GHQ-12 was shown to be multidimensional. Hankins (208) went on to claim that one must question whether it would be appropriate to use this measure in a clinical setting until more clarity is obtained.

When considering the regional and cultural variation of the performance of the measure, it has been shown that GHQ-12 responses vary depending on the population tested (Van Hermet et al., 1995). These regional and cultural variations have led to different cut-offs for various populations as well as various factor structures depending on population (see chapter 1- Cultural Variations). A meta-study by Molina and Rodriguez (2013) analysed multiple populations from across the globe to test factor structures. While not stated by the researchers, it would appear that there were differences in the most appropriate dimensional representation of GHQ-12 responses depending on where the participants were from. Western populations tended to favour Hankins' (2008) model (see table3.4), whereas eastern populations tended to favour Graetz' (1991). As many of the studies into the factor structure of the GHQ-12 were conducted using differing populations, this may, in turn, explain why different studies found varying dimensional representations to be the most appropriate for their respective populations. It is also important to note that while the GHQ-12 was designed to be unidimensional, it was derived from the GHQ-140, which was acknowledged as measuring different factors (GL Education group, 2018).

3.2.2- Proposed Factor Structures of the GHQ-12

Below is an overview of the studies that were discussed in this chapter. This table details the factor structure of each of the models within the literature. As can be seen, similar items load onto the various factors in some models such as Politi (1994) and Andrich and Schoubroeck (1989), albeit, in this case, these factors are differently labelled. Other structures such as Hankins' (2008) model represent a different conceptual approach, with method effects being simulated through the use of correlated errors (see section 3.2.5). For convenience, there are also tables which detail the structure of each of the models when they are discussed later in this chapter.

Table 3. 1

Summary of Model Structures for the GHQ 12

Item	1 factor	Politi (1994)		Andrich and Schoubroeck (1989)		Schmitz (1999)		Martin (1999)			Graetz (1991)			Worsely and Gribbin (1997)			Hankins (2008)	Ye (2009)	
		Dysphoria	Social dysfunction	Positively worded	Negatively worded	Anxiety/ depression	Social performance	cope	Stress	depression	Anxiety/ depression	Social dysfunction	Loss of confidence	Social Performance	Anhedonia	Loss of confidence	1 factor	1 factor	Met fac
1	*		*		*	*		*				*		*			()	*	*
2	*	*		*		*			*		*			*			*	*	
3	*		*		*	*	*	*				*		*			()	*	*
4	*		*		*		*	*				*		*			()	*	*
5	*	*		*			*		*		*				*		*	*	
6	*	*		*		*				*	*				*		*	*	
7	*		*		*	*			*			*			*		()	*	*
8	*		*		*		*	*				*			*		()	*	*
9	*	*		*			*			*	*					*	*	*	
10	*	*		*		*				*			*			*	*	*	
11	*	*			*	*				*			*	*			()	*	*
12	*	*	*	*			*			*		*			*		*	*	

*- Signifies that the item marked belongs to that factor

() - Signifies that the item belongs to that factor but includes a correlated error

3.2.3- Factor Structures Overview

From looking at the dimensional representations proposed above, it could be concluded that there are a number of contradictions in how these different models are structured. The most obvious example relates to a '*loss of confidence'* factor. Both Graetz' (1991) and Worsely and Gribbin's (1997) models, have a factor named '*loss of confidence*'. The different authors, however, attribute different items to represent this factor with Graetz (1991), stating that items 10 and 11 represent a loss of confidence, while Worsely and Gribbin (1997) attribute it to items 9 and 10. This could represent a labelling issue, or it may be indicative that the different populations used in the various studies (see 3.2.6) interpret the questions differently.

If one looks at the other models, a factor that frequently appears is that of '*social performance*' or '*social dysfunction*'. One could very easily assume that these factors, being the inverse of each other could be represented by the same items, however as one can see when looking at the Schmitz (1999), Graetz (1991) and Worsely and Gribbin (1997) model, the different authors represent these factors using a variety of items. Finally, when looking at items that mention depression, Martin (1999), Schmitz (1999), and Graetz (1991) mention this construct. They do, however, come to significantly divergent conclusions as to which items represent depression within their respective models. A list of reasons as to why the various researchers have come to such varying conclusions was compiled and is detailed below. These were based on the theories put forward by Rey (2014) but include other possibilities put forward by other researchers (Hankins 2008; Ye, 2009; Gao et al., 2006). This list was compiled to summarise the salient points within the literature.

That the various populations tested, interpret the items differently. It must be noted that this would contradict a number of studies which have found the GHQ-12 to be a valid test in a number of different populations and occupations, for example, Del Pillar, Sanchez-Lopez and Dresch, (2008) in Spanish populations, Hoeymans et al. (2004) in Dutch populations, Holli et all (2003) in Finnish populations, Monterezzi et al. (2003) in Iranian populations. It would, however, be in line with contrasting literature such as Benítez (2017) that found that specific items performed differently between native Spanish participants and immigrants. Furthermore, Siraram et al. (1989) found that when investigating Iranian populations, while the test continued to exhibit high levels of reliability and internal consistency, analysis of individual item analysis showed a difference in the performance of some items in an Iranian sample and an English one. The dispute in the literature would suggest that different populations can respond differently to similarly worded items, which would be likely to affect dimensionality. Finally, Smith et al. (2013) posited that item phrasing and even the individual characteristics of participants are significant predictors of responses, which may manifest as spurious factor structures.

That the different statistical techniques used, for example, Principal Component Analysis, Exploratory Factor Analysis, Confirmatory Factor Analysis, may have, as a product of the methods employed, yielded different responses. It is also important to note that when investigating factor structures, some researchers place more importance upon fit statistics. In contrast, others place more emphasis on the properties of the factors. An example of this can be seen in Shevlin and Adamson (2004), where the lack of an individual factor's predictive utility led to a conclusion of a single factor structure, despite fit statistics indicating a three-factor solution.

That different scoring methods used could distort data. As previously mentioned (see Chapter 1), a wide range of scoring mechanisms exist in the literature for this questionnaire with different researchers favouring either the Likert, collapsed scoring or a number of other variations primarily proposed by Duncan-Jones (1989). As different researchers have used various scoring techniques, this may have led to different conclusions.

That the factors that various authors have proposed may be incorrectly labelled. The labelling of factors is inherently subjective, and the clusters of items that compose a factor may be nothing more than statistical phenomena (Atchely, 2019). If factors are incorrectly labelled, then this should become apparent when analysing the relationships between covariates and the factors. If the factors do not behave in a way suggested by previous literature, then it could be argued that they are incorrectly labelled.

That models fail to account for method effects or other such biases properly. Hankins (2008) described multidimensional representations of the GHQ-12 as "*artefacts of the method of analysis, rather than aspects of the GHQ-12 itself.*" (p. 355) He went on to detail numerous occasions where CFA and EFA analyses had suggested that unidimensional scales were multidimensional, (Schmit & Schultz, 1985; Cordery & Sevestos, 1993; Marsh 1996) and that once wording effects were taken into account, the model was shown to be unidimensional (Strathman et al., 1994). The methods that can be used to model these effects are detailed later in this section.

Within the GHQ-12, it has been suggested that some statistical techniques such as Cronbach's test of internal consistency and exploratory factor analysis have been distorted by biases caused by method effects (Hankins, 2008). One of these method effects is that people are more likely to respond differently to negatively worded questions than they do for positively worded questions (Lindwall, 2012). It has been suggested that some of the reasons for this are related to a participant's inattention and carelessness (Schmitt and Stults, 1985), their educational background (Bagozzi, 1993) or an aversion to emotional content which is perceived as unfavourable by the respondent (Cordery & Sevastos, 1993).

EFA cannot distinguish between genuine factors and those generated by method effects. Hankins (2008) found that it is possible to encounter spurious results indicating a two-factor structure if wording effects are not taken into account, and furthermore, reliability coefficients such as Cronbach's Alpha tests of internal consistency assume no such biases.

Greenberger et al. (2003) tested the impact of wording effects by using a scale with similar properties as the GHQ-12, called the Rosenberg self-esteem scale (RSE). This scale was of comparable length and contained positively and negatively worded items. It also had come under similar scrutiny as the GHQ-12, with factor analyses extracting numerous factors, despite the measure being described as unidimensional. A meta-analysis of 23 studies showed that generally a two-factor solution was supported, however, once method effects were accounted for, a single factor solution was more appropriate (Huang & Dong, 2012). Greenberger constructed a new set of questions, using similar questions as the original RSE but reworded items into either exclusively positively or negatively phrased items. When the phrasing was harmonised, only one factor was extracted. Marsh (1996) also suggested that the two-factor model of the RSE could be represented by a general factor and a method factor. Horan et al. (2003) found that the wording effects above were present in many other scales that utilised both positive and negative items. Ye (2009) suggested that the GHQ, sharing many characteristics with the RSE would be equally susceptible to wording effects and proposed a similar approach to Marsh (1996) in the GHQ-12 to address it.

Andrich and van Schoubroeck (1989) suggested that the GHQ was comprised of two factors, one characterised by positively worded items and another by negatively worded items. Further analysis, however, questioned the authenticity of this finding. Using CFA, Hankins, (2008) suggested that the best fitting factor structure for the GHQ 12 was a unidimensional model with substantial response bias on the negatively worded items using the correlated traits correlated uniqueness (CTCU) method detailed below. Hankins (2008) claimed that the two factors uncovered by exploratory factor analysis were entirely a consequence of method effects. In order to investigate wording effects further, Wang and Lin (2011) constructed two versions of the GHQ and tested each against the original version, which contained six positively and six negatively worded items. The two altered versions were either comprised of entirely positively or negatively worded items. Using these models to control for wording effects, Wang and Lin found that a unidimensional model was the best representation of the data for both exclusively positively and negatively worded versions. They claimed that negatively worded items were found to require more cognitive resource to process. Marsh (1996) said that agreeing or disagreeing with negatively phrased questions increases the complexity of the task of responding to a questionnaire. The work completed by Wang and Lin (2011) would suggest that there is significant evidence that wording effects can affect how CFA and EFA analyses interpret the dimensionality of the GHQ. This work was further supported by Hankins (2008), Aguardo (2012) and Smith et al. (2013), who modelled wording effects by correlating the errors of the negatively worded items (see below). It was found that this model also provided a better fit than other more conventional models with the populations that the studies used. From these studies, it can reasonably be assumed that some of the debate around the dimensionality of the GHQ-12 could have been avoided if wording effects were taken into account from the start (Molina et al., 2014).

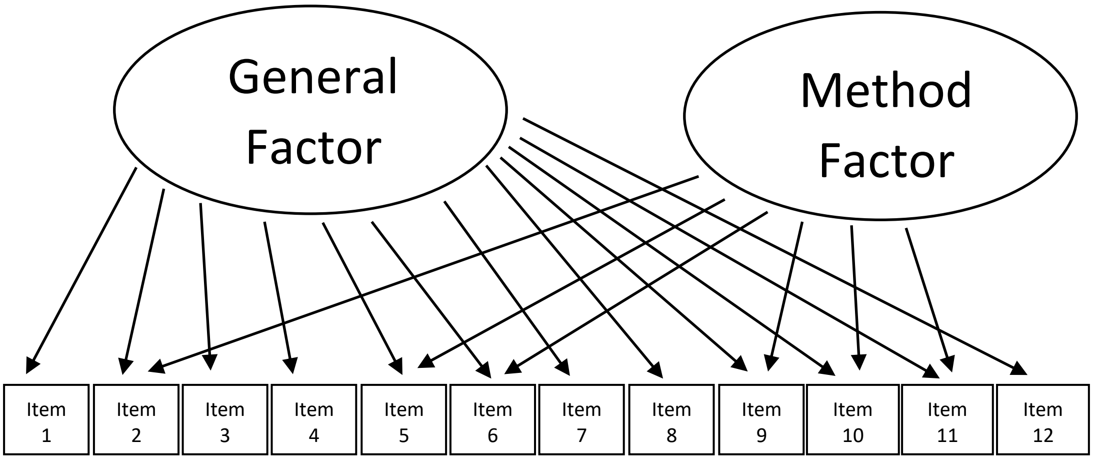
Three proposed models attempt to model method effects, these being Andrich and Schoubroeck's (1989), the so-called method factor model (Ye, 2009) and Hankins (2008). Andrich and Schoubroeck's (1989) model attempted to model method effects by splitting the data into two factors, one representing positively worded items and the other negatively worded items.

The so-called method factor (Ye, 2009) attempts to account for method-based variability by proposing a structure which encompasses a single factor that encompasses all items and a further factor only relating to the negatively worded items. This method factor would not represent variability caused in the data by any psychological phenomena but simply the way respondents react differently to negatively worded items.

Mollina and Rodrigo (2014) state that at time of writing, two methods appear widely in the literature to model wording effects, namely 'correlated traits, correlated methods' (CTCM) and 'correlated traits correlated uniqueness' (CTCU) methods of confirmatory factor analysis. CTCM refers to a statistical model that incorporates a single factor representing the latent variable in question and a second, methods factor which should capture any variance that is caused by method effects while maintaining the integrity of the original single factor. An example of a CTCM model is shown below (see figure 3.1).

Figure 3. 1

Visual Representation of the CTCM Method of Modelling Method Effects in the GHQ-12 (positively and negatively worded items noted as +VE and -VE respectively)

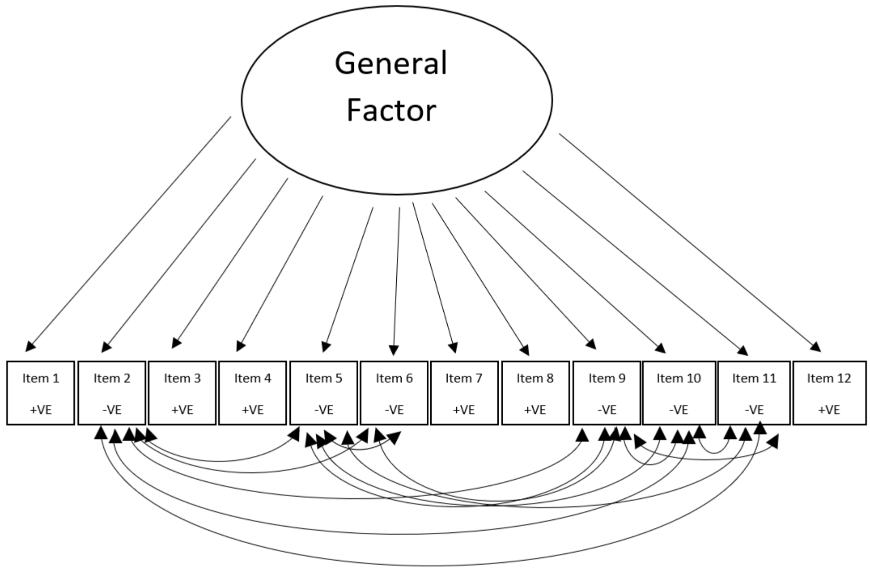


Note. The table demonstrates how the CTCM approach could be applied to the GHQ-12. Positively and negatively worded items are noted as +VE and -VE respectively.

CTCU models do not include a methods factor, but instead, items that share a common methodology, such as in the case of the GHQ, negative wording, have their errors, also known as uniqueness correlated. An example of a model using this method is shown below (see figure 3.2).

Figure 3. 2

Visual Representation of the CTCU Method of Modelling Method Effects in the GHQ-12



Note. The table demonstrates how the CTCU approach could be applied to the GHQ-12. Positively and negatively worded items are noted as +VE and -VE respectively.

Lance, Noble and Scullen (2002) recommended the use of the CTCM method because of identified 'theoretical and substantive' shortcomings in the CTCU. They argued that CTCU methods were not firmly rooted in theory and represented a way of improving fit in a method they described as ad-hoc. Furthermore, they argued that CTCM methods offered opportunities for a large range of models to be tested. As the range of models increases, the likelihood that researchers may simply be '*capitalising on chance*' increases. Finally, they argued that CTCU approaches might return admissible models in situations where the model should not converge under other circumstances. In conclusion, the authors argue that CTCU approaches may be too permissive and offer researchers the opportunity to improve fit in ways that may not be methodologically or conceptually sound. The researchers suggested that CTCU only be used in cases of inadmissible or non-convergent solutions returned by CTCU. When modelling method effects, researchers have used both methods. Ye (2009) using CTCU, whereas Hankins (2008), Aguado et al. (2012), and Smith et al. (2013) used the CTCM method instead. Abubakar and Fischer (2012) used both methods in their research, however, they stopped short of analysing any strengths and weaknesses of the two methods.

3.2.4- Review of Specific Factor Structures

In this section, a brief overview of the various models proposed was provided to aid in understanding why each researcher arrived at the dimensional representation that they did. Of particular note were the statistical techniques employed, the characteristics of the sample tested and any other issues which may have influenced the analyses. A more succinct list of the various factor structures is given in Table 3.1.

Unidimensional

Table 3. 2

Factor structure of the Unidimensional model of the GHQ-12

ITEM	Dimensions
1. Able to concentrate	Psychological Distress
2. Lost much sleep	Psychological Distress
3. Playing a useful part	Psychological Distress
4. Capable of making decisions	Psychological Distress
5. Under stress	Psychological Distress
6. Could not overcome difficulties	Psychological Distress
7. Enjoy your day-to-day activities	Psychological Distress
8. Face up to problems	Psychological Distress
9. Feeling unhappy and depressed	Psychological Distress
10. Losing confidence	Psychological Distress
11. Thinking of self as worthless	Psychological Distress
12. Feeling reasonably happy	Psychological Distress

As previously stated in this thesis, the scoring method used in the GHQ whereby individual items are scored equally and contribute to a final score implies that the GHQ was intended to measure a unidimensional construct, that of psychological distress or vulnerability.

When one looks at a number of studies that have investigated the GHQ's specificity and sensitivity (Blaxter, 1990, Stansfield & Marmot, 1992), or those which compared GHQ-12 scores with medical examinations (Layton & Rust, 1960), these researchers, by virtue of using the suggested scoring mechanism, implicitly accepted the unidimensionality of the GHQ-12. As previously stated, these studies may be rendered invalid if the GHQ-12 was subsequently shown to be multidimensional.

Analysis conducted by the GHQ's creator, (Goldberg, 1988) initially supported the concept that the GHQ-12 was unidimensional. Further studies have replicated his findings, sometimes using various statistical techniques. Such techniques provide subtly different solutions as they place greater emphasis on one desired property of the solution. PCA using Varimax rotation (see statistical procedures), for example, is a technique which attempts to account for variance in the most parsimonious solution. Studies using this method (Mallet, 2000; Lewis, 1992; Lewis 1991; Goldberg & Huxley, 1992; Kendler et al., 1987) suggested that large proportions of variance, around 35%-50%, could be accounted for by a single factor. This factor was labelled as illness severity or general dysphoria.

Finally, research by Shelvin and Adamson (2005) suggested that some analyses using confirmatory factor analysis showed that a three-factor model was the best fit for the data. This research posited, however, that the factor correlations were so high that none of the factors in themselves offered any unique predictive utility. Shevlin and

Adamson stated that in light of this it might not be appropriate to treat the GHQ-12 as multidimensional and while the data may present as multidimensional, it was more practical to treat the GHQ-12 as unidimensional.

Politi et al. (1994)

Table 3. 3

The Factor Structure of Politi et al.'s (1994) Representation of the GHQ-12

ITEM	Dimensions	
1. Able to concentrate		Social Dysfunction
2. Lost much sleep	Dysphoria	
3. Playing a useful part		Social Dysfunction
4. Capable of making decisions		Social Dysfunction
5. Under stress	Dysphoria	
6. Could not overcome difficulties	Dysphoria	
7. Enjoy your day-to-day activities		Social Dysfunction
8. Face up to problems		Social Dysfunction
9. Feeling unhappy and depressed	Dysphoria	
10. Losing confidence	Dysphoria	
11. Thinking of self as worthless	Dysphoria	
12. Feeling reasonably happy	Dysphoria	Social Dysfunction

This model proposed a two-factor solution with items divided amongst dysphoria and social dysfunction factors. The Oxford English Dictionary defines dysphoria as 'A state of unease or generalized dissatisfaction with life.' Items associated with this item were most commonly associated with anxiety and depression, whereas Social Dysfunction items tended to focus on an individual's inability to perform everyday tasks (Politi et al., 1997). In this model, item 12 cross-loaded on both factors.

Politi (1997) investigated the factor structure, internal consistency and validity of the GHQ-12 in a sample of 18-year-old Italian males. Using Cronbach's Alpha test of internal consistency, a figure of 0.81 was calculated, implying a reliable measure, however, it is a common misconception that a high value of Cronbach's Alpha confirms unidimensionality (Green, Lissitz & Muliak, 1997). When subjected to Principal Components Analysis, with a varimax (and Oblim) rotation, two dimensions were extracted from the data (a detailed explanation of the effects of different types of rotation is given later in this chapter under 'Statistical Procedures'). When looking at the items which loaded onto dysphoria, Politi et al. (1997) stated that there was a clear relationship between the two factors and whether or not an individual had emotional disturbance. This emotional disturbance was assessed according to an independent psychiatrist's assessment and was found to have relatively low, with a misclassification rate of 0.40 (Politi et al., 1994). Politi suggested that further diagnostic inferences, i.e. the detection of so-called 'cases,' could be improved if a scoring matrix was established that took account of the factors listed in his model rather than a single summed score, which accounted for the multidimensional properties of the GHQ-12 (Politi, 1994).

Andrich and van Schoubroeck (1989)

Table 3. 4

The Factor Structure of Andrich and van Schoubroeck's (1989) Representation of the GHQ-12

ITEM	Dimensions	
1. Able to concentrate	Positive	
2. Lost much sleep		Negative
3. Playing a useful part	Positive	
4. Capable of making decisions	Positive	
5. Under stress		Negative
6. Could not overcome difficulties		Negative
7. Enjoy your day-to-day activities	Positive	
8. Face up to problems	Positive	
9. Feeling unhappy and depressed		Negative
10. Losing confidence		Negative
11. Thinking of self as worthless		Negative
12. Feeling reasonably happy	Positive	

This model, which was proposed in both Andrich and van Schoubroeck (1989), was one of the first to attempt to model method effects. As is shown above, it does this by proposing two factors, one comprised of positively worded items and the other comprising negatively worded items. To generate this factor structure, Andrich and van Schoubroeck (1989) used a technique called latent trait modelling, which they adapted from work by Christofferson (1975). They then converted findings obtained via this technique to construct factor loadings, i.e. the correlational relationship between the items and a factor. It must be noted in Duncan-Jones et al. (1986) while using the GHQ-30, proposed a similar dimensional representation and found that factor correlations were high (Φ=0.88). This implied that both factors appeared to measure very similar concepts. Furthermore, it must be noted that a test of a general factor model with two 'method' factors produced a much lower inter factor correlation (Φ=0.35) between the factors but did not return acceptable model fit. While both studies utilised data obtained from Australian participants, Duncan-Jones conducted research using general population data from New South Wales while Andrich and van Schoubroeck's (1989) analyses were conducted using data from teachers in western Australia. While Duncan-Jones et al. (1986) primarily investigated the thresholds and behaviour of different scoring mechanisms, both investigators noticed that positively and negatively worded items in the GHQ behaved differently when investigating the "rate of change by an item as a function of the latent trait," which they described as discrimination. They noticed that negatively worded items discriminated "more sharply" than positively worded items and it was even noticed that only 50% of the common variance was shared between positively and negatively worded items (Andrich & van Shoubreck, 1989). It was generally concluded that the factors identified in this research were the result of method effects. While statistical tests such as those used in the above do not distinguish between genuine factors and those generated through method effects, these findings spurred Duncan-Jones to investigate if different scoring techniques could be developed that would overcome these obstacles. These have been detailed in the 'Scoring Methods' section earlier in the chapter.

Schmitz et al. (1999)

Table 3. 5

The Factor Structure of Schmitz et al.'s (1999) Representation of the GHQ-12

ITEM	Dimensions	
1. Able to concentrate	Anxiety and depression	
2. Lost much sleep	Anxiety and depression	
3. Playing a useful part	Anxiety and depression	Social Performance
4. Capable of making decisions		Social Performance
5. Under stress		Social Performance
6. Could not overcome difficulties	Anxiety and depression	
7. Enjoy your day-to-day activities	Anxiety and depression	
8. Face up to problems		Social Performance
9. Feeling unhappy and depressed		Social Performance
10. Losing confidence	Anxiety and depression	
11. Thinking of self as worthless	Anxiety and depression	
12. Feeling reasonably happy		Social Performance

The model (shown in Table 3.5) proposed two factors with one cross-loaded item. Schmitz's (1999) model comprises a social performance factor in a similar manner to Andrich and van Schoubroeck's (1989) model, however, associated different items with these factors. Schmitz' (1999) study was conducted upon 572 outpatients of 18 randomly selected primary care clinics in the German city of Dusseldorf and was intended amongst other things to validate the German version of the GHQ-12. Patients were asked to complete both the GHQ-12 and a number of clinical instruments, including examination by a mental health professional, who had no prior knowledge of the patient. The mental health professional used the SCID (Structured Clinical Interview for Diagnostic Statistics Manual) and the Impairment Rating Scale (IRS), and these were considered the gold standard against which both the GHQ was to be measured. Data reduction was conducted using PCA using oblique rotation. This rotation method was selected as Schmitz (1999) had stated that this rotation method was the most appropriate when items were linearly dependant, which Schmitz argued that they were. All items with the exception of item 3 loaded onto a single principal component by <0.5, with item 3 demonstrating acceptable fit on both. The resulting Principal Components were checked for reliability using Cronbach's alpha, yielding a result of 0.91 indicating a high degree of internal consistency. What Schmitz described as factor scores were compared with the already established subscales in the competing measures (SLC-90 and IS), and found to have exhibited correlations, ranging from 0.37-0.73, which Schmitz (1999) described as high. In this analysis, two principal components were extracted, and Schmitz referred to these as 'anxiety and depression' and 'social performance'. These principal components accounted for 59.6% of the variance in the data. In conclusion, Schmitz found strong internal consistency, sensitivity, and specificity and proposed the above factor structure for the GHQ-12 when translated into German.

Martin (1999)

Table 3. 6

The Factor Structure of Martin' (1999) Representation of the GHQ-12

ITEM	Dimensions	
1. Able to concentrate	Cope	
2. Lost much sleep		Stress
3. Playing a useful part	Cope	
4. Capable of making decisions	Cope	
5. Under stress		Stress

6. Could not overcome difficulties			Depression
7. Enjoy your day-to-day activities		Stress	
8. Face up to problems	Cope		
9. Feeling unhappy and depressed			Depression
10. Losing confidence			Depression
11. Thinking of self as worthless			Depression
12. Feeling reasonably happy			Depression

This three-factor model contains factors labelled as ‘coping’, ‘stress’ and ‘depression’. These three factors represented three distinct factors but did not account for method effects within the GHQ-12. The purpose of Martin’s study (1999) was to investigate whether Worsely and Gribbin’s (1997) and Graetz’s (1991) represented a good fit of their population and to compare these models against one which Martin proposed (see table 3.6). In this study the sample population consisted of Australians, and their friends enrolled at major universities in Sydney resulting in 169 participants in total. This non-representative sample would likely be subject to the biases associated with student samples and the inclusion of ‘friends’ make any generalisation to the broader population difficult. Confirmatory Factor Analysis was conducted using LISREL 7.2 (Joreskog & Sorbom, 1989). This statistical technique tests a sample covariance matrix against a hypothesised matrix (Martin, 1999). Martin’s analysis did not suggest that a unidimensional model adequately represented the factor structure in the GHQ. In fact, the unidimensional model was the least well-fitting. Martin’s model exhibited the best fit, expressed through the Tucker-Lewis Scale of 0.912. Interestingly Martin raises serious concerns about the robustness of data collection measures employed. This was because students were left to administrate the distribution of the questionnaires themselves and Martin (1999) felt that this opened the distinct possibility of data fabrication and random responses. It was proposed that more rigorous analysis of the proposed models should be conducted using more controlled data collection methods. While the sample used in this study is not of the standard in other tests, the model was not discounted on the basis that other researchers (Gao et al., 2006; Shevlin & Adamson, 2006) included it in their analyses and therefore it was not felt necessary to discount the model at this stage.

Graetz (1991)

Table 3. 7

The Factor Structure of Graetz' (1991) Representation of the GHQ-12

ITEM	Dimension		
1. Able to concentrate		Social dysfunction	
2. Lost much sleep	Anxiety/depression		
3. Playing a useful part		Social dysfunction	
4. Capable of making decisions		Social dysfunction	
5. Under stress	Anxiety/ depression		
6. Could not overcome difficulties	Anxiety/ depression		
7. Enjoy your day-to-day activities		Social dysfunction	
8. Face up to problems		Social dysfunction	
9. Feeling unhappy and depressed	Anxiety/ depression		
10. Losing confidence			Loss of confidence
11. Thinking of self as worthless			Loss of confidence
12. Feeling reasonably happy		Social Dysfunction	

This model proposed a three-factor solution and identified the factors of 'anxiety/depression,' 'social dysfunction' and 'loss of confidence.' All items were associated exclusively with a single factor. The data was obtained via the Australian Longitudinal Study, which was conducted by the Department of Education and Training. Scoring was conducted by the use of the Likert method to "produce a more acceptable distribution of scores for parametric analysis" (Graetz, 1991). Data reduction was performed using PCA with various rotation methods to gauge their effectiveness. Graetz (1991) also noted that different techniques tended to yield different results, therefore conducted a battery of statistical techniques to allow a more holistic picture to be gained. It was found that PCA with oblique rotation, was, in the view of the researcher, the best method, as it generated the most parsimonious solution as compared to other rotation methods. In order to test the validity of the model, Graetz (1991) regressed covariates onto factors individually to test if they behaved uniformly to these covariates or provided unique predictive ability to specific covariates. He found that the three factors did exhibit variance in relation to the covariates and that this effect was present in both cross-sectional and longitudinal analyses, suggesting that the factors were measuring distinct concepts. Finally, he suggested that the GHQ may be of more use as a diagnostic tool if this multidimensional approach was adopted.

Worsley and Gribbin (1977)

Table 3. 8

The Factor Structure of Worsely and Gribbin's (1997) Representation of the GHQ-12

ITEM	Dimension	
1. Able to concentrate	Social performance	
2. Lost much sleep	Social performance	
3. Playing a useful part	Social performance	
4. Capable of making decisions	Social performance	
5. Under stress		Anhedonia
6. Could not overcome difficulties		Anhedonia
7. Enjoy your day-to-day activities		Anhedonia
8. Face up to problems		Anhedonia

9. Feeling unhappy and depressed			Loss of confidence
10. Losing confidence			Loss of confidence
11. Thinking of self as worthless	Social performance		
12. Feeling reasonably happy		Anhedonia	

Worsely and Gribbin (1977) conducted a study upon randomised households in ‘door to door’ interviews in two areas of Australia, namely Melbourne and a number of small mining towns in northwest Australia. They conducted what they described as combined factor analysis on this sample and subsequently proposed a three-factor model consisting of factors relating to ‘social performance’, ‘anhedonia’ and ‘loss of confidence.’ They acknowledged that the factor structure extracted was similar to that which was extracted by Goldberg (1976) and Worsely, Walton and Wood (1997) when larger variations of the GHQ such as the 30 and sixty item variations were examined. Anhedonia was found to be the primary source of variance within the data accounting for 41.7%. In total, all three factors were responsible for 61.7% of the variance within the data. It is therefore stated that according to the researchers, the GHQ is not unidimensional, and the exact structure is much more complex. It must be noted in the paper that tests of internal consistency were noticeably absent, which may damage the credibility of this model.

Hankins (2008)

Table 3. 9

The Factor Structure of Hankins’ (2008) Representation of the GHQ-12

ITEM	Dimension	
1. Able to concentrate	Unidimensional	
2. Lost much sleep		Unidimensional with correlated errors
3. Playing a useful part	Unidimensional	

4. Capable of making decisions	Unidimensional	
5. Under stress		Unidimensional with correlated errors
6. Could not overcome difficulties		Unidimensional with correlated errors
7. Enjoy your day-to-day activities	Unidimensional	
8. Face up to problems	Unidimensional	
9. Feeling unhappy and depressed		Unidimensional with correlated errors
10. Losing confidence		Unidimensional with correlated errors
11. Thinking of self as worthless		Unidimensional with correlated errors
12. Feeling reasonably happy	Unidimensional	

This model proposed that the GHQ-12 was unidimensional and that the multidimensional results encountered by previous researchers were spurious and caused by method effects. In an attempt to model these method effects, Hankins (2008) correlated the errors on the negatively worded items using the CTCU technique explained previously in this chapter.

Hankins (2008) conducted research upon the 2004 cohort of the Health Survey for England. This data source is a large-scale questionnaire with approximately 8,000 adults and 2,000 children every year. Using exploratory factor analysis (EFA) followed by structural equation modelling using the package AMOS 6.0, Hankins proposed that the GHQ-12 be modelled as a single dimension, but in order to represent response bias, correlated errors on all negatively worded items were included. This CMCU technique is described earlier in this chapter. Hankins' twofold analysis yielded the following results, firstly EFA yielded two factors, focused on the negatively and positively worded items. Confirmatory factor analysis using structural equation modelling, however, suggested that the model associated with wording effects was not appropriate.

Root mean square error of approximation results (RMSEA = 0.068, 90%CL (0.064, 0.073; ECVI = 0.214, 90%CL (0.191, 0.238)) suggested that the model with correlated errors had the best fit for the data, further suggesting that a failure to account for method effects had led to spurious findings in earlier studies. Hankins (2008) warned that in the use of reliability coefficients which assume no such response bias, care must be taken in interpreting their results. Interestingly, while this research was conducted upon the UK population, its results were replicated in Northern Iran (Motamed et al., 2018), where a representative population cohort's GHQ-12 scores were subject to confirmatory factor analysis against other well-established items. Following this, a model identical to Hankins (2008) was found to be the best fitting model for this data.

Ye (2009)

Table 3. 10

The Factor Structure of Ye's (2009) Representation of the GHQ-12

ITEM	Dimension	
1. Able to concentrate	Unidimensional	
2. Lost much sleep	Unidimensional	Method Factor
3. Playing a useful part	Unidimensional	
4. Capable of making decisions	Unidimensional	
5. Under stress	Unidimensional	Method Factor
6. Could not overcome difficulties	Unidimensional	Method Factor
7. Enjoy your day-to-day activities	Unidimensional	
8. Face up to problems	Unidimensional	
9. Feeling unhappy and depressed	Unidimensional	Method Factor
10. Losing confidence	Unidimensional	Method Factor
11. Thinking of self as worthless	Unidimensional	Method Factor
12. Feeling reasonably happy	Unidimensional	

This model attempted to model method effects using the CTCM technique (see statistical procedures 3.2.5). It comprised a factor which is associated with all items which it is claimed measured psychological morbidity and a method factor which is associated with the negatively worded items only. Ye (2009) was inspired to apply this technique to the GHQ-12 after observing a similar situation relating to the Rosenberg Self Esteem Scale. This scale would frequently yield a two-factor solution, which was thought to be due to method effects. Marsch (1996) proposed that a single factor encompassing all items, and a second incorporating negatively worded items was an appropriate way to model these effects and found it to exhibit good fit.

Ye (2009) analysed a population of 348 Chinese students' GHQ-12 scores and compared this model against other established models (Graetz, 1991; Andrich & van Shoeubroeck, 1989). He also regressed the respective factors from these models onto measures of extraversion and neuroticism from the NEO five-factor inventory (Costa & McCrae, 1992) and the Satisfaction With Life Scale (Dinerner et al., 1985) to test if these factors varied uniformly. He also split extraversion and neuroticism into positively and negatively worded items. He argued that should the factors vary uniformly, then there would be little practical benefit in treating the factors separately. He found that the method factor method did provide 'goodness of fit' and that it exhibited similar fit with the competing models (see table 3.11). While all models showed acceptable fit with Graetz (1991) performing the best in RMSEA analyses (RMSEA=0.054), Ye (2009) performed the best SRMR analysis (SRMR=0.051). When regressing the factors onto the various covariates mentioned above, it was found that multidimensional representations of the GHQ-12 were similarly correlated with them. As a result, Ye (2009) concluded that treating the GHQ-12 as multidimensional was of little benefit. Interestingly, the method effects variable was found to correlate with life satisfaction and extraversion, but only when worded positively. It was not shown to correlate with negatively worded items relating to extraversion nor with scores of neuroticism regardless of how the items were worded. Ye (2009) mentioned limitations in his research relating to the population tested. Firstly there was a concern, as his sample population consisted of undergraduate students, that his results may not be generalisable to the wider population and secondly that his population of exclusively Chinese participants may not be generalisable to western populations. He mentioned previous literature that suggested that wording effects could be influenced by the population tested specifically relating to reading ability (Marsch, 1996) and therefore queried if results could be generalised to western populations where average reading ability may be different.

Table 3. 11

Fit Statistics for Competing Models in Ye (2009)

	χ^2	Df	RMSEA	SRMR	CFI	NFI	NNFI
Graetz (1991)	102.46	51	0.054	0.057	0.99	0.97	0.98
Andritch and van Shoeubroek (1989)	109.81	53	0.056	0.057	0.98	0.97	0.98
Ye (2009)	98	48	0.055	0.051	0.99	0.97	0.98

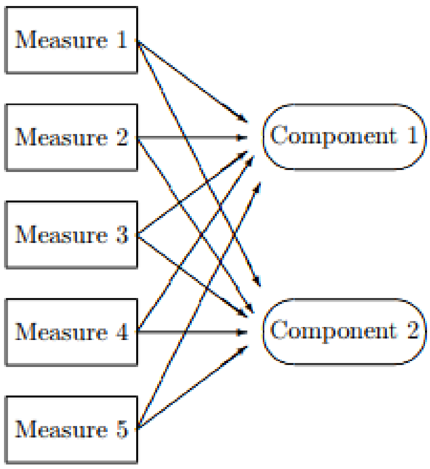
3.2.5- Statistical Procedures

As can be seen from the models detailed above, different researchers have utilised a multitude of different techniques when conducting their analyses. These techniques have been elaborated upon briefly, earlier in this chapter. This section of the thesis, however, will provide a more in-depth description of the techniques used and how this may have affected the results of these studies.

Both Politi (1994) and Graetz (1991) used PCA, which is a data reduction technique developed by Karl Pearson (1901), and is based on a model detailed in DeCoster (1998; see figure 3.3). As has been described earlier in this chapter, PCA was a predecessor to FA and used different underlying assumptions and models. Importantly this model assumes that so-called ‘principal components’ are based on observed variables. PCA does not assume an underlying model and, consequently, is described as a data reduction technique. Furthermore, PCA based principal components are composed of linear combinations of all observed variables and, as a result, will contain both common and unique variance, whereas FA will contain only unique variance (see below). Principal components are generated algorithmically and as a result, may not be grounded in theory. As a result, principal components may not be interpretable as anything more than statistical phenomena (Atchley, 2019).

Figure 3. 3

Model for Principal Component Analysis



Note. The above figure is a representation of the model for principal component analysis. Arrows denote the direction of the relationship.

Measure = Observed Variable

Component= Principal Component

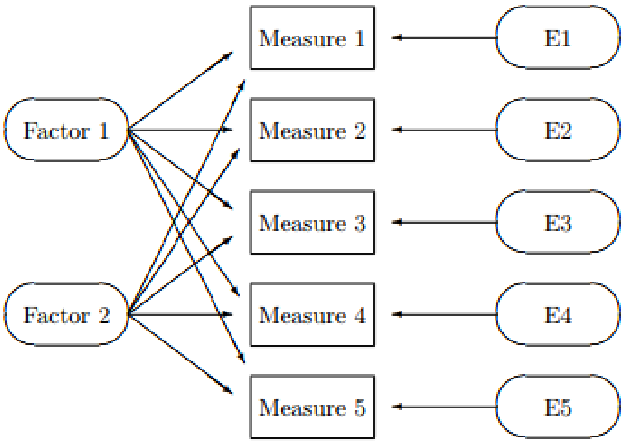
Factor analysis (FA), refers to a family of commonly used data reduction techniques which aim to simplify data from a large number of items to a smaller and more manageable format (Frucher, 1954). In contrast to PCA, FA, adopts a very different underlying assumption, that measures are influenced by latent variables called factors, (see figure 3.2).

PCA uses a technique referred to as rotation to maximise the amount of variance that can be attributed to the Principal Components. There are a number of different rotation techniques which yield slightly different results, and these are detailed below. A varimax solution will tend to load individual items more readily onto a single factor, however, this can be seen as unrealistic (Russel, 2002). Quartimax rotation tends to focus on minimising the number of factors extracted. Frequently the over-simplified factor structure is of minimal use in research purposes and, therefore, may not always be used, especially if small quantities of factors are expected to be extracted. Equimax rotation is a compromise between the two. Direct Oblimin rotation will tend to generate solutions that have higher eigenvalues. However, the trade-off is that factors produced will generally have compromised interpretability. As a result of the differences in output that each rotation type can generate, it is important to note which rotation method is used and how the rotation method may have affected the generated factor structure.

There are two types of FA, and these are exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Both of these techniques are based on the common factor model (Williams, 1978; see figure 3.2). This model proposes that observed variables (measures 1 -5) share common variance with an underlying model, consisting of latent or unobserved variables. Variables that are highly correlated are more likely to be influenced by a latent variable than those which are not (DeCoster, 1998). By examining the correlations between observed variables, latent variables can be derived from these relationships.

Figure 3. 4

The Common Factor Model



Note. The above figure is a graphical representation of the common factor model as shown in DeCoster (1998). The arrows denote the direction of the relationship. Measures 1-5 represent observed variables which DeCoster argue are influenced by both common underlying factors, represented as Factors 1 and 2 and unique factors or error as denoted by E1-5.

When using structural equation modelling techniques estimators are used as a way of utilising partially complete data and avoiding the use of listwise deletion (Peters & Enders, 2002). A review of the different techniques used to estimate data are reviewed below.

Maximum Likelihood (ML) is one if the most commonly used estimators and demonstrates characteristics such as asymptotic unbiasedness, normality, consistency, and maximal efficiency (Li, 2016), but due to using Pearson's Correlations operates under the assumption that the data being investigated is collected at the interval level i.e. nominal or ratio data (Bollen, 1989; Satorra 1990). Previous research has identified that using ML estimators when investigating categorical data can result in overstated chi square (Muthen & Kaplan, 1985) and depressed factor loadings (Beauducel & Yorck

Herzberg, 2009). MLR represents a version of ML which is statistically corrected for non-normality using what Muthen and Muthen (2009) described as a sandwich-type estimator as opposed to the traditional inverse Fischer Information Matrix and as a result is not as dependent on multivariate normal distribution. Satorra and Bentler (1994) have found this robust version to perform better when estimating nonnormal data. When conducting this analysis, consideration was given as to whether the MLR or Robust Weighted Least Squares (WLSMV) would be more appropriate. WLSMV is recommended for use in categorical data (Browne, 1984) and while initial parsing of the literature may suggest that WLS would be the most appropriate estimator to use for categorical data such as Likert scales (Li, 2016; Rodrigo, Navarro & Alvarado, 2015), Li, (2014) conducted in depth research into 4 item Likert scales such as used in this research and found the following

- WLSMV were more likely to provide inadmissible solutions, especially when sample size was small
- MLR gave more accurate put less precise standard errors than WLS
- WLS was likely to over-reject the hypothesised model compared with MLR and that MLR's rejection rate steadily increased with sample size

In conclusion, Li recommended that MLR be used "*when structural relationships are of primary concern*" (Li, 2014) and while concerns when using small sample were mentioned, the UKHLS sample of over 100,000 would be sufficient to mitigate this concern. He also noted that when missing data was incorporated into the model, MLR outperformed WLS, and he described missing data handling techniques for WLS as "*underdeveloped*". It was also noted that MLR performs better when the number of observed variables was large, in the case of their study 9 or above. Given the 12 items of the GHQ-12 this was important and supports the use of the MLR estimation method in this analysis.

Li (2006) concluded that all estimation methods had their benefits and depending on the data being investigated and purpose of the research, each could theoretically be justified. In the case of this analysis, MLR was adopted as per Li's recommendation that MLR provides the bet estimator when investigating structural relationships among latent constructs.

3.2.6.- Hypotheses

As was stated in Chapter 1, it was hypothesised that once wording effects were accounted for, a unidimensional model would be the best fit for the data. This was based on the findings of Ye (2009) and Hankins (2008) which suggested that using various methods described in the relevant section below, that models which utilised methods to account for the method effects generated by the wording of the items would demonstrate superior or comparable fit to multidimensional representations. This hypothesis was also supported by a meta-analysis by Molina and Rodrigo (2013), which showed that in western populations, simulated method effects models demonstrated superior fit to multidimensional models.

3.3.- Methods

3.3.1- Data

The data for this chapter's analysis was drawn from Wave 1 of the Understanding Society (UKHLS) database. This contained 39700 participants who at least partially completed the GHQ-12 and was weighted, clustered and stratified as directed in the UKHLS user guide (Knies, 2017). UKHLS provided participant's scores in both a caseness and Likert format (see Chapter 2), however, Likert scores were used as they provided richer data than a caseness approach.

A comprehensive review of this data is given in Chapter 2, however, in summary, it was provided on licence from the UK Data and was representative of the UK population. The analysis used maximum likelihood parameter estimates with robust standard errors (MLR) as an method of handling missing data and a justification for such is given in 3.2.5.

3.3.2- Analysis

The aim of this chapter's analysis was to identify using CFA, the most appropriate model to represent the data in UKHLS. This chapter's analysis focused on the fit and characteristics of the various models, i.e., factor loadings and correlations, whereas later chapters investigated conceptual appropriateness and validity.

Firstly, model fit was investigated using a battery of fit statistics. As there was no agreement in the literature as to the most appropriate technique to use in various situations, Kenny (2005) suggested that the simultaneous use of multiple fit statistics mitigates against the inappropriateness or limitations inherent in each of the various techniques. An in-depth discussion about the characteristics of each of these tests is given in appendix 1, however, a summary was provided below.

The fit statistics used are provided below alongside suggested interpretation guidelines where appropriate. Chi-square was reported as suggested in the literature, however, was not used for interpretative purposes as Kenny (2020) suggested that it was inappropriate for analyses of over 200 participants. Root Mean Square Error of Approximation (RMSEA), a fit statistic generated from the Chi-Squared value was reported. Callum, Browne, and Sugawara (1996) suggested that values of 0.01, 0.05, and 0.08 indicated excellent, good, and mediocre fit, respectively and these criteria were adopted. Two incremental fit indexes were reported, the Comparative Fit Index (CFI) and the Tucker Lewis Index (TLI) and Awang, (2012) suggested that values over 0.9 represented good fit for both these fit statistics. Hu and Bentler (1999) suggested that a value of 0.9 was too low a threshold to indicate a good fit. They suggested that a cut off of 0.95 would be more appropriate, and this higher threshold was adopted to demonstrate good fit. Standardised Root Mean Square Residual (SRMR) described as an absolute measure of fit (Kenny, 2020) was reported, and a value of less than 0.08 was considered sufficient to denote a good fit for the data (Hu & Bentler, 1999).

Once the model fit had been established, all models were investigated in relation to factor loadings and factor correlations in order of fit. By doing this, it was possible to ascertain if items were associated with appropriate factors. If an item was appropriately associated, one would expect to see a strong, statistically significant correlation between the item and the factor. Weak or non-statistically significant relationships would imply that the item is not associated with the correct factor. While no agreed interpretation of factor loadings exist, a number of rules of thumb exist. Stevens (1992) suggests a cut off of 0.4, whereas MacCalum et al. (1999) suggest that all items should have an average score of 0.7. Both of the above rules of thumb were thought to be too arbitrary and in line with the limitations described in Kenny, (2005) it was decided not to interpret findings using cut-offs. Instead, guidelines suggested in Comrey and Lee (2013) were used. These guidelines ascribed values to the strength of relationships based on the factor loadings by grouping results into the following categories 0.32-0.49 (*poor*), 0.45-

0.54 (*fair*), 0.55-0.62 (*good*), 0.63-0.70 (*very good*) or >0.71 (*excellent*) (Comrey & Lee, 2013).

Inter-factor relationships were investigated by analysing the correlation between them. While highly correlated factors may not be a problem in itself, if factors are highly correlated, it may imply that they measure similar concepts and consequently, the usefulness of treating the factors as separate may be called into question (Shevlin & Adamson, 2005).

3.4.- Results

3.4.1 Fit Statistics

Table 3. 12

Summary of Fit statistics for Competing Factor Structures of the GHQ-12

	1 factor	Politi et al 1994	Andrich & van Schoubroeck (1989)	Schmitz et al 1999	Graetz 1991	Martin 1999	Worsley and Gribbin (1977)	Hankins 2008	Ye (2008)
Df	54	52	53	52	51	50	51	39	48
Chi-square	12290.568	5587.068	5587.068	12223.288	4026.797	8402.851	11812.043	2459.046	5824.084
P	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
RMSEA	0.075	0.051	0.052	0.076	0.044	0.064	0.076	**0.039**	0.055
90% CI	0.074-0.076	0.050-0.052	0.051-0.054	0.075-0.077	0.043-0.045	0.063-0.065	0.074-0.077	0.038-0.4	0.053-0.056
CFI	0.877	0.944	0.941	0.878	0.960	0.916	0.882	**0.976**	0.942
TLI	0.580	0.929	0.926	0.845	0.948	0.886	0.847	**0.959**	0.920
SRMR	0.057	0.032	0.034	0.056	0.030	0.049	0.056	**0.026**	0.033

90% CI= confidence intervals at 90%

RMSEA= Root Mean Square Error of Approximation

CFI= Comparative Fit Index

TLI= Tucker Lewis Index

SRMR= Standardised Root Mean Square Residual

*Values in bold represent the highest

As suggested in Kenny (2020), a multitude of fit statistics were reported and generally, these various techniques yielded similar results. From the results reported in Table 3.12, Hankins (2008), according to the fit statistics listed, demonstrated the best fit for the data (CFI=0.976, TLI=0.959, RMSEA=0.039 & SRMR=0.026) and Graetz' (1991) model demonstrated the second-best fit for the data (CFI=0.960, TLI=0.948, RMSEA=0.044 & SRMR=0.030). Ye's (2008) model also demonstrated good fit (CFI=0.942, TLI=0.920, RMSEA=0.055 & SRMR=0.033), as did Andrich and van Shoeubroeck's (1989) model (CFI=0.941, TLI=0.929, RMSEA=0.052 & SRMR=0.034) and Politi's (CFI=0.944, TLI=0.929, RMSEA=0.051 & SRMR=0.032).

A number of models failed to demonstrate what would be considered '*acceptable fit*', namely the unidimensional model, Schmitz et al. (1999) and Worsely and Gribbin's (1999) model. These models failed to exceed some of the commonly accepted guidelines such as CFI and TLI scores exceeding 0.9 (Awang, 2012). Martin's (1999) model exceeded suggested guidelines (Kenny, 2020) when using the CFI but not when using the TLI (CFI=0.916 TLI=0.886).

3.4.2- Factor loadings and Correlations

Below, the results of each model's factor loadings are detailed in order of fit. Within each table, the estimate or correlation, the standard error (S.E.) and the probability of the results being a consequence of chance alone (P-Value) are reported.

Hankins (2008)

Table 3. 13

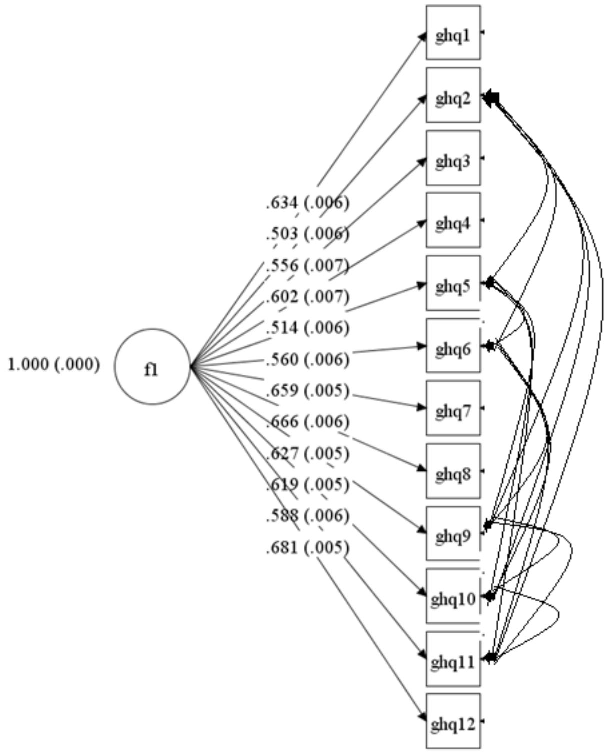
Standardised Factor Loadings of the General Mental Health Factor of Hankins' Representation of the GHQ-12

	Estimate	S.E.	Est./S.E.	P-Value
GHQ1	0.634	0.006	106.456	<0.001
GHQ2	0.503	0.006	89.091	<0.001
GHQ3	0.556	0.007	82.777	<0.001
GHQ4	0.602	0.007	89.166	<0.001
GHQ5	0.514	0.006	88.728	<0.001
GHQ6	0.56	0.006	93.432	<0.001
GHQ7	0.659	0.005	124.481	<0.001
GHQ8	0.666	0.006	109.665	<0.001
GHQ9	0.627	0.005	133.781	<0.001
GHQ10	0.619	0.005	125.383	<0.001
GHQ11	0.588	0.006	90.723	<0.001
GHQ12	0.681	0.005	127.795	<0.001

All items in this model exhibited moderately strong statistically significant correlations with the general mental health factor, with items exhibiting an average correlation of 0.6. The model performed slightly better than the standard unidimensional model, with the lowest value recorded being 0.503. In contrast, the unidimensional model had a score of 0.474. While some items performed better in a standard unidimensional model, overall, the '*correlated errors model*' displayed superior factor loadings than its most similar rival, the unidimensional model.

Figure 3. 5

A Graphical Depiction with Attached Factor Loadings and Inter-factor Correlations for Hankins' Representation of the GHQ-12



Note. This figure is a graphical representation of the unidimensional factor model of the GHQ-12 using correlated errors. F1 represents the unidimensional factor and the various items marked 'ghq' represent the items. In order to avoid the diagram looking cluttered, a table was detailed below which details the correlated errors

Table 3.13.

Correlations among error variances from the unidimensional model shown in Figure 3.5.

	1	2	3	4	5	6	7	8	9	10	11
1. ghq1	-										
2. ghq2	-	-									
3. ghq3	-	-	-								
4. ghq4	-	-	-	-							
5. ghq5	-	.17**	-	-	-						
6. ghq6	-	.11**	-	-	.15**	-					

7. ghq7	-	-	-	-	-	-	-				
8. ghq8	-	-	-	-	-	-	-	-			
9. ghq9	-	.16**	-	-	.17**	.13**	-	-	-		
10. ghq10	-	.10**	-	-	.11**	.12**	-	-	.18**	-	
11. ghq11	-	.06**	-	-	.06**	.08**	-	-	.12**	.17**	-
12. ghq12	-	-	-	-	-	-	-	-	-	-	-

* $p < 0.05$, ** $p < 0.01$,

Graetz (1991)

Table 3. 14

Standardised Factor Loadings of the Anxiety/Depression Factor of Graetz' Representation of the GHQ-12

	Estimate	S.E.	Est./S.E.	P-Value
GHQ2	0.68	0.004	159.885	<0.001
GHQ5	0.714	0.004	181.31	<0.001
GHQ6	0.717	0.004	160.981	<0.001
GHQ9	0.83	0.003	293.665	<0.001

Table 3. 15

Standardised Factor Loadings of the Social Dysfunction Factor of Graetz' Representation of the GHQ-12

	Estimate	S.E.	Est./S.E.	P-Value
GHQ1	0.636	0.006	107.609	<0.001
GHQ3	0.555	0.007	82.289	<0.001
GHQ4	0.601	0.007	89.208	<0.001
GHQ7	0.661	0.005	125.079	<0.001
GHQ8	0.665	0.006	109.295	<0.001
GHQ12	0.680	0.005	127.771	<0.001

Table 3. 16

Standardised Factor Loadings of the Loss of confidence Factor of Graetz' Representation of the GHQ-12

	Estimate	S.E.	Est./S.E.	P-Value
GHQ10	0.869	0.003	271.271	<0.001
GHQ11	0.767	0.004	186.103	<0.001

This model has yielded strong, statistically significant relationships between the factors and their component items. All relationships ranged between 0.555, representing a moderately strong correlation to 0.869, which represented a very strong correlation. The 'Loss of Confidence' factor was especially strongly correlated with its items with an average factor loading of 0.818. Anxiety and depression averaged correlations of 0.735 between that factor and its respective items, whereas Social Dysfunction only averaged 0.633. While this is still above what is recommended by MacCallum (1999), it does not score highly above this threshold.

Table 3. 17

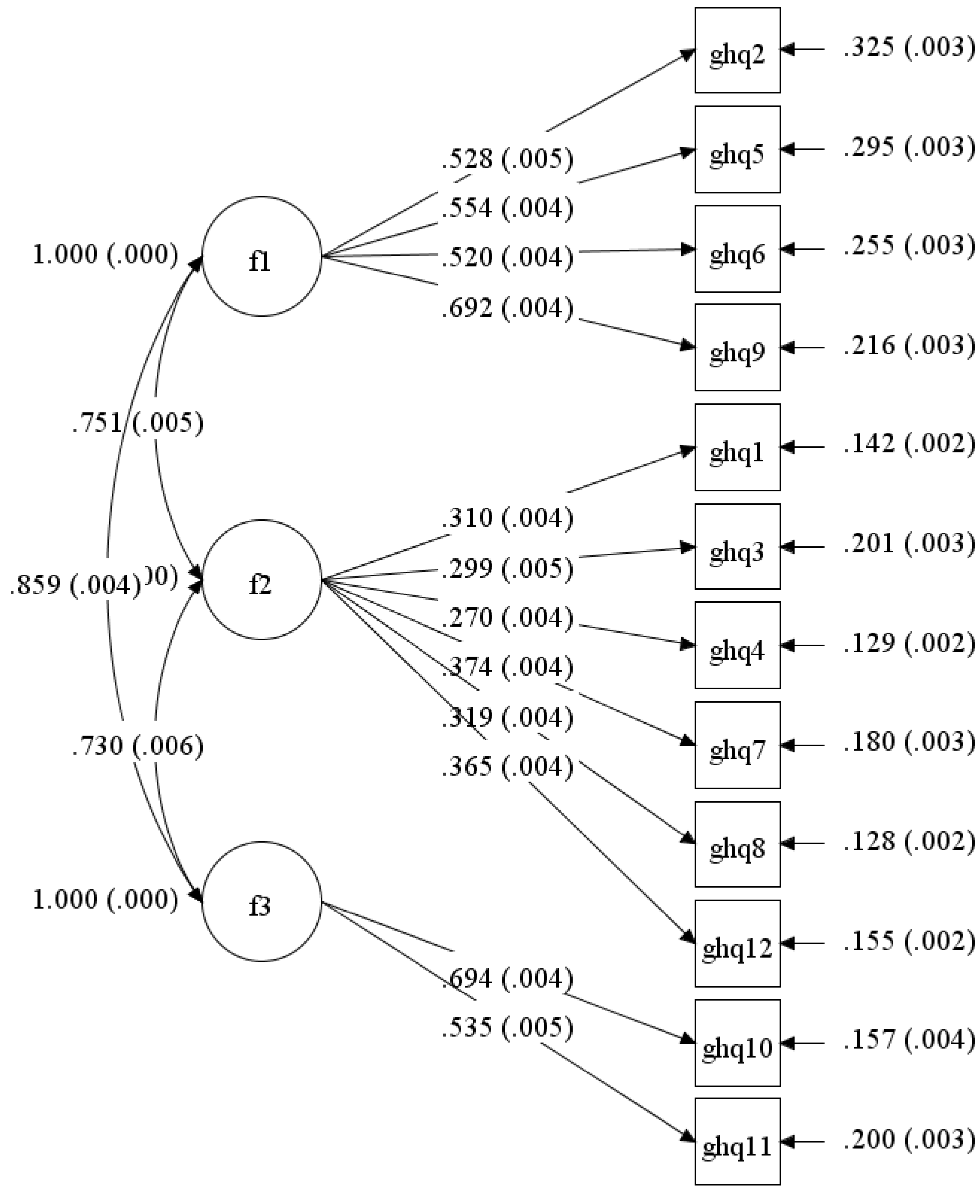
Standardised Inter-factor Correlations Between Factors Proposed in Graetz' Representation of the GHQ-12

	STDYX	S.E.	Est./S.E.	P-Value
Anxiety and Depression with Social Dysfunction	0.751	0.005	164.473	<0.001
Anxiety and Depression with Loss of Confidence	0.859	0.004	224.345	<0.001
Social Dysfunction with Loss of Confidence	0.730	0.006	130.418	<0.001

All factors were strongly correlated with each other, however, 'Loss of Confidence' correlated with 'Anxiety and Depression with a value of 0.859, which was strong enough to warrant concerns that the factors were so highly correlated as to indicate that they measured very similar concepts.

Figure 3. 6

A Graphical Depiction with Attached Factor Loadings and Inter-factor Correlations for Graetz' Representation of the GHQ-12



Note. This figure is a graphical representation of Graetz' (1991) multidimensional representation of the GHQ-12. F1 represents the 'Anxiety and Depression' Factor, F2 represents the 'Social Dysfunction' factor and F3, the 'Loss of Confidence' factor. The various items marked 'ghq' represent the items.

Ye (2009)

Table 3. 18

Standardised Factor Loadings of the General Mental Health Factor of Ye's Representation of the GHQ-12

	Estimate	S.E.	Est./S.E.	P-Value
GHQ1	0.631	0.006	104.376	<0.001
GHQ2	0.495	0.006	83.854	<0.001
GHQ3	0.559	0.007	83.546	<0.001
GHQ4	0.603	0.007	89.77	<0.001
GHQ5	0.506	0.006	82.383	<0.001
GHQ6	0.557	0.006	92.181	<0.001
GHQ7	0.656	0.005	120.296	<0.001
GHQ8	0.667	0.006	110.158	<0.001
GHQ9	0.626	0.005	132.627	<0.001
GHQ10	0.630	0.006	112.059	<0.001
GHQ11	0.596	0.007	82.829	<0.001
GHQ12	0.681	0.005	128.386	<0.001

Table 3. 19

Standardised Factor Loadings of the Methods Factor of Ye's Representation of the GHQ-12

	Estimate	S.E.	Est./S.E.	P-Value
GHQ2	0.448	0.008	58.261	<0.001
GHQ5	0.484	0.009	56.661	<0.001
GHQ6	0.437	0.007	62.825	<0.001
GHQ9	0.536	0.006	93.948	<0.001
GHQ10	0.476	0.009	53.82	<0.001
GHQ11	0.377	0.009	39.834	<0.001

All items exhibited statistically significant relationships with the primary factor and that these relationships ranged from weak to moderate. These scores averaged 6.001, which only just exceeded the suggested cut-off of 0.6 suggested in MacCallum (1999). Items 2, 5, and 6 exhibited relatively weak relationships, however, these items all exhibited relationships with the methods factor. In relation to the methods factor, all items exhibited a moderately strong, statistically significant relationship with standardised scores ranging from 0.401 to 0.570.

Table 3. 20

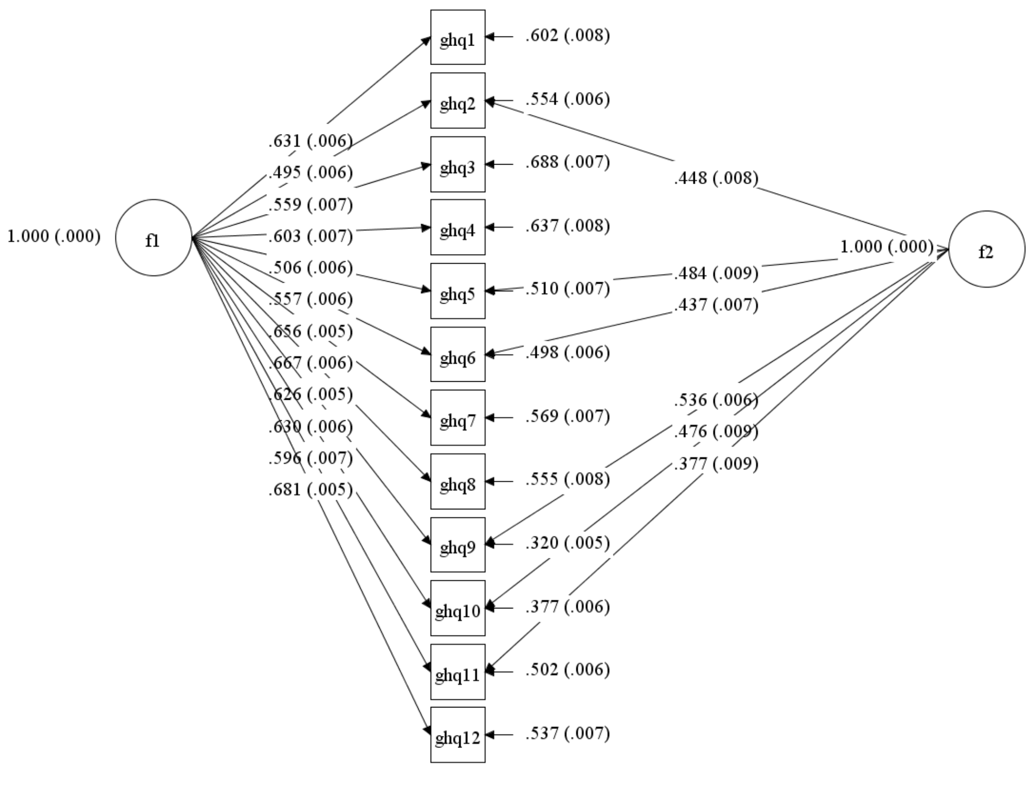
Standardised Inter-factor Correlations Between Factors Proposed in Ye's Representation of the GHQ-12

	STDYX	S.E.	Est./S.E.	P-Value
Method with General Mental Health factor	N/A	N/A	N/A	N/A

The relationship between the two factors went out of bounds, however given that conceptually the two factors were not expected to exhibit any relationships this was not deemed as indicative of poor fit.

Figure 3. 7

A Graphical Depiction with Attached Factor Loadings and Inter-factor Correlations for Ye's Representation of the GHQ-12



Note. This figure is a graphical representation of the Ye's (2009) factor model of the GHQ-12. F1 represents the unidimensional factor and F2 represents the 'method factor'. The various items marked 'ghq' represent the items.

Politi (1994)

Table 3. 21

Standardised Factor Loadings of the Social Dysfunction Factor of Politi's Representation of the GHQ-12

	Estimate	S.E.	Est./S.E.	P-Value
GHQ1	0.643	0.006	107.711	<0.001
GHQ3	0.561	0.007	82.552	<0.001
GHQ4	0.614	0.007	90.351	<0.001
GHQ7	0.664	0.005	124.23	<0.001
GHQ8	0.673	0.006	109.512	<0.001
GHQ12	0.484	0.012	41.192	<0.001

Table 3. 22 - *Standardised Factor Loadings of the Dysphoria Factor of Politi's Representation of the GHQ-12*

	Estimate	S.E.	Est./S.E.	P-Value
GHQ2	0.657	0.004	149.511	<0.001
GHQ5	0.685	0.004	168.112	<0.001
GHQ6	0.706	0.004	161.856	<0.001
GHQ9	0.821	0.003	298.87	<0.001
GHQ10	0.798	0.003	256.3	<0.001
GHQ11	0.716	0.004	159.839	<0.001
GHQ12	0.22	0.011	20.593	<0.001

All items yielded statistically significant relationships with their respective factors; however, from the tables, it was clear that the only cross-loaded item, HQ12 scores considerably poorer in both cases than the other items, with estimates of 0.484 and 0.220 respectively. The loading of GHQ-12 onto the Dysphoria factor could be described, using Comrey and Lee's (2013) classifications, as 'poor'. This would suggest that this item is incorrectly placed. This outlier would also make the investigation of averages meaningless. With the exception of the cross-loaded item, all other items displayed a strong correlation between the items.

Table 3. 23

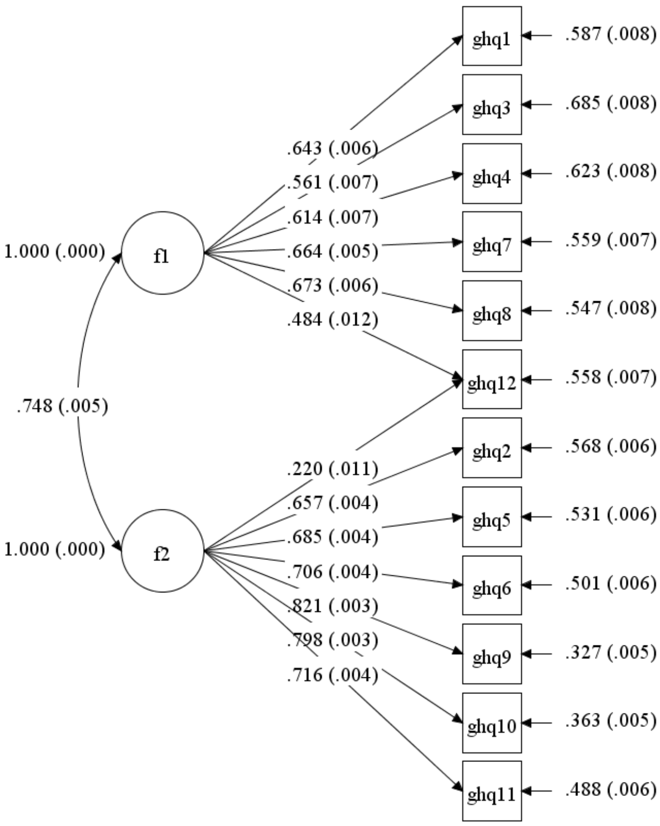
Standardised Inter-factor Correlations Between Factors Proposed by in Politi's Representation of the GHQ-12

	STDYX	STDY	S.E.	Est./S.E.	P-Value
Social Performance with General Dysphoria	0.748	0.748	0.005	154.810	<0.001

The inter-factor relationship between the two items is both strong and statistically significant while not so high as to imply that factors were measuring the same construct.

Figure 3. 8

A Graphical Depiction with Attached Factor Loadings and Inter-factor Correlations for Politi's Representation of the GHQ-12



Note. This figure is a graphical representation of the Politi's representation of the GHQ-12. F1 represents the 'Social Dysfunction' factor and F2 represents the 'General Dysphoria' factor. The various items marked 'ghq' represent the items.

Andrich and van Schoubroeck (1989)

Table 3. 24

Standardised Factor Loadings of the Positively Worded Items Factor of Andrich and von Schoubroeck's Representation of the GHQ-12

	Estimate	S.E.	Est./S.E.	P-Value
GHQ1	0.636	0.006	107.668	<0.001
GHQ3	0.554	0.007	82.362	<0.001
GHQ4	0.601	0.007	89.047	<0.001
GHQ7	0.662	0.005	126.009	<0.001
GHQ8	0.665	0.006	109.316	<0.001
GHQ12	0.681	0.005	128.173	<0.001

Table 3. 25

Standardised Factor Loadings of the Negatively Worded Items Factor of Andrich and von Schoubroeck's Representation of the GHQ-12

	Estimate	S.E.	Est./S.E.	P-Value
GHQ2	0.657	0.004	149.457	<0.001
GHQ5	0.686	0.004	168.073	<0.001
GHQ6	0.708	0.004	162.4	<0.001
GHQ9	0.819	0.003	296.007	<0.001
GHQ10	0.799	0.003	257.181	<0.001
GHQ11	0.715	0.004	159.656	<0.001

All items are strongly correlated with their respective factors in a statistically significant way with all items scoring between 'good' and 'excellent' according to Comrey and Lee's (2013) classifications. The average correlation between the items and their factor was 0.633 for positively worded items and 0.731 for negatively worded items. As with the 1-factor model, GHQ3 was the most weakly correlated item, scoring a correlational relationship of 0.554, however, this still constituted a 'good' relationship. It is also interesting to note that item 12 correlated strongly with its respective factor with a score of 0.681, suggesting a 'very good' relationship with the factor.

Table 3. 26

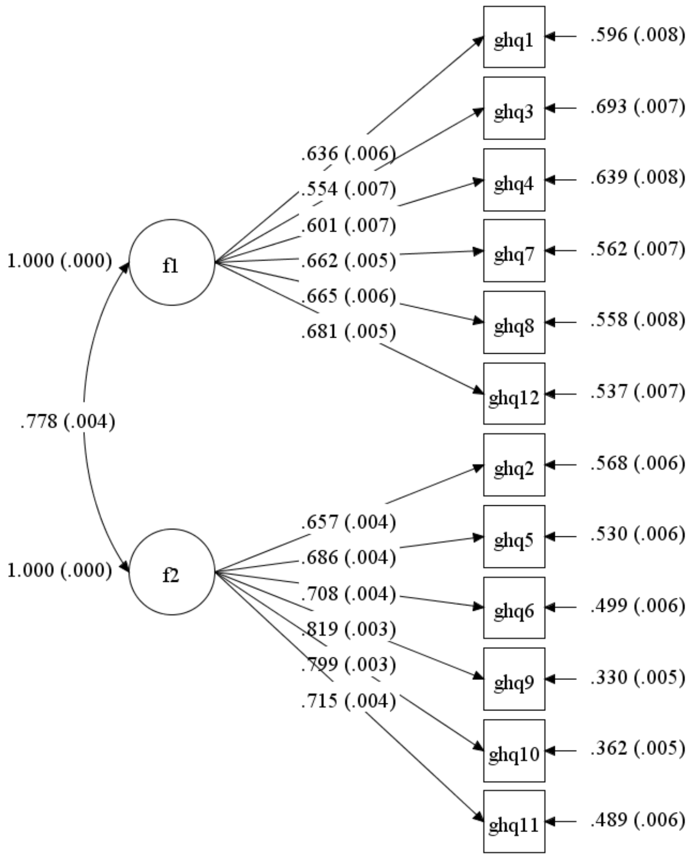
Standardised Inter-factor Correlations Between Factors Proposed in Andrich and von Schoubroeck's Representation of the GHQ-12

	STDYX	S.E.	Est./S.E.	P-Value
Positively Worded Items with Negatively Worded Items	0.778	0.004	184.384	<0.001

The two factors exhibited, a statistically significant relationship 0.778 when correlated, which was not so high as to warrant concern that the two factors were measuring similar constructs.

Figure 3. 9

A Graphical Depiction with Attached Factor Loadings and Inter-factor Correlations for Andrich and von Schoubroeck's Representation of the GHQ-12



Note. This figure is a graphical representation of Andrich and von Schoubroeck's representation of the GHQ-12. F1 and F2 represent the positively and negatively worded items respectively. The various items marked 'ghq' represent the items.

Martin (1999)

Table 3. 27

Standardised Factor Loadings of the Cope Factor of Martin's Representation of the GHQ-12

	Estimate	S.E.	Est./S.E.	P-Value
GHQ1	0.655	0.007	98.906	<0.001
GHQ3	0.57	0.007	80.144	<0.001
GHQ4	0.647	0.007	92.902	<0.001
GHQ8	0.597	0.017	35.196	<0.001

Table 3. 28

Standardised Factor Loadings of the Stress Factor of Martin's Representation of the GHQ-12

	Estimate	S.E.	Est./S.E.	P-Value
GHQ2	0.688	0.005	146.435	<0.001
GHQ5	0.719	0.005	150.784	<0.001
GHQ7	0.59	0.006	97.76	<0.001

Table 3. 29

Standardised Factor Loadings of the Depression Factor of Martin's Representation of the GHQ-12

	Estimate	S.E.	Est./S.E.	P-Value
GHQ6	0.7	0.004	158.756	<0.001
GHQ8	0.083	0.016	5.29	<0.001
GHQ9	0.813	0.003	287.256	<0.001
GHQ10	0.8	0.003	252.772	<0.001
GHQ11	0.724	0.004	161.772	<0.001
GHQ12	0.618	0.005	120.408	<0.001

This model displays sufficiently strong and statistically significant relationships between the factors and their component items with the exception of item 8, which was cross-loaded between the depression and cope factors. This variable displayed an extremely weak correlation between the item and the 'Depression' factor, but with a fair relationship with the 'Cope' factor.

With the exception of item 8, ‘Depression’ displayed a very strong relationship with its component items, whereas ‘Cope’ displayed a weaker set of relationships. The average correlation between ‘Cope’ and its items is 0.617, which just meets MacCallum’s (1999) threshold of an average factor loading of 0.6 on all items. The exhibition of such a poor loading on item 8 would suggest that this model does not adequately describe the data and, therefore, in this case, is not appropriate.

Table 3. 30

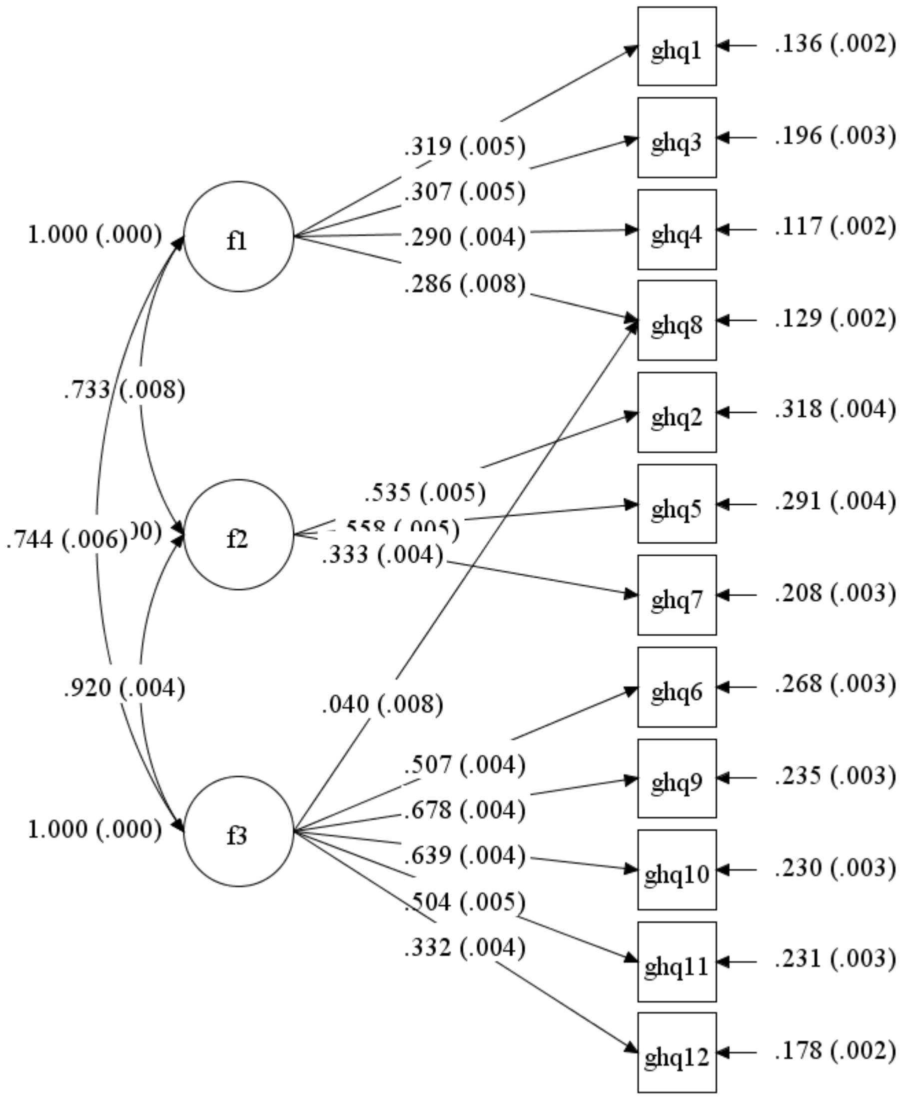
Standardised Inter-factor Correlations Between Factors Proposed in Martin’s Representation of the GHQ-12

	Estimate	S.E.	Est./S.E.	P-Value
Cope with Stress	0.733	0.008	88.143	<0.001
Cope with Depression	0.744	0.006	114.649	<0.001
Stress with Depression	0.920	0.004	235.318	<0.001

The standardised score for depression correlated with stress is noticeably high at 0.920, which would represent abnormally high factor correlations and may limit the unique predictive power of these items. Furthermore, the high degree of factor correlations may indicate that this factor was redundant. That said, literature would suggest that there should be a strong link between stress and depression, especially in a workplace setting (Melchior et al., 2007). While particularly high factor correlations may suggest that two factors may be measuring the same thing, it may also be that one of the factors is derived from the other, thus implying two distinct concepts but strong relationships between them (Carter & Garber, 2011). Regardless of the reason behind the relationship, it must be questioned whether it is appropriate for these two factors to be treated separately, given their strong correlation. The relationship between cope with depression and stress is a strong correlation but not so strong as to imply that the factors are too strongly related to measure distinct concepts.

Figure 3. 10

A Graphical Depiction with Attached Factor Loadings and Inter-factor Correlations for Martin's Representation of the GHQ-12



Note. This figure is a graphical representation of Martin's Representation of the GHQ-12. F1, F2 and F3 represent the 'Stress', 'Cope' and 'Depression' factors respectively. The various items marked 'ghq' represent the items.

Worsley and Gribbin (1997)

Table 3. 31

Standardised Factor Loadings of the Social Performance Factor of Worsely and Gribbin's Representation of the GHQ-12

	Estimate	S.E.	Est./S.E.	P-Value
GHQ1	0.566	0.006	88.731	<0.001
GHQ2	0.639	0.004	146.126	<0.001
GHQ3	0.471	0.007	65.613	<0.001
GHQ4	0.495	0.007	66.57	<0.001
GHQ11	0.7	0.005	145.289	<0.001

Table 3. 32

Standardised Factor Loadings of the Anhedonia Factor of Worsely and Gribbin's Representation of the GHQ-12

	Estimate	S.E.	Est./S.E.	P-Value
GHQ5	0.668	0.004	154.594	<0.001
GHQ6	0.697	0.005	153.742	<0.001
GHQ7	0.607	0.005	110.364	<0.001
GHQ8	0.59	0.007	90.328	<0.001
GHQ12	0.645	0.005	124.864	<0.001

Table 3. 33

Standardised Factor Loadings of the Loss of Confidence of Worsely and Gribbin's Representation of the GHQ-12

	Estimate	S.E.	Est./S.E.	P-Value
GHQ9	0.817	0.003	258.622	<0.001
GHQ10	0.806	0.003	241.575	<0.001

The 'Anhedonia' and 'Loss of Confidence' factors both exhibit strong statistically significant relationships with their component factors however, the 'Social Performance' factor does not exhibit such desirable characteristics. While possessing statistically significant relationships, these do not match the strength of relationships that the other factors do. Social Performance has an average factor loading of 0.574, which is below the suggested cut-off by MacCallum. The weakness of the relationships to this factor may raise some questions as to the appropriateness of this model, however, it would not represent a terminal failure at this stage.

Table 3. 34

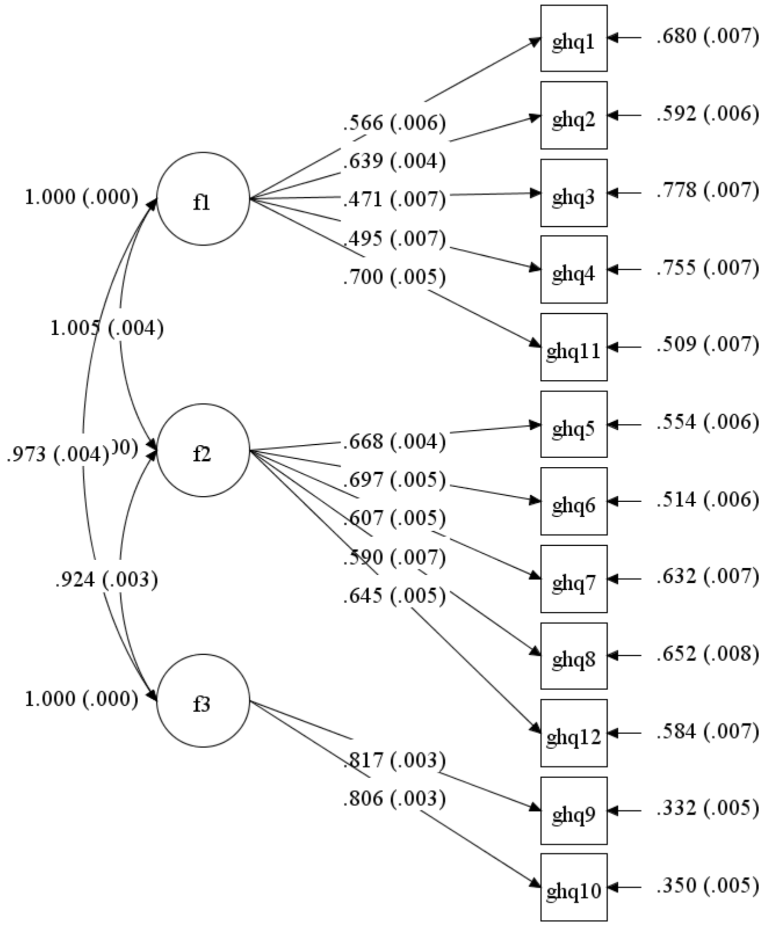
Standardised Inter-factor Correlations Between Factors Proposed in of Worsely and Gribbin's Representation of the GHQ-12

	STDYX	S.E.	Est./S.E.	P-Value
Social Performance with Anhedonia	**1.005**	**0.004**	**286.828**	**<0.001**
Social Performance with Loss of Confidence	0.973	0.004	242.942	<0.001
Anhedonia with Loss of Confidence	0.924	0.003	267.726	<0.001

The relationship between the 'Anhedonia' and 'Social Performance' factors, which is highlighted, is not 'positive definite.' This, as previously stated, is likely to indicate that the model is inappropriate to the data.

Figure 3. 11

A Graphical Depiction with Attached Factor Loadings and Inter-factor Correlations for Worsely and Gribbin's Representation of the GHQ-12



Note. This figure is a graphical representation of Worsely and Gribbin's representation of the GHQ-12. F1, F2 and F3 represent the 'Social Performance', 'Anhedonia' and 'Loss of Confidence' factors respectively. The various items marked 'ghq' represent the items.

Schmitz (1999)

Table 3. 35

Standardised Factor Loadings of the Anxiety/Depression Factor of Schmitz's Representation of the GHQ-12

	Estimate	S.E.	Est./S.E.	P-Value
GHQ1	0.571	0.006	97.306	<0.001
GHQ2	0.641	0.004	148.522	<0.001
GHQ3	0.149	0.092	1.617	0.106
GHQ6	0.691	0.004	158.749	<0.001
GHQ7	0.593	0.005	108.879	<0.001
GHQ10	0.769	0.003	226.471	<0.001
GHQ11	0.699	0.005	144.175	<0.001

Table 3. 36

Standardised Factor Loadings of the Social Performance Factor (F1) of Schmitz's Representation of the GHQ-12

	Estimate	S.E.	Est./S.E.	P-Value
GHQ3	0.323	0.091	3.542	<0.001
GHQ4	0.494	0.007	71.949	<0.001
GHQ5	0.664	0.004	159.588	<0.001
GHQ8	0.569	0.006	88.073	<0.001
GHQ9	0.787	0.003	265.936	<0.001
GHQ12	0.626	0.005	121.916	<0.001

All the items load in a statistically significant way with the exception of item GHQ3. This item has a non-significant relationship with the anxiety/depression factor and has the weakest of all relationships in the social performance factor. The performance of GHQ3 remains a problem in all two-factor models in this analysis. With the exception of GHQ3 other items score averagely with all items associated with Anxiety/depression scoring 'good' according to Cromley and Lee's (2013) classifications and all items in Social performance scoring at least 'fair'. The presence of a non-significant relationship, however, would cast doubt on the validity of the model. Overall the model does not score well enough to suggest that it represents a good explanation of the data.

Table 3. 37

Standardised Inter-factor Correlations Between Factors Proposed in Schmitz's Representation of the GHQ-12

	STDYX	S.E.	Est./S.E.	P-Value
Social performance with Anxiety/Depression	**1.024**	**0.002**	**426.509**	**<0.001**

The relationship between Social Performance and Anxiety/Depression was described as not being 'positive definite'. This occurs when a correlation (highlighted) is greater than 1. Muthen (2016) described models having non 'positive definite' relationships as inappropriate for the data at hand and in need of modification.

Unidimensional

Table 3. 38

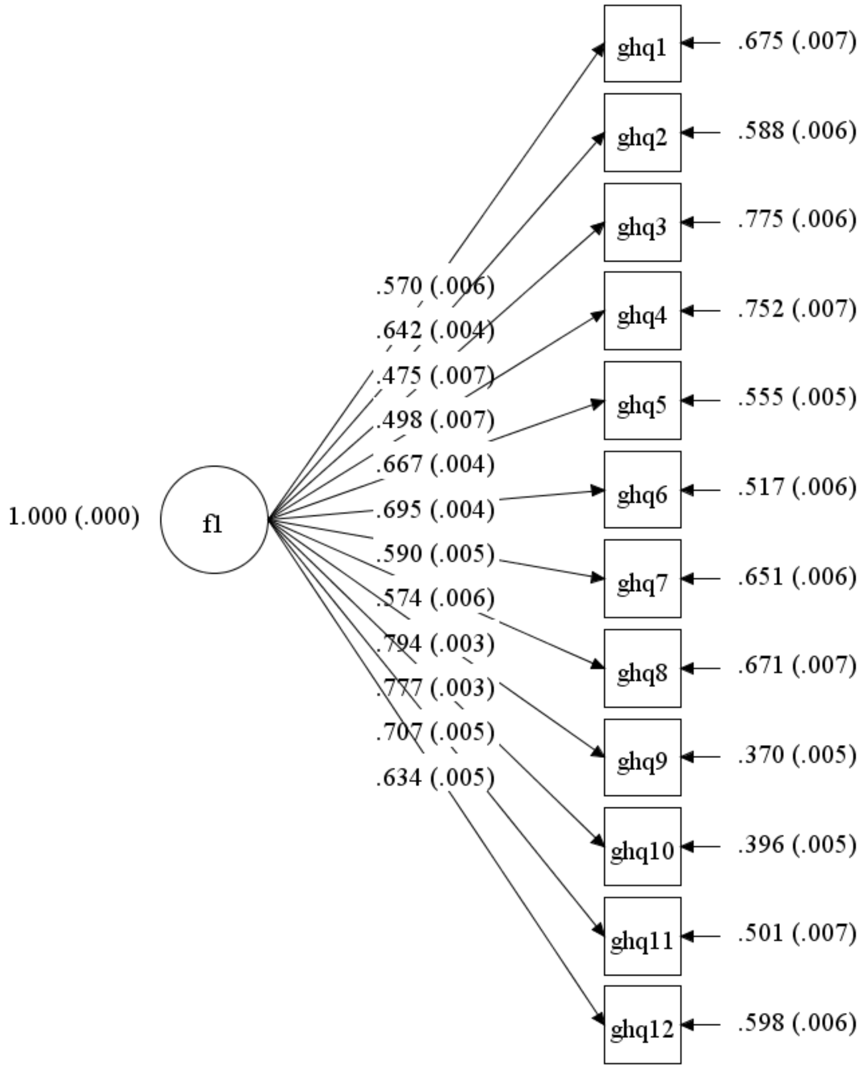
Standardised Factor Loadings of a Unidimensional Representation of the GHQ-12

	Estimate	S.E.	Est./S.E.	P-Value
GHQ1	0.57	0.006	96.459	<0.001
GHQ2	0.642	0.004	146.771	<0.001
GHQ3	0.475	0.007	70.028	<0.001
GHQ4	0.498	0.007	72.265	<0.001
GHQ5	0.667	0.004	162.615	<0.001
GHQ6	0.695	0.004	159.416	<0.001
GHQ7	0.59	0.005	107.652	<0.001
GHQ8	0.574	0.006	88.778	<0.001
GHQ9	0.794	0.003	272.13	<0.001
GHQ10	0.777	0.003	244.527	<0.001
GHQ11	0.707	0.005	153.186	<0.001
GHQ12	0.634	0.005	125.168	<0.001

All items except GHQ3 and GHQ4 exhibited strong/ moderate, statistically significant factor loadings with the single factor according to Comrey and Lee's (2013) thresholds. The average factor loading was 0.635, which fulfilled MacCalum's (1999) threshold. The weakest item, GHQ3 still yielded a result of 0.475, which according to Comrey and Lee's (2013) definition does still constitute a 'fair' relationship.

Figure 3. 12

A Graphical Depiction with Attached Factor Loadings for a Unidimensional Representation of the GHQ-12



Note. This figure is a graphical representation of the unidimensional representation of the GHQ-12. F1 represents the unidimensional factor. The various items marked 'ghq' represent the items.

3.5.- Discussion

One issue that must be taken into account when comparing the literature around factor structure and this chapter's findings, is the evolution of the statistical techniques into dimensionality. As research into the factor structure of the GHQ-12 has taken place over 30 years, the techniques that researchers have used have developed over time. A number of the studies which verified the GHQ-12 in different locations (Garyfallos et al., 1991) have used techniques such as PCA or internal consistency analysis and have not used the relatively modern approach of analysing complex and unorthodox models through Confirmatory Factor Analysis. Many of the early studies into the GHQ-12 did not investigate more complex structures such as were proposed in Hankins (2008) and Ye's (2009). Studies pre-dating 2006 only looked at conventional factor models, which separated the data into two or three factors without specific methods to model method effects. Even Andrich & van Schoubroeck (1989), a study that attempted to explain method effects, treated these as if they were two factors that behave similarly to as any other factor would. Bearing this in mind, it was important to acknowledge that much of the early research was conducted without including Hankins (2008) and Ye's (2009) model so it was possible that more complex dimensional representations would have demonstrated better fit than the models proposed if they were investigated.

This chapter's aim was to test an exhaustive list of the models that were prominent in the literature and to identify those which exhibited acceptable fit through a number of analyses detailed above. This chapter focused on model fit and model characteristics such as factor loadings and factor correlations, whereas chapter 4 will focus on the conceptual soundness and validity of these models. This section will discuss the interpretation of the analyses conducted and what implications that these may have for subsequent chapters. The analysis was conducted in two stages, model fit was investigated initially with model characteristics, i.e. factor loadings and correlations being subsequently investigated. Results relating to each stage of these analyses were discussed in turn.

Initially, when discussing model fit, the most pertinent finding was that the unidimensional model did not fit the UKHLS data well, exhibiting the poorest fit of all the models tested. The model which displayed the best fit was that of Hankins (2008). This model utilised correlated errors to model the effect of method factors. Graetz's (1991) three-factor model demonstrated good fit, performing second-best of the models tested. Ye's model also performed strongly with good fit being demonstrated by the battery of fit statistics.

While exhibiting strong fit, the factor loadings on Hankins' (2008) model were not as strong relative to other models and using the guidelines outlined in Comrey and Lee (2013) they range from '*fair*' to '*good*' relationships. Graetz' (1991) model exhibited strong factor loadings, but factor correlations were high enough between two of the factors as to indicate a likelihood that the factors may have been measuring similar concepts. Ye's (2009) model exhibited factor loadings which were noticeably weak relative to other well-fitting models. These loadings were according to Comrey and Lee's classifications (2013) classified as '*fair*'. It is possible that the nature of this model, with a number of cross-loaded items, may have adversely affected these results.

Several models had items loading onto factors with weak relationships. These weak relationships would imply that the item is inappropriately associated with the factor. Of particular note was Politi (1991) and Martin's (1999) model, which included cross-loaded items. Both cross-loaded items exhibited particularly weak, although statistically significant, relationships with their respective factors. Also, of note was

Schmitz's (1999) model, which contained a non-statistically significant factor loading, implying that the respective item, GHQ3 was not associated with the correct factor. GHQ3 performed particularly poorly in all 2-factor solutions with associations ranging from 'fair' to 'weak' according to Comrey and Lee's (2013) classifications.

In relation to inter-factor correlations, generally, those factors which attempted to simulate method effects displayed moderately strong relationships with each other. Of particular note, Worsely and Gribbin's three-factor model (1997), and Martin's two-factor model (1999) which were found to have particularly high factor correlations. As previously mentioned, high factor correlations are not necessarily a problem, however, it might imply that the factors were measuring similar concepts. Consequently, as discussed in Gao et al. (2004) and Shevlin and Adamson (2005), there may be no practical benefit in treating the factors separately.

Molina and Rodrigo (2013) conducted a meta-analysis that did include these modern models. They detailed the populations tested and the techniques that various researchers had used in their respective studies. They found that within African populations, three-factor models tend to score higher (Abubakar & Fischer, 2012, Abubakar & Fischer, 2012), and this was similar within Chinese participants (Ye, 2009). However, Smith et al. (2013), who investigated English participants, suggested that a correlated errors approach was more appropriate, and this conclusion was corroborated by Aguardo (2012) and Molina and Rodrigo (2013) in Spanish populations. It appeared that within the sources listed, there appeared to be a trend where Western European populations gravitate towards a correlated errors dimensional representation of the GHQ-12. In contrast, when investigating African and Asian populations a three-factor model was more appropriate. This research was further supported in subsequent research in 2015 where Angolan participant's responses demonstrated good fit with a 3-factor model (Tomas, Gutierrez & Sancho, 2015), and the researchers noted that method effects did not appear to affect results significantly.

The analysis conducted in this chapter showed that the GHQ responses from the UKHLS questionnaire displayed results in line with the trend identified namely, that Western European populations tended to demonstrate the best fit with Hankins' dimensional representation of the GHQ-12. This research raised questions as to the generalisability of GHQ-12 worldwide. As a result, researcher may wish to research the GHQ-12 using modern techniques worldwide in order to test its behaviour in a global context and to further investigate if dimensional representations are more appropriate in some populations than others.

Going forward in this thesis, initial findings suggested that the appropriate dimensional representation for the wave 1 data was that of Hankins (2008), however strong performances by other models would suggest that this is still not resolved. The three-factor solution of Graetz (1991) performed well, as did Ye's (2009) however both of these models had demonstrated characteristics which may give rise to issues in later analysis, namely the factor loadings in Ye (2009) and the high levels of factor correlation in Graetz (1991). It, therefore, was deemed useful to conduct research into the utility and validity of the factors of each of the models before progressing to longitudinal research, which is further detailed in Chapter 4.

3.5.1- Research Implications

While this chapter has not been able to provide a clear direction as to the most appropriate dimensional representation of the GHQ-12 in the UKHLS data, it has however conducted analyses in a representative UK population and detailed a number of potential models which represent a good fit of the data. In chapter 4, it is hoped that those findings of this chapter combined with that of a more conceptual and validity based nature, will provide a more definitive answer to the model which best represents the data. In terms of research impact, this chapter has corroborated the results of Smith et al. (2013) which found Hankins' (2008) model to be the most appropriate for a representative population within the UK, albeit this study only used English participants.

References

Abubakar, A., & Fischer, R. (2012). The factor structure of the 12-item General Health Questionnaire in a literate Kenyan population. *Stress and Health*, *28*(3), 248-254.

Aguado, J., Campbell, A., Ascaso, C., Navarro, P., Garcia-Esteve, L., & Luciano, J. V. (2012). Examining the factor structure and discriminant validity of the 12-item General Health Questionnaire (GHQ-12) among Spanish postpartum women. *Assessment*, *19*(4), 517-525.

Andrich, D., & Van Schoubroeck, L. (1989). The General Health Questionnaire: A psychometric analysis using latent trait theory. Psychological Medicine, 19, 469 – 485.

Atchley, B. (2019). *Introduction to Principal Components and Factor Analysis*.

Awang, Z. (2012). Structural equation modeling using AMOS graphic. Penerbit Universiti Teknologi MAR

Bagozzi RP: An examination of the psychometric properties of measures of negative affect in the PANAS scales. Journal of Personality and Social Psychology. 1993, 65: 836-851. 10.1037/0022-3514.65.4.836.

Barrett, P. (2007). Structural equation modelling: adjudging model fit. *Personality and Individual Differences, 42*, 815–824.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13*, 186–203. doi:10.1207/s15328007sem1302_2

Benítez, I., Van de Vijver, F., & Padilla, J. L. (2017). An integrated approach to bias in the Spanish version of the General Health Questionnaire (GHQ-12).

Bentler, P. M., & Chou, C. P. (1987) Practical issues in structural modeling. *Sociological Methods & Research, 16*, 78-117.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588-600.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, *40*(1), 5-32.

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis*. Psychology press.

Cordery JL, Sevastos PP: Responses to the original and revised job diagnostic survey: Is education a factor in responses to negatively worded items?. Journal of Applied Psychology. 1993, 78: 141-143. 10.1037/0021-9010.78.1.141.

Cordery, J. L., & Sevastos, P. P. (1993). Responses to the original and revised job diagnostic survey: Is education a factor in responses to negatively worded items?. *Journal of Applied Psychology*, *78*(1), 141.

del Pilar Sánchez-López, M., & Dresch, V. (2008). The 12-Item General Health Questionnaire (GHQ-12): reliability, external validity and factor structure in the Spanish population. *Psicothema*, *20*(4), 839-843.

Dohoo, I., Ducrot, C., Fourichon, C., Donald, A. and Hurnik, D. (1997), “An overview of techniques for dealing with large numbers of independent variables in epidemiologic studies”, Preventive Veterinary Medicine, Vol. 29 No. 3, pp. 221-239.

Fruchter, B. (1954). Introduction to factor analysis.

Galbraith, J. I., Moustaki, I., Bartholomew, D. J., & Steele, F. (2002). *The analysis and interpretation of multivariate data for social scientists*. Chapman and Hall/CRC.

Galbraith, J. I., Moustaki, I., Bartholomew, D. J., & Steele, F. (2002). *The analysis and interpretation of multivariate data for social scientists*. Chapman and Hall/CRC.

Garyfallos, G., Karastergiou, A., Adamopoulou, A., Moutzoukis, C., Alagiozidou, E., Mala, D., & Garyfallos, A. (1991). Greek version of the General Health Questionnaire: accuracy of translation and validity. *Acta Psychiatrica Scandinavica*, *84*(4), 371-378.

Goldberg DP, Gater R, Sartorius N et al. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. Psychological Medicine; 27:191–197

Goldberg DP, Williams P: A User's Guide to the General Health Questionnaire. 1988, Windsor: nferNelson

Graetz, B. (1991). Multidimensional properties of the general health questionnaire. *Social psychiatry and psychiatric epidemiology, 26*(3), 132-138.

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality1. *Educational and Psychological Measurement*, *37*(4), 827-838.

Greenberger, E., Chen, C., Dmitrieva, J., & Farruggia, S. P. (2003). Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: Do they matter?. *Personality and individual differences, 35*(6), 1241-1254.

Hankins, M. (2008). The factor structure of the twelve item General Health Questionnaire (GHQ-12): the result of negative phrasing?. *Clinical Practice and Epidemiology in Mental Health*, *4*(1), 10.

Hoeymans, N., Garssen, A. A., Westert, G. P., & Verhaak, P. F. (2004). Measuring mental health of the Dutch population: a comparison of the GHQ-12 and the MHI-5. *Health and quality of life outcomes*, *2*(1), 23.

Holi, M. M., Marttunen, M., & Aalberg, V. (2003). Comparison of the GHQ-36, the GHQ-12 and the SCL-90 as psychiatric screening instruments in the Finnish population. *Nordic journal of psychiatry*, *57*(3), 233-238.

Horan, P. M., DiStefano, C., & Motl, R. W. (2003). Wording effects in self-esteem scales: Methodological artifact or response style?. *Structural Equation Modeling*, *10*(3), 435-455.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424–453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Huang, C., & Dong, N. (2012). Factor structures of the Rosenberg self-esteem scale. *European Journal of Psychological Assessment*.

Kenny, D. (2019). *SEM: Fit (David A. Kenny)*. [online] Davidakenny.net. Available at: http://davidakenny.net/cm/fit.htm [Accessed 29 Aug. 2019].

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2014). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, in press.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling, 10*, 333-3511.

Knies, G. (2017). Understanding society: The UK household longitudinal study, Waves 1–7 (User Guide). *Institute for Social and Economic Research: University of Essex:: Colchester*.

Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological methods*, *7*(2), 228.

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior research methods*, *48*(3), 936-949.

Li, C. H. (2014). *The performance of MLR, USLMV, and WLSMV estimation in structural regression models with ordinal variables*. Michigan State University.

Lindwall, M., Barkoukis, V., Grano, C., Lucidi, F., Raudsepp, L., Liukkonen, J., & Thøgersen-Ntoumani, C. (2012). Method effects: The problem with negatively versus positively keyed items. *Journal of personality assessment*, *94*(2), 196-204.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130-149.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological methods*, *4*(1), 84.

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors?. *Journal of personality and social psychology*, *70*(4), 810.

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors?. *Journal of personality and social psychology*, *70*(4), 810.

Martin, A. J. (1999). Assessing the multidimensionality of the 12-item General Health Questionnaire. Psychological Reports, 84, 927–935

Melchior, M., Caspi, A., Milne, B. J., Danese, A., Poulton, R., & Moffitt, T. E. (2007). Work stress precipitates depression and anxiety in young, working women and men. *Psychological medicine*, *37*(8), 1119-1129.

Molina, J. G., Rodrigo, M. F., Losilla, J. M., & Vives, J. (2014). Wording effects and the factor structure of the 12-item General Health Questionnaire (GHQ-12). *Psychological assessment*, *26*(3), 1031.

Montazeri, A., Harirchi, A. M., Shariati, M., Garmaroudi, G., Ebadi, M., & Fateh, A. (2003). The 12-item General Health Questionnaire (GHQ-12): translation and validation study of the Iranian version. *Health and quality of life outcomes*, *1*(1), 66.

Motamed, N., Zakeri, S. E., Rabiee, B., Maadi, M., Khonsari, M. R., Keyvani, H., ... & Zamani, F. (2018). The Factor Structure of the Twelve Items General Health Questionnaire (GHQ-12): a Population Based Study. *Applied Research in Quality of Life*, *13*(2), 303-316.

Motamed, N., Zakeri, S. E., Rabiee, B., Maadi, M., Khonsari, M. R., Keyvani, H., ... & Zamani, F. (2018). The Factor Structure of the Twelve Items General Health Questionnaire (GHQ-12): a Population Based Study. *Applied Research in Quality of Life*, *13*(2), 303-316.

Muthen, B. (2019). *Mplus Discussion >> Non Positive Definite*. [online] Available at: http://www.statmodel.com/discussion/messages/9/22538.html?1458260110 [Accessed 20 Aug. 2019].

Muthén, B. O., & Kaplan, D. (1985). A comparison of some methodologies for the factor-analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38*, 171–180.

Navarro, P., Ascaso, C., Garcia-Esteve, L., Aguado, J., Torres, A., & Martín-Santos, R. (2007). Postnatal psychiatric morbidity: a validation study of the GHQ-12 and the EPDS as screening tools. *General Hospital Psychiatry*, *29*(1), 1-7.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559-572.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559-572.

Politi, P. L., Piccenelli, M., & Wilkinson, G. (1994). Reliability, validity and factor structure of the twelve item General Health Questionnaire among young males in Italy. Acta Psychiatrica Scandinavica, 90, 432– 437.

Rey, J. J., Abad, F. J., Barrada, J. R., Garrido, L. E., & Ponsoda, V. (2014). The impact of ambiguous response categories on the factor structure of the GHQ–12. *Psychological assessment*, *26*(3), 1021.

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and social psychology bulletin*, *28*(12), 1629-1646.

Schmitt, N., & Stuits, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents?. *Applied Psychological Measurement*, *9*(4), 367-373.

Schmitz, N., Kruse, J., & Tress, W. (1999). Psychometric properties of the General Health Questionnaire (GHQ–12) in a German primary care sample. Acta Psychiatrica Scandinavica, 100, 462– 468.

Shevlin, M., & Adamson, G. (2005). Alternative factor models and factorial invariance of the GHQ-12: a large sample analysis using confirmatory factor analysis. *Psychological assessment*, *17*(2), 231.

Satorra, A. (1990). Robustness issues in structural equation modeling: A review of recent developments. *Quality and Quantity, 24,* 367–386.

Satorra, C., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variable analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Sriram, T. G., Chandrashekar, C. R., Isaac, M. K., & Shanmugham, V. (1989). The general health questionnaire (GHQ). *Social Psychiatry and Psychiatric Epidemiology*, *24*(6), 317-320.

Stevens JP (1992) Applied multivariate statistics for the social sciences (2nd edition). Hillsdale, NJ:Erlbaum.

Strathman, A., Gleicher, F., Boninger, D. S., & Edwards, C. S. (1994). The consideration of future consequences: weighing immediate and distant outcomes of behavior. *Journal of personality and social psychology*, *66*(4), 742.

Tanaka, J. S. (1987). " How big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. *Child development*, 134-146.

Tomás, J. M., Gutiérrez, M., & Sancho, P. (2015). Factorial validity of the General Health Questionnaire 12 in an Angolan sample. *European Journal of Psychological Assessment*.

Van Hemert, A. M., Den Heijer, M., Vorstenbosch, M., & Bolk, J. H. (1995). Detecting psychiatric disorders in medical practice using the General Health Questionnaire. Why do cut-off scores vary?. *Psychological Medicine*, *25*(1), 165-170.

Wang, L., & Lin, W. (2011). Wording effects and the dimensionality of the General Health Questionnaire (GHQ-12). *Personality and Individual Differences*, *50*(7), 1056-1061.

Worsley, A., & Gribben, C. C. (1977). A factor analytic study of the twelve item General Health Questionnaire. Australian and New Zealand Journal of Psychiatry, 11, 269 – 272.

Worsley, A., & Gribbin, C. C. (1977). A factor analytic study of the twelve item general health questionnaire. *Australian & New Zealand Journal of Psychiatry*, *11*(4), 269-272.

Ye, S. (2009). Factor structure of the General Health Questionnaire (GHQ-12): The role of wording effects. *Personality and Individual Differences*, *46*(2), 197-201.

Chapter 4 - Determining and Validating the Dimensionality of the GHQ: Model Selection Using Covariates

4.1- Abstract

Introduction

In Chapter 3, Confirmatory Factor Analysis techniques within a structural equation framework were used to identify a number of dimensional representations of the GHQ-12 which demonstrated good fit. A unidimensional model did not represent a good fit for the data, however a number of multidimensional representations and those which through various techniques attempted to model method effects did. To determine the most appropriate dimensional representation of the GHQ for later longitudinal change modelling, a selection of covariates and mental health validating variables were regressed on the dimensions of the competing models.

Method

Factors from models which exhibited acceptable fit in chapter 3 were regressed onto covariates in a structural equation modelling framework relating to demography, diagnosed health conditions, similar tests of mental health to the GHQ-12, such as the SF12, and variables relating to one's perception of their neighbourhood. These analyses were employed to examine the multivariate associations between the identified dimensions of the GHQ-12 and a range of covariates that could help to determine dimensional validity.

Results

It was found that a multidimensional representation of the GHQ-12 did not offer any unique predictive utility as, generally, there was little variability between the factors

and the covariates modelled. As a result, it was decided appropriate to treat the GHQ-12 as unidimensional. Ye's (2009) model, consisted of a factor which claimed to measure mental health and a method factor which captured variability caused by the positively and negatively worded items contained in the GHQ-12. The mental health factor was much more highly associated with diagnoses of clinical depression, and with measures of wellbeing than the method factor.

Conclusion

Ye's (2009) model was deemed as the most appropriate dimensional representation of the data. This was because the model's mental health factor displayed strong relationships with similar measures such as clinical depression diagnoses and tests of wellbeing, and it demonstrated concurrent validity with covariates such as age, rurality and sex. There was also clear variability between the mental health and method factor in relation to clinical depression diagnoses and tests of wellbeing, with the mental health factor displaying noticeably stronger relationships with these covariates.

4.2.- Introduction

In the previous chapter, both unidimensional models which accounted for method effects and multidimensional models were found to exhibit acceptable fit. A number of the models also performed strongly only in certain aspects of the analyses, such as Hankins, (2008) which performed well in relation to model fit statistics but poorly in relation to model factor loadings. The models with acceptable fit varied considerably in relation to their conceptual and methodological underpinnings and are discussed in detail in Chapter 3. Some models such as Graetz (1991) and Politi (1994) suggested that the GHQ-12 may be comprised of a number of distinct dimensions of mental health, while others such as Ye (2009), Hankins (2008) and Andrich and von

Shoebroeck (1989) suggested that the multiple factors extracted from the GHQ-12 were consequences of method effects in the questionnaire and that a unidimensional structure was the optimal representation of the data.

While the analyses in Chapter 3 have provided a statistical basis for model selection, there is no consensus as to the appropriate emphasis to place upon fit statistics compared to other methods within the literature. Some papers have relied exclusively on fit statistics (Martin, 1999) whereas other studies have disregarded superior fit statistics of specific models on the basis of parsimony (Vanheule & Bogaerts, 2005) or predictive ability (Ip & Martin, 2006). On the opposite side of the spectrum to studies such as Martin (1999), some researchers have claimed that fit statistics add little to any analysis (Barret, 2007). It has even been claimed within the literature that they could be counter-productive. Hayduck et al. (2007) argued that an overreliance on arbitrary cut-offs may be misleading and recommended that fit statistics may be more appropriate when viewed as part of a wide range of statistical techniques.

Within the previous chapter, model fit and properties such as factor loadings of various models were investigated, however, it was decided that investigations into the validity the factors produced would be meritorious as generated factors may have been statistical phenomena not necessarily grounded in reality (Atchley, 2019). One technique that can be used to aid researchers in the identification of appropriate models is that of covariate analysis, which refers to examining a psychological construct's associations with a selection of covariates with which it should be meaningfully associated. Covariates are defined as "a broad term for variables which are neither dependent nor independent and are often used as a supplementary form of analysis"

(Dempster & Hannah, 2012). For this analysis, they will refer to supplementary data collected at the same time point, such as medical history.

Covariate analysis affords researches the opportunity to test two important properties of dimensional representations. In relation to multidimensional representations of data, covariate analysis allows researchers to test if there is variability between the factors in relation to covariates. Should this not occur then previous research has questioned the utility of treating the factors separately (Shevlin & Adamson, 2005). It also allows the validity of the factors to be investigated. Depending on the covariate that each factor is being regressed onto, a number of types of validity can be tested. Convergent validity refers to when a measure is strongly correlated with a construct that is assumed to measure the same or a similar concept, whereas discriminant validity refers to the absence of correlations where they should not occur or the presence of negative correlations where they would be expected to (Trochin, 2020). While the analysis in the previous chapter provides a statistical basis for model selection, validity testing in this manner is useful as it allows researchers to test if the factors measure what they are supposed to. Should this not occur, the factor may be incorrectly labelled. The labelling of factors is an inherently subjective process as researchers attempt to identify that which their factors represent. Validity testing provides a method of testing the meaningfulness of these ascribed labels. Incorrect labelling of factors was identified in Rey, (2014) as a possible explanation for divergences in the literature over factor structure.

The purpose of these analyses and expected results varied depending on the covariate and factor being compared, however broadly speaking, one of three things should occur:

1. If the factor and covariate measure the same or a very similar concept, one could expect a strong correlational relationship between the two. For example, GHQ-12 scores should demonstrate strong correlations with other measures of mental health such as the Short Form-12 Mental Health Component (SF-12 MHC) or Short Warwick Edinburgh Test of Mental Wellbeing
2. If the factor and covariate have an established relationship within the literature, one could reasonably expect that relationship to be mirrored in these analyses. For example, female participants should display higher GHQ-12 scores than male participants.
3. If the factor and covariate are not expected to exhibit a relationship, then one could expect that any relationship between the two constructs would be non-significant. This is known as discriminant validity. No variables were included in this analysis which were chosen explicitly because they did not have relationships with GHQ-12 scores, therefore discriminant validity should not be present.

Previous literature has used covariate analysis in model selection. Two studies (Shevlin & Adamson, 2005; Gao et al., 2004) Were particularly relevant as they investigated the relationship between covariates and factors following the use of fit statistics. Both concluded that while multidimensional models of the GHQ, specifically Graetz (1991), often provide the ‘better fitting’ models, the factors contained therein may not provide researchers with any extra utility over treating the GHQ-12 as unidimensional as the factors did not display variability in relation to a wide range of covariates. These studies will be expanded upon in the literature review below.

Building on these findings, it was considered beneficial to investigate the relationships between the factors contained within the models that had performed sufficiently well in Chapter 3. These covariates will be analysed in clusters with variables measuring broadly similar concepts being analysed together. These clusters will be listed below.

4.2.1- Covariates

Self-reported health

This cluster of variables referred to those that measured an aspect of self-reported health. The tests that were included were the Short Form 12 item version (SF-12) and the Short Warwick Edinburgh Scale of Mental Wellbeing (SWE). Both these scales are detailed in Chapter 1 (Section 1.10). Also included was a measure of general health. This 5 point Likert Scale measures participant's response to the question "how would you rate your health generally?" While crude in comparison to other variables in this cluster, it has been included in research similar to the current study (Shevlin & Adamson, 2004). It also provides an albeit crude but general statement as to one's health, both physical and mental which other measures do not.

These variables were included as a test of validity as factors which claim to measure mental health should exhibit strong relationships with SF-12 mental components and SWE scales due to the measurement of similar concepts. It must be noted that previous research has found that the relationship between GHQ-12 and SF-12 responses are significantly more highly correlated than GHQ12 and SWE (Burton, Laurie and Lynn, 2012) and this may be because the SWE measures mental wellbeing which they argue is not the same as the GHQ-12, which measures psychological vulnerability.

Diagnosed health conditions

The cluster of diagnosed health conditions was included in this analysis for a number of reasons, firstly, due to the wide range of variables included, multidimensional representations of the GHQ-12 would be expected to display variability between the conditions. Secondly, one of the variables in this list was a clinical depression diagnosis. This variable is particularly prevalent as many of the multidimensional representations of the GHQ-12 include a factor that is labelled depression. Depression is the most common mental illness worldwide (Vos et al., 2013) and contributes to the poor mental health of 19.7% of the UK population (Evans, Macrory & Randall, 2016). It must be noted that within the literature, the presence of a mental illness is not universally agreed to be synonymous with mental health. Westerhof and Keyes (2010) argued that mental health was a much broader concept than simply the absence of diagnosable mental health conditions and argued that the concepts of mental health and mental illness were '*distinct but related*'. Westerhof and Keyes (2010) found older people, were less likely to have diagnosed mental illnesses, however, they had poorer mental health than younger populations who were more likely to have diagnosed mental illnesses. As a result, he proposed what he referred to as a ' two continuum model' which separated mental health and mental illness as an appropriate framework to conceptualise the relationship between mental health and illness. While Westerof and Keyes (2010) did differentiate between the presence of mental health illnesses and mental health, they found them to exhibit a correlational relationship.

As a result of these arguments, unidimensional models which include a factor which claimed to measure general mental health would be expected to demonstrate concurrent validity, whereas multidimensional representations which include a factor labelled '*depression*' would be expected to demonstrate a strong correlation with depression diagnoses.

The other variables in this cluster were more physical in nature, therefore, the expected relationship between them and mental health may be more complex to interpret. The relationship between physical health and mental health is complex, but generally, it is accepted by authorities such as the World Health Organisation that mental and physical health are 'intimately linked' and were described as 'frequently accompanying, succeeding or follow each other' (Herman & Jane-Lopez, 2005). It is also important to not view physical health in a simplistic manner, as it has been shown that physical activity has a moderating effect on mental health deterioration (Fox, 2007; Abu-Omar et al., 2004), however, in general, any condition which impairs physical functioning would likely be associated with poorer mental health.

The research would generally indicate that specific conditions, such as cancer would have a determinantal effect on an individual's mental health (Hewitt & Roland, 2002). It must be noted that this is somewhat disputed with Booker and Sacker (2011) finding that cancer patients had decreased physical health, but around normal mental health, however, a meta-analysis did suggest that as many as a third of cancer sufferers had impaired mental health, which is considerably higher than the general population. This relationship has been mirrored in other serious health conditions such as heart attack survivors who were shown to be three times more likely to have mental health issues (Williams, 2011) and stroke survivors reporting considerably lower mental health following the stroke (Carod-Artal, 2000).

The physical health diagnosis covariates (High blood pressure, Arthritis, Asthma, Diabetes, Cancer, Angina, Hypothyroidism, Coronary heart disease, Chronic bronchitis, Congestive heart failure, Emphysema, Hyperthyroidism, Liver conditions, Epilepsy) were included in the analysis as it would be expected that various factors in a multidimensional representation of the GHQ-12 would demonstrate variability in their relationships with these covariates if they measured different concepts. They were also included as previously mentioned, literature did suggest that measures of physical health have established relationships with mental health within the literature and these should be replicated in this analysis if concurrent validity was to be demonstrated.

Demographics

Mental health has been found to exhibit relationships with demographic variables such as age, sex and rurality. In relation to age, the relationship between age and mental health is difficult to measure. However, Aldwin et al. (1989) suggested that while a slight increase in mental health difficulties as participants aged, a U shaped relationship may be more appropriate. These results were corroborated by Mackenzie et al. (2011), who spoke of a ‘hill shaped’ influence of age on mental health. Regardless of the linearity of said relationship, both studies suggested that when averaged over a lifespan, a positive linear relationship could be observed. Age was selected for inclusion in the analysis as unidimensional representations of age should display a relationship which indicated poorer mental health as a participant aged in order to demonstrate concurrent validity. In multidimensional representations of the GHQ-12, if all the various factors exhibited different relationships with this covariate, then it can be determined that the various factors are measuring different concepts.

In relation to sex, females have been shown to self-report higher levels of depression (Gater et al., 1998) and anxiety (Seedat et al., 2009) than males. This has been suggested to be a function of male reluctance to self-report mental illness (Dindia

& Allen, 1992) or male reluctance to seek help or treatment in relation to psychological conditions (Kessler, Brown & Broman, 1981). The variable of sex facilitates dimensional representations of the HQ-12 to demonstrate concurrent validity as unidimensional representations of metal health should display stronger relationships with females than men, as should multidimensional representations which claim to measure anxiety and depression.

Rurality was included as while it is generally accepted that rural participants have greater difficulty accessing mental health services (Human & Wassem, 1991), they tend to have better mental health (Nicholson, 2008). Paykel et al. (2000) found that significant differences were evident in the UK population between rural and urban citizens in relation to how they experienced mental health issues. However, he also said that these differences could be mitigated when a number of social factors such as deprivation were taken into account. The inclusion of this variable was deemed as appropriate as unidimensional representations of the GHQ-12 should display relationships which indicated better mental health in rural participants in order to demonstrate concurrent validity.

Neighbourhood Variables

This cluster of variables was included as many of the factors within the models brought forward from Chapter 3 included a factor which referred to social performance or social dysfunction. UKHLS included two relevant variables in this database, those being '*social cohesion*' and '*trustworthiness of others*'.

Neighbourhood Cohesion is defined in Buckner (1988) "*as a synthesis of psychological sense of community, attraction-to-neighbourhood, and social interaction within a neighbourhood*," and was measured using a truncated version of the scale developed in Buckner (1988). A lack of social cohesion was found to be associated with poor mental health (Fone et al., 2007) and was also suggested to have a mediating effect on the association between deprivation and mental health (Fone et al., 2014).

'*Trustworthiness of others*' while crudely captured was investigated primarily as a similar concept to social dysfunction. Research into the effects of ones perceived trustworthiness of their peers has shown that individuals who view their peers as trustworthy exhibit higher levels of wellbeing than people who don't (Poulin & Haase, 2015) and Poulin (2008) found that one's world benevolence beliefs, i.e. how benevolent they viewed humanity at large was linked to individuals mental health, with people who believed that humanity was ultimately benevolent experiencing greater wellbeing than people who did not. Meltzer et al. (2007) investigated this longitudinally and found that poor levels of trustworthiness of others in childhood could lead to mental health disorders in later life.

Both of these variables should demonstrate strong correlations with factors in multidimensional representations of the GHQ-12 which refer to social performance or a similar concept and should facilitate these multidimensional representations the opportunity to demonstrate variability in the performance of the factors. In unidimensional representations, these variables should demonstrate concurrent validity and should replicate the relationships mentioned above in their analyses.

4.2.2- Literature Review

Within the literature, a number of researchers have compared GHQ-12 scores with covariates in order to test validity. It was felt beneficial to conduct a review of these studies to contextualise the analyses of this chapter.

A number of studies have conducted analyses utilising covariates as supporting evidence of the GHQ-12. A number of these were conducted before the debate over the factor structure of the GHQ-12 arose, and as a result, they simply investigated the correlation of summed scores of GHQ-12 responses with identified covariates. Studies which draw comparisons between GHQ-12 summed scores and various covariates, fundamentally assumed unidimensionality and generally were conducted before advanced statistical techniques became prevalent, whereas others which adopt a multidimensional approach were generally conducted more recently and adopted a structural equation modelling analytic framework. At the instrument's inception, GHQ scores were compared with clinical assessments in an effort to confirm the measure's validity (Goldberg & Blackwell, 1970). This method of comparing GHQ scores with clinical assessments has been used within the literature on numerous occasions (Goldberg et al., 1997; Goldberg, Oldehinkel, & Ormel, 1998; Tait, Hulse, & Robertson, 2002). Two examples of this included comparing it against the primary care version of the Composite International Diagnostic Instrument (CIDI-PC) (Goldberg et al., 1997) or psychiatrist's assessments using the Clinical Interview Schedule (Lobo et al., 2009). These studies have shown the GHQ to be highly correlated with these clinical assessments and were used as evidence of its validity as a measure of mental health.

Burton, Laurie and Lynn (2004) compared and contrasted the relationships of different mental health measures as captured in wave one of UKHLS. While the purpose of this research was not to validate the GHQ-12 their findings do show significant differences between GHQ-12 scores and those of other mental health measures such as the SF-12 (Short Form-12) and the SWEMWS (Short Warwick Edinburgh Scale of Mental Wellbeing). The correlation between the GHQ-12 and SF-12 was considered by the authors to be strong. More interestingly, however, they found a much weaker correlation between the SWEMBS and GHQ-12, than between the SF-12 and GHQ-12. Using figures obtained from Tennant et al. (2007), they found a correlational relationship of -0.53 between the SWE and GHQ-12, which they described as relatively weak. They argued that the GHQ-12, measured psychological distress, and the SWEMBS measured psychological wellbeing, are distinct but related concepts. Burton, Laurie and Lynn (2004) argue that it is possible for participants to report high levels of stress and anxiety, but still report positive mental health.

Gao et al. (2004) used a similar technique but compared covariate’s scores against a multidimensional representation of the GHQ rather than the assumed unidimensionality of the above studies. Gao examined the three factors of Graetz’s model (1991), that of ‘Anxiety and Depression’, ‘Loss of Confidence’ and ‘Social Dysfunction’ against standardised clinical assessments, namely the Beck Anxiety Inventory (BAI) and the Short Form-36 (SF-36). The BAI is a self-report checklist for symptoms related to anxiety that demonstrated reliability and validity (Beck et al., 1988). This scale consists of 21 items relating to the presence of symptoms of anxiety and uses summed scores. In this instance, higher scores indicating high levels of anxiety. The SF-36 is a 36-item questionnaire that is acknowledged to be multidimensional with the following factors: physical functioning, role limitations due to physical problems, bodily pain, general health, vitality, social functioning, role limitations due to emotional problems, and mental health. This measure was also shown to be reliable (McHorney et al., 1994) and had the added benefit of comprising a number of items similar to that which Graetz claimed his factors measured. Gao (2004) found that the factors performed uniformly in relation to the covariates mentioned. Consequently, he concluded that the three-factor model offered no practical advantages or utility to the measure and that unless one had specific questions which were best answered by one of Graetz's (1991) factors, it was safe to ignore the multidimensionality of the GHQ-12. Shevlin and Adamson (2004) performed similar research. They compared the same three factors against four variables; these were 'General Health', 'Stress and Worry', 'Effect of the Troubles (A period of civil unrest in Northern Ireland)', and 'Social Support'. These variables are assessed via singular self-report, 4 item Likert scale based questions from participants in Northern Ireland. This research found that despite the three-factor model exhibiting the best model fit, the factors generated were highly correlated, and the factors did not vary in relation to the covariates tested. As a result, they also suggested that the GHQ-12 should be treated as a unidimensional instrument. Furthermore, both Shevlin and Admson's (2004) and Gao (2004) studies found inter-factor correlations of similar values as to findings in Chapter 3. Gao et al. (2004) reported correlations in the range of 0.83-0.9, while Shevlin and Adamson (2004) reported a range of 0.727-0.829.

Penninkilampi-Kerola, Miettuen and Ebeling (2006) also conducted research on the discriminant validity (see above) of several factor models of both the GHQ-12 and GHQ-20, by comparing the factors against demographic data. Specifically, they looked at living arrangements and employment status. These variables have been shown to have established relationships with psychological distress (e.g., Whelan, 1994; Bjarnason & Sigurdardottir, 2003). Using data from a Finnish population, they found that in relation to the GHQ-20, a four-factor model was likely to yield unique discriminant validity, whereas a three-factor model was not in the GHQ-20. In relation to the GHQ-12, multidimensional representations did not yield discriminant validity.

In conclusion, the literature suggests that Graetz's (1991) model is likely to demonstrate good fit in relation to fit statistics, however, its component factors generally fail to demonstrate any unique predictive utility, and its factors are likely to demonstrate considerable multicollinearity. As a result, a number of researchers suggest that it is more appropriate to treat the GHQ as unidimensional (Shevlin & Adamson, 2005; Gao, 2004).

4.2.3- Hypotheses

The hypotheses within this chapter varied depending on whether the model being investigated was unidimensional or multidimensional. In relation to multidimensional representations, this chapter's hypotheses are informed by previous research in the field particularly Shevlin and Adamson (2004) and Gao et al. (2004) who found that, despite a three-factor model demonstrating good fit in relation to fit statistics, multidimensional representation's factors did not vary in relation to a number of covariates and as a result, they questioned the utility of treating the GHQ-12 as multidimensional. Consequently, this chapter hypothesised that the factors of multidimensional representations of the GHQ-12 would not vary in relation to a range of covariates.

A number of variables were included in the analysis, which were directly comparable with what the factors of multidimensional representations claimed to measure. An example of this would be the inclusion of a clinical diagnosis variable, which could be compared with the depression factor present in many of the multidimensional representations of the GHQ-12. It was therefore hypothesised that factors of multidimensional representations of the GHQ-12 would demonstrate concurrent validity with covariates which measured similar concepts.

In relation to unidimensional dimensional representations, covariate analysis was used to investigate their validity. A number of similar measures of mental health such as the SWE and SF-12 were included in the analysis, and these facilitated the opportunity for dimensional representations of the GHQ-12 to demonstrate concurrent validity with these measures. Consequently, it was hypothesised that unidimensional representations of the GHQ-12 would exhibit strong correlational relationships with other measures of self-reported mental health.

A number of variables which had established relationships with mental health were included in the analysis. These variables such as age and sex and rurality. While these variables afforded the opportunity for factors to demonstrate variability, they also have established relationships with mental health and therefore offer the opportunity for the unidimensional representations of the GHQ-12 to demonstrate concurrent validity. It is therefore hypothesised that unidimensional representations of the GHQ-12 would exhibit relationships with covariates in a similar manner to those relationships previously mentioned in the literature.

4.3.- Methods

4.3.1- Data

The data for this chapter's analysis was drawn from Wave 1 of the Understanding Society (UKHLS) database. This contained 39700 participants who at least partially completed the GHQ-12 and was weighted, clustered and stratified as directed in the UKHLS user guide (Knies, 2017). UKHLS provided participant’s scores in both a caseness and Likert format (see chapter 2), and Likert scores were used as they provided richer data than a caseness approach.

A comprehensive review of this data is given in Chapter 2, however, in summary, it was provided on license from the UK Data Service and was representative of the UK population. The analysis used maximum likelihood parameter estimates with robust standard errors (MLR) as an estimator as a method of handling missing data (see 3.3.1 for justification for doing so).

4.3.1.1- Covariates

Diagnosed Conditions

This variable was provided in the INDRESP file of UKHLS. UKHLS asked a number of questions related to the participant having a diagnosis of certain medical conditions (High blood pressure, Arthritis, Asthma, Clinical depression, Diabetes, Cancer, Angina, Hypothyroidism, Coronary heart disease, Chronic bronchitis, Congestive heart failure, Emphysema, Hyperthyroidism, Liver conditions, Epilepsy), see table 4.2. The first question asked if a participant had ever suffered from a certain condition, the second asked if the participant still suffered, and the third asks the age of diagnosis. By utilising these questions, it was possible to ascertain a number of things ranging from the presence of a condition, the length of time that a participant has suffered from a specific condition and if a participant has recovered from a condition. It was decided to focus on whether a participant currently suffered from a condition as the GHQ-12 is very specific, that it measures the current state of participants, not historical difficulties (Goldberg, 1988). From the various responses to the question, these were collapsed into the following responses

- Missing data (coded -9)
- Currently has a condition (coded 1)
- Currently do not have a condition (coded 2)

Certain conditions such as stroke and heart attacks are not lingering conditions and therefore were therefore not considered in this analysis.

Demography data

Sex and rurality were taken from the 'individual response' (INDRESP) file of UKHLS, and both consisted of dichotomous variables where values of 1 represent male participants and urban habitation respectively and two representing female and rural habitation. The age variable consisted of the age of each participant in years and was generated from the Xdata file (see Chapter 2).

Self-reported health measures

This cluster of variables consisted of 4 variables which were listed as general health, SF-12 physical competent, SF-12 mental component and the Short Warwick Edinburgh scale of Mental Wellbeing (SWEMBS). Each of these are detailed below. Each of the variables are provided in the INDRESP file of the UKHLS.

General Health consisted of a 5 item Likert scale which asked participants to rate their general health from 1 to 5 representing very poor health and very good health respectively. Responses of 'don't know' were coded as missing.

The SF-12 mental and physical components are taken from the Short Form 12, which is a 12 item measure of mental and physical health. The responses to these items are inputted into a weighting algorithm, which computes a score out of 100. Depending on whether the researcher is primarily concerned with mental or physical health, scores are inputted into different scoring algorithms that ascribe different weights to the various items. A low score indicates poor mental or physical health.

The SWEMBS consisted of seven items which claim to measure an individual's psychological wellbeing. Scores were summed, and responses ranged between 7 and 35. In this score, a high value represented good health. All items were positively worded, so no reverse scoring was required.

Social Cohesion

This measure was provided in the INDRESP file and represents a truncated version of Buckner's social cohesion instrument (1998). This measure was tested for internal consistency, retest stability and discriminatory power on 206 participants from three neighbourhoods. This scale was originally validated as a 19 item questionnaire of which four items measured 'attraction to one's neighbourhood', six measuring 'neighbouring' and 9 measuring 'sense of community'. Each item asked participants to rate the extent to which they agreed or disagreed with various statements on a scale of 1 to 5. Individual's social cohesion was then expressed as an averaged score across all responses. Participants who failed to answer over half of the items associated with this measure were recoded as missing on all values as suggested in UKHLS user guide (GL, education Group, 2012).

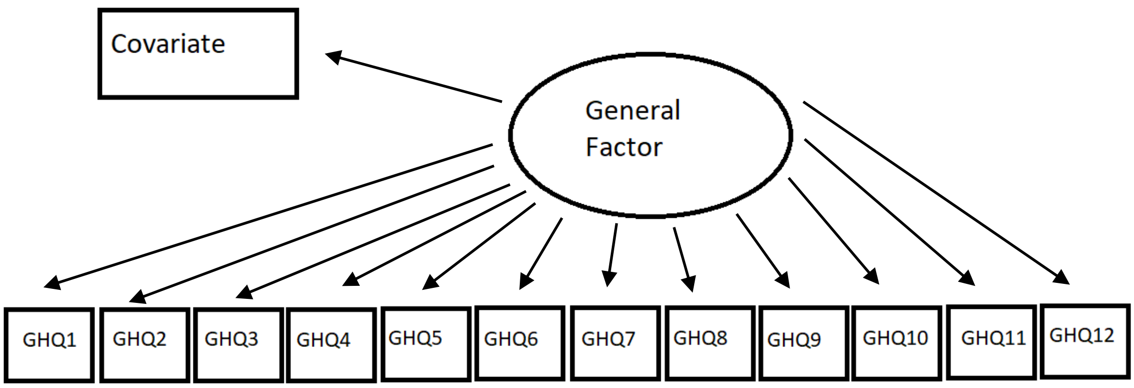
4.3.2- Analyses

Within a structural equation modelling (SEM) framework, multivariate associations between the dimensions and covariates in models which performed sufficiently well in the previous chapter as to warrant further analyses were investigated. These models were Ye's (2009), Hankins (2008), Graetz (1991), Politi (1994) and Andrich and van Shoeubroek (1989), furthermore, a unidimensional model was analysed as a comparison. A graphical representation of the respective factor structures used in this analysis is provided in table 4.8.

In this instance, the analysis involved regressing a factor onto covariates within a structural equation model in order to investigate the relationship between the two. A diagrammatical representation of a covariate analysis in a unidimensional model is shown below.

Figure 4. 1

A Graphical Representation of Covariate Analysis of the Unidimensional Representation of the GHQ-12 and Covariates



Note. This graph represents a graphical representation of covariate analysis as was conducted within this chapter's analysis.

This technique required that cases did not have missing data on either the covariate or the GHQ scores. Any case which displayed missingness was removed from the analysis. It is also important to note that regression coefficients can only be compared to judge effect size within the same model, therefore unless the covariates are run in the same model, they could not be compared. In an effort to reduce the effect of listwise deletion, analyses was run in batches, with covariates that belong to the same module (see Chapter 2) or which practically could be compared, such as mental health measures analysed simultaneously.

MPLUS provides a number of outputs with varying degrees of standardisation. It labels these as 'Model results' (unstandardised), 'STDXY' (completely standardised), 'STDY' (standardised along the Y-axis- GHQ scores) 'STDX' (Standardised along the X-axis- Covariates). While it is generally desirable to report standardised results, in cases with dichotomous variables such as gender, it is not appropriate conceptually to standardise these results, and consequently, only STDY results were reported in this case. The degree of standardisation that the figures in the 'Results' section have undergone were communicated through explanatory notes below each table.

Table 4. 1

Factor Structures to be Analysed in Chapter 4

		Politi (1994)		Andrich and Schoubroeck (1989)		Graetz (1991)			Hankins (2008)	Ye (2009)	
	1 factor	Dysphoria	Social dysfunction	Positively worded	Negatively worded	Anxiety/ depression	Social dysfunction	Loss of confidence	1 factor	1 factor	Method factor
Item											
1	*		*		*		*		()	*	*
2	*	*		*		*			*	*	
3	*		*		*		*		()	*	*
4	*		*		*		*		()	*	*
5	*	*		*		*			*	*	
6	*	*		*		*			*	*	
7	*		*		*		*		()	*	*
8	*		*		*		*		()	*	*
9	*	*		*		*			*	*	
10	*	*		*				*	*	*	
11	*	*			*			*	()	*	*
12	*	*	*	*			*		*	*	

Note. * represents items which are associated with the various factors in each dimensional representation

4.4- Results

4.4.1 Descriptive Statistics

Descriptive statistics were presented below for each of the covariates that were used in the analyses. The figures presented in table 4.2 demonstrated that high blood pressure, arthritis and asthma affected a greater proportion of the population than (8.37%, 6.72% & 5.76%) than other conditions. These three conditions all had approximately double the number of participants than the next most common condition, clinical depression, which has 1537 (3.01%). As previously mentioned, heart attack and stroke do not have a lingering effect and therefore do not have values listed under the 'still have' column. A number of conditions had negligible numbers of participants with a condition such as liver conditions only being present in 1% of the population. The small number of participants in these groups may reduce the statistical power of any analyses and therefore may reduce the chances of displaying statistically significant results, however, the size of the sample still ensured that even these proportionally small samples have populations of over 100 participants.

Table 4. 2

Descriptive Statistics relating to Medical Conditions in Wave One of Understanding Society

	Ever Diagnosed		Still Have	
Condition	%	N		N
High blood pressure	8.37	4,269	6.52	3,326
Arthritis	6.72	3,428	6.55	3,338
Asthma	5.76	2,936	4.26	2,171
Clinical depression	3.01	1,537	2.10	1,070
Diabetes	2.45	1,249	2.34	1,195
Cancer	1.64	837	0.48	245
Angina	1.41	717	1.13	575
Hypothyroidism	1.29	660	1.23	630
Coronary heart disease	0.88	450	0.74	377
Heart attack	1.02	525		–
Stroke	0.85	431		–
Chronic bronchitis	0.90	458	0.56	258
Congestive heart failure	0.26	134	0.20	106

Emphysema	0.31	162	0.30	149
Hyperthyroidism	0.41	209	0.22	116
Liver conditions	0.57	290	0.35	183
Epilepsy	0.51	263	0.32	168

N= 50994

Table 4. 3

Descriptive Statistics relating to Sex in Wave One of Understanding Society Compared with UK Government Figures (Gov.UK, 2018)

Sex	UKHLS (N)	UKHLS (%)	UK Population (%)	Variance
Male (1)	23208	45.5	49	-3.5
Female (2)	27786	54.5	51	3.5
Total	50994	100		

UKHLS = Understanding Society participants

UK= United Kingdom population

Table 4. 4

Descriptive Statistics relating to Age in Wave One of Understanding Society Compared with UK Population Data (Gov.uk, 2018)

	Mean	Median	Standard deviation	Range	Max
UKHLS	45.64	44	18.187	87	104
UK	-	40	-	-	-

UKHLS = Understanding Society participants

UK= United Kingdom population

Table 4. 5

Descriptive Statistics relating to Rurality in Wave One of Understanding Society Compared with UK Population Data (Gov.uk, 2018)

	UKHLS (N)	UKHLS (%)	UK (%)	Variance (%)
Urban (1)	40447	79.3	82.9	3.6

Rural (2)	10547	20.7	17.1	-3.6
Total	50994	100	100	

Valid=50994/50994

UKHLS = Understanding Society participants

UK= United Kingdom population

Demographic characteristics of the data were detailed in chapter 2, however, the three tables above showed that the sample was relatively in line with the population of the UK, which it attempts to emulate. The tables also demonstrated the coding of dichotomous variables which was essential in the interpretation of results.

Table 4. 6

Descriptive Statistics relating to SF-12 and SWE in Wave One of Understanding Society

	SF-12 Physical Component Summary (PCS)	SF-12 Mental Component Summary (PCS)	SWE
Mean	49.4927	50.4855	25.18
Median	53.4700	53.0400	26
Std. Deviation	11.48891	10.11810	4.544
Range	70.57	77.11	28
Minimum	4.33	.00	7
Maximum	74.90	77.11	35
Valid responses	47400	47400	38395
Total responses	50994	50994	50994

These self-reported health variables which are detailed in the above two tables, consisted of the two scoring metrics of the SF-12 and the SWE and were presented together for ease of comparison. The SF-12 based variables had a slightly higher response rate than the SWE, with 9005 fewer responses to the SWE. The SF-12 scored participants on a scale of 1-100 while the SWE was scored 7-35. The means and medians of SF-12 MCS and SWE scores are relatively comparable, suggesting that outliers have not significantly affected the distribution of the data.

Table 4. 7

Descriptive Statistics for Social Cohesion and Trustworthiness of Others in Wave 1 Understanding Society

	Social Cohesion	Trustworthiness of Other
Mean	19.24	1.9
Median	19.00	2
Std. Deviation	5.99	0.759
Range	32	2
Minimum	8	1
Maximum	40	3
Valid responses	38028	39676
Total responses	50994	50994

The descriptive statistics presented above show the descriptive statistics for participants who responded to the social cohesion questions which were adopted from Buckner's scale of social cohesion (1988) and the trustworthiness question, which is analysed individually. The results show relatively similar levels of missingness for each item with the slight variation being accounted for by responses of ‘don’t know’ being treated as missing. Social cohesion values were summed scores from 8 items, and average values were 19.00 with a standard deviation of 5.99. ‘Trustworthiness of others’ was a single item variable with scores from 1-3, and the average value was 1.9, with a standard deviation of 0.759.

4.4.2- Regression Analysis Results

Below, the results of the regression analyses for each of the covariates identified earlier were presented. Results are presented for each cluster of covariates for each model in order of fit however, a unidimensional model is included for comparison purposes and will be presented first. Results are standardised, however, as standardised binary values are not meaningful (Muthen & Muthen, 2012), different methods of standardisation are necessary depending on the characteristics of the analysis conducted.

Unidimensional model

Table 4. 8

Medical Condition Covariates Regressed onto a Unidimensional model

	Estimate	S.E.	Est/S.E.	P
Asthma	**-0.07**	0.019	-3.781	<0.001
Arthritis	**-0.214**	0.018	-11.767	<0.001
Congestive heart failure	-0.139	0.12	-1.155	0.248
Coronary heart disease	**-0.143**	0.051	-2.821	0.005
Angina	**-0.183**	0.048	-3.831	<0.001
Emphysema	**-0.253**	0.079	-3.212	0.001
Hyperthyroidism	-0.139	0.082	-1.692	0.091
Hypothyroidism	**-0.06**	0.035	-1.693	0.09
Chronic bronchitis	**-0.227**	0.066	-3.418	0.001
Liver condition	**-0.343**	0.083	-4.115	<0.001
Cancer or malignancy	**-0.385**	0.065	-5.949	<0.001
Diabetes	**-0.059**	0.027	-2.148	0.032
Epilepsy	**-0.235**	0.073	-3.202	0.001
High blood pressure	**-0.078**	0.018	-4.384	<0.001
Clinical depression	**-1.584**	0.035	-45.833	<0.001

Note. All values are standardised on the Y axis only (the unidimensional factor)

*N= 40259

**Statistically significant associations in bold

The results detailed in table 4.8 showed that clinical depression diagnoses exhibited a much stronger relationship than any other covariate (-1.584), however statistically significant relationships were observed in all conditions except those related to the thyroid and congestive heart failure. This model's results established a baseline from where all other models could be measured against.

Table 4. 9

Demographic Covariates Regressed onto a Unidimensional model

	Estimate	S.E.	Est/S.E.	P
Sex	**0.191**	0.011	17.18	<0.001
Age	0.001	0.006	0.167	0.868
Rurality	**-0.092**	0.014	-6.646	<0.001

Note. Sex and Rurality are standardised on the Y axis only (the unidimensional factor), whereas Age is standardised on both axis

*N=40452

**Statistically significant associations in bold

Significant relationships were observed between sex and rurality but not age in table 4.9. The relationship between sex and the single factor of the unidimensional model suggested that females reported higher levels of psychological distress, than males did, with a regression coefficient of 0.191. Rurality displayed a weaker relationship with the factors (-0.092), and rural dwellers were found to report less psychological distress than urban dwellers in this sample.

Table 4. 10

Self Reported Health Covariates Regressed onto a Unidimensional model

	Estimate	S.E.	Est/S.E.	P
General Health	**-0.02**	0.006	-3.226	0.001
SF 12 Physical	**-0.151**	0.007	-22.477	<0.001
SF 12 Mental	**-0.514**	0.006	-79.433	<0.001
SWEMBS	**-0.373**	0.006	-59.028	<0.001

Note. All values are standardised on both axis.

*N= 38256

**Significant Associations in Bold

Statistically significant relationships were observed in table 4.10 between the unidimensional factor of the GHQ and all other mental health scores collected within this dataset. Relationships were negative as high scores in SF-12, and SWEMBS tests indicated poor mental health, whereas the opposite was true for GHQ-12. Regression coefficients between the model and SF mental component (-0.541) were higher than that of its physical component (-0.151) and of the SWEMBS (-0.373). As previously mentioned, these results were intended to be used as a baseline against which other models could be compared.

Table 4. 11

Neighbourhood variables Regressed onto a Unidimensional model

	Estimate	S.E.	Est/S.E.	P
Social Cohesion	**-0.040**	0.003	15.508	<0.001
Trustworthiness of others	**-0.027**	0.002	12.494	<0.001

Note. Both variables are standardised on both axis

*N= 37722

**Statistically significant associations in bold

Social cohesion variables exhibited a stronger relationship (0.40) than trustworthiness of others (0.027), however, the results shown in table 4.11 indicated that in both cases there was a statistically significant relationship with the unidimensional model's factor and the covariate. As 'social cohesion' variables and 'trustworthiness of others' variables were reverse-scored, the results indicated that higher GHQ-12 scores were associated with a decrease in social cohesion and an individual's trust in other people.

Graetz (3-factor model)

Medical conditions

Table 4. 12

Anxiety and Depression, Social Performance, Loss of Confidence, regressed onto Medical Conditions Covariates

	Anxiety and Depression			Social Performance			Loss of Confidence		
	Estimate	S.E.	Est/S.E.	Estimate	S.E.	Est/S.E.	Estimate	S.E. F3	Est/S.E
Asthma	**0.101*** **	0.02	-5.158	**0.055** **	0.020	-2.759	**0.075*** **	0.021	-3.605
Arthritis	**0.113*** **	0.017	-6.445	**0.275*** **	0.019	-14.796	**0.153*** **	0.018	-8.303
Congestive heart failure	0.056	0.107	-0.525	0.198	0.140	-1.415	0.045	0.121	-0.37
Coronary heart disease	0.041	0.047	-0.87	**0.162** **	0.059	-2.749	0.078	0.054	-1.441
Angina	0.047	0.044	-1.069	**0.287*** **	0.052	-5.550	**0.11***	0.056	-1.975
Emphysema	0.108	0.078	-1.389	**0.278** **	0.082	-3.376	**0.214***	0.084	-2.554
Hyperthyroidism	0.158	0.082	-1.931	0.120	0.093	-1.294	**0.169***	0.081	-2.093
Hypothyroidism	**0.08** **	0.037	-2.156	**0.106***	0.035	-2.998	0.046	0.038	-1.188
Chronic bronchitis	**0.186** **	0.063	-2.956	**0.242** **	0.076	-3.195	**0.139***	0.066	-2.091
Liver condition	**0.326*** **	0.077	-4.223	**-0.355*** **	0.097	-3.676	**0.221*** **	0.079	-2.812
Cancer or malignancy	**0.253*** **	0.066	-3.817	**0.506*** **	0.068	-7.424	**0.192** **	0.064	-3.014
Diabetes	-0.043	0.027	1.627	**0.083** **	0.029	2.840	0.038	0.03	-1.248
Epilepsy	**0.153***	0.073	-2.088	**0.237** **	0.088	-2.690	**0.324*** **	0.076	-4.267
High blood pressure	-0.031	0.018	1.765	**0.128*** **	0.018	-7.144	0.007	0.018	-0.371
Clinical depression	**1.481*** **	0.03	-50.194	**1.433*** **	0.043	-33.439	**1.523*** **	0.035	-43.197

Note. Variables are standardised on the Y axis (the factors) only

*N= 40259

[b] Statistically significant association in bold, *<0.05, **<0.01, ***<0.001

The conditions which displayed statistically significant relationships with the various factors of this model have been presented in bold text in table 4.12 to ease comparison. Noticeable variation was found between the factors. Asthma, arthritis, bronchitis, liver conditions, cancer, epilepsy and clinical depression exhibited statistically significant relationships with all factors of the model. Only congestive heart failure did not display a statistically significant relationship with any of the factors. Of particular note, the ‘anxiety and depression’ factor did not display statistically significant relationships with ‘high blood pressure’, ‘congestive heart failure’ and ‘angina’. ‘Social performance’ exhibited statistically significant relationships with all but two of the variables, hyperthyroidism and congestive heart failure. Finally, ‘loss of confidence’ displayed statistically significant relationships with approximately half of the covariates.

All factors displayed a statistically significant relationship with the clinical depression covariate, however, the ‘loss of confidence’ factor displayed a stronger relationship, with a regression coefficient of 1.523 compared with the ‘anxiety and depression’ factor’s coefficient of 1.481 and social performance displaying a regression coefficient of 1.433. None of the factors displayed as strong a relationship between the factors and a clinical depression diagnosis as the unidimensional model did, furthermore the two conditions that did not display statistically significant relationships with the unidimensional model, i.e. diabetes and congestive heart failure, only diabetes displayed a statistically significant relationship with any of the factors, namely ‘social performance’

Demographics

Table 4. 13

Anxiety and Depression, Social Performance and Loss of Confidence with Demographic Covariates

	Anxiety and Depression			Social Performance			Loss of Confidence		
	Estimate	S.E.	Est/S.E.	Estimate	S.E.	Est/S.E.	Estimate	S.E.	Est/S.E.
Sex	**0.179*** **	0.011	15.741	**0.143*** **	0.012	12.113	**0.179*** **	0.011	15.741
Age	**-0.075*** **	0.007	-11.533	**-0.075*** **	0.007	-11.533	**-0.025*** **	0.007	-3.85
Rurality	**-0.11*** **	0.015	-7.378	**-0.064*** **	0.014	-4.56	**-0.11*** **	0.015	-7.378

Note. Sex and Rurality are standardised on the y axis (the factors) only, whereas Age is standardised on both axis.

*N=40452

[b] Statistically significant association in bold, *<0.05, **<0.01, ***<0.001

All demography based covariates in this section exhibited statistically significant relationships with all factors in the model. With the exception of rurality, the regression coefficients of each relationship, regression coefficients in this cluster are all of similar direction and magnitude, regardless of the factor that the covariate is being regressed onto. Each of the three factors exhibited a range of 0.036 for sex, 0.05 for age and 0.079 for rurality respectively. In comparison with the unidimensional model, age was shown to have a statistically significant relationship with the various factors, whereas the unidimensional model's factor did not.

Mental Health Measures

Table 4. 14

Anxiety and Depression, Social Performance and Loss of Confidence with Mental Health Covariates

	Anxiety and Depression			Social Performance			Loss of Confidence		
	Estimate	S.E.	Est/S.E.	Estimate	S.E.	Est/S.E.	Estimate	S.E.	Est/S.E.
General Health	-0.003	0.007	-0.494	**-0.052*** **	0.008	-6.662	-0.006	0.008	-0.73
SF 12 Physical	**-0.07*** **	0.007	-10.074	**-0.26*** **	0.009	-30.22	**-0.113*** **	0.008	-14.28
SF 12 Mental	**-0.538*** **	0.006	-84.856	**-0.428*** **	0.009	-50.17	**-0.421*** **	0.008	-55.723
SWE	**-0.327*** **	0.007	-49.605	**-0.328*** **	0.007	-43.915	**-0.379*** **	0.007	-53.293

Note. All variables are standardised on both axis

*N= 38256

[b] Statistically significant association in bold, *<0.05, **<0.01, ***<0.001

Statistically significant relationships were observed between all factors and all mental health measures with the exception of the ‘general health’ variable. ‘General health’ only exhibited statistically significant relationships with factor two, ‘social performance’. The SF physical and mental components both displayed a small degree of variability between the various factors, with a range of 0.19 and 0.117 between the factors respectively, whereas the SWE displayed less variability, only 0.052 between the highest and lowest ranging factors. The SF-12 mental component did display much stronger relationships with the various factors than the physical component, with the average regression coefficient being 0.462 for the mental component and 0.148 for the physical component, respectively. Similarly to the unidimensional model, all relationships were negative as the GHQ-12 uses a reversed scoring metric compared to all covariates investigated. In comparison with the unidimensional model, all covariates displayed weaker relationships that the factor of the unidimensional model with the exceptions of the ‘loss of confidence’ factors relationship with the SWE and the anxiety and depressions relationship with the SF-12 mental component.

Social Networks Variables

Table 4.14

Anxiety and Depression, Social Performance and Loss of Confidence with Social Networks Covariates

	Anxiety and Depression			Social Confidence			Loss of Confidence		
	Estimate	S.E.	Est/S.E.	Estimate	S.E.	Est/S.E.	Estimate	S.E.	Est/S.E.
Social Cohesion	0.115***	0.007	16.611	0.08***	0.008	10.506	0.101***	0.007	14.161
Trust in others	0.083***	0.006	13.812	0.042***	0.006	6.97	0.068***	0.006	11.035

Note. All variables standardised on both axis

*N= 37722

[b] Statistically significant association in bold, *<0.05, **<0.01, ***<0.001

Statistically significant relationships were observed between all covariates in the Neighbourhood Covariates cluster and all factors. Social cohesion and trust in others displayed the weakest relationships with the social performance factor with a regression coefficient of 0.08 and 0.042, respectively. Both variables exhibited stronger relationships with the ‘anxiety and depression’ factor with regression coefficients of 0.115 and 0.083, respectively

Hankins (correlated errors model)

Table 4. 15

Psychological Distress with Medical Conditions Covariates

	Estimate	S.E.	Est/S.E.	P
Asthma	**-0.021**	0.006	-3.57	<0.001
Arthritis	**-0.088**	0.006	-13.841	<0.001
Congestive heart failure	-0.011	0.009	-1.266	0.205
Coronary heart disease	**-0.018**	0.007	-2.56	0.010
Angina	**-0.036**	0.008	-4.725	<0.001
Emphysema	**-0.022**	0.007	-3.226	<0.001
Hyperthyroidism	-0.011	0.007	-1.606	0.108
Hypothyroidism	**-0.016**	0.006	-2.817	0.005
Chronic bronchitis	**-0.025**	0.008	-3.26	0.001
Liver condition	**-0.033**	0.008	-3.863	<0.001
Cancer or malignancy	**-0.046**	0.007	-6.718	<0.001
Diabetes	**-0.015**	0.006	-2.268	0.023
Epilepsy	**-0.021**	0.007	-3.065	0.002
High blood pressure	**-0.033**	0.006	-5.323	<0.001
Clinical depression	**-0.325**	0.009	-35.674	<0.001

Note. All variables are standardised on both Axis

*N= 40259

[b] Statistically significant association in bold, *<0.05, **<0.01, ***<0.001

Every medical condition, with the exception of congestive heart failure and hyperthyroidism, displayed statistically significant relationships with the general mental health factor of this model. Some relationships such as those for diabetes and asthma were negligibly weak, with relationships of -0.015 and -0.201, respectively. Clinical depression was the most strongly associated covariate with the factor in this model, with a regression coefficient of -0.325, which is considerably weaker than other models.

Table 4. 16

Psychological Distress with Demographic Covariates

	Estimate	S.E.	Est/S.E.	P
Sex	**0.167**	0.012	14.361	<0.001

Age	**0.004**	0	13.022	<0.001
Rurality	**-0.082**	0.014	-5.825	<0.001

Note. Sex and Rurality are standardised on the Y axis (the Factors) only. Age is standardised on both axis

*N=40452

[b] Statistically significant association in bold

Weak but statistically significant relationships (0.167, 0.004 and -0.082 respectively) were observed between the above demography data covariates and the general factor of Hankins' (2008) model. The results displayed a similar pattern to the unidimensional model, displaying a similar relationship between sex and rurality. Age, however, exhibited a statistically significant relationship with the factor in Hankins' (2008) model were it did not with the unidimensional model.

Table 4. 17

Psychological Distress with Mental Health Covariates

	Estimate	S.E.	Est/S.E.	P
General Health	**-0.037**	0.007	-5.357	<0.001
SF 12 Physical	**-0.213**	0.008	-28.139	<0.001
SF 12 Mental	**-0.496**	0.008	-65.875	<0.001
SWE	**-0.37**	0.007	-54.553	<0.001

Note. All variables are standardised on both Axis.

*N= 38256

[b] Statistically significant association in bold

The results presented in table 4.17 demonstrated that Hankins model expressed strong relationships with all measures of mental health and was more strongly associated with the SF-12 mental component than the SF-12 physical component (-0.496 and -0.213 respectively). A relatively weak relationship with general health (-0.037) was observed between general health and the general factor, however generally, this model performed similarly to the unidimensional model. In comparison with the unidimensional model, a weaker relationship was observed between the SF-12 mental and SWE, however, the difference between the SWE was minimal, with the unidimensional model exhibiting a relationship of -0.373 and this model exhibiting -0.370. Finally, stronger relationships were observed between the general health factor and the SF-12 physical component than were present in the unidimensional model.

Table 4. 18

Psychological Distress with Social Networks Covariates

	Estimate	S.E.	Est/S.E.	P
Social Cohesion	**0.045**	0.004	12.458	<0.001
Trust in others	**0.027**	0.003	9.296	<0.001

Note. All Variables are standardised on both Axis

*N= 37722

[b] Statistically significant association in bold.

As per figures presented in table 4.18, statistically significant relationships were observed between the social cohesion and trust in others covariates, however, they were relatively weakly associated. Social cohesion yielded a regression coefficient of 0.045, whereas trust in others yielded a regression coefficient of 0.027.

Ye (method effect model)

Medical Conditions

Table 4. 19

Psychological Distress and the Method Factor with Medical Condition Covariates

	Psychological Distress			Method Factor		
	Estimate	S.E.	Est/S.E.	Estimate	S.E.	Est/S.E.
Asthma	**-0.053****	0.02	-2.625	**-0.089*****	0.024	-3.637
Arthritis	**-0.277*****	0.019	-14.917	**0.124*****	0.021	5.979
Congestive heart failure	-0.196	0.139	-1.412	0.147	0.114	1.292
Coronary heart disease	**-0.166****	0.059	-2.805	0.11	0.059	1.861
Angina	**-0.287*****	0.052	-5.474	**0.233*****	0.049	4.715
Emphysema	**-0.289*****	0.082	-3.511	0.11	0.088	1.249
Hyperthyroidism	-0.125	0.092	-1.354	-0.115	0.092	-1.247
Hypothyroidism	**-0.101****	0.035	-2.847	0.008	0.042	0.182
Chronic bronchitis	**-0.241****	0.076	-3.186	0.009	0.069	0.127
Liver condition	**-0.35*****	0.096	-3.643	-0.06	0.073	-0.815
Cancer or malignancy	**-0.499*****	0.068	-7.369	**0.217****	0.07	3.099
Diabetes	**-0.094****	0.03	-3.169	**0.141*****	0.032	4.407
Epilepsy	**-0.256****	0.088	-2.918	-0.037	0.094	-0.399
High blood pressure	**-0.129*****	0.018	-7.184	**0.185*****	0.02	9.34
Clinical depression	**-1.455*****	0.044	-33.406	**-0.722*****	0.037	-19.447

Note. All variables are standardised on the Y axis (the factors) only.

*N= 40259

[b] Statistically significant association in bold, *<0.05, **<0.01, ***<0.001

In this model, the two factors, one which represented mental health and the other which represented the variance caused by method effects exhibited a number of statistically significant relationships with various medical conditions. The psychological factor exhibited statistically significant relationships with all of the medical conditions, with the exception of congestive heart failure and hyperthyroidism. The method factor however, only displayed relationships with half of the conditions investigated.

There were significant differences in the relationships exhibited between the method factor and the general factor. In relation to clinical depression, the general factor exhibited a much stronger relationship than the method factor, with regression coefficients of -1.455 and -0.722 respectively, whereas, in relation to conditions such as cancer, the relationship was not unidimensional, with a negative coefficient of -0.499 observed for the general factor, while 0.217 was observed for the method factor. Generally, the relationships between the general factor and medical conditions were negative, whereas the method factor generally exhibited positive relationships.

Demographics

Table 4. 20

Psychological Distress and the Method Factor with Demographic Covariates

	Psychological Distress			Method Factor		
	Estimate	S.E.	Est/S.E.	Estimate	S.E.	Est/S.E.
Sex (STDY)	**0.144*** **	0.012	12.066	**0.132*** **	0.014	9.356
Age(STDYX)	**0.132*** **	0.006	21.34	**-0.264*** **	0.008	-33.516
Rurality (STDY)	**-0.063*** **	0.014	-4.501	**-0.078*** **	0.017	-4.448

Note. Sex and rurality are standardised on the Y axis (the factors) only, whereas age is standardised on both axis.

*N=40452

[b] Statistically significant association in bold, *<0.05, **<0.01, ***<0.001

All demographic covariates exhibited statistically significant relationships with both the factors of this model. Age displayed a statistically significant relationship with age, whereas the unidimensional model did not. Sex exhibited positive relationships with both factors, with relationships of 0.144 and 0.132, respectively and rurality exhibited mutually negative relationships of -0.063 and -0.078, respectively. Only age exhibited relationships of different directions between the two factors, with relationships of 0.132 and -0.264, respectively.

Mental Health Measures

Table 4. 21

Psychological Distress and the Method Factor with Mental Health Covariates

	Psychological Distress			Method Factor		
	Estimate	S.E.	Est/S.E.	Estimate	S.E.	Est/S.E.
General Health	**-0.051*** **	0.008	-6.581	**0.03** **	0.009	3.384
SF 12 Physical	**-0.262*** **	0.009	-30.662	**0.06*** **	0.009	6.443
SF 12 Mental	**-0.419*** **	0.008	-49.395	**-0.454*** **	0.008	-56.505
SWE	**-0.338*** **	0.008	-45.021	**-0.274*** **	0.008	-32.739

Note. All variables are standardised on both axis

*N= 38256

[b] Statistically significant association in bold, *<0.05, **<0.01, ***<0.001

Statistically significant relationships were observed in table 4.21 between the measures of mental health and both factors. The general factor was more strongly associated with all covariates in this cluster than the method factor, with the exception of the SF-12 mental, where the method factor was slightly more strongly associated with a regression coefficient of -0.454 compared to -0.419. Stronger relationships were observed between the general factor and the covariates when regressed onto general health and the SF-12 physical, whereas weaker relationships were observed between the SF-12 mental and the SWE when the factors were regressed onto them.

Social Networks

Table 4. 22

Psychological Distress and the Method Factor with Social Networks Covariates

	Psychological Distress			Method Factor		
	Estimate	S.E.	Est/S.E.	Estimate	S.E.	Est/S.E.
Social Cohesion	**0.08*** **	0.008	10.525	**0.085*** **	0.008	10.257
Trust in others	**0.043*** **	0.006	7.033	**0.077*** **	0.007	10.528

Note. All variables are standardised on both axis.

*N= 37722

[b] Statistically significant association in bold, *<0.05, **<0.01, ***<0.001

All relationships in this cluster were statistically significant but were weak in nature, with no relationships exceeding 0.1 in strength. The method factor was more strongly associated with the variables in this cluster than the general factor with regression coefficients of 0.85 compared to 0.8 in relation to social cohesion and 0.077 compared to 0.043 respectively.

Politi (2 Factor Model)

Medical Conditions

Table 4. 23

Medical Conditions Regressed onto Social Dysfunction and General Dysphoria

	Social Dysfunction			General Dysphoria		
	Estimate	S.E.	Est/S.E.	Estimate	S.E.	Est/S.E.
Asthma	**0.094*** **	0.019	4.909	**0.052** **	0.02	2.562
Arthritis	**0.132*** **	0.017	7.655	**0.286*** **	0.019	15.098
Congestive heart failure	0.052	0.111	0.473	0.212	0.142	1.492
Coronary heart disease	0.058	0.049	1.198	**0.169** **	0.06	2.799
Angina	0.069	0.047	1.481	**0.305*** **	0.053	5.817
Emphysema	**0.154* **	0.078	1.97	**0.288** **	0.084	3.439
Hyperthyroidism	**0.17* **	0.079	2.139	0.115	0.095	1.216
Hypothyroidism	0.07	0.036	1.939	**0.105** **	0.036	2.946
Chronic bronchitis	**0.177** **	0.062	2.841	**0.243** **	0.077	3.148
Liver condition	**0.302*** **	0.076	3.999	**0.348*** **	0.098	3.55
Cancer or malignancy	**0.237*** **	0.064	3.699	**0.52*** **	0.069	7.546
Diabetes	-0.011	0.027	-0.413	**0.091** **	0.03	3.034
Epilepsy	**0.23** **	0.071	3.215	**0.236** **	0.09	2.614
High blood pressure	-0.019	0.017	-1.118	**0.139*** **	0.018	7.608

Clinical depression	**1.573*** **	0.031	51.264	**1.397*** **	0.043	32.115

Note. All variables standardised on Y axis (the factors) only.

*N= 40259

[b] Statistically significant association in bold, *<0.05, **<0.01, ***<0.001

The two factors displayed a range of statistically significant relationships with the medical conditions investigated. The general dysphoria factor displayed more relationships with conditions than the social dysfunction variable did. With the exception of hyperthyroidism, there were no conditions where the social dysfunction variable exhibited a relationship, and the general dysphoria factor did not. Generally, in the cases where both factors exhibited statistically significant relationships with the covariate, the relationships were quite similar, with clinical depression exhibiting a relationship of 1.573 and 1.397 with the respective factors, epilepsy exhibiting relationships of 0.26 and .0.269 respectively and asthma exhibiting relationships of 0.094 and 0.052 respectively between the two factors of this model. The relationship between the social performance factor and a clinical depression diagnosis was of similar magnitude to that of the multidimensional model, whereas the general dysphoria factor was slightly weaker.

Demographics

Table 4. 24

Demographic Covariates Regressed onto Social Dysfunction and General Dysphoria

	Social Dysfunction			General Dysphoria		
	Estimate	S.E.	Est/S.E.	Estimate	S.E.	Est/S.E.
Sex	0.197***	0.011	17.513	0.141***	0.012	11.772
Age	-0.058***	0.006	-9.237	0.141***	0.006	22.31
Rurality	-0.098***	0.014	-6.86	-0.061***	0.014	-4.3

Note. Sex and rurality are standardised on the Y axis (the factors) only whereas Age is standardised on both axis

*N=40452

[b] Statistically significant association in bold, *<0.05, **<0.01, ***<0.001

Statistically significant relationships were observed in table 4.21 for the relationship between all demographic covariates and factors. When regressed onto the factors, the covariates exhibited positive regression coefficients for sex and negative coefficients for rurality. Age displayed a positive coefficient of 0.141 in relation to general dysphoria, and a negative coefficient of -0.058 was observed in relation to social dysfunction. In contrast to the unidimensional model, age showed a statistically significant relationship with the two factors, where it did not with the factor of a unidimensional model.

Mental Health Measures

Table 4. 25

Mental Health Covariates Regressed onto Social Dysfunction and General Dysphoria

	Social Dysfunction			General Dysphoria		
	Estimate	S.E.	Est/S.E.	Estimate	S.E.	Est/S.E.
General Health	**-0.057*****	0.008	-7.129	-0.003	0.007	-0.504
SF 12 Physical	**-0.277*****	0.009	-30.883	**-0.086*****	0.007	-12.901
SF 12 Mental	**-0.416*****	0.009	-47.159	**-0.514*****	0.006	-82.434
SWE	**-0.318*****	0.008	-41.707	**-0.364*****	0.006	-56.847

Note. All variables are standardised on both axis.

*N= 38256

[b] Statistically significant association in bold, *<0.05, **<0.01, ***<0.001

With the exception of the ‘General Health’ question, both factors displayed a statistically significant relationship with the mental health measures covariates. Relationships between SWE and SF mental components ranged from -0.318 to -0.514. These can generally be described as moderately strong relationships, with small differences in regression coefficients observed between the factors. The relationships presented would suggest that social performance was associated with physical health, while general dysphoria is more strongly associated with mental health.

Social Networks

Table 4. 26

Neighbourhood Variables Regressed onto Social Dysfunction and General Dysphoria

	Social Dysfunction			General Dysphoria		
	Estimate F1	S.E. F1	Est/S.E. F1	Estimate F2	S.E. F2	Est/S.E. F2
SOCIAL COHESION	**0.031*** **	0.003	9.571	**0.077*** **	0.005	16.678
TRUSTWORTHINESS OF OTHERS	**0.016*** **	0.003	6.213	**0.055*** **	0.004	13.713

Note. All variables standardised on both axis

*N= 37722

[b] Statistically significant association in bold, *<0.05, **<0.01, ***<0.001

All relationships between the factors and the social networking questions were, while generally statistically significant, of a weak nature. Social dysfunction was found to be more weakly associated with the two variables in this cluster than general dysphoria, with regression coefficients of 0.077 and 0.031 for social cohesion and 0.016 and 0.055 in relation to ‘trustworthiness of others’.

Andrich and van Shoeubreck (2-factor model)

Medical Conditions

Table 4. 27

Medical Conditions Covariates Regressed onto Positively Worded Items and Negatively Worded Items

	Positively Worded Items			Negatively Worded Items		
Condition	Estimate	S.E.	Est/S.E.	Estimate	S.E	Est/S.E.
Asthma	**0.064****	0.02	3.141	**0.095*****	0.019	4.956
Arthritis	**0.284*****	0.019	14.655	**0.133*****	0.017	7.789
Congestive heart failure	0.211	0.14	1.507	0.049	0.11	0.448
Coronary heart disease	**0.164****	0.059	2.754	0.058	0.048	1.191
Angina	**0.275*****	0.055	5.021	0.08	0.046	1.747
Emphysema	**0.32*****	0.087	3.668	0.137	0.077	1.786
Hyperthyroidism	0.137	0.097	1.405	**0.167***	0.079	2.105
Hypothyroidism	**0.081****	0.037	2.173	**0.086***	0.036	2.366
Chronic bronchitis	**0.244****	0.078	3.136	**0.185****	0.063	2.955
Liver condition	**0.353*****	0.098	3.589	**0.313*****	0.076	4.094
Cancer or malignancy	**0.476*****	0.069	6.845	**0.266*****	0.065	4.075
Diabetes	**0.096****	0.031	3.085	-0.021	0.027	-0.795
Epilepsy	**0.286*****	0.088	3.269	**0.206****	0.071	2.892
High blood pressure	**0.111*****	0.019	5.999	-0.011	0.017	-0.654
Clinical depression	**1.599*****	0.043	37.038	**1.533*****	0.030	49.663

Note. All variables are standardised on the Y axis (the factors) only

*N= 40259

[b] Statistically significant association in bold, *<0.05, **<0.01, ***<0.001

Positively and negatively worded factors displayed statistically significant relationships with most of the medical conditions investigated, and only congestive heart failure failed to display statistically significant with either of the factors. A number of conditions only displayed statistically significant relationships with one of the factors such as coronary heart disease, angina and emphysema having a statistically significant relationship with positively worded items, however, only hyperthyroidism was associated in a statistically significant manner with negatively worded items. Generally, when both items were statistically significantly associated, the regression coefficient was relatively similar between the two factors, for example, when investigating clinical depression, the regression coefficient was 1.599 and 1.533 respectively, and regarding hypothyroidism, the coefficients are 0.081 and 0.086 respectively. Relationships were generally of similar strength to those demonstrated in the unidimensional model (see table 4.8)

Demographics

Table 4. 28

Demographic Covariates Regressed onto Positively Worded Items and Negatively Worded Items

	Positively Worded Items			Negatively Worded Items		
	Estimate	S.E.	Est/S.E.	Estimate	S.E.	Est/S.E.
Sex	**0.162*** **	0.012	13.559	**0.196*** **	0.011	17.386
Age	**0.108*** **	0.006	16.744	**-0.051*** **	0.006	-8.068
Rurality	**-0.069*** **	0.014	-4.827	**-0.1*** **	0.014	-7

Note Sex and Rurality are standardised on the Y axis only whereas Age is standardised on both axis

*N=40452

[b] Statistically significant association in bold, *<0.05, **<0.01, ***<0.001

All covariates displayed statistically significant relationships with the various factors in this model. In a similar pattern to other models, sex was positively associated with the two factors and rurality was negatively associated. Age displayed a mixed relationship with the two factors with a regression coefficient of 0.108 and -0.051, respectively. There were small differences in the regression coefficient of the factors being regressed onto the rurality and sex covariates, with coefficients of 0.162 and 0.196 for sex and -0.069 and -0.1 respectively.

Mental Health Scores

Table 4. 29

Mental Health Covariates Regressed onto Positively Worded Items and Negatively Worded Items

	Positively Worded Items			Negatively Worded Items		
	Estimate	S.E.	Est/S.E.	Estimate	S.E.	Est/S.E.
General Health	**-0.051*** **	0.007	-6.974	-0.005	0.007	-0.73
SF 12 Physical	**-0.271*** **	0.008	-32.137	**-0.087*** **	0.007	-12.99
SF 12 Mental	**-0.452*** **	0.008	-55.581	**-0.524*** **	0.006	-83.485
SWEMWBS	**-0.36*** **	0.007	-50.414	**-0.356*** **	0.006	-54.839

Note. All variables standardised on both Axis

*N= 38256

[b] Statistically significant association in bold, *<0.05, **<0.01, ***<0.001

General health was the only covariate that failed to demonstrate a statistically significant relationship when the factors of this model were regressed onto it. While a statistically significant relationship was observed between positively worded items (p=<0.001), it was not evident for negatively worded items (p=0.465). The regression coefficient of the SF-12 mental component was of greater magnitude than the physical component. The difference between the factors was most evident in the relation with the SF-12 physical component where regression coefficients were -0.271 and -0.087 respectively, however, generally, similar relationships were observed between covariates and both factors.

Social Networks

Table 4. 30

Neighbourhood Covariates regressed onto Positively Worded Items and Negatively Worded Items

	Positively Worded Items			Negatively Worded items		
	Estimate	S.E.	Est/S.E.	Estimate	S.E.	Est/S.E.
SOCIAL COHESION	0.032***	0.003	10.323	0.077***	0.005	16.589
TRUSTWORTHINESS OF OTHERS	0.017***	0.002	6.922	0.055***	0.004	13.678

Note. All variables are standardised on both axis

*N= 37722

[b] Statistically significant association in bold, *<0.05, **<0.01, ***<0.001

Both social cohesion and trustworthiness of others variables were found to demonstrate a statistically significant relationship when factors were regressed onto the neighbourhood variables. Negatively worded items were shown to have greater regression coefficients than positively worded items in relation to both variables. All relationships were relatively weak, with social cohesion displaying relationships displaying regression coefficients of 0.031 and 0.77 respectively for social cohesion and 0.017 and 0.055 respectively for ‘trustworthiness of others’. Differences between the regression coefficients depending on each factor were evident with negatively worded items displaying significantly higher regression coefficients than positively worded items.

4.5- Discussion

The purpose of this section of the chapter was to provide both a discussion of the results obtained in this analysis and, in conjunction with analyses conducted in chapter 3, to conclude which model was the most appropriate dimensional representation of the data. In line with the procedure detailed in the methods section, the best fitting model which performed sufficiently well in covariate analysis would be selected as the most appropriate to bring forward for longitudinal analyses. A summary of each of the dimensional representations which were brought forward to this chapters analysis was provided in table 4.1.

Prior to the covariate analysis of models that had demonstrated sufficient fit in Chapter 3, a unidimensional model was regressed onto all covariates to provide a baseline from which other models could be compared. Regressions with the unidimensional model showed statistically significant relationships with a number of medical conditions; however, the strongest relationship was between the unidimensional model and the clinical depression covariate. While statistically significant relationships were observed for other conditions, this was not unexpected as other conditions have been shown to have a relationship with mental health, such as cancer, which has been shown to result in elevated prevalence of mental health conditions relative to the general population (Singer, Das Munshi & Brahler, 2010).

Demographic covariates displayed statistically significant relationships with the unidimensional model’s factor when it was regressed onto rurality and sex, however not with age. While age did not display a statistically significant relationship with the unidimensional models’ factor, it did with every other model investigated. These relationships generally contained a mixture of positive and negative relationships in multidimensional representations. This was incongruent with previous literature which suggested that age would have an effect on mental health and elderly participants would report higher levels of phycological distress (Aldwin et al., 1998), however, it was acknowledged that this relationship was reported as U shaped in the literature (Mackenzie et al., 2011) and therefore may not have been properly captured using linear regressions. The fact that basic unidimensional models did not display a statistically significant relationship with age, whereas other unidimensional representations did was indicative of those representations demonstrating concurrent validity, whereas the simple unidimensional model did not.

All mental health covariates displayed significant relationships with the unidimensional model. Multidimensional models tended to have factors which displayed strong relationships with those that were physical in nature such as the SF-12 and general health variables and others which displayed stronger relationships with the SWE and SF mental components which were more psychological in nature.

Following analysis in chapter 3, a unidimensional model did not demonstrate sufficient fit as to suggest that it was an appropriate dimensional representation of the data. While it was included in chapter 4 analyses, this was for the purposes of establishing a baseline, for how a unidimensional model was associated with the covariates investigated and from which the relationships that the factors of other models could be compared.

A number of models, listed in table 4.31 were discounted during chapter 3, due to not demonstrating appropriate fit during these analyses. The remaining models were then investigated in order of fit and will be discussed as such.

While the best fit for the data, Graetz's (1991) model was discounted due to concerns about the validity of the factors. The three factors did demonstrate variability in relation to the covariates, especially in relation to medical conditions, but the validity of these relationships was of concern. The social performance factor was primarily associated with conditions which affected the heart and blood pressure. While it was expected that anxiety and depression would have a stronger relationship with these covariates, research suggested that the relationship between anxiety and high blood pressure would not always manifest (Jones- Webb et al., 1996; Shinn et al., 2001) and that social support may affect blood pressure responses in some circumstances (Gallagher & Whitney, 2012). Of greater concern was the performance of the '*anxiety and depression*' which was not the most strongly associated factor with clinical depression, and while this factor did exhibit a statistically significant relationship with the covariate mentioned, one could reasonably expect a factor which explicitly claimed to measure depression to be the most strongly associated with having a clinical depression diagnosis. Similarly, when investigating the 'social performance' factor, one could reasonably expect it to be strongly associated with the neighbourhood based variables investigated. When analysed, however, social performance was found to be the most weakly associated of the factors with these covariates.

Variability was also not demonstrated in relation to demographic variables with similar regression coefficients for all factors in relation to the demographic and social variables.

The lack of concurrent validity demonstrated in these analyses would suggest that the factors were not measuring what they claimed to measure. Furthermore, in Chapter 3, the three factors were found to be highly correlated, and in chapter 4, this may have contributed to the similarity in the relationships demonstrated between the various factors and the covariates, especially demographic covariates and those of mental health measures. In line with previous findings in Gao et al. (2004) and Shevlin and Adamson (2004), given the similarity in the relationships between the factors, then treating the GHQ-12 as multidimensional in this way, would not provide any benefit to researchers or clinicians.

Hankins model was the second-best fit for the data. This model makes extensive use of correlated errors which are used to simulate wording effects. As detailed in the introduction of this chapter, this method has been extensively criticised, and while a more detailed discussion of the historical debate around correlated errors inclusion in analyses is detailed in the introduction to this chapter, the most pertinent conclusions were that of Lance, Noble and Cullen (2002) who compared the use of CTCU and CTCM and identified a number of theoretical shortcomings with correlating errors as used in the CTCU approach. From these finding, it was suggested that the CTCU method relied upon in Hankins (2008) should only ever be used as a last resort and only when CTCM methods and other dimensional representations had been shown to be ineffective. Instead of discounting the model based on the concerns expressed in the above studies, this model was retained as a 'last resort', which would be brought forward if no other dimensional representation was found to be appropriate.

Ye's (2009) model, which modelled wording effects using the CTCM method was the third best-fitting model. It demonstrated sufficient fit as to indicate that it was an appropriate dimensional representation for the data, and when subjected to covariate analysis, the general mental health factor was associated in a statistically significant manner with many more medical conditions than the methods factor. The general factor also was much more strongly associated with having a clinical depression diagnosis and other conditions such as cancer than the method effects factor. While demographic covariates such as sex and rurality exhibited similar relationships with both factors, a distinction could be made when the factors were regressed onto age.

Finally, the general factor exhibited stronger relationships with all covariates in the mental health cluster, with the exception of the SF-12 mental component. While the SF-12 correlation was smaller than that of the method factor, it was still statistically significant and was not thought sufficient as to fatally undermine the validity of the model. While little theoretical justification could be found for the relationships between the method factor it may be that the factor was inappropriately labelled and was actually capturing a distinct construct such as physical health, a sub dimensional of mental health. It may also have been the case that the factor captured relatively meaningless variation in the data and that this variance could not be attributably to any latent construct. In the absence of any discernible findings which could identify what the factor captured, it was decided to disregard it in further analyses.

Ye's (2008) model was the best fitting model that was deemed to have performed sufficiently well in the supplementary analysis as to warrant its inclusion as the most appropriate model for longitudinal analysis. However, it was deemed appropriate to conduct the analysis on all models which were identified in Chapter 3 as representing a good fit of the data, in order to provide a more complete view of model performance.

The two models which were also included were Politi's (1991) and Andrich and van Shoeubroek's (1989) model. The structure of these two models is extremely similar, with the only difference being a cross-loaded item in Politi's (1991) model being exclusively associated with a factor in Andrich and van Shoeubroek's (1989) model. Due to the similarity of the models, they performed similarly in all analyses and therefore, will be discussed together. While demonstrating appropriate fit in analyses conducted in chapter 3, the two factors did not display a significant difference between the associations between factors and many of the covariates. In relation to positively worded items and anxiety and depression, these factors exhibited a statistically significant relationship with general health, whereas social dysfunction and negatively worded items did not. While this did demonstrate variability in these factors, the meaningfulness of these variations was not apparent. When investigating medical conditions, similar regression coefficients were evident for clinical depression and other conditions. Politi's model's factors did not perform as was expected, with neighbourhood variables exhibiting weaker regression coefficients when the social performance factor was regressed onto it, and general dysphoria displaying weaker relationships with conditions which one could reasonably expect it to outperform social performance, such as depression diagnoses. While research would suggest that general dysphoria is associated with social performance (Unger, 1999), these terms are rarely used in modern research, and consequently, this relationship is not well researched. It was concluded that the performance of these factors in covariate analysis was not supportive of the validity of this model.

While the aforementioned models are similarly structured, the different labelling of the factors necessitated a different interpretation of the findings. Research into the behaviour of factors comprised of positively and negatively worded items is far from conclusive, Mook found results which directly contradicted each other in the space of a year (Mook et al., 1992; Mook et al., 1991) when he evaluated two different scales which had similar characteristics to the GHQ-12, i.e. positively and negatively worded items. In one study (Mook et al., 1992) he found that positively worded items had a significantly higher mean than negatively worded items, whereas a year later, he found the opposite. In relation to predictive ability, Lia (1994) suggested that positively worded items were more likely to yield statistically significant relationships with medical conditions diagnoses than negatively worded items. This research was conducted using a scale which was similar to the GHQ-12 called the Life Orientation Test on a sample in Hong Kong. The analysis conducted in this chapter showed little variation in the factors suggesting little benefit to researchers to treat the GHQ-12 in this way.

Table 4.31

Summary of Findings in Chapter 4

Model	Reason for rejection
Unidimensional	Poor fit for data (see chapter 3)
Politi (1994)	Factors invalid (see chapter 4)
Andrich and Schoubroeck (1989)	Superior performance by other models
Schmitz (1999)	Poor fit for data (see chapter 3)
Martin (1999)	Poor fit for data (see chapter 3)
Graetz (1991)	Invalid Factors (see chapter 4)
Worsely and Gribbin (1997)	Poor fit for data (see chapter 3)
Hankins (2008)	Conceptual concerns about CTCU
Ye (2009)	Deemed most appropriate.

4.5.1- Limitations

The above research was not without limitations which are detailed below. Firstly when investigating relationships between covariates, there is an implicit acceptance that the measure is reliable and valid. While some covariates such as the SWEMBS and SF-12 have been extensively tested in relation to their reliability and validity, others, have not. For example, the Social Cohesion instrument proposed by Buckner (1998) was validated, however, the truncated version used in this analysis was not. Responses from the Understanding Society User Support Group could not provide scientific underpinnings as to why these particular variables were chosen and why the instrument was truncated in such a manner, other than concerns around the practicality of using the complete version. The study also relies on self-report of data and has been shown in Hunt, Auriemma and Cashaw (2010), underreporting of depressive symptoms, especially in males, is apparent in self-report research. There were also concerns about the use of single-item variables such as the general health variable and trust in others as these represented crude measures of the variable that they claimed to measure. These variables were included as they were present in other research (Shevlin and Adamson, 2004), however, in all cases, these single-item variables were included alongside more complex multi-item scales.

There were also concerns about the conceptual basis of how models attempted to model method effects. Several models attempted to model method effects in one way or another, either through correlated errors or multiple factors. Some of these methods have associated methodological shortcomings, which will be detailed below.

The use of correlated errors has been extensively used throughout the literature but has also been criticised. As early as 1983, researchers have cautioned against the use of correlated errors (Cliff, 1983), with subsequent papers urging a cautionary approach due to a number of methodological shortcomings (Shah and Goldstein, 2006; Tomanken and Waller, 2003; MacCallum, Roznowski and Waller, 1992). Generally, these authors mention how correlated errors can be used to improve model fit primarily on a post hoc basis, without necessarily having theoretical underpinnings.

While some authors have laid out criticisms of this approach, others have attempted to outline situations where it would be appropriate to use correlated errors. Cortina (2009) stated that this technique should only be used when it is unavoidable. Examples of such a situation were listed as when research consisted of multiple uses of the same instrument were used longitudinally and when indicator variables share components. Papers have also ardently argued against using this technique on a post hoc basis as it may simply improve model fit with no underlying conceptual basis (Landis et al., 2009). Gerbing and Anderson (1984) go so far as to argue that the inclusion of correlated errors can mask the underlying relationships between components in a model.

One must also look at the effect that correlating errors in a model has on its degrees of freedom. As correlating errors will inevitably reduce the number of free parameters within a model, it is likely that this will inevitably improve model fit. Forster and Sober (1994) found stated that all things being equal, degrees of freedom can significantly affect a model's ability to provide a good fit of the data at hand.

More recently, Herminda (2015) provided an almost unequivocal rejection of the use of correlated errors in a meta-analysis into the practice where it was stated that "with few exceptions, there is no theoretically defensible reason for the practice of error correlation."

This statement above all is of great concern as methodologically, Hankin's (2008) study relies heavily on correlated errors and while a rationale for doing so does exist, there remains a significant concern within the scientific community about their inclusion and effectiveness.

4.6- Conclusion

The analysis in this chapter identified Ye's (2009) dimensional representation to be the most appropriate to bring forward into later chapter's longitudinal analysis. This dimensional representation of the data posits that the GHQ-12 is unidimensional, however, accounts for method effects by the inclusion of a factor which captures the variance of negatively worded items. This model demonstrated variance in relation to a number of covariates and demonstrated expected relationships with relevant covariates included in this analysis. As a result, Ye's (2009) model will be used in all longitudinal analyses hereon.

References

Abu-Omar, K., Rütten, A., & Lehtinen, V. (2004). Mental health and physical activity in the European Union. *Sozial-und Präventivmedizin*, *49*(5), 301-309.

Aldwin, C. M., Spiro, A., Levenson, M. R., & Bossé, R. (1989). Longitudinal findings from the normative aging study: I. Does mental health change with age?. *Psychology and Aging*, *4*(3), 295.

Andrich, D., & Van Schoubroeck, L. (1989). The General Health Questionnaire: a psychometric analysis using latent trait theory. *Psychological Medicine*, *19*(2), 469-485.

Barrett, P. (2007). Structural equation modelling: adjudging model fit. *Personality and Individual Differences, 42*, 815–824.

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: psychometric properties. *Journal of consulting and clinical psychology*, *56*(6), 893.

Bjarnason, T., & Sigurdardottir, T. J. (2003). Psychological distress during unemployment and beyond: social support and material deprivation among youth in six northern European countries. *Social Science & Medicine*, *56*(5), 973-985.

Booker, C. L., & Sacker, A. (2011). Health over the life course: associations between age, employment status and well-being. *Understanding Society*, *2*.

Buck, N., & McFall, S. (2011). Understanding Society: design overview. *Longitudinal and Life Course Studies*, *3*(1), 5-17.

Buckner, J. C. (1988). The development of an instrument to measure neighborhood cohesion. *American journal of community psychology*, *16*(6), 771-791.

Burton, J., Laurie, H., Lynn, P. (2011). Appendix: Understanding society design overview. In S. McFall (Ed.), Understanding society: Early findings from the first wave of the UK's household longitudinal study (pp. 129–140). Colchester, Essex, UK: Understanding Society, Institute for Social and Economic Research, University of Essex.

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate behavioral research*, *18*(1), 115-126.

Cortina, J. M. (2002). Big things have small beginnings: An assortment of "minor" methodological misunderstandings. *Journal of Management*, *28*(3), 339-362.

Dempster, M., & Hanna, D. (2016). *Research methods in psychology for dummies*. New York, NY: John Wiley et Sons.

Evans, J., Macrory, I., & Randall, C. (2016). Measuring national wellbeing: Life in the UK, 2016. ONS. Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/articles/measuringnationalwellbeing/2016#how-good-is-our-health.

Fat, L. N., Scholes, S., Boniface, S., Mindell, J., & Stewart-Brown, S. (2017). Evaluating and establishing national norms for mental wellbeing using the Short

Warwick–Edinburgh mental well-being scale (SWEMWBS): findings from the health survey for England. *Quality of Life Research*, *26*(5), 1129-1144.

Forster, M., & Sober, E. (1994). How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *The British Journal for the Philosophy of Science*, *45*(1), 1-35.

Gao, F., Luo, N., Thumboo, J., Fones, C., Li, S. C., & Cheung, Y. B. (2004). Does the 12-item General Health Questionnaire contain multiple factors and do we need them?. *Health and Quality of Life Outcomes*, *2*(1), 63.

Gerbing, D. W., & Anderson, J. C. (1984). On the meaning of within-factor correlated measurement errors. *Journal of Consumer Research*, *11*(1), 572-580.

Gl-assessment.co.uk. (2019). *General Health Questionnaire (GHQ)*. [online] Available at: https://www.gl-assessment.co.uk/products/general-health-questionnaire-ghq/ [Accessed 7 Jun. 2019].

Goldberg DP, Gater R, Sartorius N et al. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. Psychological Medicine; 27:191–197

Goldberg, D. P., Oldehinkel, T., & Ormel, J. (1998). Why GHQ threshold varies from one place to another. *Psychological medicine*, *28*(4), 915-921.

Graetz, B. (1991). Multidimensional properties of the general health questionnaire. *Social psychiatry and psychiatric epidemiology*, *26*(3), 132-138.

Hankins, M. (2008). The factor structure of the twelve item General Health Questionnaire (GHQ-12): the result of negative phrasing?. *Clinical Practice and Epidemiology in Mental Health*, *4*(1), 10.

Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! testing! one, two, three–Testing the theory in structural equation models!. *Personality and Individual Differences*, *42*(5), 841-850.

Hewitt, M., & Rowland, J. H. (2002). Mental health service use among adult cancer survivors: analyses of the National Health Interview Survey. *Journal of Clinical Oncology*, *20*(23), 4581-4590.

Human, J., & Wasem, C. (1991). Rural mental health in America. *American Psychologist*, *46*(3), 232.

Ip, W. Y., & Martin, C. R. (2006). Factor structure of the Chinese version of the 12-item General Health Questionnaire (GHQ-12) in pregnancy. *Journal of Reproductive and Infant Psychology*, *24*(02), 87-98.

Jones-Webb, R., Jacobs Jr, D. R., Flack, J. M., & Liu, K. (1996). Relationships between depressive symptoms, anxiety, alcohol consumption, and blood pressure: results from the CARDIA study. *Alcoholism: Clinical and Experimental Research*, *20*(3), 420-427.

Lai, J. C. (1994). Differential predictive power of the positively versus the negatively worded items of the Life Orientation Test. *Psychological Reports*, *75*(3_suppl), 1507-1515.

Lawrence, V., Murray, J., Banerjee, S., Turner, S., Sangha, K., Byng, R., ... & Macdonald, A. (2006). Concepts and causation of depression: A cross-cultural study of the beliefs of older adults. *The Gerontologist*, *46*(1), 23-32.

Lobo, A., Pérez-Echeverría, M. J., & Artal, J. (1986). Validity of the scaled version of the General Health Questionnaire (GHQ-28) in a Spanish population. *Psychological medicine*, *16*(1), 135-140.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychological bulletin*, *111*(3), 490.

Mackenzie, C. S., Reynolds, K., Cairney, J., Streiner, D. L., & Sareen, J. (2012). Disorder-specific mental health service use for mood and anxiety disorders: Associations with age, sex, and psychiatric comorbidity. *Depression and anxiety*, *29*(3), 234-242.

Marston, G. M., Perry, D. W., & Roy, A. (1997). Manifestations of depression in people with intellectual disability. *Journal of Intellectual Disability Research*, *41*(6), 476-480.

Martin, A.J. (1999). Assessing the multidimensionality of the 12-Item General Health Questionnaire. Psychological Reports, 84, 927-935

McHorney, C. A., Ware Jr, J. E., Lu, J. R., & Sherbourne, C. D. (1994). The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical care*, 40-66.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus: Statistical analysis with latent variables; user's guide;[version 7]*. Muthén et Muthén.

Nicholson, L. A. (2008). Rural mental health. *Advances in Psychiatric Treatment*, *14*(4), 302-311.

Office for National Statistics, 2018, accessed from https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/male-and-female-populations/latest#main-facts-and-figures [accessed 16/09/2016]

Paykel, E. S., Abbott, R., Jenkins, R., Brugha, T. S., & Meltzer, H. (2000). Urban–rural mental health differences in Great Britain: findings from the National Morbidity Survey. *Psychological medicine*, *30*(2), 269-280.

PENNINKILAMPI-KEROLA, V. A. R. P. U., Miettunen, J., & Ebeling, H. (2006). A comparative assessment of the factor structures and psychometric properties of the GHQ-12 and the GHQ-20 based on data from a Finnish population-based sample. *Scandinavian Journal of Psychology*, *47*(5), 431-440.

Rey, J. J., Abad, F. J., Barrada, J. R., Garrido, L. E., & Ponsoda, V. (2014). The impact of ambiguous response categories on the factor structure of the GHQ–12. *Psychological assessment*, *26*(3), 1021.

Shah, R., & Goldstein, S. M. (2006). Use of structural equation modeling in operations management research: Looking back and forward. *Journal of Operations management*, *24*(2), 148-169.

Shevlin, M., & Adamson, G. (2005). Alternative factor models and factorial invariance of the GHQ-12: a large sample analysis using confirmatory factor analysis. *Psychological assessment*, *17*(2), 231.

Shinn, E. H., Poston, W. S. C., Kimball, K. T., St. Jeor, S. T., & Foreyt, J. P. (2001). Blood pressure and symptoms of depression and anxiety: a prospective study. *American Journal of Hypertension*, *14*(7), 660-664.

Singer, S., Das-Munshi, J., & Brähler, E. (2010). Prevalence of mental health conditions in cancer patients in acute care—a meta-analysis. *Annals of oncology*, *21*(5), 925-930.

Sriram, T. G., Chandrashekar, C. R., Isaac, M. K., & Shanmugham, V. (1989). The general health questionnaire (GHQ). *Social Psychiatry and Psychiatric Epidemiology*, *24*(6), 317-320.

Tait, R. J., Hulse, G. K., & Robertson, S. I. (2002). A review of the validity of the General Health Questionnaire in adolescent populations. *Australian & New Zealand Journal of Psychiatry*, *36*(4), 550-557.

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., ... & Stewart-Brown, S. (2007). Health and Quality of Life Outcomes. *Health and Quality of Life Outcomes, 5*, 63.

Tomarken, A. J., & Waller, N. G. (2005). Structural equation modeling: Strengths, limitations, and misconceptions. *Annu. Rev. Clin. Psychol.*, *1*, 31-65.

Unger, J. B., McAvay, G., Bruce, M. L., Berkman, L., & Seeman, T. (1999). Variation in the impact of social network characteristics on physical functioning in elderly persons: MacArthur Studies of Successful Aging. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *54*(5), S245-S251.

Vanheule, S., & Bogaerts, S. (2005). The factorial structure of the GHQ-12. *Stress and Health: Journal of the International Society for the Investigation of Stress*, *21*(4), 217-222.

Vos, T., Barber, RM., Bell, B., Bertozzi-Villa, A., Biruyukov, S., Bollinger, I., ...Murray, CJ.. (2013). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: A systematic analysis for the Global Burden of Disease study. The Lancet, 386(9995), 743-800.

Westerhof, G. J., & Keyes, C. L. (2010). Mental illness and mental health: The two continua model across the lifespan. *Journal of adult development*, *17*(2), 110-119.

Westerhof, G. J., & Keyes, C. L. (2010). Mental illness and mental health: The two continua model across the lifespan. *Journal of adult development*, *17*(2), 110-119.

Whelan, C. T. (1994). Social class, unemployment, and psychological distress. *European Sociological Review*, *10*(1), 49-61.

Ye, S. (2009). Factor structure of the General Health Questionnaire (GHQ-12): The role of wording effects. *Personality and Individual Differences*, *46*(2), 197-201.

Chapter 5 - Factorial Invariance of the GHQ-12

5.1- Abstract

Introduction

During Chapter 4, Ye's (2009) model was deemed the most appropriate dimensional representation of the data. This conclusion, however, was based on cross-sectional data. Should longitudinal analyses be conducted, researchers need to be relatively sure that important characteristics of the data remain constant over time. If the stability of a model cannot be established, or important characteristics of the model change over time, longitudinal data may be difficult or impossible to analyse in a structural equation framework.

Methods

The stability of the best performing model in previous chapters was tested using guidelines described in Widaman and Reisse (1997). This framework consists of running increasingly constrained confirmatory factor analyses and investigating fit statistics as these constraints were applied. Finally, modification indices were examined to ascertain if any inter-item effects were present within the data

Results

'*Strong*' measurement invariance, as defined in literature guidelines, was found to be present in the data, which would suggest that the model remains stable enough to warrant further longitudinal analyses. Furthermore, some items were found to demonstrate a longitudinal effect, but it was not deemed severe enough to affect the further analysis.

Conclusions

The stability of the model shows that the relationship between the characteristics of the model changed little over time. This implied a stable model resilient to retest effect and extraneous variables. Furthermore, this degree of invariance will allow for a full range of longitudinal analysis in later chapters.

5.2- Introduction

In order to conduct longitudinal analysis using structural equation techniques, researchers must be confident that their underlying model remains stable at all time points. Furthermore, they must whether the model changes, and if it does, to what degree. The extent to which these changes occur, or do not, is referred to as 'Factorial Invariance.'

Factorial invariance refers to whether a factor's characteristics remain the same under different conditions. These conditions can refer to subgroups of a population, occasions of measurement, and different test settings (Meade & Wright, 2012). There are two types of factorial invariance, multigroup invariance, which tests whether a model remains constant between groups and longitudinal invariance, which tests model properties across time (Bialosiewicz, Murphy & Kelly, 2013).

For the purposes of this research, analyses focused on the longitudinal aspects of the GHQ-12's factorial invariance, i.e., how its characteristics changed over time.

In order to do this, numerous CFA models were estimated with increasingly more stringent constraints placed upon Ye's (2009) model to detect if it remained a good fit. Widaman and Reise (1997) proposed a structure of these increasingly demanding constraints (see table 5.1). By compiling fit statistics from these analyses, researchers can observe the effects that constraining each characteristic have on a model's overall fit. By determining the point at which a model no longer represented an acceptable fit, it is possible to determine which characteristics remain constant and which do not.

Through the use of modification indices the effect to which small changes in the model can improve the overall fit was investigated. Through this analysis, it was possible to investigate how individual items performed over time and to investigate temporal phenomena, which may have affected model fit over time. The use of modification indices has prompted a discussion in the literature about the appropriateness of their use, however, these are explained later in the chapter, and efforts were made to ensure that they were used in a way that was methodologically and conceptually sound.

5.2.1- Review of Significant Studies into Factorial Invariance of the GHQ

In order to gain an understanding of the context of the literature surrounding the factorial invariance of the GHQ-12, a review of relevant studies was completed. When investigating factorial invariance of the GHQ, it was important to note that the underlying factor structure model that the researchers used was not the same. This may, in turn, have led researchers to come to differing conclusions. As a result, the underlying model that each researcher used was detailed.

Generally, it has been claimed that researchers have focused on invariance across groups but have neglected invariance across time, with only a small number of studies investigating the GHQ-12's longitudinal characteristics (Mäkikangas et al., 2006). Early research into longitudinal factorial invariance was conducted by Graetz (1991), where he found a three-factor solution to be the most appropriate dimensional representation to describe the data using maximum likelihood factor analysis. He found the structure to be stable over four years of longitudinal data, however, he did not use a representative sample and instead focused on young people. Mäkikangas et al. (2006) tested participants who undertook the Jyva¨skyla¨ Longitudinal Study of Personality and Social Development, which provided continuous data across six years with all participants being aged 36 at the first wave and 42 at the studies completion. This research also found a three-factor structure to be stable across time.

More recently, Smith et al. (2012) analysed longitudinal factorial invariance using Hankins' (2008) model. This model was unidimensional but accounted for method effects using the '*correlated traits correlated uniqueness*' (CMCU) method (see chapter 3). Alongside this dimensional representation, a unidimensional model and two and three-factor models were tested. This study compared waves 1 and 3 of the English Longitudinal Study of Aging, which consisted of participants aged 50 and over. Data was collected in waves commencing between March 2002 and March 2003, while wave three was collected between May 2006 and August 2017. This research was slightly simplistic as rather than look at a number of time points, it simply treated the two waves as start and end-points and investigated change between them. It did, however, separate participants into those who scored highly and those who scored poorly using a technique known as Cluster Analysis. It found that Hankins' (2008) model was the most appropriate model at the two time points, suggesting stability over time. This study did not use a representative sample, instead focusing on elderly participants. Nonetheless, the analysis did suggest that Hankins' (2008) model demonstrated stability and good fit longitudinally within the population tested.

The most representative evaluation across time was conducted by Hammarström et al. (2016). This research was conducted over two time points, one in 1981 and another in 2016. They attempted to avoid the method effects that have been proposed to be present in the GHQ-12 by modelling the positive and negative items separately, which they referred to as GHQ-6-P and GHQ-6-N. They found that the GHQ-6-N was invariant over time, however, the GHQ-6-P was not. This would suggest that the differently worded items of the GHQ behave differently over time, which had previously not been suggested.

5.2.2- Statistical Procedures

The studies mentioned above used a number of statistical techniques to undertake their analysis. It was considered important to investigate the statistical procedures that each researcher used in order to understand why they obtained the findings that they did.

The first investigation into temporal invariance was conducted in 1991. Graetz' (1991) research used principal component analysis (PCA) with a variety of rotation methods. Firstly, an orthogonal (varimax) was utilised, however, Graetz suggests that this rotation method yielded too complex solutions, and as a result, oblique rotations were used instead. A more in-depth description of PCA rotation methods is given in Chapter 3. It must be noted that this approach is not advised as it could be considered to be cherry-picking statistical techniques to find the desired outcome (Kenny, 2015). Instead, a rationale for a particular rotation method should be outlined before analysis is conducted, and the results reported accordingly. PCA is a much less advanced technique than CFA, and operates under different mathematical and theoretical assumptions. A detailed explanation of the differences between the techniques is given in Chapter 3 (section 1.10).

More recently, Mäkikangas' (2006) research used a CFA within a structural equation modeling framework. This method was similar to the methods employed thus far within this thesis. Mäkikangas' (2006) research investigated four models, two of which were three factors model (Martin, 1999; Graetz, 1991), a two-factor (Schmitz, 1999) and a single factor model. More recent additions to the literature, such as Hankins (2008) and Ye (2008), were not investigated. Three-factor solutions were found to provide the best fit of the data and were found to be “relatively stable across the two time points.” Mäkikangas’s (2006) study did not investigate temporal invariance as is proposed in this chapter and instead analysed the GHQ-12 scores across two time points and compared the results.

Smith et al. (2010) utilised a technique known as a Rasch Analysis alongside conventional CFA techniques. Rasch models are latent trait models that investigate the probabilistic relationship between individuals and items. Importantly, Rasch models require a large number of assumptions to be fulfilled. These assumptions include population response invariance, i.e. the entire population must perform uniformly and model unidimensionality. While research has shown that certain characteristics are invariant across subgroups such as sex and age (Shevlin & Adamson, 2005; Cheung, 2002) and even between clinical and non-clinical subgroups (Fernandes & Vasconcelos-Raposo, 2013), it cannot be said with confidence that the general population would behave uniformly in relation to GHQ-12 scores, nor can it be said that GHQ-12 responses are unidimensional. As Smith’s sample was fairly niche, that of cancer patients, it can be assumed that these would be relatively heterogeneous compared to general population data and therefore while perhaps not appropriate to a general population sample, it may have been appropriate for Smith’s sample.

Overall it can be summarised that the large number of statistical procedures were likely to have affected the results obtained by each researcher, however techniques used in Mäkikangas’ study (2006) were similar to those which were employed in this analysis so were of particular relevance.

5.2.3- Hypotheses

In line with Chapter 1's thesis structure, this chapter investigated the hypotheses That the model will demonstrate sufficient factorial invariance as to merit further longitudinal research.

The extent to which the first hypothesis was fulfilled would have an impact on the nature of longitudinal research on the model. Depending on the level of factorial invariance demonstrated, the extent of longitudinal analysis on future analysis would be either enabled or curtailed. The second hypothesis related to whether individual items contributed disproportionately to poor fit, which was supplementary and exploratory in nature and would be fulfilled if no clear pattern emerged, nor were there abnormally ill-fitting relationships between items in the longitudinal model. Furthermore, should factorial invariance not be found, analysing the individual items may provide an opportunity to explain this.

5.3- Methods

5.3.1- Data

The data for this analysis were obtained from a merged dataset of Waves 1 to 5 of the Understanding Society database. While not every participant completed every wave of the GHQ-12, 'Maximum Likelihood Estimator with Robust Standard Errors' (MLR) was used to compute probable values for missing data (see section 3.2.5 for justification). This was necessary as SEM cannot be conducted with missing data in the analyses (Muthen & Muthen, 2012). MLR can only estimate missing values in cases where data is partially completed, and consequently, participants who never completed GHQ-12 were discounted. MLR is also considered resistant to data not being normally distributed (Muthen & Muthen, 2012). Following the discounting of participants with no GHQ-12 data, 65568 participants completed GHQ-12 scores at least once during the five waves and were included in the analysis. Proxy responses were not collected for the GHQ-12 responses, which contributed to the large levels of attrition suffered by participants in this process.

For more information on the merged dataset, including the numbers of participants who completed each wave and pattern analysis of responses, see Chapter 2. The dataset uses weights clustering and stratification variables in order to retain its representativeness of the UK population. These are detailed in Chapter 2, however, this research required the longitudinal weighting variable to be generated as laid down in the UKHLS user guide (GL Fumagali, Knies & Buck, 2017).

5.3.2- Analyses

The analysis conducted in this chapter consisted of two parts, a CFA to investigate measurement invariance of the model in its entirety and secondly the use of modification indices to investigate individual items relationships across time. Both of these analyses were detailed below.

5.3.2.1- Confirmatory Factor Analyses

Measurement invariance was established by conducting a number of CFA sequentially with increasingly constrained parameters on Ye’s (2009) model, which is graphically represented in figure 5.1. This model incorporated a single factor which claimed to measure psychological disturbance and another to simulate method effects and was found to be the most appropriate dimensional representation in previous chapters. This was conducted by fixing certain characteristics of the model across time. Widaman and Reise (1997) proposed a structure of these increasingly demanding constraints, shown below.

Table 5.1

Guidelines for Levels of Factorial Invariance

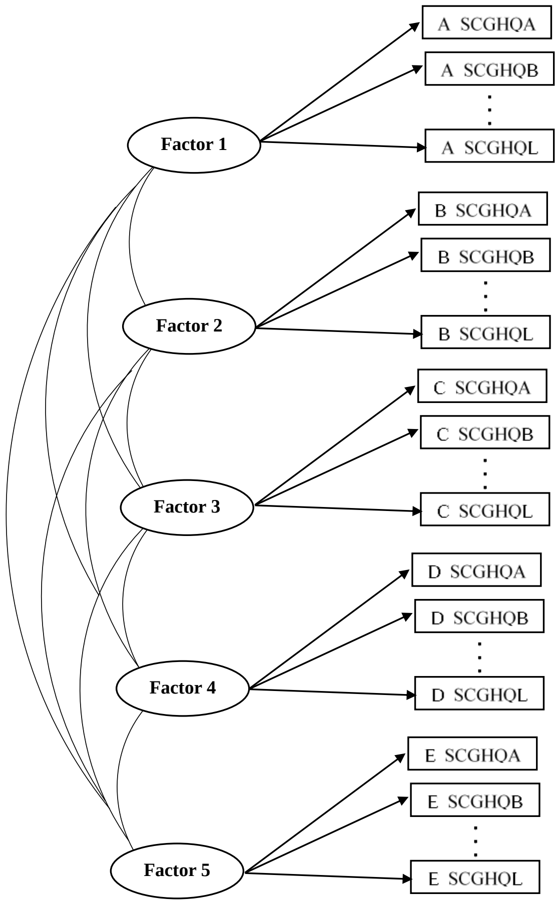
	Factor loadings	Intercepts	Residuals	Factor Variances
Configural				
Weak	✓			
Strong	✓	✓		
Strict	✓	✓	✓	
Very Strict	✓	✓	✓	✓

Should the model continue to demonstrate good fit, despite the constraints placed upon it, then it can be assumed that the characteristic that has just been constrained did not vary to a large degree over time. Once fit statistics were collated, the most restrictive model which still performed acceptably well in relation to fit statistics was selected as it represents the most parsimonious model that still adequately represents the data (Geiser 2013). Alternatively, a chi-square difference test can be used to ascertain if two groups are significantly different from each other (Werner & Schermelleh-Engel, 2010). This approach, however, has been viewed as inappropriate as Chi-squared based tests tend to be influenced by sample size (Kelloway, 1995). Due to the large sample of this data, this influence would likely be pronounced. While Chi-squared results were reported, their influence on interpretation was limited. Instead, greater emphasis was placed on the Comparative Fit Index (CFI) and the Tucker Lewis Index (TLI) as these were shown to be less affected by sample size in this context (Cheung & Resvold, 2002). Furthermore, Root Mean Square Error of Approximation (RMSEA) scores were shown to be resilient to model complexity penalties and, as a result, were given emphasis (Vandenburg & Vance, 2000). For a more detailed analysis of fit statistics, see chapter 3. Unfortunately, no guidelines were apparent within the literature regarding how to compare differences in TLI and CFI scores in relation to different levels of measurement invariance (Vandenberg & Lance 2000). Cheung and Renswood (2002) suggested, as a rule of thumb, however, that CFI and TLI should not fall below commonly accepted guidelines of good fit, namely 0.9, and these recommendations were adopted in this analysis for all fit statistics reported.

If '*configural invariance*' was established, it could be asserted that the structure of the model remained constant over time, however, the characteristics within that model may not. '*Weak invariance*' tests not only if the structure remains constant but also the extent to which factor loadings remained constant over time. If '*weak invariance*' was established, it could be said that not only does the structure of the model remain stable, but the relationship between the factors and the various items did too. Failure to establish '*weak invariance*' would imply that the extent to which various items load onto a factor changes, however the factor that they primarily load onto did not. '*Strong invariance*' tests the extent to which intercepts are constant. A change in intercepts would suggest that participants' average scores were either increasing or decreasing across time, usually down to a global effect outside the scope of analysis. This level of invariance was necessary to conduct analyses of means across time points. '*Strict*' and '*Very Strict*' levels of invariance are rare, and while not absolutely necessary to be established for the purposes of this research, it was felt important to test for them to determine the extent of stability that the model demonstrated over time. '*Strict*' invariance refers to the extent that residuals remain constant over time and should these not remain constant, the degree of variation that participants responses deviate from the best fit line would be subject to change.

Figure 5. 1

A Graphical Representation of Ye's (2009) Model



Note. The arrows indicate the regression based relationships between the factors and their respective items.

* Not all items are included for visual reasons however the presence of three consecutive dots represents items C through K.

** Factor loadings shown in Table 5.2, error variances shown in Table 5.3 and inter-factor correlations shown in Table 5.4.

*** correlated errors were not shown for visual reasons but were present in a manne

Table 5.2

Factor loadings (unstandardized) for the five-factor model shown in figure 5.1.

Factor	*Item*	*Factor loading*	*SE*	*P*
1	A_SCGHQA	1.00	0.00	-
1	A_SCGHQB	1.25	0.02	<0.001
1	A_SCGHQC	0.98	0.02	<0.001
1	A_SCGHQD	0.88	0.01	<0.001
1	A_SCGHQE	1.29	0.02	<0.001
1	A_SCGHQF	1.33	0.02	<0.001
1	A_SCGHQG	1.21	0.02	<0.001
1	A_SCGHQH	1.05	0.02	<0.001
1	A_SCGHQI	1.70	0.03	<0.001
1	A_SCGHQJ	1.63	0.03	<0.001
1	A_SCGHQK	1.36	0.03	<0.001
1	A_SCGHQL	1.20	0.02	<0.001
2	B_SCGHQA	1.00	0.00	-
2	B_SCGHQB	0.98	0.00	<0.001
2	B_SCGHQC	1.00	0.00	<0.001
2	B_SCGHQD	0.99	0.00	<0.001
2	B_SCGHQE	0.99	0.00	<0.001
2	B_SCGHQF	0.97	0.00	<0.001
2	B_SCGHQG	1.00	0.00	<0.001
2	B_SCGHQH	0.99	0.00	<0.001
2	B_SCGHQI	0.97	0.00	<0.001
2	B_SCGHQJ	0.96	0.00	<0.001
2	B_SCGHQK	0.93	0.00	<0.001
2	B_SCGHQL	0.99	0.00	<0.001
3	C_SCGHQA	1.00	0.00	-
3	C_SCGHQB	1.30	0.02	<0.001
3	C_SCGHQC	1.10	0.02	<0.001
3	C_SCGHQD	0.90	0.02	<0.001
3	C_SCGHQE	1.33	0.02	<0.001
3	C_SCGHQF	1.39	0.02	<0.001
3	C_SCGHQG	1.18	0.02	<0.001
3	C_SCGHQH	1.00	0.02	<0.001
3	C_SCGHQI	1.69	0.03	<0.001
3	C_SCGHQJ	1.61	0.03	<0.001
3	C_SCGHQK	1.36	0.03	<0.001
3	C_SCGHQL	1.24	0.02	<0.001
4	D_SCGHQA	1.00	0.00	-
4	D_SCGHQB	1.2	0.02	<0.001
4	D_SCGHQC	1.12	0.02	<0.001
4	D_SCGHQD	0.9	0.02	<0.001
4	D_SCGHQE	1.3	0.02	<0.001
4	D_SCGHQF	1.36	0.02	<0.001

4	D_SCGHQG	1.21	0.02	<0.001
4	D_SCGHQH	1.02	0.02	<0.001
4	D_SCGHQI	1.61	0.02	<0.001
4	D_SCGHQJ	1.57	0.02	<0.001
4	D_SCGHQK	1.34	0.02	<0.001
4	D_SCGHQL	1.23	0.02	<0.001
5	E_SCGHQA	1.00	0.00	-
5	E_SCGHQB	0.97	0.00	<0.001
5	E_SCGHQC	0.99	0.00	<0.001
5	E_SCGHQD	0.99	0.00	<0.001
5	E_SCGHQE	0.99	0.00	<0.001
5	E_SCGHQF	0.97	0.00	<0.001
5	E_SCGHQG	1	0.00	<0.001
5	E_SCGHQH	0.99	0.00	<0.001
5	E_SCGHQI	0.98	0.00	<0.001
5	E_SCGHQJ	0.96	0.00	<0.001
5	E_SCGHQK	0.93	0.00	<0.001
5	E_SCGHQL	0.99	0.00	<0.001

Table 5.3

Error variances for the five-factor model shown in figure 5.1.

Item	*Unstandardized Estimate*	*SE*	*P*
A_SCGHQA	0.15	0.002	<0.001
A_SCGHQB	0.33	0.004	<0.001
A_SCGHQC	0.20	0.003	<0.001
A_SCGHQD	0.13	0.002	<0.001
A_SCGHQE	0.31	0.004	<0.001
A_SCGHQF	0.27	0.004	<0.001
A_SCGHQG	0.19	0.003	<0.001
A_SCGHQH	0.13	0.002	<0.001
A_SCGHQI	0.23	0.003	<0.001
A_SCGHQJ	0.24	0.004	<0.001
A_SCGHQK	0.24	0.003	<0.001
A_SCGHQL	0.16	0.002	<0.001
B_SCGHQA	0.13	0.003	<0.001
B_SCGHQB	0.32	0.004	<0.001
B_SCGHQC	0.19	0.004	<0.001
B_SCGHQD	0.12	0.003	<0.001
B_SCGHQE	0.28	0.004	<0.001
B_SCGHQF	0.24	0.004	<0.001
B_SCGHQG	0.14	0.003	<0.001
B_SCGHQH	0.11	0.002	<0.001
B_SCGHQI	0.21	0.003	<0.001
B_SCGHQJ	0.20	0.003	<0.001

B_SCGHQK	0.22	0.003	<0.001
B_SCGHQL	0.18	0.004	<0.001
C_SCGHQA	0.17	0.003	<0.001
C_SCGHQB	0.37	0.004	<0.001
C_SCGHQC	0.20	0.004	<0.001
C_SCGHQD	0.13	0.002	<0.001
C_SCGHQE	0.32	0.004	<0.001
C_SCGHQF	0.27	0.004	<0.001
C_SCGHQG	0.16	0.003	<0.001
C_SCGHQH	0.12	0.002	<0.001
C_SCGHQI	0.22	0.003	<0.001
C_SCGHQJ	0.24	0.003	<0.001
C_SCGHQK	0.24	0.003	<0.001
C_SCGHQL	0.16	0.003	<0.001
D_SCGHQA	0.17	0.003	<0.001
D_SCGHQB	0.35	0.004	<0.001
D_SCGHQC	0.20	0.004	<0.001
D_SCGHQD	0.13	0.002	<0.001
D_SCGHQE	0.31	0.004	<0.001
D_SCGHQF	0.25	0.004	<0.001
D_SCGHQG	0.15	0.003	<0.001
D_SCGHQH	0.12	0.002	0.00
D_SCGHQI	0.22	0.003	0.00
D_SCGHQJ	0.24	0.004	0.00
D_SCGHQK	0.24	0.003	0.00
D_SCGHQL	0.16	0.003	0.00
E_SCGHQA	0.15	0.003	0.00
E_SCGHQB	0.32	0.004	0.00
E_SCGHQC	0.18	0.003	0.00
E_SCGHQD	0.12	0.002	0.00
E_SCGHQE	0.28	0.004	0.00
E_SCGHQF	0.23	0.004	0.00
E_SCGHQG	0.14	0.003	0.00
E_SCGHQH	0.11	0.002	0.00
E_SCGHQI	0.19	0.003	0.00
E_SCGHQJ	0.20	0.003	0.00
E_SCGHQK	0.22	0.003	0.00
E_SCGHQL	0.15	0.003	0.00

Table 5.4

Inter-factor correlations for the five-factor model shown in figure 5.1.

		1	2	3	4
1.	Factor 1	-			
2.	Factor 2	.06**	-		

3.	Factor 3	.44**	.46**	-	
4.	Factor 4	.42**	.38**	.52**	-
5.	Factor 5	.02	.04**	.04**	.05**

* $p < 0.05$, ** $p < 0.01$

5.3.2.2- Modification Indices

When investigating longitudinal structural equation models, a number of inter-item, phenomena such as '*retest effects*' may become apparent and must be accounted for. Modification indices refer to a command available in MPLUS which investigates how correlating errors within a model can improve overall fit statistics. It achieves this by displaying the reduction in Chi-Squared values if two variables were correlated. Correlated errors have already been used in Chapter 3 to simulate model effects relating to the wording of the items (Hankins, 2008). While it was intended to use this method in a different way than Hankins (2008) did, it was considered prudent to investigate the effects of correlating errors across items. This was conducted in a '*configural invariance*' test, as this was the most unconstrained model.

It is important to note that while modification indices provide a method for improving the fit of a model, Kenny (2011) warns that this method should be used sparingly. He warns that one must only correlate errors on items where there is a specific rationale for doing so and that if errors are correlated, it is important that any other items which fulfil this rationale must also be correlated. Herminda (2015) went further and stated that the application of correlated errors on a post hoc basis fundamentally changed the nature of research from confirmatory to exploratory. This paper argued that there was no justification for using modification indices to improve model fit on a post hoc basis and that doing so may mask the underlying structure of the data. This analysis used modification indices to investigate if a discernible pattern of model fit reduction was present and not to try and improve fit statistics results for the reasons outlined in Chapter 4. It is accepted that by using modification indices, this portion of the research does adopt hallmarks of exploratory analysis, however, this is mitigated by the novel way that the analysis is used. Modification indices were used as an indicator of poor fit, rather than as a tool to alter the model. As the criticisms of this technique referred to altering models, it was felt that a large number of the criticisms of Herminda (2015) did not apply. The use of modification indices in this way was not noticed in the literature previously and may provide an alternative method of their use that avoids the pitfalls of post hoc application, while also providing useful information on inter-item interactions.

5.4- Results

5.4.1- CFA

Table 5.5

Fit Statistics for Different Levels of Measurement Invariance in Ye's (2008) Proposed Factor Structure of the GHQ-12

	Configural	Weak	Strong	Strict
Df	1640	1704	1752	1800
Chi-square	60603.369	57246.949	73847.743	74268.845
P	<0.001	<0.001	<0.001	<0.001
RMSEA	0.023	0.024	0.025	0.025
90% CI	0.023-0.024	0.024-0.024	0.025-0.025	0.025-0.025
CFI	0.973	0.970	0.967	0.967
TLI	0.971	0.968	0.966	0.967
SRMR	0.020	0.047	0.068	0.068

Configural – No parameters constrained

Weak – Factor loadings

Strong- Factor loadings and intercepts

Strict – Factor loadings, intercepts and residuals

Very strict - Factor loadings, intercepts and residuals and factor variances

90% CI= confidence intervals at 90%

RMSEA= Root Mean Square Error of Approximation

CFI= Comparative Fit Index

TLI= Tucker Lewis Index

SRMR= Standardised Root Mean Square Residual

*Values in bold represent the highest

Table 5.5 shows the various fit statistics generated from the different levels of invariance testing. The results show that the model was successfully run at all stages, with the exception of '*very strict*' when data failed to converge. The results suggest that '*strict invariance*' was demonstrated within the data for this model.

Given that Cheung and Resvold (2002) suggested that CFI and TLI were the most appropriate metrics to use in this setting, these were given the greatest emphasis in analysis and were accordingly reported first, with fit statistics reported in order of emphasis placed.

At all stages of the analysis, neither CFI nor TLI scores dropped below the commonly accepted levels of good fit suggested in Renswood (2002) of 0.9 nor failed to exceed the more stringent cut off of 0.95. Differences between the '*strict'* and '*configural*' test were minimal with fit ranging from 0.973 and 0.971 in the CFI and TLI respectively in a '*configural*' test to 0.967 for both tests in the '*strict*' test.

RMSEA scores were relatively stable, with scores ranging from 0.023 to 0.025 between ‘*configural*’ and ‘*strict*’ invariance tests respectively, suggesting that according to this statistic, the model remains invariant over time.

Other fit statistics do suggest some variation in model fit depending on the condition investigated. SRMR increased from 0.02 to 0.068 from ‘*configural*’ to ‘*strict*’, while chi-squared fluctuated, with ‘*weak*’ invariance actually demonstrating lower chi-squared results when compared with ‘*configural*’, but ‘*strong*’ and ‘*strict*’ results demonstrating higher results than both ‘*configural*’ and ‘*weak*’ tests.

5.4.2- Modification Indices

Table 5.6- Modification Indices for Yee’s (2008) Model on a Configural Measurement Invariance CFA

Variable relationships			Modification effect	Expected parameter change	Standardised Expected parameter change	Standardised on XY axis EPC
WAVE 2 ITEM 11	WITH	WAVE 2 ITEM 10	2360.787	0.086	0.086	0.412
WAVE 3 ITEM 11	WITH	WAVE 3 ITEM 10	1799.714	0.088	0.088	0.367
WAVE 2 ITEM 12	WITH	Factor F2A	1784.704	0.04	0.127	0.299
WAVE 5 ITEM 11	WITH	WAVE 5 ITEM 10	1735.474	0.078	0.078	0.38
WAVE 4 ITEM 11	WITH	WAVE 4 ITEM 10	1695.543	0.087	0.087	0.366
WAVE 1 ITEM 11	WITH	WAVE 1 ITEM 10	1641.988	0.084	0.084	0.347
WAVE 5 ITEM 11	WITH	WAVE 4 ITEM 11	1451.907	0.075	0.075	0.329
WAVE 4 ITEM 11	WITH	WAVE 3 ITEM 11	1317.768	0.075	0.075	0.315
WAVE 2 ITEM 12	WITH	WAVE 2 ITEM 9	1041.192	0.048	0.048	0.247

WAVE 2 ITEM 6	WITH	WAVE 2 ITEM 5	1008.09	0.059	0.059	0.231
WAVE 3 ITEM 11	WITH	WAVE 2 ITEM 11	1007.488	0.06	0.06	0.261
WAVE 5 ITEM 2	WITH	WAVE 4 ITEM 2	974.317	0.089	0.089	0.264
WAVE 2 ITEM 11	WITH	WAVE 1 ITEM 11	950.984	0.061	0.061	0.264
WAVE 2 ITEM 11	WITH	WAVE 2 ITEM 5	936.425	-0.056	-0.056	-0.227
WAVE 4 ITEM 2	WITH	WAVE 3 ITEM 2	929.264	0.093	0.093	0.261
WAVE 5 ITEM 11	WITH	WAVE 3 ITEM 11	921.841	0.061	0.061	0.267
WAVE 3 ITEM 2	WITH	WAVE 2 ITEM 2	892.917	0.081	0.081	0.239
WAVE 4 ITEM 5	WITH	WAVE 3 ITEM 5	823.216	0.079	0.079	0.251
WAVE 2 ITEM 10	WITH	WAVE 2 ITEM 5	808.207	-0.055	-0.055	-0.23
WAVE 2 ITEM 2	WITH	WAVE 1 ITEM 2	768.683	0.075	0.075	0.232
WAVE 2 ITEM 5	WITH	WAVE 2 ITEM 2	761.283	0.059	0.059	0.197
WAVE 4 ITEM 4	WITH	WAVE 4 ITEM 3	740.898	0.037	0.037	0.23
WAVE 3 ITEM 5	WITH	WAVE 2 ITEM 5	736.112	0.066	0.066	0.221
WAVE 4 ITEM 11	WITH	WAVE 2 ITEM 11	735.313	0.053	0.053	0.232
WAVE 5 ITEM 4	WITH	WAVE 5 ITEM 3	731.532	0.034	0.034	0.236
WAVE 5 ITEM 11	WITH	WAVE 5 ITEM 5	730.289	-0.056	-0.056	-0.226
WAVE 5 ITEM 5	WITH	WAVE 4 ITEM 5	722.46	0.068	0.068	0.232
WAVE 3 ITEM 11	WITH	WAVE 3 ITEM 5	713.86	-0.061	-0.061	-0.222

WAVE 3 ITEM 11	WITH	WAVE 1 ITEM 11	695.52	0.062	0.062	0.258
WAVE 4 ITEM 11	WITH	WAVE 4 ITEM 5	694.519	-0.061	-0.061	-0.224
WAVE 1 ITEM 11	WITH	WAVE 1 ITEM 5	688.494	-0.059	-0.059	-0.215
WAVE 2 ITEM 5	WITH	WAVE 1 ITEM 5	636.885	0.063	0.063	0.215
WAVE 5 ITEM 6	WITH	WAVE 5 ITEM 5	629.128	0.054	0.054	0.21
WAVE 5 ITEM 11	WITH	WAVE 2 ITEM 11	628.577	0.046	0.046	0.213
WAVE 5 ITEM 2	WITH	WAVE 3 ITEM 2	626.708	0.074	0.074	0.216
WAVE 4 ITEM 2	WITH	WAVE 2 ITEM 2	620.092	0.069	0.069	0.207
WAVE 5 ITEM 5	WITH	WAVE 3 ITEM 5	608.607	0.066	0.066	0.217
WAVE 1 ITEM 5	WITH	WAVE 1 ITEM 2	598.579	0.065	0.065	0.203
WAVE 4 ITEM 10	WITH	WAVE 3 ITEM 10	592.31	0.054	0.054	0.225
WAVE 5 ITEM 5	WITH	WAVE 5 ITEM 2	565.204	0.059	0.059	0.194
WAVE 4 ITEM 11	WITH	WAVE 1 ITEM 11	564.953	0.059	0.059	0.242
WAVE 2 ITEM 12	WITH	WAVE 2 ITEM 11	554.724	0.034	0.034	0.171
WAVE 2 ITEM 11	WITH	WAVE 2 ITEM 2	545.439	-0.045	-0.045	-0.171
WAVE 3 ITEM 5	WITH	WAVE 3 ITEM 2	524.127	0.065	0.065	0.19
WAVE 5 ITEM 10	WITH	WAVE 4 ITEM 10	517.578	0.046	0.046	0.212
WAVE 5 ITEM 10	WITH	WAVE 5 ITEM 5	515.613	-0.049	-0.049	-0.206

WAVE 1 ITEM 10	WITH	WAVE 1 ITEM 5	513.677	-0.055	-0.055	-0.202
WAVE 4 ITEM 5	WITH	WAVE 2 ITEM 5	512.111	0.056	0.056	0.191
WAVE 5 ITEM 12	WITH	Factor F5A	500.676	0.024	0.069	0.177
WAVE 3 ITEM 10	WITH	WAVE 3 ITEM 5	499.621	-0.055	-0.055	-0.197
WAVE 2 ITEM 10	WITH	WAVE 2 ITEM 2	483.237	-0.044	-0.044	-0.175
WAVE 3 ITEM 4	WITH	WAVE 3 ITEM 3	478.106	0.029	0.029	0.18
WAVE 5 ITEM 11	WITH	WAVE 1 ITEM 11	469.141	0.05	0.05	0.219
WAVE 3 ITEM 2	WITH	WAVE 1 ITEM 2	466.222	0.073	0.073	0.21
WAVE 5 ITEM 2	WITH	WAVE 2 ITEM 2	461.69	0.056	0.056	0.176
WAVE 5 ITEM 10	WITH	WAVE 3 ITEM 10	461.054	0.044	0.044	0.204
WAVE 4 ITEM 6	WITH	WAVE 4 ITEM 5	455.641	0.052	0.052	0.185
WAVE 4 ITEM 2	WITH	WAVE 1 ITEM 2	453.622	0.073	0.073	0.216
WAVE 4 ITEM 10	WITH	WAVE 4 ITEM 5	450.304	-0.052	-0.052	-0.192
WAVE 5 ITEM 12	WITH	WAVE 5 ITEM 9	447.613	0.032	0.032	0.183

Note. Factors are denoted by factor 1A. The numerical designation denotes the wave that the factor was generated from and factor A refers to the general mental health factor, whereas B refers to the method factor

Table 5.6 shows the top 60 relationships that would have the greatest effect on chi-square values if they were correlated. This table was then interrogated to determine any patterns of relationships which were contributing to poor fit in the longitudinal model. The reductions in chi-square were placed on a graph to determine the overall pattern of model fit reduction and shown below.

Figure 5. 2

A graph showing the magnitude of chi-square reduction for the top 60 relationships shown in table 5.3

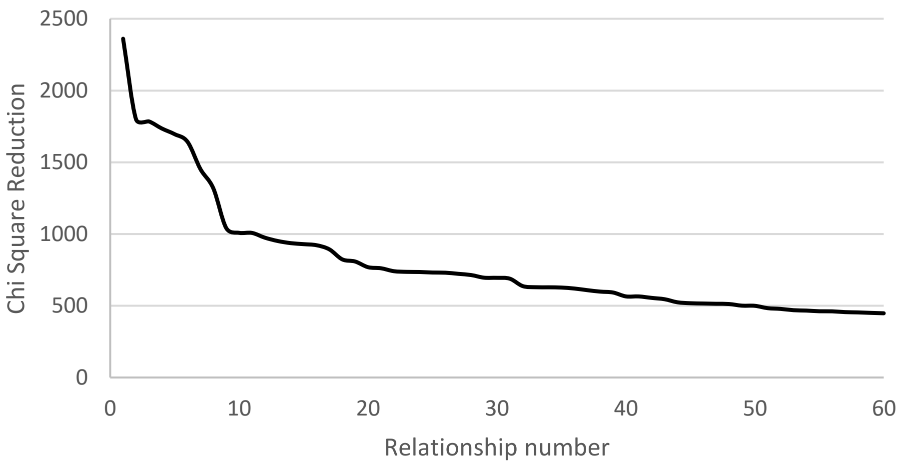


Figure 5.2 shows that the pattern of chi-square reduction was exponential in nature and that a small number of the most ill-fitting relationships contributed disproportionately to poor fit. The magnitude of chi-square reduction plateaus after the first nine relationships and thereafter, the degree of chi-squared reduction between items reduced at a much slower pace.

The relationship between items 10 and 11 in each wave appeared to have abnormally large Chi-squared reduction values. Of the 60 relationships displayed, 5 of the top 6 consisted of these relationships with Chi-squared reductions ranging from 2360.787 to 1641.988. These were noticeably inflated in comparison to the other relationships shown in this table.

The inter wave relationships between item 11 featured prominently in the table. These items occupied positions 7, 8, 11, and 14 and 58. There was, however, a wide range of chi-squared reductions within these relationships with the effect of correlating waves 5 and 4 resulting in 1451.907 reductions in chi-squared results, however, correlating waves 1 and 5 only reduced chi-square by 469.141.

Overall, with the exception of items 10 and 11, it was decided that there was no discernible pattern of poor fitting relationships present in the longitudinal data. Consequently, the relationship between 10 and 11 would be investigated in the discussion section to determine if there was a reason for such disproportionately poor fit.

5.5- Discussion

As the analysis into factorial invariance was divided into two parts throughout the course of this chapter the discussion of the two analyses will be conducted separately, and a more general discussion about the chapter will follow, including research and clinical implications and an overall summary of the chapter’s findings.

5.5.1- CFA

The interpretation of fit statistics in relation to factorial invariance was particularly challenging because as previously mentioned, there is no consensus between researchers as to which of the fit statistics reported is superior (Thompson & Daniel, 1996). Cheung and Resvold (2002) evaluated the use of a number of fit statistics within a measurement invariance context and tentatively suggested that the CFI was the most appropriate test to apply greater weight to when conducting an analysis of this type. He claimed, contrary to other research, that CFI tests were relatively robust against misleading results as a function of sample size and complexity. It was also noted in the literature that RMSEA scores should decrease with large sample sizes and may be artificially distorted in models with low degrees of freedom (Kenny, Kaniskan & McCoach 2014). Due to the lack of consensus, all fit statistics were reported, however, in line with Cheung and Resvold's (2002) findings, greater emphasis was placed upon CFI results, and during the results, section results were reported in order of the emphasis placed upon them.

The model was successfully run on all conditions with the exception of '*very strict'*, which was not completed due to a failure of the data converging. This suggested that factor variances varied to a significant degree between the model at various time points.

CFI and TLI scores were found to be relatively stable across all conditions, with little change between the '*configural'* and '*strict*' constraints. Fit statistics exceed the guidelines laid down in Kenny (2015), which stated that CFI and TLI scores should not fall below 0.9.

CFI and TLI scores inflict a penalty on models for each parameter estimated (Kenny, 2015), however, the exact nature of these penalties are different and are detailed in appendix 1. Due to the complex nature of this model (see figure 5.1), it is reasonable to state that this complexity would have an impact on performance. The results indicating good fit are particularly important given the complexity penalties that such a complex model would inevitably attract. Given that Cheung and Renswold (2000) suggested that these fit statistics were particularly appropriate within this context, these findings would support the hypothesis that measurement invariance was present within this data. Conversely, it must also be noted that research by Taguma (2001) showed that CFI, while described as relatively stable, was shown to alter with sample size. It is also important to note that Taguma's research was conducted on sample sizes ranging from 50 to 1000, this analysis comprises 65568 participants. It is, therefore, reasonable to assert that any instability relating to sample size uncovered by Taguma would be amplified in this abnormally large sample.

RMSEA scores remained stable, with only minimal differences between weak and strict conditions. RMSEA scores have been described as the most informative of the numerous fit statistics (Diamantopoulos & Siguaw, 2000); however, they generally tend to reward parsimony over model fit. Given the complexity of the model (see figure 5.1), such a strong performance was viewed positively. SRMR results did vary to a larger degree, however, were consistently below the suggested cut-off of <0.08 as proposed in Hu & Bentler (1999).

Overall, it was decided that a '*strong*' level of measurement invariance was present in this model over time, and as a result, a full range of longitudinal analysis could be conducted in later chapters.

5.5.2- Modification Indices

The use of correlated errors has been criticised in the literature for being used simply to improve model fit, and it has been claimed that generated models may have no theoretical or conceptual basis. (Herminda, 2015). While Kenny (2015) suggested a number of safeguards which would ensure that models which contained correlated errors remained meaningful, the use of correlated errors is not looked upon favourably within the literature. In this instance, however, correlated errors were used for the purposes of identifying poor-fitting relationships between items and factors, not to improve model fit on a post hoc basis and as such, it was felt that the criticisms directed at the technique did not fully apply.

From the analysis, it became apparent that a small number of relationships in the data accounted for a disproportionate amount of 'poor fit' within the model. Of these high scoring relationships, the interaction between item 10 and 11 was frequently present and accounted for a disproportionate amount of poor fit in the model than other relationships. No other relationship was deemed to have a significant impact on the fit of the model. The interaction of items 10 and 11 was unexpected as, when one looks at the content of these items, item 10 refers to self-confidence, and 11 refers to feelings of worthlessness (see table 3.1), which one would expect to be highly correlated. These items were viewed by Graetz (1991) as measuring the same factor, that of 'loss confidence' and therefore it is implicit that Graetz (1991) viewed them as contributing to a similar latent variable. The items were inversely worded, and it was not clear from the literature as to why this poor fit would be apparent. With the notable exception of items 10 and 11, given the lack of coherent patterns of the modification indices, it was concluded that inter-item relationships within the model were sufficiently random as not to highlight problems with the model, which would cause difficulties in later longitudinal analyses.

5.5.3- Research implications

The research implications for this chapter relate to the stability of Ye's (2009) model over time. At the time of writing, Ye's (2009) model had not been tested in relation to its invariance, and this chapter's analysis have shown it to be stable over time. Should measurement invariance not have been established or only to a partial degree, then longitudinal research may not have been possible or certain aspects of the GHQ-12 may not have been comparable over time. For example, if only Configural invariance was established, then the scores between waves could not be directly compared (Geiser, 2012). As strict measurement invariance was demonstrated, it could be said that this dimensional representation remained stable enough to facilitate longitudinal analysis. Further research could be conducted into what causes the relationship between items 10 and 11 to be so incongruous with the model and what causes the large chi-square implications of relationships between certain items.

5.5.4- Clinical implications

Within a clinical setting, the demonstration of strict measurement invariance suggested that the GHQ-12 was resistant to the retest effect when conducted annually. This was consistent with the established literature (Pevalin, 2000). This allows the GHQ-12 to be utilised multiple times on the same participant without experiencing a steady increase or decrease in scores through a practice or fatigue effect.

5.6.5- Limitations

In terms of limitations, this research was considered methodologically robust. Statistically, the limitations of CFA as laid down in Kenny (2005) were addressed in appendix 1, and it was deemed as an appropriate methodological framework for the analysis in question. While the statistical techniques provided an objective method of analysing fit, the interpretation of them was much more subjective. As mentioned earlier, there are no established rules for measuring the difference between fit statistics to determine if they are statistically significant (Vandenberg and Lance 2000), therefore an element of subjectivity is introduced when comparing the CFI, TLI RMSEA and SRMR between the various groups. While rules of thumb have been used from the literature, these are undeniably crude and do not capture the change between groups but simply place an arbitrary baseline for which certain fit statistics should not fail to exceed.

5.6- Summary

In conclusion, '*strict*' measurement invariance was established in Ye's (2009) model in the population of UKHLS respondents. This level of invariance permitted a full range of longitudinal statistical analysis to be conducted on the model in subsequent analyses. The analysis showed, not only that the factor structure remained relatively stable over time, but so too did the factor loadings, and the means of participants' responses. While '*very strict*' invariance was not established, and that may imply small changes in how participants respond over time, it was not so great as to limit the range of longitudinal analysis that could be conducted on the data.

Finally, the two-stage analysis has shown that, Ye's (2009) model has demonstrated stability in relation to the entirety of the model and that, inter-item relationships were not judged to be problematic. As later chapters depended on the degree of measurement invariance that exists in the model, the fact that '*strict'* invariance was found allowed unhindered analysis in later chapters (Geiser, 2012).

References

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual differences*, *42*(5), 815-824.

Bialosiewicz, S., Murphy, K., & Berry, T. (2013). Do our measures measure up? The critical role of measurement invariance. *Demonstration Session at American Evaluation Association*.

Bun Cheung, Y. (2002). A confirmatory factor analysis of the 12-item General Health Questionnaire among older people. *International journal of geriatric psychiatry*, *17*(8), 739-744.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equation modeling. Journal of Cross-Cultural Psychology, 31, 187–212.

del Pilar Sánchez-López, M., & Dresch, V. (2008). The 12-Item General Health Questionnaire (GHQ-12): reliability, external validity and factor structure in the Spanish population. *Psicothema*, *20*(4), 839-843.

Diamantopoulos, A., Siguaw, J. A., & Siguaw, J. A. (2000). *Introducing LISREL: A guide for the uninitiated*. Sage.

Fernandes, H. M., & Vasconcelos-Raposo, J. (2013). Factorial validity and invariance of the GHQ-12 among clinical and non-clinical samples. *Assessment*, *20*(2), 219-229.

Fumagalli, L., Knies, G., & Buck, N. (2017). Understanding Society, The UK Household Longitudinal Study, Harmonised British Household Panel Survey (BHPS) User Guide.

Geiser, C. (2012). *Data analysis with Mplus*. Guilford press.

Graetz, B. (1991). Multidimensional properties of the general health questionnaire. *Social psychiatry and psychiatric epidemiology*, *26*(3), 132-138.

Hammarström, A., Westerlund, H., Kirves, K., Nygren, K., Virtanen, P., & Hägglöf, B. (2016). Addressing challenges of validity and internal consistency of mental health measures in a 27-year longitudinal cohort study–the Northern Swedish Cohort study. *BMC medical research methodology*, *16*(1), 4.

Hankins, M. (2008). The factor structure of the twelve item General Health Questionnaire (GHQ-12): the result of negative phrasing?. *Clinical Practice and Epidemiology in Mental Health*, *4*(1), 10.

Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! testing! one, two, three–Testing the theory in structural equation models!. *Personality and Individual Differences*, *42*(5), 841-850.

Hermida, R. (2015). The problem of allowing correlated errors in structural equation modeling: concerns and considerations. *Computational Methods in Social Sciences*, *3*(1), 5-17.

Kelloway, E. K. (1995). Structural equation modelling in perspective. Journal of Organizational Behaviour, 16, 215–224

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2014). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, in press.

Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological methods*, *7*(2), 228.

Mäkikangas, A., Feldt, T., Kinnunen, U., Tolvanen, A., Kinnunen, M. L., & Pulkkinen, L. (2006). The factor structure and factorial invariance of the 12-item General Health Questionnaire (GHQ-12) across time: evidence from two community-based samples. *Psychological Assessment*, *18*(4), 444.

Meade, A. W., & Wright, N. A. (2012). Solving the measurement invariance anchor item problem in item response theory. *Journal of Applied Psychology*, *97*(5), 1016.

Muthén, L. K., & Muthén, B. O. (1998-2012). Mplus User's Guide: Statistical Analysis with Latent Variables (7th ed.). Los Angeles, CA: Muthén & Muthén

Pevalin, D. J. (2000). Multiple applications of the GHQ-12 in a general population sample: an investigation of long-term retest effects. *Social psychiatry and psychiatric epidemiology*, *35*(11), 508-512.

Shevlin, M., & Adamson, G. (2005). Alternative factor models and factorial invariance of the GHQ-12: a large sample analysis using confirmatory factor analysis. *Psychological assessment*, *17*(2), 231.

Smith, A. B., Oluboyede, Y., West, R., Hewison, J., & House, A. O. (2013). The factor structure of the GHQ-12: the interaction between item phrasing, variance and levels of distress. *Quality of Life Research*, *22*(1), 145-152.

Tanguma, J. (2001). Effects of sample size on the distribution of selected fit indices: A graphical approach. *Educational and Psychological Measurement*, *61*(5), 759-776.

Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines.

Werner, C., & Schermelleh-Engel, K. (2010). Deciding between competing models: Chi-square difference tests. *Goethe University. Available online: https://perma. cc/2RTR-8XPZ (accessed on 21 July 2017)*.

Widaman , K. F. , & Reise, S . P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain . In K .] . Bryant , M. Windle, &: S. G. West (Eds.), The science of prevention: Methodological advances from alcohol and substance abuse research (pp. 281-324).Washington, DC: American Psychological Association .

Ye, S. (2009). Factor structure of the General Health Questionnaire (GHQ-12): The role of wording effects. *Personality and Individual Differences*, *46*(2), 197-201.

Chapter 6 - Longitudinal Heterogeneity- Investigating Trajectories of Mental Health Longitudinally

6.1- Abstract

Introduction

Due to ‘strict’ measurement invariance demonstrated in Chapter 5, Ye’s (2009) dimensional representation of the GHQ-12 data was shown to be appropriate for longitudinal analysis. The data was further analysed in this chapter to ascertain whether different sub-populations of participant exhibited different trajectories of GHQ-12 scores over time using growth mixture modelling techniques.

Methods

A growth mixture model was conducted on waves 1-5 of the Understanding Society data to determine if sub-populations were present. Missing data were accounted for using the MLR technique. Generated fit statistics were compared to determine which class solution was the most appropriate representation of the data.

Results

The results suggested that both a four and five class solution would constitute an appropriate fit for the data. The four-class solution was comprised of classes which represented low stable scores, steadily increasing scores, steadily decreasing scores and high stable scores. The five-class solution was similar, with an additional class which represented participants with stable scores that were lower than the low, stable group.

Discussion

It was demonstrated that a number of subpopulations were present in the data and that a 4-class solution was the most appropriate as it provided a more parsimonious, meaningful and interpretable solution.

6.2- Introduction

As factorial invariance was established in Chapter 5, it was possible to conduct longitudinal analysis. The 'strict' nature of invariance permitted unimpeded longitudinal analysis to be conducted. An explanation of the difficulties of longitudinal analyses in models which fail to demonstrate measurement invariance is provided in the previous chapter.

Longitudinal analysis consisted in the first instance of a form of latent class modelling, known as a growth mixture model (GMM) to identify if the data were homogeneous or heterogeneous. More information on this statistical technique is given in the methods section. Homogeneity refers to whether participants within a dataset behave uniformly and heterogeneity refers to whether they have a number of sub-populations which behave differently. By performing this analysis, it was possible to identify if these sub-populations exist and if so, how they behave and what characterises those within them.

6.2.1- Literature Review

In relation to mental health, it is well understood in the literature that different groups of people behave in different ways, especially when a longitudinal aspect is introduced. For example, when investigating demographic characteristics, gender and age have been found to have a complex relationship with mental health over time. Afifi (2008) conducted research into gender differences over time and found that gender differences were much more complex than simply comparing incident rates between genders. He highlighted that specific conditions had periods of time when they were particularly prevalent throughout an individual’s development, such as the increased incidence of ‘conduct disorder’ being three times as likely in young males than similarly aged females, with this gap closing in later life (Scott, 1998). During adolescence, females are more likely to suffer from depression (Parker and Roy, 2001), adolescent males are much more likely to commit suicide (Hawton et al., 2002) which could suggest increased severity of depressive symptoms or decreased resilience in boys and this pattern continues into adulthood (Hawton et al., 2002).

One way to investigate subpopulations is through techniques which seek to separate the population being investigated into groups based on characteristics not observed at the time of data collection, known as classes. This analysis can be conducted either cross-sectionally or longitudinally, and while this chapter’s analysis investigated the longitudinal aspects of the data, a number of relevant studies which use cross-sectional data are discussed below. Both longitudinal and cross-sectional studies have generated a number of classes of participants based on the pattern of the participant’s responses to mental health questionnaires. Studies which use latent class modelling techniques on GHQ scores are limited, but a number of papers were identified and are detailed below.

Chronologically, the first study to use latent class modelling techniques on GHQ-12 scores, was conducted in the very specific population of individuals who had experienced a natural disaster (Høyer Holgersen et al., 2011). This study investigated how GHQ-12 scores of survivors of natural disasters would change over time from the date of the disaster. Data was collected 5 and 25 years after the incident and researchers identified a four-class solution as the most appropriate representation of the data.

Consequently, these researchers suggested that there were four distinct patterns of behaviour after the disaster.

The extracted classes consisted of a low, stable group which consisted of approximately 61% of the sample and three relatively equally proportioned classes of approximately 14% of the sample. These classes consisted of an increasing trajectory, a decreasing trajectory and a high stable group. This research was relatively niche, as it was not conducted on a general population sample, and was collected over a considerably longer period of time than was available in Understanding Society. This study was relevant as, despite the differences mentioned, the specific trajectories and proportions of these classes could be used as a comparison to the general population analysis that will be conducted in this chapter.

A study using a more representative sample than above was conducted in Northern Ireland (Mahedy et al., 2013) and utilised 5000 randomly selected participants from the general Northern Irish population. This analysis was cross-sectional in nature and had specific research aims relating to the period of time known as 'the troubles' which are unique to the Northern Irish population. This study identified five classes, which were described as Neurotic-depressed, high, medium and low risk and finally a reference group. While using a different methodological framework to what is proposed in this chapter, i.e. using a cross-sectional approach, it also suggested that various subpopulations existed when investigating GHQ-12 scores in a sub-section of the UK population.

Funderbunk et al. (2008) utilised LCA to investigate risk factors for patients who used primary care services in America. This analysis aimed to identify subgroups of individuals based on a wide range of risk factors such as psychological distress as measured by the GHQ-12 (Goldberg, 1978), the Alcohol Use Disorder Identification Test (Dawson et al., 2005), Posttraumatic Stress Disorder Screen- Primary Care test (Prins et al., 2003) and BMI. Importantly this analysis attempted to identify classes based on the probability of the above tests, not on the likelihood of displaying health distress. The analysis uncovered three classes of individuals who were likely to use primary care. The classes demonstrated that two of the three classes of primary care users were likely to exhibit numerous risk factors simultaneously and that frequently risk factors aligned to mental health such as the GHQ-12 were reported alongside physical risk factors such as smoking and BMI. As with other analyses mentioned above, this research was cross-sectional in nature and used LCA in such a way as to identify clusters of symptoms rather than trajectories. It did however demonstrate that use of primary care was predisposed by a number of clusters of covariates and that these clusters were not universal for all users. It also highlighted how GHQ-12 scores could be associated with other predictors of primary care usage.

Jamali and Ayatollahi (2015) used LCA in a novel way to investigate mental health in 771 Iranian nurses, selected using multi-stage cluster sampling. This research identified a two-class solution based on interpretability. They identified these classes as representing the presence of mental disorders and used the class structures to determine appropriate cut-offs for GHQ-12. In this study, the class structure was relatively simplistic and was chosen for a specific purpose, however, it does demonstrate a novel way that LCA can be used in mental health research.

In summary, the literature surrounding this chapter was scarce and all studies identified had significant differences to the research methods employed in this chapter. Generally speaking, however, the literature suggested that within the populations they investigated, researchers were able to identify subgroups within the population.

6.3- Methods

6.3.1- Statistical Techniques

As the scope of the above literature review was very narrow, the range of statistical techniques employed by the studies mentioned was limited. While an in-depth description of Growth Mixture Modelling (GMM) specifically, is provided in the methods section, this section details the similarities and differences between the statistical techniques used in studies mentioned in the above review, rather than an in-depth analysis of each technique.

Studies in the review above mentioned three statistical techniques. Latent Class Analysis (LCA), Latent Profile Analysis (LPA) and Growth Mixture Modelling (GMM), all of which fall under the umbrella term of Latent Variable Mixture Modelling. Terminology around this type of analysis can be confusing as a different name for LPA is a “gaussian (finite) mixture model”. LCA is also referred to as “binomial (finite) mixture models” (Obserki, 2006). Furthermore, Latent Class Growth Modelling is a term for a specific type of statistical analysis which falls under the umbrella definition of Growth Mixture Modelling (Jung & Wickrama, 2008) based on whether the researcher permits variability within the latent classes.

While the terminology surrounding these techniques can be confusing, there are significant differences and similarities between the various techniques which are detailed below.

In terms of similarity, all the techniques above have been described as the art of unscrambling eggs (Oberski, 2016). They estimate one or numerous multinomial latent variables, and using this, they assign participants to a finite number of classes which are exhaustive and mutually exclusive (Nylund-Gibson, Grimm & Masyn, 2019).

Importantly these techniques separate the participants into discrete subgroups which were not identified when the data was collected.

While LCA and LPA are relatively similar, they differ based upon the type of data that are used as indicators. LPA divides heterogenous participants responses into relatively homogeneous subgroups, based on the participant's responses to continuous variables whereas LCA utilises categorical indicators (Berlin, Williams and Parra, 2014). This difference will allow LPA analyses to be unconstrained by the limitations of categorical data, in that categorical data is cruder than continuous data. Growth mixture models are described in the Methods section, and therefore, in the interests of avoiding duplication, they were not detailed in this section.

6.3.2 Data

The analysis was conducted using participants of waves 1 – 5 of Understanding Society. For a more detailed analysis of this merged dataset refer to Chapter 3 '*methods*' section. Participant's GHQ-12 scores have been converted to standardised measures, known as F-scores and as previously mentioned, this process removed participants who did not answer any of the GHQ-12 questions reducing the original 104814 participants to 65568. Participants have been weighted, clustered and stratified to ensure that they remain representative of the general population using the UKHLS 'adult self-completion' longitudinal weighting variable, which was computed in the previous chapter. Proxy responses were not allowed for the GHQ-12 responses, which contributed to the high levels of attrition suffered by participants in this process.

6.3.3- Analysis

Data were analysed to identify subpopulations within the data. These sub-populations will be referred to as classes. By doing this, it was possible to determine if the data were hetero or homogeneous. If data were found to be heterogeneous, it would have been necessary to treat these subgroups differently and to conduct analyses accordingly. Most analyses that investigate class differences compare the difference between various classes and a reference class. This will be explained in later chapters, and, for the purposes of this chapter, a reference group is the class which all other classes will be compared against.

The heterogeneity of data was determined by conducting a growth mixture model in MPLUS (see below). Through this analysis, MPLUS generated fit statistics which were compared in order to ascertain if there were distinct sub-populations within the database and if so, how many were present. The fit statistics generated and suggested interpretation are detailed in appendix 1. Missing data will be managed using MLR, which is detailed in Chapter 3. In order to make interpretation easier, it was decided to impose that all class trajectories would be linear.

6.3.3.1- Growth Mixture Modelling (GMM)

This form of modelling is used to identify multiple subpopulations in a dataset. (Grimm & Ram, 2013). It provides extra utility over its predecessor, growth curve modelling, which previously provided a way of investigating, changes between and within participant's behaviour (Bryk & Raudenbush, 1987). GMM, however, allows for simultaneous modelling of change among multiple populations, not specified prior to data collection. This approach allows researchers to determine how these groups behave over time. Importantly, it does not require prior knowledge of group membership and was described by Ram and Grim (2013) as providing "a framework for *post-hoc* identification and description of group differences in change". Within an MPLUS framework, the process of identifying the appropriate number of classes is conducted by sequential analyses, adding an extra class at each stage. Fit statistics (see appendix 1 for detailed description) are then compared to determine the best fitting class solution. Morgan (2015) stresses the need for findings to be parsimonious and interpretable, however as a rule of thumb, when a class solution returns a non-significant Lo Mendel Rubin (LMR) result, the previous class solution may be the most appropriate.

6.4- Results

6.4.1- Growth Mixture Modelling

Below is a table detailing the fit statistics generated by the growth mixture model. The fit statistics indicated that either 4 or 5 class solution might be appropriate. The LMR returned a non statistically significant result when a 6 class solution was run suggesting the appropriateness of a 5 class solution. AIC, BIC and SABIC continuously decrease with each added class, suggesting that the most appropriate model was not identified by these statistics.

The entropy figure for a class 5 solution is higher than that of the 4 class solution (entropy =0.899 & 0.882). Entropy is a measure of class delineation with values approaching 1 indicating clear delineation of classes (Celeux & Soromenho, 1996). Generally, the more classes that are added to a model, the less distinct each class becomes, however, in this case, a 5 class solution has more delineated classes than a 4 class solution. It must be noted, however, that this difference is minimal. Given the similarity of the two classes, it was considered prudent to investigate the graphs of both 4 and 5 class solutions to investigate the trajectories and proportions of each class, which is shown in figure 6.1 and 6.2.

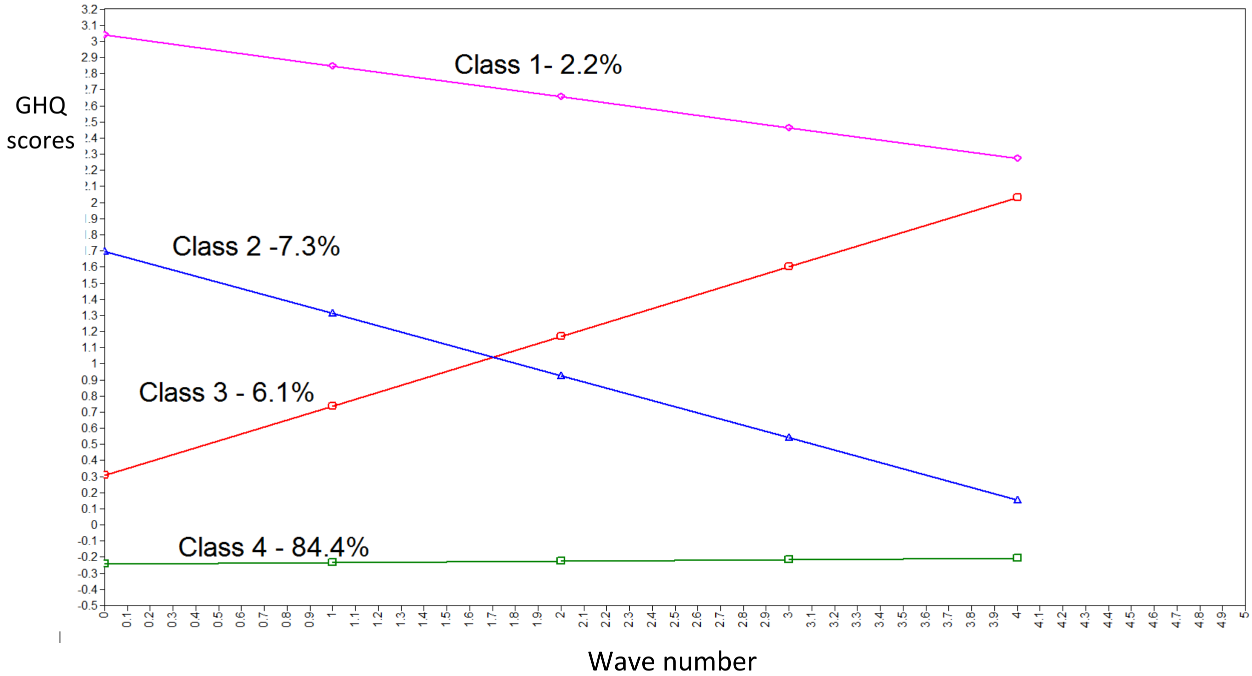
Table 6. 1

Fit Statistics from Growth Mixture Model

Class number	Replicated	AIC	BIC	Sample Adjusted BIC	Adjusted LMR	P=	Entropy
1	Yes	551528.094	551600.821	551575.396	N/A	N/A	N/A
2	Yes	515820.148	515920.147	515885.189	34671.888	<0.000	0.928
3	Yes	509544.920	509672.192	509627.699	6097.955	<0.000	0.899
4	Yes	503405.892	503560.436	503506.410	5965.729	<0.000	0.882
5	Yes	501181.297	501363.114	501299.554	2165.511	<0.000	0.885
6	Yes	498973.649	499182.738	499109.644	2149.059	0.0498	0.877
7	Yes	496663.482	496899.844	496817.215	2248.586	0.2201	0.857

Figure 6. 1

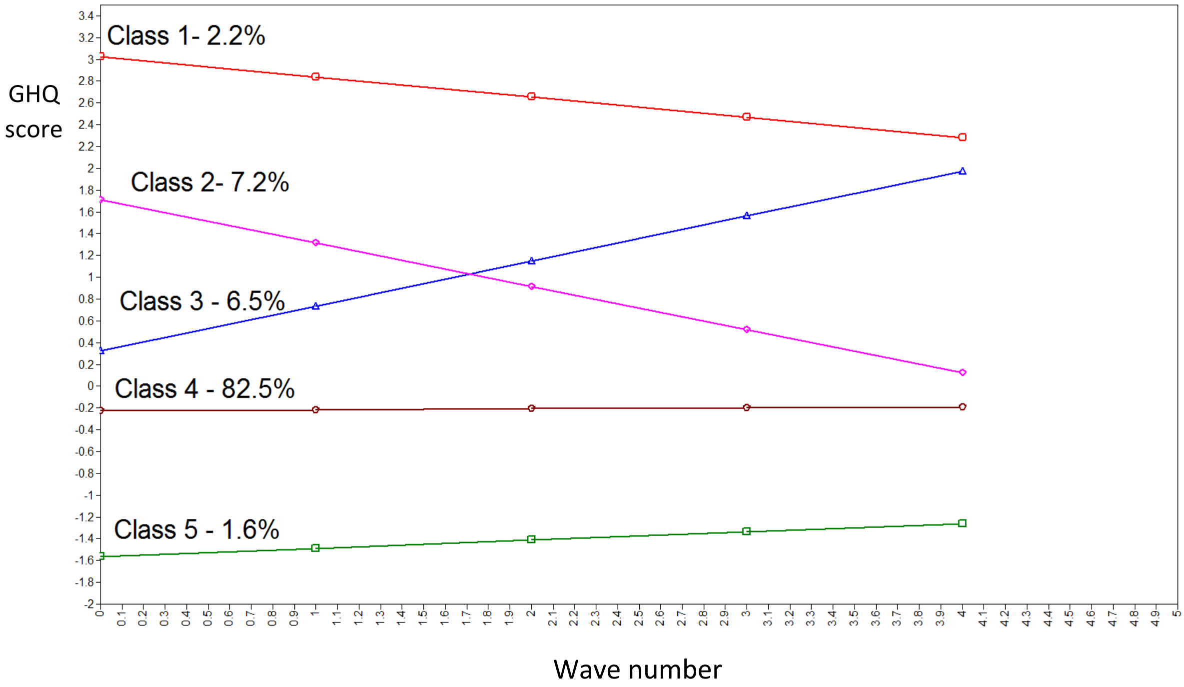
A line graph showing trajectories of GHQ scores in a 4 Class Solution



Class 4 comprised the majority of participants in the data accounting for 84.4%. The volume of participants and the low, stable nature of this class made it an obvious reference group in this scenario. Class 3 consists of 6.1% 9f participants who exhibited steadily increasing scores and could be conceptualised as participants whose mental health is steadily improving over time. Conversely, Class 2, comprises 7.3% and is characterised as a downward slope, representing participants who exhibit steadily decreasing mental health. Finally, class 1 comprising only 2.2% of the participants represents participants who have consistently elevated GHQ-12 scores, and while these scores are steadily decreasing, they are doing so in a manner which is less than class 2.

Figure 6. 2

A line graph showing trajectories of GHQ scores in a 5 Class Solution



This graph shows the class structure of a 5 class solution. The reference group is class 4 as it contains 82.5% of participants, and they exhibit relatively stable and low scores. Classes performed relatively similarly to the 4 class model, with the exception of class 5. This class was added and represented those participants who exhibit stable and exceptionally low GHQ-12 scores throughout, representing consistently good mental health. It represents 1.6% of the data and considering that the reference group has decreased in size by 1.9% and many other classes have stayed relatively similar, and it is likely that a class 5 solution represents a very similar class structure to class 4, with the reference group separated into two groups. The relatively similar fit statistics and structure of the four and five class solutions made the decision of which is the most appropriate solution difficult. This is explained in detail in the discussion section.

6.5 -Discussion

As with other chapters in this thesis, the interpretation of results was a balancing act between fit statistics, parsimony, interpretability, utility and validity. When determining the most appropriate and meaningful class solution for this analysis, this was no exception. While the battery of fit statistics conducted indicated that a 5 class solution was appropriate for this data, other factors had to be taken into account.

Given the relatively similar structure of the 4 and 5 class solutions and the urging of Morgan (2015) to be cognisant of parsimony and interpretability, it was decided that the 4 class solution would be more meaningful. The small number of participants that comprised class 3 in the 5 class solution were unlikely to be large enough to yield statistically significant results and that the other classes remained largely unaffected by the increase from 4 to 5 class solutions. Furthermore, class 3, i.e. the notably lower class may make it difficult to discover statistically significant relationships between the classes and any covariates as it effectively removes the extreme values from the reference group and reduces the difference between it and the other classes. This decision resulted in 4 clearly defined classes with imposed linear trajectories. The linearity of relationships was necessary to aid in interpretation for intended future analysis.

The four classes represented clearly distinct trajectories of mental health over time. As previously mentioned, the reference class represented the majority of participants, with stable low GHQ-12 scores. Other classes displayed either steadily deteriorating or improving mental health with a final class showing substantially elevated and only slightly improving GQH-12 scores over time.

The extracted classes were consistent with previous literature identified in 6.3, exhibiting relatively similar class structures from longitudinal research using GHQ when investigating the disaster survivor population (Høyer Holgersen, 2011). Noticeably, the reference group which represented participants with stable and low scores was considerably larger in this general population sample than in other research, however, this was attributed to the other studies being conducted in populations with abnormal levels of trauma and therefore was not unexpected.

The act of separating the population into four distinct classes of clearly differentiating behaviours was conducted for a number of reasons, most obviously to ascertain if subpopulations did exist within the population albeit it would have been unexpected if they did not. It was also done to inform further analysis of what predisposes individuals to be a member of these classes. In subsequent chapters, appropriate covariates that would be likely to affect one's mental health will be identified to ascertain if they display relationships with class membership, and if so, to what degree.

6.5.1- Limitations

The research used fit statistics in the determination of the appropriate class structure. Morgan (2015) has shown that fit statistics can, in some circumstances, be a poor indicator of the correct class structure. In this research, under certain circumstances, fit statistics were effective at determining the correct class solution only around 35% of the time, and this fell as low as <1% in some circumstances. In fact, instances, where fit indices were more than 50% accurate, were rare. Literature guidelines suggested the consideration of factors such as parsimony and interpretability, therefore, the subjective judgement of the individual interpreting the data was required to weigh these competing considerations. While research by Morgan (2015) has shown the weaknesses in relying on fit statistics alone, in this analysis they were used as an initial indicator and not the sole criterion for class selection, therefore the limitations were mitigated.

The research was also limited through the imposition of linear trajectories. While linear trajectories provide a clear and interpretable class solution, they may mask either quadratic (U-shaped) or exponential trajectories (an ever increasing slope). Non linear trajectories may also have provided alternative class solutions but require an element of subjectivity in interpretation which linear models avoid.

6.5.2- Clinical and Research Implications

The implications for both clinical and research from these findings are similar and are therefore presented together. This research suggested that participants exhibited a number of trajectories of GHQ-12 scores over time with some demonstrating stable scores and others changing over time. A solution of four distinct trajectories of mental health for the UK population was selected on the basis of parsimony and interpretability, however, the results also showed that a five-class solution could also have been selected and in different circumstances may have been a more meaningful solution. It may be meritorious to investigate, as is proposed in subsequent chapters, what predisposes individuals to membership of the various groups and what covariates could explain the various trajectories exhibited. Clinicians may be particularly interested in what predisposes an individual to membership of the class which represent improving mental health and how they can encourage their clients to adopt these behaviours. They may also be interested in what predisposes individuals to display increasing GHQ-12 scores, representing deteriorating mental health. Identification of these variables could facilitate clinicians to identify ‘at risk’ individuals based on the exhibition of variables which were associated with this class.

Researchers should be aware that the data was rigorously collected to ensure that it was representative of the UK population at large. If these results are a true reflection of the UK, population, then it may not be appropriate to measure the effects of various covariates on mental health for the population at large and that individuals who exhibit various trajectories may be more appropriately analysed separately.

References

Afifi, T. O., Enns, M. W., Cox, B. J., Asmundson, G. J., Stein, M. B., & Sareen, J. (2008). Population attributable fractions of psychiatric disorders and suicide ideation and attempts associated with adverse childhood experiences. American journal of public health, 98(5), 946-952.

Akaike, H. (1987). Factor analysis and AIC. In *Selected papers of hirotugu akaike* (pp. 371-386). Springer, New York, NY.

Asparouhov, T., & Muthén, B. (2018). Variable-specific entropy contribution. *Los Angeles, Muthén & Muthén*.

Berlin, K. S., Williams, N. A., & Parra, G. R. (2014). An introduction to latent variable mixture modeling (part 1): Overview and cross-sectional latent class and latent profile analyses. *Journal of pediatric psychology*, *39*(2), 174-187.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological bulletin*, *101*(1), 147.

Buitenweg, D. C., Bongers, I. L., van de Mheen, D., van Oers, H. A., & van Nieuwenhuizen, C. (2018). Subjectively different but objectively the same? Three profiles of QoL in people with severe mental health problems. *Quality of Life Research*, *27*(11), 2965-2974.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of classification*, *13*(2), 195-212.

Grimm, K. J., Ram, N., & Shiyko, M. P. (2013). A simulation study of the ability of growth mixture models to uncover growth heterogeneity. In Contemporary Issues in Exploratory Data Mining in the Behavioral Sciences (pp. 194-211). Routledge.

Hawton, K., Rodham, K., Evans, E., & Weatherall, R. (2002). Deliberate self harm in adolescents: self report survey in schools in England. *BMJ*, *325*(7374), 1207-1211.

Holgersen, K. H., Klöckner, C. A., Jakob Boe, H., Weisæth, L., & Holen, A. (2011). Disaster survivors in their third decade: Trajectories of initial stress responses and long-term course of mental health. *Journal of traumatic stress*, *24*(3), 334-341.

Jones, J. W., Ledermann, T., & Fauth, E. B. (2018). Self-rated health and depressive symptoms in older adults: A growth mixture modeling approach. *Archives of gerontology and geriatrics*, *79*, 137-144.

Kazlauskas, E., Gegieckaite, G., Hyland, P., Zelviene, P., & Cloitre, M. (2018). The structure of ICD-11 PTSD and complex PTSD in Lithuanian mental health services. *European journal of psychotraumatology*, *9*(1), 1414559.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*(3), 767-778.

Mahedy, L., Todaro-Luck, F., Bunting, B., Murphy, S., & Kirby, K. (2013). Risk factors for psychological distress in Northern Ireland. *International journal of social psychiatry*, *59*(7), 646-654.

Morgan, G. B. (2015). Mixed mode latent class analysis: An examination of fit index performance for classification. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(1), 76-86.

Morgan, G. B.(2012). *Mixed Mode Latent Class Clustering: An Examination of Fit Index Performance for Identifying Latent Classes.* (Doctoral dissertation). Retrieved from https://scholarcommons.sc.edu/etd/1031

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Structural Equation Modeling, 14, 535–569.

Dawson, D. A., Grant, B. F., Stinson, F. S., & Zhou, Y. (2005). Effectiveness of the derived Alcohol Use Disorders Identification Test (AUDIT-C) in screening for alcohol use disorders and risk drinking in the US general population. *Alcoholism: Clinical and Experimental Research*, *29*(5), 844-854.

Prins, A., Bovin, M. J., Smolenski, D. J., Marx, B. P., Kimerling, R., Jenkins-Guarnieri, M. A., ... & Tiet, Q. Q. (2016). The primary care PTSD screen for DSM-5 (PC-PTSD-5): development and evaluation within a veteran primary care sample. *Journal of General Internal Medicine*, *31*(10), 1206-1211.

Goldberg, D. (1978). *Manual of the general health questionnaire*. Nfer Nelson.

Oberski, D. (2016). Mixture models: Latent profile and latent class analysis. In *Modern statistical methods for HCI* (pp. 275-287). Springer, Cham.

Parker, G., & Roy, K. (2001). Adolescent depression: a review. *Australian & New Zealand Journal of Psychiatry*, *35*(5), 572-580.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*(3), 333-343.

Scott, S. (1998). Conduct disorders. *CLINICS IN DEVELOPMENTAL MEDICINE*, 1-27.

Ye, S. (2009). Factor structure of the General Health Questionnaire (GHQ-12): The role of wording effects. *Personality and Individual Differences*, *46*(2), 197-201.

Jamali, J., & Ayatollahi, S. M. T. (2015). Classification of Iranian nurses according to their mental health outcomes using GHQ-12 questionnaire: a comparison between latent class analysis and K-means clustering with traditional scoring method. *Materia socio-medica*, *27*(5), 337.

Chapter 7 - Reviewing Explanatory Variables Which Explain Change in Mental Health Over Time

7.1- Abstract

Introduction

In the previous chapter, growth mixture modelling techniques were used to ascertain if participants exhibited various trajectories over time and if they did, to identify an appropriate class structure. Following this analysis, it was established that individuals exhibited four different trajectories of mental health over time. Over the course of the next two chapters', analyses will attempt to explain why participants exhibit different trajectories through the modelling of latent classes extracted in Chapter 6 onto covariates. Prior to this, a number of variables were selected using the biopsychosocial model as a framework for selection, the literature concerning them and their impact on mental health over time was reviewed with particular emphasis placed upon those which had available data in the Understanding Society database.

Methods

A broad array of biological, social and psychological covariates were selected including sex, age income and personality traits. The data for these covariates were drawn from waves one through five of the Understanding Society Database, and descriptive statistics were given. Some of these covariates remained stable over time, whereas others changed, referred to as time-invariant and time-varying respectively. In order to prepare them for further analysis, time-variant covariates were converted into slopes and intercepts to represent change over time and initial values respectively, whereas it was not necessary to convert time-invariant covariates. The appropriates of a linear interpretation of time-invariant covariates was tested to assess if it was appropriate for each time-varying covariate.

Results

Significant levels of missingness were found in a number of the social variables such as job satisfaction and leisure satisfaction, however generally these were much lower in psychological and biological covariates. In relation to those variables which were time-varying, all demonstrated good fit with a linear model.

Conclusion

A broad array of variables which aligned with the biopsychosocial model were selected and have had their appropriateness for inclusion in future analysis established. Where appropriate these covariates have either been recoded or transformed into a form which was compatible with future regression analysis proposed in chapter 8. The high levels of missingness in some of the social covariates was so high as to be likely to bring about a number of methodical issues in future analyses, therefore methodical techniques which mitigate the effects of missing data would have to be considered in later analyses. The good fit of time-invariant covariates with a linear model indicated that a linear interpretation in future analyses was appropriate for these covariates and that they could be modelled using this interpretation.

7.2- Introduction

The analysis completed in Chapter 6 demonstrated that different participants in the Understanding Society (UKHLS) database exhibited different trajectories of mental health, as measured by the GHQ-12, over time. The previous analysis conceptualised these different trajectories as membership of different classes. The previous chapter identified four trajectories of participants in the UKHLS database. These were interpreted as a reference group of low, stable scores, a group of participants who exhibited high stable scores and two more groups representing increasing and decreasing scores, respectively. In an effort to explain the reasons that participant's displayed these various trajectories, it is intended to conduct regression analysis in chapter 8, modelling the classes extracted in chapter 6 on a broad array of covariates. Prior to this analysis, the selection and review of covariates was completed in this chapter to establish their appropriateness and to prepare them in such a format as to be compatible with regression analysis. Covariates were selected from available variables in the UKHLS which corresponded with elements from the biopsychosocial model of mental health (Gathchel et al., 1996), which provided a useful framework for covariate selection (see figure 6.1). This model proposed that mental health is affected by the interaction of one's genetic predispositions, social environment and psychological characteristics and is commonly adopted by clinical psychologists when treating patients.

Covariates which vary with time presented distinct methodical difficulties and as such required transformation into a format which was compatible with the proposed regression analysis in Chapter 8. The values collected at each wave were transformed into individual slopes and intercepts for each participant and were subsequently tested to ensure that a linear interpretation of these variables change over time was an appropriate interpretation.

7.2.1- Review of the Biopsychosocial Model of Mental Health

As stated above, variables were selected from those available in the UKHLS which align with the biopsychosocial model of mental health (see figure 6.1) and in this section, the model is discussed to provide context.

The biopsychosocial model of mental health was the successor to the biomedical model of mental health. This model emphasised a biological approach to mental health disorders and was primarily a consequence of the fact that many early interventions for mental health illness were performed by psychiatrists (Deacon, 2013). This model posited that mental illness was primarily caused by abnormalities in the brain and therefore could be treated by interventions which attempted to address these abnormalities (Andreasen, 1985). The shortcomings of this approach were eloquently described by Engel (1977) as below.

"The dominant model of disease today is biomedical, with molecular biology its basic scientific discipline. It assumes diseases to be fully accounted for by deviations from the norm of measurable biological (somatic) variables. It leaves no room within its framework for the social, psychological, and behavioural dimensions of illness. The biomedical model not only requires that disease be dealt with as an entity independent of social behaviour, it also demands that behavioural aberrations be explained on the basis of disordered somatic (biochemical or neurophysiological) processes (p. 130)."

In summary, the biomedical model was criticised for failing to account for the psychological and social aspects of mental health, which were ill-understood or accounted for at the time. Following calls by George Engle for the need of medicine to adopt an "integrative, non-reductionist clinical and theoretical perspective in biomedicine" (Engle, 1997), the biopsychosocial model saw extensive use as an appropriate model for investigating mental health and subsequently has formed the basis of modern psychiatry and clinical psychology (Ghaemi, 2009).

The biopsychosocial model was proposed by Gathchel et al. (1996) as a model of explaining chronic pain. This approach suggested that chronic pain was a consequence of a number of interrelated factors incorporating an individual's biological predispositions, their psychological profile and their social environment. It was revolutionary as it reintroduced the concept of a separation of body and mind which dates back to early writings by Descartes, but had been largely overlooked in medical practices at that time (Dombeck, 2019). Over time it has become adopted as a multidisciplinary model which has applications for a wide range of areas of study ranging from mental health, pain, and human development.

The model suggested that the biological, psychological and social aspects of an individual had to be considered when determining causes of mental illness. It is usually represented diagrammatically as three overlapping circles with various covariates positioned according to their correspondence to the biological, social or psychological components of the model.

While not without its critics, the specifics of which go beyond the scope of this thesis, many of the criticisms treat the model as an attempt to encapsulate the entirety of mental illness within a simplistic framework, and this may be both unreasonable and unrealistic. Rather than treat the model in that way, this thesis will use the model as described in Borrell-Carrió et al. (2004) as a way of '*organising one's thoughts*' when investigating mental health, informing covariate selection and of conceptualising the interactive nature of the various predictor variables which impact mental health.

7.2.2- Covariates

Variables were selected from those available in the UKHLS, which aligned with the biopsychosocial model. The rationale for their selection in the analysis was included below, alongside pertinent literature that may provide insight into how these variables should relate to mental health change over time. As previous analyses extracted classes which denoted both stable and increasing or decreasing scores over time, the literature referenced was widened to encompass terms such as reliance and recovery. Mental resilience was investigated as it was likely that increased mental resilience was likely to correspond with stable scores. Recovery was also investigated as, by its very nature, covariates which encourage mental health recovery should correlate with the membership of the steadily improving GHQ-12 scores class.

Resilience is a commonly used term in psychology (Bonanno, 2004; Carver, 1998; Garmezy, 1991; Kaplan, 1999; Luthar, Cicchetti, & Becker, 2000; O'Leary & Ickovics, 1995; Rutter, 2006), however different researchers place different emphasis on its definition. Masten (2001) and Bonano et al. (2010) focused on participant's outcome following stressful events with the latter defining the concept as "*an outcome pattern following a potentially traumatic event*". Others, however, approach the concept from a homeostatic standpoint defining it in terms of an individual's ability to return to their original state in response to stressors (Neuman and Faucet, 2002). Finally, Fredrikson (2001) and Steinhardt and Dolbier (2008) define it in terms of '*protective factors*' which enable adaption to stressful environments. Regardless of the definition adopted, resilient individuals should be unlikely to be part of the deteriorating GHQ-12 scores class, which was extracted in the previous chapter. Resilience can be measured either by using a number of standardised tests such as the 'Resilience Scale' (Wagnild & Young, 1993) or by measuring fluctuations in scores form mental health instruments and both will be used in the section below.

Recovery was defined as "*The interval wherein a subject displays steady improvement with regard to quantifiable rebound of capabilities and dexterity following severe health issues or trauma*" (Pam, 2013). This concept should correspond to membership of the recovery class identified in the previous chapter, where participants exhibit steadily increasing GHQ scores over time.

Variables are ordered according to the subheading of the biopsychosocial model that they relate to.

7.2.2.1- Biological

These covariates relate to those which are biological in nature and tend to be invariant. The specifics of the variables included are listed below.

Age

This variable has already been detailed in cross-sectional analysis in chapter 4. The entry in Chapter 4 focused on the cross-sectional aspects of the relationship and given the longitudinal nature of this analysis, it was felt important to elaborate further. Adding to the literature identified in that chapter, the concept of recovery was deemed important to ascertain if age was likely to affect an individuals ability to recover from mental health difficulties. Hinrichsen (1992) investigated the recovery and relapse rates of participants from the National Institute of Mental Health study (Katz, 1980). He concluded that in these participants, recovery rates for participants who were over 60 years old did not differ significantly from those of the general population. Corrigan et al. (1999) however found that age was inversely proportional to recovery, suggesting that as individuals age, they become less likely to recover from poor mental health. This research did not meet certain levels of statistical significance and was described by the author as possibly representing artefact. As a result, the soundness of this research may be called into question. While this research, albeit dated, may suggest that older people are less likely to recover from mental illness, research also suggests that they may be more resilient to mental illness. Netuveli (2008) found increased levels of resilience in participants of the British Household Panel Survey, which is the precursor to the database used in this analysis. This was supported by Glonti et al. (2015), who found the same relationship when conducting a systematic review. It must be noted however that Economou (2013) suggested a more nuanced interpretation finding that specific age groups (<34, 25–34, 45–54, 55–64) were more susceptible to changes in their mental health and that, consequently it may not be appropriate to view age as having a linear relationship with resilience. This approach was supported by Hauksdottir et al. (2000), who also suggested that there were critical periods of susceptibility to mental health changes.

While the evidence would suggest a general increase in resilience with age, Mehta et al. (2008) found that as participants aged, resilience appeared to become less relevant relating to the onset of late-life depression. Due to the complexity of this relationship, it may be possible that this is a simplistic interpretation of the data and that consequently, a linear relationship between age and resilience may not manifest.

Sex

Generally, research has found that females tend to exhibit poorer self-reported mental health than males (Gili et al. 2013; Katikireddi, Niedzwiedz & Popham, 2012; Economou et al. 2013) however when investigating change over time the research is less conclusive.

One method of investigating change over time, especially resilience was investigating the changes in the average self-reported change in a time of turbulence and investigating if there was a significant difference in self-reported mental health changes between the sexes.

Hauksdottir et al. (2013) investigated changes in self-reported mental health in Iceland following the global recession. In this paper, it was noted that self-reported mental health deteriorated in a statistically significant manner for females but not males.

This may suggest increased resilience in males relative to females, however, given the very specific time period that this research took place in, this resilience may have been specific to economic factors.

Conversely, however, the opposite finding was found in England (Katikireddi et al. 2008) and Spain (Agudelo-Suarez et al. 2013) suggesting that these findings may be specific to the population that is being investigated or a consequence of the cultural nuances of the various populations.

Peng et al. (2012) investigated the resilience of Chinese medical students and found that male students were more resilient to the stress that undertaking a medical degree entailed. This study was conducted in a sample of students, and usually, this population is not generalisable to the general population as students tend to be of a specific age cohort and medical students would tend to be more intelligent than the general population.

In a study primarily investigating the effects of mental health recovery programmes within the USA criminal justice system, Kothari et al. (2014) found that females were more likely to benefit from recovery programmes than males suggesting that within this specific cohort, that females were more likely to display an improving mental health trajectory than males if they had access to appropriate mental health support programmes. In relation to specific conditions such as depression, gender differences have been observed, with females experiencing depressive symptoms for longer than males and were more likely to suffer from relapses, suggesting that in the case of depression, recovery was less common in females (Lewinsohn et al., 1989).

In conclusion, the literature surrounding resilience and recovery in relation to gender differences was unclear, with numerous studies suggesting conflicting results.

The most relevant study identified was that of Katikireddi et al. (2008) as this study was conducted in a UK population.

Physical Health

Physical health has been detailed in chapter 4 in relation to its impact on mental health from a cross-sectional standpoint. This section will examine the extent to which physical health predisposes individuals to change over time.

Physical activity was found to be closely associated with resilience in relation to mental health in a Chinese population of young people. This study used the mental health component of the SF-12 which is detailed in chapter 4 to measure mental health in this population and found that those who had higher levels of physical activity were more likely to exhibit stability in their SF-12 scores in this study. This implies that physical activity was likely to protect individuals from fluctuations in their GHQ-12 scores. A systematic review of American studies showed that this relationship was also mirrored in American school-age participants (Strong et al. 2005).

While these studies focused on young people, research has also suggested that these relationships exist in elderly people. Wells (2012) found that New York residents displayed statistically significant relationships between resilience and physical health as measured by the Resilience Scale (Wagnild & Young, 1993) and SF-12 scores. Felton (2000) supported these findings with participants from the American Midwest with what they defined as ‘frailty’ being associated with resilience in a statistically significant manner. Hardy, Concato and Gill (2002) operationalised physical health using a number of variables such as grip strength, the presence of chronic conditions, ability to perform daily tasks, physical activity and self-reported health. This study found a statistically significant relationship between these measures of physical health and resilience in over 70's in the Connecticut area using a self-report questionnaire of their own devising. This study was more comprehensive than Felton's as there were more variables included and found similar results suggesting that physical health was likely to predispose individuals to greater resilience. Finally, Lamond et al. (2008) utilised a different measure of resilience, that of the CD-RISC (Connor & Davidson, 2003) and found that this measure of resilience was statistically significantly associated with what they referred to as freedom from disability.

Only one study could be identified, which suggested that physical health did not display a statistically significant relationship with age (Nygren, Alex, Jonsen, Gustafson, Norberg, & Lundman, 2005). This study was conducted in a Swedish sample of participants of over 85 years old and therefore may not be relevant to general population research.

While the study mentioned above casts some doubt on the relationship between physical health and resilience the literature overwhelmingly supports the concept that physical health is a significant predictor of resilience and this relationship should be mirrored in analysis conducted in subsequent chapters.

7.2.2.2- Psychological Variables

Personality

Numerous definitions of personality exist, however many refer to the semi-permanent nature of characteristics which combine to form an individual's personality. For example, Mischel (1999) defined personality as "*The distinctive patterns of behaviour (including thoughts and well as 'affects,' that is feelings, and emotions and actions that characterise each individual enduringly*." Feist and Feist (2009) stated that "personality is a pattern of relatively permanent traits and unique characteristics that give both consistency and individuality to a person's behaviour." Cattell (1950) described it as, "That which permits a prediction of what a person will do in a given situation." Given the consensus within the literature as to the relative stability of personality, and the fact that it was only collected at one time point, it was viewed as appropriate to treat this variable as time-invariant, however, this will be explained in the methods section. It is also important to note that recent research which utilised waves 1-6 of the understanding society tested the stability of personality traits between waves three and six where they were collected. Busic-Sontic, Czap & Fuerst, (2017) found that personality traits remained stable over time.

Personality is operationalised in the UKHLS using the 'Big 5 model' of personality. This model breaks the concept of personality into five distinct factors defined as 'openness', 'conscientiousness', 'extraversion', 'agreeableness' and 'neuroticism'. The origin of this model is unclear as Tupes and Christal (1961) first proposed this model in 1961, however, the model failed to reach prominence until the 1980's when numerous researchers (Goldberg,1982; Costa and McCrea, 1976 & Tupes and Christal, 1961) had independently come to similar conclusions as to a 5 factor model of personality.

Literature into the relationship between personality traits as operationalised by the 'Big 5' model suggested that some of the factors have a greater effect than others. Early research into this area suggested that Extraversion and Neuroticism were the greatest predictors of happiness, with greater levels of extraversion and lower levels of neuroticism displaying the strongest relationships (Costa & McCrae, 1980). This relationship has frequently been corroborated in the literature (Brebner et al., 1995; Chan & Joseph, 2000; Furnham & Brewin, 1990).

Hayes and Joseph (2003) performed regression analyses on three measures of subjective well-being, the Oxford Happiness Inventory (Argyle, Martin, & Crossland, 1989), the Depression–Happiness Scale (Joseph & Lewis, 1998), and the Satisfaction With Life Scale (Diener, Emmons, Larsen, & Griffin, 1985). Extraversion and Neuroticism were found to be the strongest predictor of the Oxford Happiness Inventory (Argyle, Martin, & Crossland, 1989), however, while extraversion remained the strongest predictor of the Satisfaction With Life Scale (Diener, Emmons, Larsen, & Griffin, 1985), conscientiousness outperformed extraversion as the primary indicator.

While the above studies utilise various measures of mental health, it was felt important to focus on a number of studies which utilised the GHQ-12 (see Chapter 1) as this is the specific instrument that was used in this thesis.

While not conducted in a representative sample, neuroticism was investigated in relation to psychological distress as defined by the GHQ-12 was found to be closely associated with having a neurotic personality in Russian nursing students. These findings were supported in Menon et al. (2017) who found that participants with lower GHQ-12 scores were likely to display high levels of neuroticism and low levels of extraversion, however, the relationship for neuroticism was stronger.

The studies mentioned above relate to the relationship between mental health and personality, they fail to encapsulate the extent to which an individual's mental health may be susceptible to change, which is investigated hereon. Horsburgh et al., (2009) conducted bivariate regression analyses into the relationship between mental toughness and the components of the 'Big 5' model and a questionnaire designed to measure 'mental toughness' (Clough et al., 2001). All aspects of the model exhibited statistically significant relationships with mental toughness, with the exception of agreeableness, which displayed no relationship. All relationships were positive except that of neuroticism which displayed a moderately strong negative relationship with ‘mental toughness’. These findings would suggest that with the exception of openness and neuroticism GHQ-12 scores should be stabilised by these factors of personality and that high levels of neuroticism should be a predictor of fluctuating scores.

7.2.2.3- Social Variables

Ethnicity

Within the UKHLS, there exists a variable which is referred to as both race and ethnicity. It is coded as ‘*RACEX*’, however, its description mentions ethnicity. Due to this discrepancy, it was felt important to differentiate between race and ethnicity prior to investigating the literature pertinent to both. Race generally refers to biological characteristics, and when investigating race, people are usually subdivided into subgroups based on physical characteristics such as skin colour and facial structure. Ethnicity refers to a more sociocultural construct which examines groups of people based on characteristics such as language, nationality and customs (Nittle, 2020).

When investigating the options that participants were asked to choose from, the variable options are identical to the list of ethnic groups mentioned on the UK government website (Gov.uk, 2020) and therefore it was decided that is was more appropriate to refer to these subgroups as ‘ethnicity’ rather than ‘race’.

Psychiatric epidemiologists have frequently proposed that minority groups experience increased stress due to the associated disadvantage that these groups frequently endure (Kleiner, Tuckman & Lavell, 1960; Fischer, 1969; Kramer, Rosen & Willis, 1973; Cannon & Locke, 1977; Mirowsky & Ross, 1980). These authors argue that as most minorities within a population exhibit a poorer quality of life as measured by lower levels of life satisfaction, happiness, marital happiness, and higher levels of anomie and mistrust than white people. Hughes and Demo (1989) argue that poorer mental health was a function of these factors. The relationship may, however, be more complex than first envisaged as certain aspects of mental health do not exhibit differences between races. A large number of researchers have reported no significant differences between self-esteem between those of different skin colours (Porter & Washington, 1979; Twenge & Crocker, 2002; Jackson, Williams & Torres, 2003), and with specific exceptions of schizophrenia and phobias, (Pinto, Ashworth & Jones, 2008) no significant differences could be identified between incidence rates of mental disorders between black and white participants of the Epidemiologic Catchment Area Study (Robins & Reiger, 1991). Conversely, the incidence of mental health disorders was frequently found to be lower in participants from an African background other participants of the 1990 National Comorbidity Study (Blazer et al., 1994; Kessler, McGonagle, Zhao, Nelson, Hughes, Eshleman, Wittchen & Kendler, 1994). Consequently, the assertion made in Jackson (2004), that incidence rates of mental disorders between black and white participants were at least comparable, if not lower in black participants, is supported in the literature.

In summary, it appears that while self-reported mental health and distress appear to be consistently poorer in black participants, the incidence rates of mental disorders are not. This may be attributed to a number of factors including access to services, however, one interpretation of these results is that while black participants are exposed to greater stress as a result of socioeconomic factors, this stress is less likely to result in an increase in the manifestation of mental disorders. This may, in turn, suggest a greater resilience in black participants.

Education

This variable has already been detailed from a cross-sectional perspective in Chapter 4 however, as this was cross-sectional in nature, it was felt important to investigate the longitudinal aspects of this relationship.

While generally it is accepted that people who have higher educational attainment report higher levels of mental health, this is somewhat difficult to isolate the effects of education from that of the effects of better education such as income and quality of life (Friedli, 2009). Friedli (2009) emphasised the reciprocity of this relationship stating that an individual's mental health was a consequence of education however also stating that people with lower emotional intelligence and those who experienced mental health problems had lower educational attainment, describing poorer educational attainment as both a cause and consequence of poor mental health.

The Effective Pre-school and Primary Education Project (EPPE) was the largest longitudinal study investigating the factors which affect resilience which is defined as '*higher than expected attainment*' in Europe. Sylva (2009) demonstrated that educational attainment was greatly impaired by emotional problems, and while this research used participants aged 5-10, the results from this study do suggest that the variables were co-dependent. Another factor which makes the relationship between education and mental health difficult is that of IQ. IQ and educational attainment are correlated, and it has been demonstrated that people with higher IQ scores are more resilient (Batty, 2006).

Allan (2015) investigated the resilience of students in Higher Education. He stated that resilience was necessary for high levels of academic achievement in a higher educational setting as a result of the drastically changed environment that students at university found them in (Sugarman, 1986). Allan’s study, however, found that resilience, as defined by the Connor Davidson Resilience Scale (Connor & Davidson, 2003), was a clear indicator for educational attainment in females however he described the relationship for males as ‘convoluted’.

Veldman et al. (2014) conducted research into educational attainment at two time points 9 years apart. This study found that mental health outcomes significantly improved in participants who had mental health conditions but who received appropriate help with these conditions in comparison to those who did not. This research suggested that improvement and recovery were unlikely while mental health conditions remained unaddressed.

In conclusion, the literature around mental health and education is difficult to differentiate as education is directly linked to a number of other pertinent variables which also affect mental health. The causality of the relationship also appears to be in question with literature suggesting that poor mental health was both a cause and consequence of education. It is for that reason that other variables such as income and subjective wellbeing will be included in this analysis, and they are detailed below.

Financial Situation/ Income

The extent to which income and mental health are linked has been a source of debate within the literature. Before a discussion can be had around the effects of income, it was felt pertinent to outline some of the difficulties that analysis in this area has faced.

Similar levels of income are not directly comparable depending on the area that an individual lives. Within a UK setting, the cost of living can be measured using a number of metrics such as house prices and cost of living indexes. The ‘Big Mac Index’

(Ong, 2003) is one such method. This albeit novel index compares the price of a Big Mac Hamburger in different areas and is a measure of purchasing power as the price of said hamburger fluctuates on the purchasing power of the area it is sold. In London, the average cost of a Big Mac is £4.49, however, the average price in Scotland is much lower at £3.39 (Global Price info, 2020). From differences in purchasing power and cost of living in different areas of the UK, it can be said with confidence that an equivalent salary in different areas of the UK would translate into different outcomes. As a result, it may not be appropriate to directly compare income levels.

While subsequently disputed ‘The Spirit Level’ (Picket & Wilkinson, 2009) suggested that the levels of income disparity were a greater determinant of health outcomes than absolute income and stated that the impact of being poor in a rich area was greater than being poor in an area of people with similar incomes. Much of the findings in this book have been disputed, most notably in the book titled ‘The Spirit Level Delusion’ (Snowdon, 2010), which levelled accusations of selective data reporting and sample biases alongside methodological and reporting errors in the previous book. Due to the thorough deconstruction of the arguments presented and the findings that in many cases, no relationship existed, it was felt that investigating inequality was unlikely to yield statistically significant results.

Furthermore, Oskrochi, Bani-Mustafa & Oskrochi (2018) found that subjective financial situation was a greater predictor of self-reported mental health scores than actual income figures. These findings informed the selection of variables that would encapsulate financial situation as detailed in the methods section.

Parental Employment Status

Research has demonstrated that the employment status of parents can have an impact on the mental health of their children and that this effect can persist into later life (Bacikova-Sleskova, Benka & Orosova, 2014). Research into this area has mostly emphasised the importance of economic and social consequences that parental unemployment had on the family unit (Strom 2003), with researchers suggesting that lower mental health scores were attributable to poorer income levels and the financial barriers to social activities that unemployment may entail. Other researchers have drawn attention to the issue of social mobility, stating that the children of unemployed adults tend to be more likely to not secure employment themselves and that this may be in part due to lower reported levels of mental health in these people (Christofferson, 1966). Additionally, Harland et al. (2002) demonstrated that the mental health impacts of short term unemployment were lesser than that of long term unemployment. It must be noted that few studies at the time of writing have distinguished between maternal and paternal unemployment (e.g. Magklara et al., 2010; Piko & Fitzpatrick, 2007; Sleskova et al., 2006) and these studies have generated inconsistent findings which has in part been attributed to the various different populations that the research was conducted in. Bacikova-Sleskova, Benka & Orosova, (2014) found that there were negative mental health impacts from paternal unemployment but not maternal. This research was particularly pertinent as it attempted to control for associated extraneous variables such as financial strain and parental relationships, which it found did not account for the relationship between paternal unemployment and poor mental health in later life. This paper also found that participants did not report differences in the relationship with their parents depending on their employment status during childhood but did report more negative feelings towards their father if they became unemployed if the father was also unemployed during their childhood. The authors suggested that this may have been consistent with findings by Elder (1974) which suggested that adolescents tend to blame their father for changes to their life circumstances, but the author admits that more research in this area was needed. This research was particularly relevant as it suggested that there were more issues at play simply financial issues which unemployment entailed and secondly that the children of unemployed parents did not report poorer relationships with their parents which other research had suggested may be responsible for poorer mental health.

Change of Marital Status

Changes in marital status can represent stressful times in peoples lives. While married people tend to report better mental health (Ueker, 2012), the act of being married has been included in the Holmes Rasche scale (1967) as a major stressful life event which can contribute to mental illness. This scale assigned values to stressful events and suggested that the scores of these events were summed. They suggested that participants who scored over 300 points were at high risk of developing an illness, and that scores between 150 and 299 indicated a moderate risk, approximately 30% lower than high-risk individuals.

Included in this scale are all marital status changes including being widowed, separated and divorced. This scale proposes that the most stressful event that can contribute to mental illness was the death of a spouse, which it scored as 100 points, divorce was attributed a score of 75 and marriage 50.

Cross-sectional research has suggested that the stress of being married does not necessarily result in a predisposition to mental health conditions such as depression with some research suggesting no significant differences in the mental health of recently married individuals (Horwitz & White 1991; Wu & Hart, 2002) whereas others found an increase in reported mental health (Simon 2002; Lamb et al., 2003). The research is relatively consistent in its assertion; however, that marital break up leads to an increase in mental illness (Hope et al., 1996; Marks & Lambert 1998; Simon, 2002).

Longitudinal research offered the opportunity to investigate marital status change in more detail, and found that women tend to be more vulnerable than men to changes in their mental health after a martial break up (Marks & Lambert 1998; Simon and Marcussen, 1999), being widowed (Williams, 2003), or remarried (Williams, 2003). Research has shown that given appropriate time, self-reported mental health usually returns to premarital change status (Lorenz et al., 1997; Booth & Amato, 1996).

Given the importance of these significant life events and the disagreement in the literature around the impacts of being recently married, these variables should provide a useful insight into the mental health trajectories of participants in the UKHLS database.

7.3- Methods

7.3.1 - Data

The data used in this chapter were acquired from a merged dataset of waves 1 through 5 of the Understanding Society Database. The dataset contained 65568 participants who completed the GHQ-12 during at least one wave of data collection and included all covariates listed below. Time-varying covariates were acquired from the various waves, whereas time-invariant were usually obtained from the WAVEX datafile which contained time-invariant data in all participants.

The participants were weighted, stratified and clustered to ensure that they remained representative of the UK population. During the model testing phase explained below, missing data were handled using the MLR technique.

7.3.2- Analysis

A wide range of biological, psychological and social variables were selected, which aligned with the biopsychosocial model of mental health. These variables were reviewed below, and descriptive statistics were provided in the results section. Some of these variables remained stable over time, whereas other varied. Those which varied had to be transformed into a format which was compatible with the regression analysis proposed in Chapter 8. This process involved converting the values collected at each wave into a series of slopes and intercepts which measured, change over time and initial value respectively. In order to assess if a linear interpretation of these variables was appropriate linear models were tested for model fit using a series of fit statistics. The fit statistics used are provided below alongside suggested interpretation guidelines where appropriate. Chi-square was reported as suggested in the literature, however, was not used for interpretative purposes, as Kenny (2020) suggested that it was inappropriate for analyses of over 200 participants. Root Mean Square Error of Approximation (RMSEA), a fit statistic generated from the Chi-Squared value was reported. Callum, Browne, and Sugawara (1996) suggested that values of 0.01, 0.05, and 0.08 indicated excellent, good, and mediocre fit, respectively and these criteria were adopted. Two incremental fit indexes were reported, the Comparative Fit Index (CFI) and the Tucker Lewis Index (TLI) and (Awang, 2012) suggested that values over 0.9 represented good fit for both these fit statistics. Hu and Bentler (1999) suggested that a value of 0.9 was too low a threshold to indicate a good fit. They suggested that a cut off of 0.95 would be more appropriate, and this higher threshold was adopted to demonstrate good fit. Standardised Root Mean Square Residual (SRMR) described as an absolute measure of fit (Kenny, 2020) was reported, and a value of less than 0.08 was considered sufficient to be considered a good fit (Hu and Bentler, 1999).

7.3.2.1- Covariates

Variables which were selected to correspond with the various aspects of the biopsychosocial model are detailed below and are introduced alongside any associated literature was felt pertinent to include.

Biological

Age

While this variable is not given in UKHLS, it was computed from birth year by subtracting the value from the year of the final wave of the study. Due to the nature of age, i.e. that everybody varies at the same rate, it is proposed to treat the variable as time-invariant. It is important to note that these are simply computed values and may not represent the age that the participant was when they completed the questionnaire, but simply the age they were when wave E began.

Sex

This dichotomous variable has complete data, with no missing values. The variable was treated as invariant and dichotomous and was located from the Xwavedat file in the UKHLS download pack.

Physical Health

Physical health is measured through the physical component of the SF-12 (Ware, Kosinski, & Keller, 1996). This variable has already been used in earlier analyses in chapter 4. This component, which claims to measure physical health, is drawn from the Short Form 12, a 12 item questionnaire, which is itself a refined version of the SF-36. The SF 12 contains items which aim to capture eight domains of health outcomes, including physical functioning (PF), role-physical (RP), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role-emotional (RE), and mental health (MH). The physical component of the SF-12 is derived from a scoring matrix which encompasses all items of the SF-12 but weights them according to the guidelines laid down in Ware, Kosinski & Keller (1998). The total number of valid responses to at least one wave of data is 65230, and this represents 99.5% of participants who also answered at least one wave of the GHQ-12. This measure is asked at every wave, and scores consist of a value between 1 and 100. As a result, it will be considered as time-varying.

Psychological variables

Personality

Personality was operationalised using the big 5 model, which was outlined above. This variable was only collected at wave C and as a result, was treated as time-invariant for the reasons outlined in the introduction section. Each participant is apportioned a value from one through 7 for each of the components of the big 5, with high values indicating a high level of that construct.

Social variables

Education

Participants were asked what the highest academic qualification they have achieved was. This variable consists of 6 possible options that participants could choose from, ranging from other higher degree (coded 1) through to no qualification (coded 6). Within the UK higher are generally accepted as anything above a bachelor's degree, and as such are likely to encapsulate both PhD and Masters Degree's, however the wording on the prompt card is unclear, and it is possible that participants who failed to understand what was being asked of them may put these degrees under the 'other qualification' category.

The data is ordinal in nature and while academic achievement is collected at various time points and can change through the course of the five waves, the variable used in this section related specifically to the highest qualification that the participant had achieved at the time of wave five being collected. As a result, the variable was treated as time-invariant.

Ethnicity

UKHLS provides a number of responses that participants can respond to when asked about their ethnicity. Only a small number (1.2%) are listed as missing, and these represent participants who refused to give their ethnicity when asked. Demographics of the entire dataset are given in chapter 2, however, the demographics of the participants who are included in this analysis, i.e. those who completed a wave of the GHQ-12 are detailed in the results section. This variable was treated as time-invariant and will be dummy coded to allow each group to be individually analysed.

Financial Situation

Understanding Society asks a number of questions relating to income and financial situation. These range from absolute income figures, also supplied as monthly and weekly income, to subjective measures of both current and expected income in the future. It was decided that absolute figures were too crude a measurement as they failed to account for regional differences in purchasing power as described in the introduction section of this chapter. It was therefore decided that subjective measures were a better measure of financial situation. UKHLS asks two subjective questions on income, one relating to current financial situation and the other relating to expected future income. Both were included as conceptually, ones current financial circumstance is more relevant to the model, but the previously mentioned research (Oskrochi, Bani-Mustafa & Oskrochi, 2018) did suggest that expected financial situation was the greater predictor of mental health. Both financial situations, that of present and anticipated financial situation were measured subjectively of a Likert scale, however, the scales differed. Subjective measures of current financial situation were measured using a 5 point Likert Scale with scores ranging from 1 representing a response of 'living comfortably' to 5 representing 'finding it very difficult financially'. Information on this variable was collected at every wave during the survey.

Job Satisfaction

Job satisfaction is one of the covariates that was selected to investigate socioeconomic circumstances. It consists of participant's responses to questions about their job satisfaction and is measured using a 7 item Likert scale, with 1 representing complete dissatisfaction with their job and 7 representing complete satisfaction. It was measured at all waves of the survey and consequently was treated as time-varying

Satisfaction with leisure time

This variable consists of participant's responses to being asked, how satisfied they are with the amount of leisure they have. It consists of a 7 point Likert scale, with responses of 1 representing complete dissatisfaction with leisure time and 7 representing complete satisfaction. This variable was collected at all stages of the survey and consequently, was treated as time-varying.

Relationships

UKHLS provides two main subsets of variables which provide insight into the relationship status of participants. One subset asks participants to rate the quality of their relationships, the other asks about relationship status, i.e. divorced, married etc. While at first, the quality of relationships variable looked like could have yielded interesting results, when further investigated, it was apparent that it had a number of shortcomings. Firstly, it was only answered by those who were currently in a relationship and secondly should the participant experience a break-up, they would move from being listed in the data as inappropriate. It would not have methodically sound to generate imputed scores for relationships quality for participants who had explicitly mentioned that they were not in a relationship. As a result, the analysis may have been difficult to interpret, and subsequently, the later variable was selected. This variable referred to whether there had been a change in marital status. By collating these changes, it was possible to ascertain if an individual had married, divorced, separated or been widowed during the course of the five waves where data was collected. The variable collected at wave B onward asks participants for change in their marital status. As there are a range of possible responses they are shown in a table below

Table 7. 1

Response Options for 'Change in Marital State' in Understanding Society

Option	Coding
Missing	-9
Inapplicable	-7
Don't know	-1
Single and never in a legally recognised partnership	1
Married	2
In a civil partnership	3
Separated but legally married	4
Divorced	5
Widowed	6
Separated from civil partner	7
Former civil partner	8
Surviving civil partner	9

It was decided that options' 2 and 3', '4 and 7', '6 and 9' and '5 and 8' were comparable and therefore were collapsed into each other. The variables were then organised, using dummy coding into whether a marriage (items 2&3), separation (items 4&7), divorce (5&8) or death (6&9) had occurred.

7.4- Results

7.4.1 Descriptive Statistics

Below the descriptive statistics for all the mentioned variables were listed. These variables are the ones that will be brought forward to analyses in later chapters, so will be referred to subsequently. Following this, fit statistics for transformed variables were presented to ascertain if a linear model was appropriate for each covariate.

Age

Table 7. 2

Descriptive Statistics for Age of Participants in the Understanding Society Database.

Descriptor	N
Mean	50.493
Median	49.000
Std. Deviation	19.080
Range	94.00
Minimum	14.00
Maximum	108.00

N=65531

Missing=37

The table shows that the number of valid participants who had ages recorded for them was 65531, with only 37 participants having no age variable recorded. As previously mentioned, the variable represents the age that participants would have been at the end of wave 5, not the age that they answered the specific questions. The mean and median age of participants was 50.49 and 49, respectively, with a SD of 19.080. Participants ages ranged from 108 to 14.

Sex

Table 7. 3

Descriptive Statistics for Sex of Participants in the Understanding Society Database.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	29725	45.3	45.3	45.3
	Female	35843	54.7	54.7	100.0
	Total	65568	100.0	100.0	

This dichotomous variable has complete data, with no missing values. Shown below are descriptive statics for participants to be analysed in the UKHLS database. There are slightly more females than males (F=54.7% M=45.3%) which is in line with the population of the UK, however, the degree of overrepresentation is such that the sample is slightly over-represented by females.

Physical Health

Table 7. 4

Descriptive Statistics for Physical Health of Participants in the Understanding Society Database.

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
N	44468	39907	40621	39222	37193
Mean	49.632	49.54	49.779	49.663	49.456
SD	11.35	11.289	10.984	11.200	11.210
Range	70.49	70.07	71.65	67.65	71.44

The total number of valid responses to SF-12 Physical component items was 65230, and this represents 99.5% of participants who also answered at least one wave of the GHQ-12. This measure is asked at every wave and scores consist of a value between 1 and 100 and results detailed above show relatively similar means, SD and ranges across the waves however it must be noted that wave one has noticeably more participants than other waves.

Personality

Table 7. 5

Descriptive Statistics for Personality of Participants in the Understanding Society Database Wave C

	Agreeableness	Conscientiousness	Extraversion	Neuroticism	Openness
Valid	40625	40618	40626	40626	40586
Missing	26047	26054	26046	26046	26086
Mean	5.63	5.46	4.59	3.57	4.54
Median	6.00	6.00	5.00	4.00	5.00
Std. Deviation	1.045	1.120	1.301	1.444	1.317

It is important to note that frequently, low scores in one characteristic does not necessarily imply the absence of that characteristic, but may, in fact, imply the presence of a mutually exclusive characteristic. For example, low scores in extroversion may actually imply introversion, rather than the lack of extroverted tendencies, which while a slight difference, should not be overlooked

As can be seen from the table, this variable was only collected at ‘Wave C’ with the numbers of participants responding to each item fluctuating slightly, but ranging between 40626 and 40586. The small discrepancy in scores was accounted for by partially completed questionnaires and refusals to answer specific questions.

High levels of agreeableness and conscientiousness were found with average scores of 5.63 and 5.46 respectively, whereas lower levels of neuroticism with levels of 3.57 were observed in this population.

Social

Education

Table 7. 6

Descriptive Statistics for Educational Attainment of Participants in the Understanding Society Database

	N	%
Degree	14595	22.3
Other higher degree	7286	11.1
A-level etc	14201	21.7
GCSE etc	13764	21.0
Other qualification	6192	9.4
No qualification	9277	14.1
Total	65425	99.8
Missing	253	.4
65568	100.0	

The table shows that only 0.4% of participants had recorded missing values and that within the participants investigated, there was a diverse range of educational attainment. The most populous group consisted of people who had achieved a degree with 22.3% of the respondents stating that this was their highest academic achievement. Participants who achieved GCSE’s and A-levels also made up a similar proportion of the sample with 21% and 21.7% respectively.

Financial Situation

Table 7. 7

Descriptive Statistics for Current Financial Situation of Participants in the Understanding Society Database

	Wave A		Wave B		Wave C		Wave D		Wave E	
	N	%	N	%	N	%	N	%	N	%
living comfortably	11741	17.9	13310	20.3	11951	18.2	12297	18.8	25102	38.3
doing alright	14205	21.7	16744	25.5	15611	23.8	14797	22.6	12292	18.7
just about getting by	12404	18.9	13394	20.4	12205	18.6	10771	16.4	14293	21.8
finding it quite difficult	4211	6.4	3997	6.1	3608	5.5	3127	4.8	9906	15.1
finding it very difficult?	2032	3.1	1744	2.7	1621	2.5	1454	2.2	2735	4.2
Missing	20975	32.0	16379	25.0	20572	31.4	23122	35.3	25102	38.3

Table 7. 8

Descriptive Statistics for Expected Financial Situation of Participants in the Understanding Society Database

	Wave A		Wave B		Wave C		Wave D		Wave E	
	N	%	N	%	N	%	N	%	N	%
better off	11909	18.2	11962	18.2	10448	15.9	10034	15.3	9847	15.0
worse off than you are now	24528	37.4	27680	42.2	24900	38.0	24868	37.9	24263	37.0
About the same?	7302	11.1	8797	13.4	8999	13.7	7038	10.7	5871	9.0
Missing	21829	33.3	17129	26.1	21221	32.4	23628	36.0	25587	39.0

The results for both anticipated future financial situation and self-report responses of current financial situation are shown above. The results for subjective current financial situation show significant levels of missingness throughout with wave E having the highest levels of missingness at 38.3%. This is due to a number of factors, with a small number of participants refusing to answer the question (.4%), however, the largest percentage was by proxy respondents (see chapter 2) who were not allowed to answer self-report questions on their partner's behalf. The largest number of participants at all stages was 'doing alright' and 'just about getting by' which together usually accounted for between 40% and 50% of all responses.

Similar levels of missingness were recorded for the anticipated financial situation; however, there was a slightly higher level of 'don't know' responses which were recorded as missing. This may suggest that a slightly higher proportion of the population misunderstood or were unwilling to provide an answer to this question.

Large numbers of participants stated that they felt that they would be less well off in the future with, depending on the wave, between 38% and 42% of participants stating this. Unexpectedly participants stating that they would be in a similar financial situation in the future were the smallest group with this group accounting for 9% to 13.7%.

Ethnicity

Table 7. 9

Descriptive Statistics for Ethnicity of Participants in the Understanding Society Database

	Frequency	Percent	Valid Percent
Missing	770	1.2	1.2
British/English/Scottish/Welsh/Northern Irish	51069	77.9	77.9
Irish	1368	2.1	2.1
Gypsy or Irish traveller	20	.0	.0
Any other white background	1832	2.8	2.8
White and black Caribbean	434	.7	.7
White and black African	185	.3	.3
White and Asian	267	.4	.4
Any other mixed background	223	.3	.3
Indian	2140	3.3	3.3
Pakistani	1654	2.5	2.5

Bangladeshi	1143	1.7	1.7
Chinese	334	.5	.5
Any other Asian background	716	1.1	1.1
Caribbean	1190	1.8	1.8
African	1553	2.4	2.4
any other black background	103	.2	.2
Arab	285	.4	.4
Any other ethnic group	282	.4	.4
Total	65568	100.0	100.0

UKHLS provides a number of responses that participants can respond to when asked about their ethnicity. Only a small number (1.2%) are listed as missing, and these represent participants who refused to give their ethnicity when asked. While demographics of the entire dataset are given in chapter 2, the demographics of the participants who are included in this analysis, i.e. those who completed a wave of the GHQ-12 are shown above. While the overwhelming number of participants identified as British, English, Scottish, Welsh or Northern Irish, accounting for 77.9% of the sample population. Some groups are so small that they are unlikely to give statistically significant results such as 'gipsy or Irish traveller', who accounted for less than 0.1% of the sample population. It is intended to retain these participants in the analysis as, while they are unlikely to give statistically significant results, results can be interpreted in the light of this. As was previously mentioned, the UK was described as having a large number of ethnic minorities, however, each of these groups only accounts for a small percentage of the total population. These groups have been purposefully oversampled in order to capture variation in these demographics and then weighted back to ensure that the sample is representative of the general UK population.

Job Satisfaction

Table 7. 10

Descriptive Statistics for Job Satisfaction of Participants in the Understanding Society Database

	Wave A		Wave B		Wave C		Wave D		Wave E	
	N	%	N	%	N	%	N	%	N	%
completely dissatisfied	776	1.2	600	.9	543	.8	471	.7	479	.7
mostly dissatisfied	999	1.5	959	1.5	800	1.2	748	1.1	722	1.1
somewhat dissatisfied	1925	2.9	2147	3.3	1966	3.0	1789	2.7	1708	2.6
neither satisfied or dissatisfied	1809	2.8	2027	3.1	2306	3.5	2212	3.4	2138	3.3
somewhat satisfied	4194	6.4	5461	8.3	5921	9.0	5639	8.6	5837	8.9
mostly satisfied	10569	16.1	11926	18.2	9489	14.5	9080	13.8	8465	12.9
completely satisfied	4615	7.0	4558	7.0	4421	6.7	4114	6.3	3863	5.9
Total	24887	38.0	27678	42.2	25446	38.8	24053	36.7	23212	35.4
Missing	40681	62.0	37890	57.8	40122	61.2	41515	63.3	42356	64.6

The results shown in the above table show the waves performed relatively uniformly and that the percentages of each response by wave were relatively similar. There were significant levels of missingness in the data for this variable, especially when compared with other measures of socioeconomic, status. Missingness was relatively high in this variable, however, it must be acknowledged that like other variables there were proxy respondents and some participants refused, however, participants who were not employed, either through retirement, unemployment or any other reason would also be counted as inapplicable and therefore listed as missing.

Satisfaction with Leisure Time

Table 7. 11

Descriptive Statistics for Satisfaction with Leisure Time of Participants in the Understanding Society Database

	Wave A		Wave B		Wave C		Wave D		Wave E	
	N	%	N	%	N	%	N	%	N	%

completely dissatisfied	1640	2.5	1639	2.5	1813	2.8	1910	2.9	1766	2.7
mostly dissatisfied	2718	4.1	3039	4.6	3796	5.8	3452	5.3	3221	4.9
somewhat dissatisfied	5708	8.7	6457	9.8	5875	9.0	5662	8.6	5659	8.6
neither satisfied or dissatisfied	5261	8.0	5764	8.8	5537	8.4	5593	8.5	5562	8.5
somewhat satisfied	8063	12.3	8420	12.8	7838	12.0	6664	10.2	6673	10.2
mostly satisfied	10285	15.7	11141	17.0	10565	16.1	10352	15.8	9227	14.1
completely satisfied	5810	8.9	6923	10.6	5219	8.0	5262	8.0	5107	7.8
Total	39485	60.2	43383	66.2	40643	62.0	38895	59.3	37215	56.8
Missing	26083	39.8	22185	33.8	24925	38.0	26673	40.7	28353	43.2

Every wave exhibited relatively similar structure to their responses with similar levels of missingness as well as similar percentages of responses for each category. Most participants responded as ‘mostly satisfied’ at each wave with responses ranging from 14.1% of responses at wave 5 to 17% at wave 2. Relatively high levels of missingness were observed, and upon further investigation, this was found to have been caused by large numbers of proxy respondents. Furthermore, large numbers of participants were judged as inappropriate for this item due to employment status.

Relationships

Table 7. 12

Descriptive Statistics for Relationships of Participants in the Understanding Society Database

Condition	Occurrences over 5 years
Married	1041
Separated	316
Divorced	660
Widowed	549

This variable identified participants who experienced a change in their marital status over the course of the five waves investigated. During this time 1041 participants got married, 316 separated for their legally married partner, 660 divorced and 549 were widowed. Of all the participants investigated, only 2,566 experienced a change in their marital status, which only accounts for less than 4% of the sample population. This may provide methodological issues which will be discussed in the discussion section.

7.4.2- Fit Statistics

Table 7. 13

Fit Statistics for a Linear Model for all Time-Varying Covariates

	Job satisfaction	Present Financial Situation	Future Financial Situation	Satisfaction with Leisure Time	Physical Health
Df	10	10	10	10	10
Chi-square	186.120	326.780	498.251	141.634	349.630
P	<0.000	<0.000	<0.000	<0.000	<0.000
RMSEA	0.021	0.022	0.027	0.014	0.025
90% CI	0.018-0.023	0.020-0.024	0.025-0.029	0.012-0.16	0.023-0.027
CFI	0.979	0.992	0.967	0.992	0.992
TLI	0.979	0.992	0.967	0.992	0.992
SRMR	0.036	0.020	0.025	0.013	0.057

90% CI= confidence intervals at 90%

RMSEA= Root Mean Square Error of Approximation

CFI= Comparative Fit Index

TLI= Tucker Lewis Index

SRMR= Standardised Root Mean Square Residual

Table 7.13 shows the fit statistics for a linear interpretation of each of the time-varying covariates. All models demonstrated good fit with CFI and TLI scores ranging from 0.967 to 0.992. RMSEA and SRMR scores also indicated good fit with low scores also fall below accepted thresholds as laid down in Jöreskog & Sörbom (1993).

7.5- Discussion

The covariates identified were chosen because of their correspondence with the biopsychosocial model of mental health (see introduction). The specific reasons for selection are detailed in each covariates section in the introduction section of this chapter, however, efforts were made to ensure that as many of the aspects of the model were represented in the covariates selected. This had to be done within the restrictions of the data available, however, and as a result, only one appropriate variable was identified correlating to the psychological aspect of the model, while four were selected for the biological aspect and six were selected for the social component.

The covariates identified have all been prepared for further analysis, and those which were treated as time-invariant have been transformed into slopes and intercepts to provide a way of measuring change over time. The descriptive statistics provided do show significant levels of missingness in a number of variables for a number of reasons, primarily proxy respondents (see chapter 2) and inapplicableness. The nature of analysis planned in the final chapter, namely covariate analysis on latent classes, means that participants who have missing data on any of these covariates will be removed from the entirety of the analysis (Muthen & Muthen, 2018). This may result in unacceptably large numbers of participants being removed from the final model as to render it unrepresentative of the general population. This will be further explained in the limitation section below.

The time-varying covariates were found to have exhibited acceptable fit to suggest that a linear model for each would be appropriate. The generated fit statics show that all unidimensional representations of the covariates exhibited good fit, with CFI and TLI scores exceeding the established threshold of 0.9 (Bentler & Bonet, 1980) and even exceeded the more stringent threshold of 0.95 (Hu & Bentler, 1999). RMSEA scores also indicated good fit as all models exceeded the suggested guidelines as laid down in Jöreskog & Sörbom (1993), which suggested that RMSEA scores of <0.08 indicated reasonable fit and <0.05 represented good fit.

7.5.1- Limitations

There are a number of limitations which must be born in mind during this and subsequent chapters. The purpose of this chapter is to prepare covariates for use in analyses which will generate the probability of various covariates having an effect on class membership of classes identified in the earlier chapter. This technique requires full data on each covariate, and if a participant has missing data on any of the covariates, they will be removed from the analysis entirely (Muthen & Muthen, 2018). Within this chapter, large proportions of missingness were identified in some covariates and the reasons for this detailed. Should these variables be included in their current form into a model as proposed, the levels of removed participants may be so large as to render to analysis meaningless.

It is also important to note that this data was collected during the time of the economic downturn of 2007 to 2009. This may have had an effect on certain behaviours such as the noticeable economic pessimism demonstrated when participants were asked if they felt they would be in a stronger or weaker financial position financially in the future. Cohen (2014) also found that levels of divorce were suppressed during the economic crisis and attributed this to the financial barriers that the recession imposed. There has also been evidence that marriage rates were similarly suppressed during this time (Payne, 2014). This may mean that the levels of marriage and divorce are artificially suppressed from what could be expected.

Finally, it is noted that during this time, there was a change in how individuals self-reported health conditions. Frank and Elgar (2014) noticed that in the Canadian population that they investigated, the financial hardship caused by the recession had a statistically significant link with self-reported health. They also found this relationship to be more pronounced in participants with lower ‘social capital’, which in this case related to support networks and socialisation. It is therefore important to note that the timeframe in which this data was collected may have had an effect on the results obtained, and this will be borne in mind in the final chapter.

References

Agudelo-Suárez, A. A., Ronda, E., Vázquez-Navarrete, M. L., García, A. M., Martínez, J. M., & Benavides, F. G. (2013). Impact of economic crisis on mental health of migrant workers: what happened with migrants who came to Spain to work?. *International journal of public health*, *58*(4), 627-631.

Allan, J. F., McKenna, J., & Dominey, S. (2014). Degrees of resilience: profiling psychological resilience and prospective academic achievement in university inductees. *British Journal of Guidance & Counselling*, *42*(1), 9-25.

Argyle, M., Martin, M., & Crossland, J. (1989). Happiness as a function of personality and social encounters. *Recent advances in social psychology: An international perspective*, 189-203.

Argyle, M., Martín, M., & Crossland, J. (1989). Happiness as a function of personality and social encounters In: Forgas JP, Innes JM, editors. Recent advances in social psychology: An international perspective.

Awang, Z. (2012). Structural equation modeling using AMOS graphic. Penerbit Universiti Teknologi MAR

Batty, G. D., Der, G., Macintyre, S., & Deary, I. J. (2006). Does IQ explain socioeconomic inequalities in health? Evidence from a population based cohort study in the west of Scotland. *BMJ*, *332*(7541), 580-584.

Benning, T. B. (2015). Limitations of the biopsychosocial model in psychiatry. *Advances in Medical Education and practice*, *6*, 347.

Bentler, P. M. (1990). Comparative fit indexes in structural models. Psychological Bulletin, 107, 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Blazer, D. G., Kessler, R. C., McGonagle, K. A., & Swartz, M. S. (1994). The prevalence and distribution of major depression in a national community sample: the National Comorbidity Survey. *The American journal of psychiatry*.

Bonanno, G. A. (2004). Loss, trauma, and human resilience: Have we underestimated the human capacity to thrive after extremely aversive events?. *American psychologist*, *59*(1), 20.

Borrell-Carrió, F., Suchman, A. L., & Epstein, R. M. (2004). The biopsychosocial model 25 years later: principles, practice, and scientific inquiry. *The Annals of Family Medicine*, *2*(6), 576-582.

Brebner, J., Donaldson, J., Kirby, N., & Ward, L. (1995). Relationships between happiness and personality. *Personality and Individual Differences*, *19*(2), 251-258.

Busic-Sontic, A., Czap, N. V., & Fuerst, F. (2017). The role of personality traits in green decision-making. *Journal of Economic Psychology, 62*, 313-328.

Butler, C. C., Evans, M., Greaves, D., & Simpson, S. (2004). Medically unexplained symptoms: the biopsychosocial model found wanting. *Journal of the Royal Society of Medicine*, *97*(5), 219-222.

Cannon, M. S., & Locke, B. Z. (1977). Being black is detrimental to one's mental health: Myth or reality?. *Phylon (1960-), 38*(4), 408-428.

Carver, C. S. (1998). Resilience and thriving: Issues, models, and linkages. *Journal of social issues*, *54*(2), 245-266.

Cattell, R. B. (1950). Personality: A systematic theoretical and factual study.

Chan, R., & Joseph, S. (2000). Dimensions of personality, domains of aspiration, and subjective well-being. *Personality and Individual differences*, *28*(2), 347-354.

Clough, P., Earle, K., & Sewell, D. (2002). Mental toughness: The concept and its measurement. *Solutions in sport psychology*, 32-43.

Cohen, P. N. (2014). Recession and divorce in the United States, 2008–2011. *Population research and policy review*, *33*(5), 615-628.

Connor, K. M., & Davidson, J. R. (2003). Development of a new resilience scale: The Connor-Davidson resilience scale (CD-RISC). *Depression and anxiety*, *18*(2), 76-82.

Corrigan, P. W., Giffort, D., Rashid, F., Leary, M., & Okeke, I. (1999). Recovery as a psychological construct. *Community mental health journal*, *35*(3), 231-239.

Costa, P. T., & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: happy and unhappy people. *Journal of personality and social psychology*, *38*(4), 668.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of personality assessment*, *49*(1), 71-75.

Economou, M., Madianos, M., Peppou, L. E., Patelakis, A., & Stefanis, C. N. (2013). Major depression in the era of economic crisis: a replication of a cross-sectional study across Greece. *Journal of affective disorders*, *145*(3), 308-314.

Engel G. L., (1977), The need for a new medical model: a challenge for biomedicine, Science 196: 129-136

Feist, J., & Feist, G. (2015). J.(2009). Theories of personality.

Fischer, J. (1969). Negroes and whites and rates of mental illness: Reconsideration of a myth. *Psychiatry*, *32*(4), 428-446.

Frank, C., Davis, C. G., & Elgar, F. J. (2014). Financial strain, social capital, and perceived health during economic recession: a longitudinal survey in rural Canada. *Anxiety, Stress, & Coping*, *27*(4), 422-438.

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American psychologist*, *56*(3), 218.

Friedli, L., & World Health Organization. (2009). *Mental health, resilience and inequalities* (No. EU/08/5087203). Copenhagen: WHO Regional Office for Europe:.

Furnham, A., & Brewin, C. R. (1990). Personality and happiness. *Personality and individual differences*, *11*(10), 1093-1096.

Garmezy, N. (1993). Children in poverty: Resilience despite risk. *Psychiatry*, *56*(1), 127-136.

Gatchel, R. J. (1996). Psychological disorders and chronic pain: cause-and-effect relationships.

Gatchel, R. J., Peng, Y. B., Peters, M. L., Fuchs, P. N., & Turk, D. C. (2007). The biopsychosocial approach to chronic pain: scientific advances and future directions. *Psychological bulletin*, *133*(4), 581.

Ghaemi, S. N. (2009). The rise and fall of the biopsychosocial model. *The British Journal of Psychiatry*, *195*(1), 3-4.

Gili, M., Roca, M., Basu, S., McKee, M., & Stuckler, D. (2013). The mental health risks of economic crisis in Spain: evidence from primary care centres, 2006 and 2010. *The European Journal of Public Health*, *23*(1), 103-108.

Glonti, K., Gordeev, V. S., Goryakin, Y., Reeves, A., Stuckler, D., McKee, M., & Roberts, B. (2015). A systematic review on health resilience to economic crises. *PloS one*, *10*(4).

Goldberg, L. R. (1990). An alternative" description of personality": the big-five factor structure. *Journal of personality and social psychology*, *59*(6), 1216.

Hardy, S. E., Concato, J., & Gill, T. M. (2004). Resilience of community-dwelling older persons. *Journal of the American Geriatrics society*, *52*(2), 257-262.

Hauksdóttir, A., McClure, C., Jonsson, S. H., Ólafsson, Ö., & Valdimarsdóttir, U. A. (2013). Increased stress among women following an economic collapse—a prospective cohort study. American journal of epidemiology, 177(9), 979-988.

Hauksdóttir, A., McClure, C., Jonsson, S. H., Ólafsson, Ö., & Valdimarsdóttir, U. A. (2013). Increased stress among women following an economic collapse—a prospective cohort study. *American journal of epidemiology*, *177*(9), 979-988.

Hayes, N., & Joseph, S. (2003). Big 5 correlates of three measures of subjective well-being. *Personality and Individual differences*, *34*(4), 723-727.

Hinrichsen, G. A. (1992). Recovery and relapse from major depressive disorder in the elderly. *The American journal of psychiatry*.

Horsburgh, V. A., Schermer, J. A., Veselka, L., & Vernon, P. A. (2009). A behavioural genetic study of mental toughness and personality. *Personality and individual differences*, *46*(2), 100-105.

Hughes, M., & Demo, D. H. (1989). Self-perceptions of Black Americans: Self-esteem and personal efficacy. *American Journal of Sociology*, *95*(1), 132-159.

Jackson, J. S., Torres, M., Caldwell, C. H., Neighbors, H. W., Nesse, R. M., Taylor, R. J., ... & Williams, D. R. (2004). The National Survey of American Life: A study of racial, ethnic and cultural influences on mental disorders and mental health. *International journal of methods in psychiatric research*, *13*(4), 196-207.

Jackson, J. S., Torres, M., Caldwell, C. H., Neighbors, H. W., Nesse, R. M., Taylor, R. J., ... & Williams, D. R. (2004). The National Survey of American Life: A study of racial, ethnic and cultural influences on mental disorders and mental health. *International journal of methods in psychiatric research*, *13*(4), 196-207.

Jöreskog, K. G., & Sörbom, D. (1993). LISREL 8: Structural equation modeling with the SIMPLIS command language. Chicago, IL: Scientific Software International.

Joseph, S., & Lewis, C. A. (1998). The Depression–Happiness Scale: Reliability and validity of a bipolar self-report scale. *Journal of clinical psychology*, *54*(4), 537-544.

Katikireddi, S. V., Niedzwiedz, C. L., & Popham, F. (2012). Trends in population mental health before and after the 2008 recession: a repeat cross-sectional analysis of the 1991–2010 Health Surveys of England. *BMJ open*, *2*(5), e001790.

Kleiner, R. J., Tuckman, J., & Lavell, M. (1960). Mental disorder and status based on race. *Psychiatry*, *23*(3), 271-274.

Kothari, C. L., Butkiewicz, R., Williams, E. R., Jacobson, C., Morse, D. S., & Cerulli, C. (2014). Does gender matter? Exploring mental health recovery court legal and health outcomes. *Health & justice*, *2*(1), 12.

Kramer, M., Rosen, B., & Willis, E. (1973). Definitions and distributions of mental disorders in a racist society. *Racism and mental health*, 353-459.

Lamond, A. J., Depp, C. A., Allison, M., Langer, R., Reichstadt, J., Moore, D. J., ... & Jeste, D. V. (2008). Measurement and predictors of resilience among community-dwelling older women. *Journal of psychiatric research*, *43*(2), 148-154.

Lewinsohn, P. M., Zeiss, A. M., & Duncan, E. M. (1989). Probability of relapse after recovery from an episode of depression. *Journal of Abnormal Psychology*, *98*(2), 107.

Luthar, S. S., Cicchetti, D., & Becker, B. (2000). Research on resilience: Response to commentaries. *Child development*, *71*(3), 573-575.

Marmor, J., & Pumpian-Mindlin, E. (1950). Toward an integrative conception of mental disorder. *Journal of Nervous and Mental Disease*.

Masten, A. S. (2001). Ordinary magic: Resilience processes in development. *American psychologist*, *56*(3), 227.

Mehta, M., Whyte, E., Lenze, E., Hardy, S., Roumani, Y., Subashan, P., ... & Studenski, S. (2008). Depressive symptoms in late life: associations with apathy, resilience and disability vary between young-old and old-old. *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences*, *23*(3), 238-243.

Mehta, M., Whyte, E., Lenze, E., Hardy, S., Roumani, Y., Subashan, P., ... & Studenski, S. (2008). Depressive symptoms in late life: associations with apathy, resilience and disability vary between young-old and old-old. *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences*, *23*(3), 238-243.

Menon, V., Shanmuganathan, B., Thamizh, J. S., Arun, A. B., Kuppili, P. P., & Sarkar, S. (2018). Personality traits such as neuroticism and disability predict psychological distress in medically unexplained symptoms: A three-year experience from a single centre. *Personality and mental health*, *12*(2), 145-154.

Mirowsky, J., & Ross, C. E. (1980). Minority status, ethnic culture, and distress: A comparison of Blacks, Whites, Mexicans, and Mexican Americans. *American Journal of Sociology*, *86*(3), 479-495.

Mischel, W., & Shoda, Y. (1999). Integrating dispositions and processing dynamics within a unified theory of personality: The cognitive-affective personality system. *Handbook of personality: Theory and research*, *2*, 197-218.

Muthén, L. K., & Muthén, B. (2016). Mplus. *The comprehensive modelling program for applied researchers: user's guide*, *5*.

Netuveli, G., Wiggins, R. D., Montgomery, S. M., Hildon, Z., & Blane, D. (2008). Mental health and resilience at older ages: Bouncing back after adversity in the British Household Panel Survey. *Journal of Epidemiology & Community Health*, *62*(11), 987-991.

Nittle, N K. (2013, March 25). Understanding the Difference Between Race and Ethnicity. Retrieved from www.thoughtco.com/difference-between-race-and-ethnicity-2834950 / (accessed May 30, 2020).

Nygren, B., Aléx, L., Jonsén, E., Gustafson, Y., Norberg, A., & Lundman, B. (2005). Resilience, sense of coherence, purpose in life and self-transcendence in relation to perceived physical and mental health among the oldest old. *Aging & mental health*, *9*(4), 354-362.

O'Leary, V. E., & Ickovics, J. R. (1995). Resilience and thriving in response to challenge: an opportunity for a paradigm shift in women's health. *Women's health (Hillsdale, NJ)*, *1*(2), 121-142.

Ong, L. (2003). *The Big Mac Index: applications of purchasing power parity*. Springer.

Oskrochi, G., Bani-Mustafa, A., & Oskrochi, Y. (2018). Factors affecting psychological well-being: Evidence from two nationally representative surveys. *PloS one*, *13*(6).

Pam M.S., “RECOVERY,” in PsychologyDictionary.org, April 28, 2013, https://psychologydictionary.org/recovery/ (accessed May 30, 2020).

Payne, K. K. (2014). *The Marriage Rate and the Great Recession*. FP-14-18. Bowling Green, OH: National Center for Family and Marriage Research. http://www. bgsu. edu/content/dam/BGSU/college-of-arts-and-sciences/NCFMR/documents/FP/FP-14-18-marriage-rate-recession. pdf.

Peng, L., Zhang, J., Li, M., Li, P., Zhang, Y., Zuo, X., ... & Xu, Y. (2012). Negative life events and mental health of Chinese medical students: the effect of resilience, personality and social support. *Psychiatry research*, *196*(1), 138-141.

Peng, L., Zhang, J., Li, M., Li, P., Zhang, Y., Zuo, X., ... & Xu, Y. (2012). Negative life events and mental health of Chinese medical students: the effect of resilience, personality and social support. *Psychiatry research*, *196*(1), 138-141.

Porter, J. R., & Washington, R. E. (1979). Black identity and self-esteem: A review of studies of Black self-concept, 1968-1978. *Annual Review of Sociology*, *5*(1), 53-74.

Robins, L. N., & Regier, D. A. (1991). *Psychiatric disorders in America: The epidemiological catchment area study*.

Rutter, M. (2006). Implications of resilience concepts for scientific understanding. *Annals of the New York Academy of Sciences*, *1094*(1), 1-12.

Steinhardt, M., & Dolbier, C. (2008). Evaluation of a resilience intervention to enhance coping strategies and protective factors and decrease symptomatology. *Journal of American college health*, *56*(4), 445-453.

Strong, W. B., Malina, R. M., Blimkie, C. J., Daniels, S. R., Dishman, R. K., Gutin, B., ... & Rowland, T. (2005). Evidence based physical activity for school-age youth. *The Journal of pediatrics*, *146*(6), 732-737.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2009). Effective pre-school and primary education 3-11 (EPPE 3-11) final report from the primary phase: Pre-school, school, and family influences on children's development during key stage 2 (age 7-11). *London, England: University of London, Institute of Education*.

Tupes, E. C. (81). i Christal, RE (1961.). *Recurrent personality factors based on trait ratings*, 61-67.

Twenge, J. M., & Crocker, J. (2002). Race and self-esteem: meta-analyses comparing whites, blacks, Hispanics, Asians, and American Indians and comment on Gray-Little and Hafdahl (2000).

Ullmann, L. P., & Krasner, L. (1975). *A psychological approach to abnormal behavior*. Prentice-Hall.

Veldman, K., Bültmann, U., Stewart, R. E., Ormel, J., Verhulst, F. C., & Reijneveld, S. A. (2014). Mental health problems and educational attainment in adolescence: 9-year follow-up of the TRAILS study. *PLoS One*, *9*(7).

Von Bertalanffy, L. (1968). General system theory: Foundations. *Development, applications*, *3*.

Wagnild, G. M., & Young, H. (1993). Development and psychometric. *Journal of nursing measurement*, *1*(2), 165-17847.

Wells, M. (2012). Resilience in older adults living in rural, suburban, and urban areas. *Online Journal of Rural Nursing and Health Care*, *10*(2), 45-54.

Chapter 8 - Explaining Mental Health Trajectories, Covariate Analysis

8.1 Abstract

Introduction

Following the identification of an appropriate class structure in Chapter 6 and the preparation of appropriate covariates in Chapter 7, this chapter's aim was to explain why various participants exhibited different trajectories over time. This was done by examining the likelihood that changes in the various covariates would have on membership of the various classes.

Methods

Participants of waves one through five were investigated in this chapter. Odds ratios of class membership for changes in a wide array of biological, social and psychological covariates were calculated using the R3step technique in order to retain the integrity of the model identified in Chapter 6. Time-varying covariates were transformed into measures of change and initial value to facilitate investigation of change over time. Missing data were handled using a mixture of MLR for the GHQ-12 classes and data imputation techniques for the covariates.

Results

The results generated indicated that variables from all aspects of the biopsychosocial model had a statistically significant impact on class membership. In relation to psychological variables, neuroticism was found to have the largest impact on class membership. Biological variables were found to have a statistically significant

impact on mental health however, these effects were generally small. Social variables were found to exhibit more pronounced impacts on class membership. The largest odds ratios were found to be the rapidity of deteriorating financial situation and job satisfaction, however, ethnicity was found not to have a statistically significant impact on class membership.

Conclusion

The results showed that all aspects of the biopsychosocial model had an impact on the class membership of participants within the sample, however, of the variables selected, social variables affected class membership the most. The pronounced impact of social variables such as income was consistent with previous literature, as was the extent to which personality traits impacted mental health, however, the results for ethnic minorities was inconsistent with much previous literature and warranted further research.

8.2- Introduction

The purpose of this chapter was to explain the reasons why different UKHLS participants displayed different patterns of GHQ-12 score over time. During Chapter 6, a four-class solution was extracted from the data, and these classes represented four distinct patterns of GHQ-12 scores over time listed below

- Stable and relatively low GHQ-12 scores- referred to as the reference group
- Stable but high GHQ-12 scores – referred to as the high stable group
- Steadily decreasing GHQ-12 scores – referred to the recovery group
- Steadily increasing GHQ-12 scores – referred to as the deteriorating group

In this chapter, the classes identified above were regressed onto a wide range of covariates in an effort to explain the reasons why participants exhibit the trajectories associated with each class. In Chapter 7, the appropriateness of various covariates was established, and time-variant covariates were converted into a form which was compatible with regression analysis. This analysis facilitated investigation into how changes in the identified covariates correlated with class membership and for time-invariant covariates to provide a method of measuring how change over time was associated with class membership.

8.2.1- Review of Relevant Literature

Relevant literature for each of the covariates is detailed in the previous chapter and in the interests of avoiding duplication will not be repeated in this chapter. The section below will review literature which has attempted to explain different mental health trajectories in an effort to contextualize findings from this chapters analysis. Much of the identified literature which investigated biological relationships focused on the recovery of a participant following a traumatic event such as an injury and used statistical techniques which investigated the discriminant properties of covariates on class membership. For an explanation of discriminant validity, see Chapter 4.

Helgeson, Snyder and Seltman (2004) conducted research into the health following a cancer diagnosis. They conducted two analyses simultaneously, using latent profile analysis to investigate mental and physical health trajectories following cancer diagnoses. The classes were not linear, but the largest class consisting of 43.2% of participants represented stable good mental health scores over time. Two other classes represented a general trend of increasing and decreasing scores respectively, and the final class represented a relatively similar trajectory to the first class but with a lower intercept. For both physical and mental health, the researchers identified four trajectories of both physical and mental health. They found, that age was a predictor of physical functioning class but not of mental class membership. They noticed variables which measured what they defined as ‘personal resources’, including concepts such as self-image and optimism alongside variables related to ones ‘social network’, were able to distinguish between the physical and mental health classes.

Strohschein (2005) investigated the effect that parental divorce had upon the mental health of children aged 4-7 in Canada. By investigating the growth curves of participants, they found that children started to exhibit poor mental health prior to the divorce of their parents and that after the divorce there was an increase in the prevalence of mental health conditions such as depression and anxiety. After controlling for socioeconomic status and psychosocial resources, the effects of pre-divorce deviations from the norm were mitigated, however, the effects post-divorce were not accounted for. This would suggest that the act of divorce itself has specific mental health effects on the children of divorced parents that cannot be explained by other factors.

Kaptein et al. (2006) investigated the trajectories of mental health following cardiovascular conditions. They identified five trajectories of depressive symptoms following a cardiovascular condition and identified that participants who exhibited ‘significant and increasing depressive symptoms’ were more likely to experience a cardiovascular event in relation to the other classes. They used sociodemographic variables as control variables and found that they did not alter participant’s trajectories when outcomes were controlled for, suggesting that these variables did not affect trajectories. This research is distinct from what is proposed in this chapter, as trajectories of participants were investigated at the individual level, and participants were not grouped based on their trajectories. It does, however, attempt to explain how sociodemographic variables can affect mental health trajectories.

Kariuki et al. (2011) conducted research into the mental health trajectories of individuals after the onset of a disability. This research utilised data from the Household, Income and Labour Dynamics in Australia dataset. The researchers identified a specific cohort of participants who reported the onset of a physical disability during waves one through seven of the study. The research used growth mixture modelling techniques to extract three trajectories from participants who experienced the onset of an illness. The largest class, consisting of 65% of respondents exhibited positive mental health both prior and post the onset of the illness, however, two other classes, one representing poor mental health both prior to the onset of illness and another representing deterioration of mental health following the onset of an illness were extracted consisting of 19% and 16% respectively. This research was notable as it was expected that the onset of an illness would act as a driver of mental health change however this class was relatively small and suggests that the effect of the onset of an illness may not be as large as previously expected.

Van Leeuwen et al. (2012) used growth mixture modelling techniques to track mental health trajectories of participants following spinal injuries. This analysis identified five trajectories of mental health following such an injury. Once the class model was established, a number of covariates were modelled using multinomial regression techniques to analyse which of the covariates were able to discriminate between the various trajectories. This research was able to identify that covariates such as gender and educational attainment were found to be predictors which could differentiate between the various classes.

They argued that the impact on mental health of having a spinal injury was a consequence of that the restrictions that this placed on an individual’s ability to participate in activities that they used to. They also found that social support was not found to be a predictor of class membership in this analysis, which the author mentioned had stood in contrast to other research (North, 1999).

Morack et al. (2013) investigated the trajectories of mental health and how they related to age and personality traits using individual growth curves. This research was pertinent as it investigated health trajectories as they related to both biological and psychological covariates. This research utilised data from waves 1-10 of the Household, Income and Labour Dynamics in Australia dataset. They reported that as participants aged, trajectories of mental health remained stable in the participants of this study, however, participants who reported higher levels of neuroticism and lower extraversion and conscientiousness were less likely to maintain the stable trajectory usually exhibited.

In relation to social covariates research, a wide range of social variables which relate to the relationships between mental health trajectories and mental health have been investigated. Studies which investigated variables relevant to the variables in this analysis are detailed below.

Bell (2014) conducted research which investigated mental health trajectories using the British Household Panel Study, the precursor to Understanding Society. While the primary purpose of this research was to counter claims that the relationship between age and mental health was U shaped, the research extracted three classes of response and modelled covariates onto these. These classes all exhibited similar trajectories but had different intercepts. Bell (2014) suggested that ethnicity and education had little effect on GHQ-12 scores and that marriage gender and age were likely to predispose individuals to better mental health, however, this effect decreased with age.

This research was particularly pertinent as it used a dataset similar to that of Understanding Society and while it focused on the presence of cohort effects to discredit the U shaped nature of mental health proposed within the literature, this research did offer explanations for various trajectories based on the covariates mentioned above.

Brisson, Lopez and Yoder (2014) conducted research into the effects of one's perception of their neighbourhood on mental health scores. They investigated the mental health of 2400 randomly selected mothers in Boston, Chicago and San Antonio over three-time points. They analysed the responses using growth curve modelling but did separate participants into groups. They found that while over time, the mental health of the participants investigated steadily improved the trajectory of participants who perceived their neighbourhood to have problems.

Veldman et al. (2014) investigated how educational attainment and employment of individuals affected their mental health scores over time. This research used growth mixture modelling to separate 2230 Dutch children into subpopulations based on their longitudinal self reported mental health and then investigated the educational and employment status of individuals to ascertain if these variables could be used to explain the different trajectories exhibited by various participants. Four trajectories of what they referred to as '*total problems*' were extracted, however different numbers of classes were extracted when the researchers looked at specific issues such as attention problems. These classes were referred to as '*high stable*', '*moderate stable*', '*low stable*' and '*decreasing*'. It was found that participants who exhibited stable and high trajectories of poor mental health were likely to have poor educational attainment and were unlikely to be in full-time education or employment.

Meyrose et al. (2018) investigated how an individual's mental health trajectories were determined by the educational attainment of their mothers. This research identified three mean trajectories of mental health based on the educational attainment of participant's mothers. While generally, the trajectories of children with high maternal academic achievement were higher than those with lower achievements, these trajectories exhibited different patterns of scores. The difference between mental health scores in children with high maternal academic achievement was most pronounced in younger children and the difference between the mean scores reduced as the participant aged. The effect of maternal educational attainment was also more pronounced in participants who were from single-parent or families where a step-parent was present. This research utilised individual growth modelling techniques and averaged the results and did not investigate the presence of latent classes of participants within the population, however, it did provide a contextual basis to how mental health scores can be affected by the circumstances of members of their family.

8.2.2- Statistical Procedures

The proposed research hinges upon statistical techniques which regress latent classes onto covariates. It was felt beneficial to provide an overview and historical context for these techniques at this juncture as it is pivotal to the proposed research in this chapter.

Initially, the accepted way of investigating class membership in relation to auxiliary variables was to combine the original latent class or GMM model which extracted the classes with a latent class regression model which could be estimated using a maximum likelihood estimator (Asparouhov & Muthen, 2014). This was later referred to as a one-step approach. Vermunt (2010) identified that this might not be optimal as the inclusion of a latent class regression model may alter the latent class formation, and the generated classes could lose their meaningfulness as a result. In response to this shortcoming, Vermunt (2010) developed a technique based on previous work by Bolck et al. (2004) which could estimate a latent class structure independently of any covariates which were added to a model.

In this three-step approach, the population is initially sorted into classes using only the latent class indicator variables. The class structure generated from this analysis forms the basis of all further analysis. The second step involves calculating the 'most likely' or reference class based on the class structure calculated in the first step. The final step involves regressing this most likely class onto the covariates while accounting for misclassification in the second step.

8.2.3- Hypotheses

This chapter's goal was to explain the mental health trajectories of participants using covariates. By using the biopsychosocial model (Gatchel, 1996) model as a framework it was implicit that all components of the model would be expected to affect trajectories to some degree, therefore, the first hypothesis of this chapter was that covariates from all aspects of the biopsychosocial model (Gatchel, 1996) would display statistically significant relationships with class membership. Research has suggested however that while all aspects of the model should display statistically significant relationships with class membership, social variables should be the strongest predictor of mental health (Oskrochi, Bani-Mustafa & Oskrochi, 2018) and it is expected that social variables will be the strongest predictor of class membership. As detailed in Oskrochi, Bani-Musafa & Oskrochi (2018), the effect that social variables predict mental health varied but if this research was to be consistent with other findings, anticipated financial state in the future should be the strongest predictor of class membership.

In relation to psychological variables, research has differed as to which personality traits were statistically significantly associated with mental health (Costa & McCrae, 1980; Furnham & Brewin, 1990; Brebner et al., 1995; Chan & Joseph, 2000; Hayes & Joseph, 2003, Menon et al., 2017), however, a number of consistent themes emerge. Neuroticism was found to be a predictor of poor mental health in all cases mentioned above and therefore would be likely to predispose individuals to be members of either consistently or deteriorating classes of mental health. All other personality traits of those investigated were found to be associated with improving mental health and therefore would be expected to predispose individuals to membership of classes which denote either stable good mental health or improving classes.

All of the biological variables investigated have been shown in the literature to display statistically significant relationships with mental health, namely age (Aldwin et al., 1989) sex, (Gili et al. 2013) and physical health (Fox, 2007; Abu-Omar, 2004). It is hypothesised that these variables will all exhibit statistically significant relationships with mental health class membership, with younger, more physically healthy and male participants being more predisposed to membership of classes which indicate good or improving mental health.

8.3- Methods

8.3.1- Data

The data in this analysis was generated from waves one through five of UKHLS, as per the previous two chapters. In total, 65558 participants who completed at least one wave of the GHQ-12 at one time point were analysed. The participants of this dataset were subdivided into four linear latent classes of responses as per the growth mixture modelling completed in Chapter 5. These four classes represented a reference group of low, stable mental health scores, two classes that represented steadily increasing and decreasing scores, respectively. A fourth class denoting high stable scores was also extracted.

This dataset had a wide range of biological, social and psychological covariates merged into it. A more detailed discussion of the variables used in this analysis was provided in Chapter 7. Some of these variables varied over time, whereas others remained stable. Methodological issues relating to time-varying and time-invariant covariates are listed below in section 8.4.3, and 8.4.4 Missing data were handled using a variety of techniques which is detailed in section 8.4.5

8.3.1.1- Covariates

The variables selected are detailed in the previous chapter and therefore, were not discussed in this section to avoid duplication. It was felt beneficial to aid interpretation to provide a table which details the directionality and range of scoring matrices used (see table 8.1).

Table 8. 1

Scoring Metrics for Each Covariate

Covariate	Scoring interpretation
Age	Continuous variable computed by (2012- year of birth)
Qualification	1=higher degree 7= no qualification
Sex	1= male 2= female
I-Job Satisfaction	**1= completely dissatisfied 7= completely satisfied**
S- Job Satisfaction	**+ve = improving satisfaction**
I-Financial Future	**1= living comfortably 5= finding it very difficult**
S-Financial Future	**+ve = deteriorating financial situation**

I-Financial Present	**1 better off 2 same 3 worse**
S-Financial Present	**+ve=deteriorating financial situation**
I-Leisure Time	**1= completely dissatisfied= 7 completely satisfied**
S-Leisure Time	**+ve= improving satisfaction**
I-Physical Health	**0= low health 100= high health**
S-Physical Health	**+ve= improving health**
Ethnicity- Irish	1= yes 0= no
Ethnicity- Gypsie or Irish Traveller	1= yes 0= no
Ethnicity- Any other White Background	1= yes 0= no
Ethnicity- White and Black Caribbean	1= yes 0= no
Ethnicity- White and Black African	1= yes 0= no
Ethnicity- White and Asian	1= yes 0= no
Ethnicity- Any other Mixed Background	1= yes 0= no
Ethnicity- Indian	1= yes 0= no
Ethnicity- Pakistani	1= yes 0= no
Ethnicity- Bangladeshi	1= yes 0= no
Ethnicity- Chinese	1= yes 0= no
Ethnicity- Any other Asian Background	1= yes 0= no
Ethnicity- Caribbean	1= yes 0= no
Ethnicity- African	1= yes 0= no
Ethnicity- Any other White Background	1= yes 0= no
Ethnicity- Arab	1= yes 0= no
Ethnicity- Any other ethnic group	1= yes 0= no
Mother not working when 14	1= yes 0= no
Mother deceased when 14	1= yes 0= no
Mother absent when 14	1= yes 0= no
Father not working when 14	1= yes 0= no
Father deceased when 14	1= yes 0= no
Father absent when 14	1= yes 0= no
Married over last 4 years	1= yes 0= no
Divorced over last 4 years	1= yes 0= no
Separated over last 4 years	1= yes 0= no
Widowed over last 4 years	1= yes 0= no
Personality – Agreeableness	1= does not apply to me at all 7= applies perfectly
Personality – Conscientiousness	1= does not apply to me at all 7= applies perfectly
Personality – Extraversion	1= does not apply to me at all 7= applies perfectly
Personality – Neuroticism	1= does not apply to me at all 7= applies perfectly
Personality – Openness	1= does not apply to me at all 7= applies perfectly

Key
+ve= increasing values
I= intercept
S= slope
Time varying covariates marked in bold

Those variables which are marked in bold were designated as time-varying and have been prepared in Chapter 7 to convert them to a format that is compatible with regression analysis. The methodological issues relating to time-varying and time-invariant covariates are detailed below.

Time invariant covariates

These variables were analysed with a structural equation modelling framework by regressing the classes onto the covariates using the r3step technique, as explained in the introduction section above. The initial class structure that represents the first stage of this process is reported in Chapter 6. The output of this analysis was converted to odds ratios which allowed it to be determined how one unit change in the time-invariant covariate affects class membership and these were reported in tables 8.2-8.4

Time-varying covariates

In Chapter 7, these variables were converted into a format that was compatible with regression analysis. This was done by converting the values collected at the various waves into measures of slope and intercept. These represented a measure of change over time and value at time point one, respectively. These models were tested for fit and demonstrated good fit with a linear model in all cases. Using the aforementioned r3tep technique, the latent classes were regressed onto the slopes and intercepts of these variables to ascertain if one unit increase in the initial starting value or the slope led to an increased or decreased likelihood of various class membership. Odds ratios were calculated and reported in tables 8.2-8.4

8.3.2- Analysis

In order to explain the trajectories displayed by participants of the UKHLS database, the four classes extracted from the growth mixture model conducted in Chapter 5 were regressed onto the covariates mentioned above using the R3step technique. This technique facilitated the retention of the original model extracted in Chapter 5, which may be altered by the inclusion of covariates if more conventional one-step techniques were used (Asparouhov & Muthen, 2014). Following the transformation that the covariates underwent, both time-invariant and varying covariates were analysed simultaneously. By using this technique in MPLUS, logits were generated, which were then transformed into odds ratios for ease of interpretation which determine the likelihood of class membership based on the variability of the auxiliary variable. (Asparouhov & Muthen, 2014). Odds ratios, with 95% confidence intervals were reported in table 8.2-8.4.

8.3.3.- Missing Data

Two methods of handling missing data were used in this technique. These were data imputation and Maximum Likelihood with Robust standard errors (MLR). MLR is discussed in previous chapters, and data imputation is described below.

In order to overcome the difficulties associated with the high levels of missingness present in some of the covariates which is discussed briefly in the 'limitations' section of the previous chapter, data were imputed using the MPLUS data imputation function.

The use of the R3step was essential to retain the integrity of the underlying model identified in Chapter 5. This process, however, requires full data on all cases, with participants who had missing values on any of the covariates being deleted entirely.

This is problematic as the R3step approach was selected partially because of the large number of covariates that would inevitably cause significant changes in the underlying model, however, all these variables have significant volumes of missingness. When analysed, less than half of all participant's responses contained no missingness, and as a result, over half of the participants would have been excluded. The large number of participants excluded raises methodological issues around random missingness, i.e. it would be unlikely that the missingness occurred randomly throughout the sample, therefore, the data would be distorted. The only way to avoid using listwise deletion in MPLUS is by using 'multiple imputation'. This process involves creating multiple datasets of 'plausible' values for missing data, generated from specified variables within the dataset. Data imputation has been described as having statistical properties comparable to maximum likelihood techniques, but having the advantage of being able to be applied to a wider range of techniques (Allison, 2003) This process would alleviate the problems that listwise deletion would bring, however, would also necessitate either rerunning all analysis with imputed data or tolerating two different strategies for the handling of missing data to be present in one analysis. It was decided that this would be tolerated in order to preserve the integrity of previous analyses, especially the latent class analysis. As a result, the auxiliary covariates will be subject to data imputation, whereas the GHQ-12 scores will be subject to MLR, as they have been in all previous chapters.

8.4- Results

While all covariates were analysed in a single model, results are presented in three separate tables relating to their correspondence to the components of the biopsychosocial model (see Chapter 6) to ease in interpretation and reporting.

Confidence intervals which indicated a statistically significant effect, i.e. those which do not encapsulate a 1 (Hicks, 2020) are marked with an asterisk.

8.4.1 Biological Variables

Table 8. 2

Biological Covariates Odds Ratios for Latent Class Membership

Covariate	Class 1 (Recovery) Odds Ratios (95%CI)	Class 2 (High Stable) Odds Ratios (95%CI)	Class 3 (Deteriorating) Odds Ratios (95%CI)
Age	1.02 (1.02-1.03)*	1.02 (1.01-1.03)*	1.01 (1-1.01)*
Sex	1.5 (1.34-1.68)*	1.52(1.27-1.82)*	1.65 (1.43-1.89)*
I-Physical Health	0.95(0.95-0.96)*	0.91 (0.9-0.92)*	0.95(0.94-0.95)*
S-Physical Health	0.94(0.84-1.04)	0.81(0.71-0.93)*	0.84(0.74-0.96)*

I= intercept
S= slope
*= values which were statistically significant at $P<0.05$

Of all the biological variables investigated, only age, sex and the initial value for physical health demonstrated statistically significant relationships with all of the latent classes. In relation to age, the odds ratios exceed 1 in all classes which suggested that with one unit increase in age, the probability of being a member of the reference group decreases and the probability of class membership of the three groups mentioned above increases, albeit by a small amount, with odds ratios of between 1.01 and 1.02. It was noted that all the latent classes examined exhibited higher GHQ-12 scores than the reference group. In relation to sex, with 1 representing male and 2 representing female, the figures suggest that females are more likely to be a member of all groups other than the reference group. In relation to physical health, with low scores indicating poor health, the figures suggest that as physical health improves, the likelihood of being in the reference group increases. It was noted however, that increases in physical health had no discernible difference between being a member of the recovery or deteriorating class with odds ratios of 0.95, respectively. The extent of which the GHQ-12 scores changed over time, i.e. the angle of the slope was found to demonstrate a significant effect of the high stable group and the deteriorating group, but not the recovery group. These relationships were 0.81 and 0.84, respectively suggesting that with one unit increase in the angle of the slope, class membership was less likely in this scenario.

8.4.2 Psychological Variables

Table 8. 3

Psychological Covariates Odds Ratios for Latent Class Membership

Covariate	Class 1 (Recovery) Odds Ratios (95%CI)	Class 2 (High Stable) Odds Ratios (95%CI)	Class 3 (Deteriorating) Odds Ratios (95%CI)
Agreeableness	0.93 (0.83-0.99)*	0.94 (0.85-1.04)	0.92 (0.86-0.99)*
Conscientiousness	0.83 (0.78-0.88)*	0.72 (0.66-0.79)*	0.8 (0.75-0.86)*
Extraversion	0.91 (0.86-0.96)*	0.81 (0.75-0.88)*	0.84 (0.79-0.9)*
Neuroticism	1.47 (1.41-1.53)*	1.87 (1.73-2.02)*	1.58(1.5-1.67)*
Openness	1.02 (0.97-1.07)	1.01 (0.93-1.1)	1.09 (1.02-1.16)*

*= values which were statistically significant

All components of personality displayed statistically significant relationships with one or more of the latent classes, and each component will be detailed in turn. Agreeableness was found to exhibit statistically significant relationships with the recovery group and the deteriorating group but not the high stable class. The odds ratios of 0.93 and 0.92 respectively suggested that participants who demonstrated high levels of agreeableness were less likely to be in classes which denoted increasing or deteriorating GHQ-12 scores.

Conscientiousness was found to exhibit statistically significant relationships with all classes with all odds ratios being less than 1. This would indicate that the more conscientious a participant was, the less likely they were to be a member of the three classes mentioned above.

Extraversion was found to exhibit relationships with all classes mentioned above and like conscientiousness, all odds ratios were less than 1. Similar interpretations of results can be drawn with extraversion as can be made with conscientiousness mentioned above.

Neuroticism displayed a statistically significant relationship with all classes, however conversely to all other components mentioned above, high levels of neuroticism were indicative of increased likelihood of class membership of all the latent classes mentioned above. It is also important to note that the magnitude of this relationship is noticeably higher than that of other components of personality with odds ratios of 1.47, 1.87 and 1.58 for each of the classes, respectively.

Openness displayed a statistically significant relationship with the deteriorating class only with odds ratios of 1.09, suggesting that with every unit increase of openness, the likelihood of membership of this class increased.

8.4.3- Social Variables

Table 8. 4

Social Covariates Odds Ratios for Latent Class Membership

Covariate	Class 1 (Recovery) Odds Ratios (95%CI)	Class 2 (High Stable) Odds Ratios (95%CI)	Class 3 (Deteriorating) Odds Ratios (95%CI)
Qualification	0.99 (0.96-1.01)	1.00 (0.96-1.04)	1.00(0.97-1.03)
I- Job Satisfaction	0.69 (0.63-0.76)*	0.66 (0.57-0.77)*	0.70 (0.62-0.79)*
S- Job Satisfaction	1.99 (1-3.96)*	0.29 (0.09-0.88)*	0.05 (0.02-0.1)*
I- Financial Future	1.11 (0.9-1.36)	1.89 (1.34-2.68)*	1.21 (0.94-1.55)
S- Financial Future	0.38 (0.06-2.33)	16.34(1.12-230.58)*	3.04 (0.38-24.16)

I- Financial Present	2.00 (1.84-2.18)*	3.61 (3.14-4.16)*	2.06 (1.86-2.28)*
S- Financial Present	0.13 (0.05-0.32)*	7.63 (2.13-27.38)*	530.6 (196.81-1430.47)*
I- Leisure Satisfaction	0.56 (0.52-0.6)*	0.33 (0.29-0.38)*	0.47 (0.43-0.51)*
S- Leisure Satisfaction	1.87 (1.04- 3.37)*	0.26 (0.07-0.89)*	0.02 (0.01-0.04)*
Mother not working when 14	1.03 (0.92-1.16)	0.93 (0.76- 1.13)	0.92 (0.8-1.07)
Mother deceased when 14	0.83 (0.53-1.33)	0.88 (0.44-1.74)	0.8 (0.46-1.42)
Mother absent when 14	1.12 (0.73-1.73)	0.60 (0.29-1.25)	0.88 (0.44-1.76)
Father not working when 14	1.36 (1.1-1.68)*	1.12 (0.82- 1.53)	1.28 (1-1.63)*
Father deceased when 14	0.76 (0.59-0.99)*	1.05 (0.73-1.51)	1.19(0.85-1.67)
Father absent when 14	1.36 (1.04-1.78)*	1.57 (1.04-2.36)*	1.16(0.84-1.62)
Married over 4 years	1.09 (0.72-1.64)	1.4 (0.77-2.53)	0.78(0.41-1.48)
Divorced over 4 years	1.76 (1.14-2.71)*	1.86 (0.97-3.57)	1.53(0.96-2.41)
Separated over 4 years	1.5 (0.76-2.96)	2.15 (1.06-4.39)*	1.81 (0.9-3.63)
Widowed over 4 years	1.72 (1.04- 2.85)*	3.36 (1.64-6.92)*	5.63(3.38-9.39)*
Ethnicity- Irish	1.05(0.72-1.54)	1.15(0.7-1.91)	1.07(0.66-1.73)
Ethnicity-Gypsie/ Irish traveller	0.07(0-1237.59)	4.7(0.49-44.79)	0(0-0)
Ethnicity-Any other white background	0.70 (0.5-1)	0.61(0.33-1.11)	0.70 (0.47-1.07)
Ethnicity-White and Black Caribbean	0.69(0.22-2.19)	2.07(0.88-4.86)	2.63(1.24-5.56)
Ethnicity-White and Black African	1.58(0.79-3.15)	0.69(0.11-4.29)	0(0-48.92)
Ethnicity-White and Asian	1.34(0.52-3.48)	1.65(0.48-5.67)	0.95(0.21-4.41)
Ethnicity-Any other mixed background	0.81(0.32-2.02)	0.20 (0.03-1.37)	0.61 (0.23-1.63)
Ethnicity-Indian	0.99(0.72-1.35)	0.81(0.51-1.29)	0.64 (0.4-1.05)
Ethnicity-Pakistani	0.78(0.53-1.17)	0.75(0.46-1.21)	1.24 (0.81-1.89)
Ethnicity-Bangladeshi	0.71(0.41-1.24)	0.82(0.4-1.65)	0.85 (0.32-2.29)
Ethnicity-Chinese	1.10(0.48-2.53)	0(0-575.04)	0.25 (0.06-1.04)
Ethnicity-Any other Asian Background	0.71(0.41-1.24)	0.04(0-1.04)	0.75 (0.39-1.46)
Ethnicity-Caribbean	0.68(0.47-0.99)*	0.66(0.36-1.21)	0.65(0.38-1.11)
Ethnicity-African	0.72 (0.49-1.06)	0.38(0.21-0.72)*	0.51(0.26-1)
Ethnicity-Any other black background	0.72(0.09-5.9)	0.23 (0-33.04)	0.69(0.08-5.61)
Ethnicity-Arab	1.32(0.53-3.28)	2.7 (0.55-13.28)	1.74(0.72-4.22)
Ethnicity-Any other ethnic background	1.00(0.53-1.91)	0.67(0.16-2.82)	0.71(0.19-2.66)

I= intercept
S= slope
*= values which were statistically significant

Qualifications were found to not exhibit statistically significant relationships with any of the classes identified. Job satisfaction was found to exhibit statistically significant relationships with all classes; however, the relationship was complicated. In this variable, higher values indicated higher levels of job satisfaction, and consequently, when investigating initial values, high initial levels of job satisfaction did not predispose individuals to membership of any of the classes mentioned above. When investigating the extent to which change over time affected class membership, with every unit increase in the angle of the slope of the linear model, individuals were more likely to be members of the recovery class but less likely to be members of the deteriorating class.

In relation to subjective income evaluations, one's interpretation of their current financial situation was found to be a statistically significant indicator of class membership in relation to both initial values and the slopes of change. This variable was scored with high scores representing financial difficulty. High initial levels of financial difficulty were associated with an increased likelihood of membership of all classes, which all represented elevated GHQ-12 scores than the reference group. The extent of change as represented by the slope of the linear model was found to have a significant effect on the likelihood of class membership. With every unit increase in the steepness of the slope of financial change, i.e. the rapidity of ones worsening financial situation, participants were 0.13 times more likely to be a member of the recovery class as opposed to the reference class whereas they were 530.6 times more likely to be members of the deteriorating class. The wide range of the confidence intervals of this variable was particularly noticeable (196.81-1430.47).

In relation to anticipated financial circumstances in the future, this variable was scored with higher values indicating a more pessimistic outlook for their financial situation. People who initially anticipated that they would be more likely to the financially worse of were more likely to be members of the three groups identified as opposed to the reference group with odds ratios exceeding 1 in all cases.

When investigating how change in current financial situation over time affected class membership, with one unit increase in the steepness of the slope, participants were less likely to be members of the recovery group and more likely to be part of the high stable and deteriorating group with odds ratios of 16.34 and 3.04 respectively. This suggests that the rapidity of ones financial outlook becoming more pessimistic was primarily an indicator of the high stable group with odds ratios of 16.34, but remained indicative of increased membership of the deteriorating group and decreased likelihood of the recovery group, but by smaller margins.

Leisure satisfaction was measured using a metric which indicated that high scores equalled high levels of satisfaction. The initial value of this variable was statistically significantly associated with membership of all the classes in the table. Each of the odds ratios were less than one indicating that with each unit increase in the initial value of leisure satisfaction, the likelihood of all classes decreased relative to the reference group. In relation to change over time, odds ratios for the high stable and deteriorating group were less than one, suggesting that as the angle of the slope increases the likelihood of class membership of these groups decreased. Odds ratios exceeded one in the case of the recovery group, indicating that for every unit increase in the angle of the slope, the likelihood of being a member of this class was 1.87 times higher.

The effect of the status of the mother when the participant was 14 was not found to exhibit statistically significant relationships with any of the classes in the table, however, the status of the father did. Father not working when the participant was 14 was found to exhibit a statistically significant relationship with the recovery and deteriorating class, but not the high stable class. It was found that having a father who was not working when an individual was 14 increased that participants likelihood of the aforementioned classes with odds ratios of 1.36 and 1.28. Participants who had a decreased father when they were 14 were less likely to be members of the recovery group, but no other relationship was apparent. Finally having an absent father was associated with class membership of the recovery and high stable group, with odds ratios indicating that participants were more likely to be members of this class.

When investigating the effect of relationships, marriage was found to exhibit no statistically significant relationship with class membership. Divorced participants were more likely to be members of the recovery class, however, no other relationships were apparent. Separated participants were more likely to be members of the high stable group, but no other relationships were apparent. Finally, being widowed increased the participant's likelihood of class membership of all classes but the odds ratios were significantly higher for the deteriorating class with reported values of 5.63.

In relation to ethnicity, generally, no statistically significant relationships were evident, however, participants from a Caribbean background were less likely to be members of the recovery group and those from an African background were less likely to be members of the high stable group.

8.5- Discussion

Variables within all components of the biopsychosocial model (Gatchel, 1996) were found to have a statistically significant effect on class membership. This was in line with previous research which suggested that all components of the model were associated with mental health.

The biological variables investigated showed that in line with the literature provided in the previous chapter, which suggested that, sex, age and physical health had statistically significant impacts on mental health, class membership was also predisposed in a statistically significant manner with these covariates.

While research had suggested that the relationship between age and mental health may not be appropriately represented by a linear model (Aldwin et al., 1989), the fit statistics generated in chapter 7 confirmed the appropriateness of a linear model in this data.

Females and older participants were found to be more likely to be members of the recovery, deteriorating and high stable group. These three groups represented increased scores in relation to the reference group. This was congruent with research detailed in Chapter 7 that suggested that females were more likely to report poorer levels of mental health (Economou et al., 2013; Katikeredi et al., 2008; Hauksdottir et al., 2013). Research had previously also reported that the recovery and relapse rates of participants who were over 60 years old did not differ significantly from those of the general population (Hinrichsen, 1992).

In conclusion, the increased likelihood of females and older participants being members of all three classes was interpreted as females and older people generally having poorer levels of self-reported mental health and was not indicative of specific trajectories of behaviour being more prevalent in either group as they were more likely to be in both the recovery and deteriorating group and all groups displayed similar odds ratio figures.

Of the various descriptions of ethnicity provided in UKHLS, only two groups were found to exhibit statistically significant relationships with ‘African’ participants being more likely to be members of the high stable group and ‘Caribbean’ participants being more likely to be members of the recovery class, albeit the relationship was close to being non-significant. The literature associated with ethnic and racial differences would have suggested that ethnic minorities would have been expected to exhibit poorer levels of mental health (Hughes & Demo, 1989) and therefore would have been more likely to be members of classes which demonstrated higher levels of GHQ-12 scores.

This was not borne out in the data investigated and may be indicative of these differences being less prevalent in a UK population than in American studies on which much of the research focused on. African participants being more likely to be members of the high stable group was indicative of sustained episodes of poorer mental health in this cohort. Research into African descended UK citizens help-seeking behaviours relating to mental health services have found that this particular community are less likely to seek help when experiencing poor mental health than other communities in the UK (Mantovani, Pizzolati & Edge, 2016) due to sociological factors relating to stigma. This may explain why the trajectories of mental health are more likely to indicate prolonged episodes of mental health.

The specific reasons why Caribbean participants were more likely to exhibit class membership which indicated improving GHQ-12 scores over time was not clear in the literature. McClean, Campbell and Cornish (2003) interviewed participants from the South of England from an African-Caribbean background. They found that participants frequently mentioned social exclusion and perceived racial mistreatment when accessing mental health services. They also mentioned the importance of religious and spiritual influences on the perception of mental health. Rabiee and Smith (2014) investigated African and African Caribbean participants in Birmingham and emphasised the spiritual and religious aspects of mental health and recovery. They also investigated the role of social support networks and family life in recovery. While this research did not suggest that participants from this cohort are more likely to experience mental health recovery, it did suggest that participants from this cohort of the population were more receptive to culturally competent healthcare professionals and mental health services which respected the cultural nuances. The reasons why this particular cohort of the population was more predisposed to membership of the recovery group warrants further investigation.

The large number of non-significant results for the recovery and deteriorating classes was in some way consistent with the previous literature which suggested that mental health trajectories were similar for people of white and black backgrounds (Robins and Reiger et al., 1991; Jackson, 2004; Twenge and Crocker, 2002; Jackson, Williams and Torres, 2003).

Physical health, a time-varying covariate, was investigated using the initial value and slope. The results show that the intercept or starting value was indicative of all three class memberships by a similar degree, with increasing levels of physical health being associated with declining likelihood of class membership of all three classes relative to the reference group. While research suggested that physical health would have been a significant predictor of mental health recovery, (Concato & Gill, 2002), Nygren et al., (2005) disputed this. While the results did suggest that physical health was a predictor of membership of the recovery class, this was to a similar degree as that of the other classes, and so the results were interpreted as indicating that poor physical health was likely to be an indicator of poor mental health in general and not to indicate a propensity to either recovery or deterioration in GHQ-12 scores.

In line with previous research which suggested that high levels of neuroticism were indicative of poor mental health (McManus, Keeling & Paice, 2004; Menon et al., 2018; Hayes & Joseph, 2003), neuroticism was found to demonstrate a strong relationship with the high stable group, recovery and deteriorating mental health scores groups. All these classes, as shown in figure 2, represent elevated GHQ-12 scores relative to the reference class. Odds ratios were highest for the high stable group, which indicated that GHQ-12 scores were highly likely to consistently report poorer mental health relative to the reference group. Extraversion was found to be the second strongest predictor of class membership which was consistent with Hayes and Joseph (2003) which found similar results when comparing personality traits with The Oxford Happiness Inventory (Argyle, Martin & Crossland, 1989).

The particularly strong relationship between neuroticism and the high stable group was interpreted as an indication that periods of mental health distress amongst neurotic individuals were more likely to be prolonged. Within the literature there have been suggestions that measures of general mental health vulnerability were simply measures of neuroticism (Brandes et al., 2019), however, the findings of this analysis would suggest that this may not be the case as while neuroticism was the strongest predictor of mental health trajectories, other personality traits were associated with class membership. The research in this chapter would suggest that clinicians and researchers would be ill-advised to focus their attention solely on neurotic personality traits to the exclusion of all others and that a wider view of personality traits may be advisable.

Agreeableness was found to be an indicator of the recovery and deterioration classes, but not the high stable class. This was interpreted as less agreeable individuals being more likely to experience fluctuations in their self reported mental health but not as likely to exhibit prolonged spells of psychological distress. These results stand in contrast to previous research by Clough et al. (2001), who was unable to uncover statistically significant relationships between what he referred to as mental toughness and mental health. Should the results in this analysis have matched with Clough's (2001) findings, it would have been expected that no relationship would have been evident from the deteriorating class and this covariate.

Openness was found to be an indicator of the deteriorating class only, with more open individuals being more likely to be members of this class. While research frequently suggested that openness was a poor predictor of mental health in comparison with other components (Hayes & Joseph, 2003), most of the studies were able to identify statistically significant albeit weak relationships with mental health in general (Menon et al., 2017) and with resilience and recovery (Clough et al., 2002).

Conscientiousness and extraversion were found to be a predictor of all three classes, and while conscientiousness was found to be a greater indicator, both personality traits were found to be indicators of less likelihood of class membership. This was interpreted as indicating a greater likelihood of being a member of the reference class, and therefore conscientious and extraverted people were less likely to exhibit psychological distress. Hayes and Joseph (2003) had found that conscientiousness was the primary predictor of ‘satisfaction with life’ and consequently, it was not unexpected that this covariate exhibited such strong relationships with GHQ-12 scores.

In relation to social variables, satisfaction with leisure and job, parental status, subjective and anticipated financial situation and parental status were found to exhibit relationships with class membership. Only qualifications were found not to exhibit any relationships. The failure of qualifications to exhibit strong relationships was unexpected as research had previously identified poor educational attainment as both a predictor and cause of poor mental health (Friedli, 2009) and of resilience (Sugarman, 1986). It is possible that these relationships did not manifest in this dataset due to how the data was coded as the inclusion of ‘other higher degree’ may have been ambiguous and therefore not understood by participants. It is also important to note that when using the UKHLS predecessor as a data source, Bell (2014) found that once an individual passed post-primary, education was not a predictor of mental health and the results obtained would correlate to these findings.

Previous research investigating the relationship between job satisfaction and mental health has shown that job satisfaction is a predictor of mental health in a number of settings including mental health staff (Prosser et al., 1999) civil servants (Bogg & Copper, 1995) and teachers (Travers & Cooper, 2007). Job satisfaction was a significant predictor of class membership of all three classes. Increased job satisfaction was found to predispose participants to not be a member of the deteriorating group with particularly small odds ratios relative to other time-varying covariates. This suggests that as the participants of the UKHLS who represented a representative sample of the UK population, experienced an increase in their job satisfaction, they were very unlikely to experience a deteriorating trajectory of mental health. The results obtained build on the research mentioned above and suggest that the relationships found in specific fields could be generalised to the wider population.

One’s subjective evaluation of their financial situation, both present and anticipated was found to exhibit relationships with class membership. The strongest relationship was observed when individuals experienced a decrease in their current financial situation, and this was linked to being a member of the deteriorating group. Oskrochi, Bani Mustafa and Oskrochi (2018) had already suggested that one’s subjective financial situation was the largest predictor of mental health, of the social variables that they tested, however, this research was not conducted in a longitudinal or latent class context. The rate of a loss in income correlating to the deteriorating group was not as prevalent, however, was still one of the largest predictors observed within the analysis. This analysis was relatively unique as it allowed it to be ascertained how the rate of change in income affected the GHQ-12 scores of participants which was not identified in the literature previously.

In line with previously discussed literature (Haar et al., 2014; Habib & Shirazi, 2003), job satisfaction was found to be a predictor of the three classes membership in relation to both its initial value and its slope. The rate of change was particularly prevalent, with individuals who experienced an increase in their job satisfaction finding themselves less likely to be members of the deteriorating group. The results also indicate that increasing levels of job satisfaction predispose participants to not be members of the high stable class suggesting that increases in job satisfaction would lead to individuals not experience prolonged episodes of poor mental health.

Leisure satisfaction was found to exhibit similar relationships with the latent classes as that of job satisfaction which is detailed above, however, the relationships were less pronounced. Relationships between mental health and satisfaction with leisure time have been observed in the literature (Pearson, 1999) and the results obtained in this research are consistent with previous research, with participants who experienced an increase in leisure satisfaction, being less likely to be a member of classes which denoted elevated GHQ-12 scores.

Parental status was found to exhibit relationships with class memberships in certain circumstances, but only concerning the participant's father. This is consistent with research by Harvey (1999), who suggested that long term outcomes of individuals were not significantly affected by the employment status of the mother.

Having a working father when the participant was 14 was linked to increased class membership of the recovery and deteriorating class, while no relationship was evident with the high stable class. The increased propensity for fluctuations in scores was unexpected as having a working father at aged 14 represented the largest response category (see Chapter 2), and it was felt likely that this would lead to it being representative of the baseline population. A possible explanation of this may be that the time commitment that working entails would suggest that the father would have spent less time with the child and this lack of paternal time being spent with the father may have resulted in attachment issues growing up. In the literature, working fathers were viewed by their children as being more supportive in comparison to non-working fathers (Bacikova-Sleskova, 2011) and this relationship was not evident in childhood perceptions of working or non-working mothers. It may be that perceived differences in parental attentiveness may have in turn been responsible for an increased propensity for fluctuating GHQ-12 scores over time.

Participants who had a deceased father when they were 14 were less likely to be in the recovery class but did not display any relationships with the other classes. While no research suggested a difference in experiences between losing a mother or a father in early childhood, research was fairly consistent in suggesting that the death of a parent was a predictor of mental health distress in both childhood and in later life (Fristad et al., 1993; Siegel, Karus & Raveis, 1996). It could be hypothesised that this increased psychological distress experienced in early adulthood may have resulted in participants developing lower levels of resilience in later life. This hypothesis was consistent with research by Kennedy et al. (2018), who suggested that the loss of a family member in childhood decreased resilience in adolescence. This research focused on the effects of adolescents and did not investigate the long term effects of resilience following parental death. This may be an area which requires further research.

It was acknowledged that parental status variables were collected in a crude manner, with no account for the time elapsed since the participant experienced the loss.

It is expected that the experience of participants who endured the relatively recent loss of a parent would be markedly different from those who had significant periods of time to process the loss. Unfortunately, the variable collected did not allow for distinctions of this type to be made, and consequently, the results from this variable may not respect the temporal differences in this relationship.

Participants who married over the course of the analysis were found to be more likely to be members of the recovery class which is congruent with research that suggests that being married was beneficial for mental health (Horwitz, White & Howel-White, 1996). Divorced participants were also more likely to be members of the recovery group, which was unexpected, however, could be an outworking of escaping a relationship which was making the individual unhappy. Research into divorce was described as problematic as it is frequently difficult to obtain consent for both participants in research (Cohen & Finzi-Dottan, 2012) however Vangelesti (2006) investigated the effect of what she described as ‘hurtful interactions’. While she acknowledged that divorce usually was associated with helplessness, aggression, sadness, guilt, and loneliness, she also suggested that an end to these hurtful interactions’ may lead to the finality of a divorce representing a relief to the people involved.

Separated participants were more likely to be members of the high stable class. It was expected that the results for divorce and separation would be similar, however, this was not borne out in the data. Vangelesti’s (2006) writings on ‘hurtful interactions’ may provide some explanation as ongoing animosity between partners was likely to differentiate separation and divorce experiences within this cohort.

Being widowed over the course of the analysis lead to an increase in the likelihood that the participant would be a member of the three classes mentioned above. While the class likelihood for all classes increased, the results showed that widowed participants were much more likely to be members of the deteriorating group. This was not unexpected, as research had suggested that individuals who lost their partner were likely to experience deteriorating mental health (Brock, 1984). The research around grieving is complex, and differentiations are made depending on the expectedness of the death (Graff et al., 2016), the cause of death (Shah et al., 2016) and the pre-death quality of life (Wright et al., 2010), however generally there is no dissent in the literature that there are severe physical and mental health impacts associated with the death of a partner and that this is most prevalent in the year following the death (Graff et al., 2016)

8.5.1- Limitations

This research had a number of limitations which are detailed below. Firstly, the methodological constraints that are mentioned earlier guided how missing data were handled during this analysis. While generally it is considered untidy to include numerous techniques for handling missing data in one analysis, this was unavoidable in order to retain the integrity of the underlying class model while also allowing all participants to be used in covariate analysis. The use of data imputation was problematic as it involved estimating values based on the other variables provided for each participant. In the case of a number of variables, such as job satisfaction, this may have involved the generation of plausible values for participants who may not have been employed or in some other way were inappropriate. It must be noted that the analysis was rerun without data imputation techniques employed, and similar results were obtained.

The second weakness was the crudeness of some of the variables selected. While analysis of this type would necessitate using information that was available, not necessarily what was ideal, a number of variables such as the parental status variables were too broad as they would have included participants who had recently lost a parent and those who would have lost them a considerable time ago. This may, in turn, have masked some of the complexities of the relationships that occurred.

While the research was not intended to be an exhaustive analysis of all the factors which affect mental health a number of important components of the biopsychosocial model (Gatchel, 1996) were not collected by the UKHLS database. In particular, there was a lack of data available for the psychological components mentioned. Coping skills are frequently mentioned in the model, and no data was available. Research has shown that interventions that seek to improve coping skills are effective in building psychological resilience and promoting good mental health across a number of population cohorts (Khanghahi, 2001; Aghajani, 200; Ramesht & Farshad, 2004). There are also references to self-esteem in the model, and this has been shown to have an effect on mental health (Mann et al., 2004). Had data for these variables been available, a much wider picture of psychological factors which affect mental health trajectories would have been obtained. It may be meritorious for future research to use a similar research framework but use a dataset which contains these variables.

8.5.2- Research Implications

This research provided an opportunity to investigate a wide range of biological, social and psychological variables in relation to how they affect the mental health trajectories of participants in a representative UK population. While considerable research exists as to the effect on mental health of covariates, most of this research investigates a relatively narrow range of covariates relative to what was included in this thesis, for example, Oskrochi, Bani-Mustafa & Oskrochi, (2018) which focused on social covariates and did not investigate them in a latent class context as was done in this chapter. Furthermore, the results in this chapter do add extra context in relation to previous literature. Oskrochi, Bani Mustafa and Oskrochi (2018) for example found that anticipated financial situation was the largest predictor of mental health within a cross-sectional context, however this research, using the same data found that rate of change of current financial situation was a larger predictor of class membership. These findings demonstrated the validity of using a latent class membership framework as a way of investigating change over time and some of the findings that this analysis facilitated.

The wide range of ethnic minority categories collected allowed an in-depth investigation of these population cohorts. The increased predisposition of participants from a Caribbean background to be members of the recovery class was a relationship which would not have been apparent through other analytic frameworks and demonstrated the utility of latent class-based analyses.

Research has suggested that general measures of mental health were actually just measures of neuroticism and depending on the measure being investigated, they identified that different personality traits displayed statistically significant relationships with the mental health measure, albeit neuroticism was the strongest predictor in all cases (Hayes & Joseph, 2003). This research has identified that a wide range of personality traits were associated with class membership and that the relationship is perhaps more complex than cross-sectional correlational analyses would have facilitated.

8.5.3- Clinical Implications

Clinically, the value of this research lies in the identification of covariates which predispose individuals to recovery or deterioration in their mental health. This research also has developed a framework for investigating time-varying and time-invariant covariates simultaneously, which could be beneficial to researchers conducting longitudinal research.

Using the results obtained in this chapters analysis, it would be possible to identify participants who were at risk of developing poor mental health based on covariates that were strong predictors of class membership. The findings in relation to social variables which were the greatest predictor of class membership are of particular note to clinicians as it may inform interventions relating to difficulties at work which were shown to impede recovery.

The findings may also inform interventions which rather than attempt to address symptoms of poor mental health, address social situations, instead of or alongside symptoms management, which may drive change in mental health. The social variables relating to financial situation and satisfaction with leisure time and job satisfaction have been identified in the literature with UK civil servants (Bogg & Cooper, 1995) and mental health professionals (Prosser et al., 1999) being found to report poorer mental health when they experienced lower job satisfaction. Clinicians may wish to investigate the employment satisfaction of those in need of their care and given the findings of Rout and Rout (1994) that changes in work conditions were a major predictor of job satisfaction and mental health, clinicians may wish to increase monitoring of service users who would be likely to experience a deterioration of their job satisfaction. Employers may also wish to consider if exposing employees to more onerous work conditions may represent a false economy as if job satisfaction deteriorates, then deterioration in mental health may negate any productivity increases projected.

The research may be of particular interest to clinicians in relation to mental health provision. While historically, conventional wisdom has suggested that ethnic minorities experienced poorer mental health as a consequence of the socioeconomic disadvantages associated (Kleiner, Tuckman & Lavell, 1960; Fischer, 1969; Kramer, Rosen & Willis, 1973; Cannon & Locke, 1977; Mirowsky & Ross, 1980), this research was able to differentiate ethnic minorities based on UK ethnic minority guidelines (Office for National Statistics, 2020) and suggested that this may be oversimplistic. This research identified that participants from an African and Caribbean background exhibited distinct trajectories from the ethnic majority in the UK and other ethnic minorities. Research has shown a reluctance from participants from an African background to engage with mental health services (McClean, Campbell & Cornish, 2013) and this may in some way explain why episodes of poor mental health were prolonged and why they were no more predisposed to recovery nor deterioration than the reference group.

References

Abu-Omar, K., Rütten, A., & Lehtinen, V. (2004). Mental health and physical activity in the European Union. *Sozial-und Präventivmedizin*, *49*(5), 301-309.

Aldwin, C. M., Spiro, A., Levenson, M. R., & Bossé, R. (1989). Longitudinal findings from the normative aging study: I. Does mental health change with age?. *Psychology and Aging*, *4*(3), 295.

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of abnormal psychology*, *112*(4), 545.

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using M plus. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(3), 329-341.

Bacikova-Sleskova, M., Geckova, A. M., van Dijk, J. P., Groothoff, J. W., & Reijneveld, S. A. (2011). Parental support and adolescents' health in the context of parental employment status. *Journal of adolescence*, *34*(1), 141-149.

Bell, A. (2014). Life-course and cohort trajectories of mental health in the UK, 1991–2008–a multilevel age–period–cohort analysis. *Social science & medicine*, *120*, 21-30.

Bogg, J., & Cooper, C. (1995). Job satisfaction, mental health, and occupational stress among senior civil servants. *Human relations*, *48*(3), 327-341.

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, 3-27.

Brandes, C. M., Herzhoff, K., Smack, A. J., & Tackett, J. L. (2019). The p factor and the n factor: Associations between the general factors of psychopathology and neuroticism in children. *Clinical Psychological Science*, *7*(6), 1266-1284.

Brisson, D., Lopez, A., & Yoder, J. (2014). neighborhoods and mental health trajectories of low-income mothers. *Journal of Community Psychology*, *42*(5), 519-529.

Chan, R., & Joseph, S. (2000). Dimensions of personality, domains of aspiration, and subjective well-being. *Personality and Individual differences*, *28*(2), 347-354.

Clough, P., Earle, K., & Sewell, D. (2002). Mental toughness: The concept and its measurement. *Solutions in sport psychology*, 32-43.

Corrigan, P. W., Giffort, D., Rashid, F., Leary, M., & Okeke, I. (1999). Recovery as a psychological construct. *Community mental health journal*, *35*(3), 231-239.

Costa, P. T., & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: happy and unhappy people. *Journal of personality and social psychology*, *38*(4), 668.

Economou, M., Madianos, M., Peppou, L. E., Patelakis, A., & Stefanis, C. N. (2013). Major depression in the era of economic crisis: a replication of a cross-sectional study across Greece. *Journal of affective disorders*, *145*(3), 308-314.

Friedli, L., & World Health Organization. (2009). *Mental health, resilience and inequalities* (No. EU/08/5087203). Copenhagen: WHO Regional Office for Europe:.

Fristad MA, Jedel R, Weller RA, Weller EB (1993) Psychosocial functioning in children after the death of a parent. Am J Psych 150:511–513

Furnham, A., & Brewin, C. R. (1990). Personality and happiness. *Personality and individual differences*, *11*(10), 1093-1096.

Gatchel, R. J., Peng, Y. B., Peters, M. L., Fuchs, P. N., & Turk, D. C. (2007). The biopsychosocial approach to chronic pain: scientific advances and future directions. *Psychological bulletin*, *133*(4), 581.

Gatchel, R. J. (1996). Psychological disorders and chronic pain: cause-and-effect relationships.

GEthnicity-Martin, S. A. S. K. (2010). Data analysis with SPSS: A first course in applied statistics. *Statistics*, *4*, 27.

Gili, M., Roca, M., Basu, S., McKee, M., & Stuckler, D. (2013). The mental health risks of economic crisis in Spain: evidence from primary care centres, 2006 and 2010. *The European Journal of Public Health*, *23*(1), 103-108.

Graff, S., Fenger-Grøn, M., Christensen, B., Pedersen, H. S., Christensen, J., Li, J., & Vestergaard, M. (2016). Long-term risk of atrial fibrillation after the death of a partner. *Open heart*, *3*(1).

Haar, J. M., Russo, M., Suñe, A., & Ollier-Malaterre, A. (2014). Outcomes of work–life balance on job satisfaction, life satisfaction and mental health: A study across seven cultures. *Journal of Vocational Behavior*, *85*(3), 361-373.

Habib, S., & Shirazi, M. A. (2003). Job satisfaction and mental health among the employees of a general hospital. *Iranian journal of psychiatry and Clinical psychology*, *8*(4), 64-73.

Harvey, E. (1999). Short-term and long-term effects of early parental employment on children of the National Longitudinal Survey of Youth. *Developmental psychology*, *35*(2), 445.

Hauksdóttir, A., McClure, C., Jonsson, S. H., Ólafsson, Ö., & Valdimarsdóttir, U. A. (2013). Increased stress among women following an economic collapse—a prospective cohort study. *American journal of epidemiology*, *177*(9), 979-988.

Helgeson, V. S., Snyder, P., & Seltman, H. (2004). Psychological and physical adjustment to breast cancer over 4 years: identifying distinct trajectories of change. *Health Psychology*, *23*(1), 3.

Hicks, T., (2020). *A Beginner's Guide To Interpreting Odds Ratios, Confidence Intervals And P-Values - Students 4 Best Evidence*. [online] Students 4 Best Evidence. Available at: <https://www.students4bestevidence.net/blog/2013/08/13/a-beginners-guide-to-interpreting-odds-ratios-confidence-intervals-and-p-values-the-nuts-and-bolts-20-minute-tutorial/> [Accessed 16 June 2020].

Horwitz, A. V., White, H. R., & Howell-White, S. (1996). Becoming married and mental health: A longitudinal study of a cohort of young adults. *Journal of Marriage and the Family*, 895-907.

Hughes, M., & Demo, D. H. (1989). Self-perceptions of Black Americans: Self-esteem and personal efficacy. *American Journal of Sociology*, *95*(1), 132-159.

Kariuki, M., Honey, A., Emerson, E., & Llewellyn, G. (2011). Mental health trajectories of young people after disability onset. *Disability and health journal*, *4*(2), 91-101.

Kaptein, K. I., De Jonge, P., Van Den Brink, R. H., & Korf, J. (2006). Course of depressive symptoms after myocardial infarction and cardiac prognosis: a latent class analysis. *Psychosomatic Medicine*, *68*(5), 662-668.

Katikireddi, S. V., Niedzwiedz, C. L., & Popham, F. (2012). Trends in population mental health before and after the 2008 recession: a repeat cross-sectional analysis of the 1991–2010 Health Surveys of England. *BMJ open*, *2*(5), e001790.

Kleiner, R. J., Tuckman, J., & Lavell, M. (1960). Mental disorder and status based on race. *Psychiatry*, *23*(3), 271-274.

Mantovani, N., Pizzolati, M., & Edge, D. (2017). Exploring the relationship between stigma and help-seeking for mental illness in African-descended faith communities in the UK. *Health Expectations*, *20*(3), 373-384.

Mclean, C., Campbell, C., & Cornish, F. (2003). African-Caribbean interactions with mental health services in the UK: experiences and expectations of exclusion as (re) productive of health inequalities. *Social science & medicine*, *56*(3), 657-669.

Menon, V., Shanmuganathan, B., Thamizh, J. S., Arun, A. B., Kuppili, P. P., & Sarkar, S. (2018). Personality traits such as neuroticism and disability predict psychological distress in medically unexplained symptoms: A three-year experience from a single centre. *Personality and mental health*, *12*(2), 145-154.

Meyrose, A. K., Klasen, F., Otto, C., Gniewosz, G., Lampert, T., & Ravens-Sieberer, U. (2018). Benefits of maternal education for mental health trajectories across childhood and adolescence. *Social Science & Medicine*, *202*, 170-178.

Morack, J., Infurna, F. J., Ram, N., & Gerstorf, D. (2013). Trajectories and personality correlates of change in perceptions of physical and mental health across adulthood and old age. *International Journal of Behavioral Development*, *37*(6), 475-484.

North, N. T. (1999). The psychological effects of spinal cord injury: a review. *Spinal cord*, *37*(10), 671-679.

Office for National Statistics (2013) 2011 Census: Key Statistics and Quick Statistics for Local Authorities in the United Kingdom. Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigratio n/populationestimates/bulletins/keystatisticsandquickstatisticsforlocalauthoritiesintheunitedkingdom/2013-10-11#ethnicity-and-country-of-birth

Oskrochi, G., Bani-Mustafa, A., & Oskrochi, Y. (2018). Factors affecting psychological well-being: Evidence from two nationally representative surveys. *PloS one, 13*(6).

Pinto, R., Ashworth, M., & Jones, R. (2008). Schizophrenia in black Caribbeans living in the UK: an exploration of underlying causes of the high incidence rate. British Journal of General Practice, 58(551), 429-434.

Prosser, D., Johnson, S., Kuipers, E., Dunn, G., Szmukler, G., Reid, Y., ... & Thornicroft, G. (1999). Mental health,"burnout" and job satisfaction in a longitudinal study of mental health staff. *Social Psychiatry and Psychiatric Epidemiology, 34*(6), 295-300.

Robins, L. N., & Regier, D. A. (1991). *Psychiatric disorders in America: The epidemiological catchment area study*.

Shah, S. M., Carey, I. M., Harris, T., DeWilde, S., Victor, C. R., & Cook, D. G. (2016). The mental health and mortality impact of death of a partner with dementia. *International journal of geriatric psychiatry*, *31*(8), 929-937.

Siegel K, Karus D, Raveis VH (1996) Adjustment of children facing the death of a parent due to cancer. J Am Acad Child Adolesc Psychiatr 35:442–450.

Strohschein, L. (2005). Parental divorce and child mental health trajectories. *Journal of Marriage and Family*, *67*(5), 1286-1300.

van Leeuwen, C. M., Hoekstra, T., van Koppenhagen, C. F., de Groot, S., & Post, M. W. (2012). Trajectories and predictors of the course of mental health after spinal cord injury. *Archives of physical medicine and rehabilitation*, *93*(12), 2170-2176.

Veldman, K., Bültmann, U., Stewart, R. E., Ormel, J., Verhulst, F. C., & Reijneveld, S. A. (2014). Mental health problems and educational attainment in adolescence: 9-year follow-up of the TRAILS study. *PLoS One*, *9*(7).

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political analysis*, 450-469.

Wright, A. A., Keating, N. L., Balboni, T. A., Matulonis, U. A., Block, S. D., & Prigerson, H. G. (2010). Place of death: correlations with quality of life of patients with cancer and predictors of bereaved caregivers' mental health. *Journal of Clinical Oncology*, *28*(29), 4457.

Chapter 9 - A Review of the Findings of this Thesis and its Implications

9.1- Introduction

The main goal of the thesis was to establish an appropriate way to model mental health within a structural equation modelling framework, to ascertain if different participants exhibited different trajectories of mental health over time and, using a wide range of covariates from a biological, social or psychological background, to attempt to explain those trajectories. In order to do this, a number of preparatory steps had to be completed to determine an appropriate dimensional representation, to test stability over time and to identify latent classes based on GHQ-12 responses.

The thesis could broadly be split into three parts which are identified below

- The utility and psychometric properties of the GHQ-12 (Chapters 3-5)
- Identification of longitudinal mental health trajectories amongst the population (Chapter 6)
- Explaining longitudinal change in mental health through covariates (Chapters 7 & 8)

While the specific discussion sections of each chapter provide comprehensive discussions on each analysis, this chapter summarised the important findings in relation to the above three points and linked them to existing literature where appropriate. Once this was done, an evaluation of the thesis was presented in relation to the value of the GHQ-12 and population data, the value of mental health modelling over time, the clinical and research implications of the findings and any limitations identified.

9.2- The Utility and Psychometric Structure of the GHQ-12

The purpose of the first chapter of this thesis was to provide a review of mental health and its measurement and to introduce the GHQ-12 as this was the primary mental health measure that was used during this thesis. The historical background of the GHQ family of mental health questionnaires was provided for context. While this chapter was not empirical, it did demonstrate the merits of measuring mental health using a continuum rather than using diagnoses prevalence rates and provided evidence as to why the GHQ-12 was an appropriate measure of the general concept of mental health. Due to the disagreement in the literature identified in the introduction chapter, i.e. that numerous researchers had proposed various dimensional representations of the GHQ-12, Confirmatory Factor Analysis was used to ascertain the most appropriate dimensional representation of the data. Researchers had suggested a number of dimensional representations that were appropriate and these represented a number of methodological standpoints and conceptual underpinnings. Graetz (1991) and others proposed that the GHQ-12 was unidimensional, whereas others such as Hankins (2008) and Ye (2009) argued that the numerous factors were a function of method effects caused by the wording of the items within the measure. All models which could be identified in the literature were tested for model fit, and factor loadings and inter-factor correlations were also investigated. This chapter identified that a number of dimensional representations demonstrated sufficient fit with the data as to be considered an appropriate dimensional representation for the data. The models that were brought forward represented a number of theoretical and methodical interpretations of the data, with multidimensional representations such as Graetz (1991) and unidimensional representations which accounted for method effects such as Ye (2009) and Hankins (2008) demonstrating sufficient fit.

In order to further inform which dimensional representation was the most appropriate for the Understanding Society (UKHLS) data, those dimensional representations which demonstrated sufficient fit for the data were investigated for validity. This was achieved by regressing the factors of the various dimensional representations onto covariates within a structural equation framework to ascertain a number of things. In relation to multidimensional representations, the analysis investigated in the factors demonstrated variability in relation to a number of covariates which they ought to. Should variability not be demonstrated, then the utility of treating the GHQ-12 as multidimensional would be called into question. The covariates included a number which were similar to those which were common themes within the multidimensional representations of the data, namely, social performance and depression. If these factors were valid, they should have demonstrated strong correlations with these covariates. Finally, in relation to unidimensional models, a number of covariates were included which had established relationships with mental health in the literature, for example, age, which was shown to have a relationship with self-reported mental health, with participants reporting poorer mental health as they age. This allowed unidimensional constructs to demonstrate strong correlations with these covariates.

This chapter found that while multidimensional representations of the GHQ-12 exhibited variability in relation to a number of covariates, they did not demonstrate concurrent validity with those which measured similar concepts as the factors claimed to. Notably, in multidimensional representations which included a depression factor, and those which included a factor which related to social function, these factors did not display stronger relationships with covariates which measured similar concepts than other factors. In fact, in all cases, other factors displayed stronger relationships with these covariates than the named factor.

As a result, unidimensional representations were deemed as the most appropriate to bring forward to future chapter's analysis. Ye's (2009) model, which included a universal factor which claimed to measure mental health and a second factor which captured the variability caused by method effects, was selected.

Ye's (2009) model was tested for measurement invariance to ensure that the dimensional representation remained stable over time. This was done using a framework detailed in Widaman and Reiuse (1997) which proposed a number of sequential analyses with ever constrained models to investigate fit. Should a model continue to demonstrate acceptable fit, despite a parameter of that model being fixed across time, then that parameter remained stable over time. This was important as should a model fail to demonstrate longitudinal invariance, then the future longitudinal analysis may not be possible. The model demonstrated strict measurement invariance, which indicated a high level of stability over time and meant that longitudinal analysis could proceed unimpeded.

In summary, this stage of the thesis identified an appropriate dimensional representation for the data at a cross-sectional point of time and then further investigated its properties over time to ensure that it remained stable enough to remain an appropriate dimensional representation in longitudinal analyses. Once this was established, analysis which identified if different participants exhibited different trajectories was conducted.

9.3- Identification of Longitudinal Trajectories of Mental Health.

This analysis utilised growth mixture modelling techniques in order to ascertain if different subpopulations of the UKHLS sample demonstrated different trajectories over time. The analysis identified that both a 4 and 5-class solution demonstrated good fit over time. A four-class solution was selected for reasons of parsimony and interpretability. These classes identified a reference class which included the majority of the population and represented those who consistently reported good mental health, a high stable group which represented those participants which consistently reported poor mental health and two classes which represented steadily increasing and decreasing scores over time.

9.4- Explaining Longitudinal Change in Mental Health

In chapter 7, a number of covariates from those available in the UKHLS were selected and arguments to establish their appropriateness for analysis in the final chapter were presented. A wide range of variables were selected using the biopsychosocial model (Gatchel et al., 1996) as a framework for model selection. The selected variables were converted into a format that was compatible with longitudinal regression analysis, i.e. those variables which were classed as time-invariant were converted into measures of slopes and intercepts and were tested for fit to ensure that a linear interpretation of these variables was appropriate. All variables which were converted were found to demonstrate good fit with a linear model, and consequently, a linear interpretation was appropriate.

In order to explain the trajectories that participants displayed, this analysis utilised the R3step regression analysis technique (Asparouhouv & Muthen, 2013) to investigate if changes in the covariates identified above related to the likelihood that participants would exhibit the various trajectories that were identified in chapter 6. This analysis showed that a wide range of variables from all components of the biopsychosocial model demonstrated statistically significant relationships with class membership. The strongest predictors of mental health trajectory were changes in financial situation and job satisfaction. A number of personality traits relating to the ‘big 5’ model of personality (Tupes & Christal, 1961) were found to also exhibit relationships with neuroticism exhibiting the strongest relationships.

9.5- Methodological Discussion

9.5.1- The Value of Using the GHQ-12 and Population Data

Within the literature, there has been disagreement as to how to appropriately measure mental health. Within a clinical setting, measurement of poor mental health has been diagnostically driven with participants receiving diagnoses based on the presence of symptoms as dictated by the Diagnostic and Statistics Manual of Mental Disorders (DSM) or a similar nosology (Caspi et al., 2014). This approach has been instrumental in framing psychiatric practice and research into mental health for decades (Kupfer, Kuhl & Regier, 2013). A failing of this approach has been the existence of multiple comorbidities, where an individual could be classified with one or a number of psychological disorders (Hasin & Kilcoyne, 2012; Kessler et al., 2005). Newman et al. (1998) suggested what they referred to as the rule of 50%, where half of those diagnosed with a mental disorder would also simultaneously exhibit symptoms or have a diagnosis of a second disorder and half of those with a second disorder would exhibit symptoms of a third and so on. In light of this, shortly after the publication of the *DSM-IV* (American Psychiatric Association, 1994), psychological scientists noted the need for research that would examine patterns of comorbidity to “*elucidate the broad,* *higher-order structure of phenotypic psychopathology*" (Clark, Watson. & Reynolds, 1995, p. 131). While self-report measures of mental health have been developed as early as 1949 (Brodman et al., 1949), more recent literature suggested that mental health, specifically within a research setting, could be reconceptualised more parsimoniously as a continuum with individuals being placed on a sliding scale between good and poor mental health (Keyes, 2002). This has more recently been expanded upon within the literature which the concept of a 'p-factor (Caspi et al., 2014). This concept posits that mental health diagnoses can be more parsimoniously measured by a single underlying latent variable which underlies many mental health disorders. It took inspiration from research into intelligence where the so-called g-factor of general intelligence has been shown as an effective way of measuring intelligence and research has demonstrated that frequently individuals who perform well in one form of intelligence test, tend to also perform well in other such tests (Deary, 2020; Jensen, 1998; Spearman, 1961).

This thesis has operationalised mental health through the identification of a single latent variable which captured general mental health from the 12 items of the GHQ-12 and a method factor which captured the variance caused by the wording of the positively and negatively worded items. In a clinical context, cut-off's have been identified for the GHQ-12 which indicate that an individual is at high risk of psychological distress (Perenboom, 2000). A caseness approach, as this is referred to, (see section 1.4) was not adopted in this thesis as Graetz (1991) argued that "a more acceptable distribution of scores" was generated by investigating GHQ-12 responses in a similar way to what was done in this thesis. This approach was beneficial in a number of other ways from diagnostic-based interpretations of mental health as it affords researchers and clinicians the opportunity to gauge severity in a way that a dichotomous diagnosis variable does not. It also affords greater sensitivity as individuals may not meet the criteria to warrant a specific diagnosis but may still be experiencing psychological distress.

Analyses conducted in Chapter 4 investigated the relationships between numerous dimensional representations of the GHQ-12 with a range of covariates, including a diagnosis of clinical depression. Research has frequently operationalised mental health as the absence of mental disorders (Keyes, 2002) however research has also proposed a two continua approach (Westerof & Keyes, 2010) which suggested that the two concepts of general mental health and diagnoses were distinct albeit highly correlated. The findings of this thesis corroborated the assertion that the two concepts were highly correlated and that a general mental health measure was a strong predictor of specific clinical diagnoses such as depression.

Self-reported mental health measures were refined after the initial shortcomings of the Cornel Medical Inventory (Brodman et al., 1949), with a number of recommendations being reported in the literature as to avoid biases and methodical shortcomings. One such recommendation which was particularly pertinent to the GHQ was the recommendation that positively and negatively worded items should be included to avoid acquiescence biases (Carr & Krause, 1978). When measures which included both were subjected to factor analytic techniques a number of self-report questionnaires exhibited numerous factors, such as the Rosenberg self-esteem scale (RSE) (Greenberger et al., 2003) and the GHQ (Graetz, 1991). Some researchers had claimed that this was a consequence of the method effects caused by the presence of positively and negatively worded items (Hankins, 2008). A meta-analysis of 23 studies investigating the RSE showed that generally a two-factor solution was supported, however, once method effects were accounted for a single factor solution was more appropriate (Huang & Dong, 2012).

The literature has not come to a definitive conclusion as to whether the multiple dimensions exhibited by the GHQ-12 were spurious and caused by method effects, however, Hankins (2008) had reported that unidimensional representations of the GHQ-12 had failed to outperform multidimensional representations of the GHQ-12 when subject to factor analytic techniques. This thesis relied heavily on the GHQ-12, and in order to use it within a structural equation framework, it was necessary to identify an appropriate dimensional representation to use in subsequent analysis and to determine if the GHQ-12 could be treated as unidimensional, representing a general factor of mental health. By subjecting all models which were identified within the literature to analyses of fit and validity this research determined that the GHQ-12 could be adequately represented by a unidimensional structure, with a second factor accounting for method effects by capturing the variance caused by negatively worded items as proposed by Ye (2009). This research was meaningful as it was in agreement with a number of previous studies which suggested that the GHQ-12 could be treated as unidimensional (Shevlin and Adamson, 2004; Gao et al., 2004). Should the GHQ-12 not have been found to be unidimensional, then its scoring matrix and use as a measure of ‘general mental health’ would have been called into question.

This dimensional representation had to demonstrate stability over time if it was to be used within a longitudinal setting, which was integral to the research aims of this thesis, i.e. to investigate trajectories over time. A high level of stability was demonstrated which in itself was a meaningful finding, as the GHQ-12 is used within a clinical and research setting numerous times (Graetz, 1991; Smith et al. 2012; Montazeri et al., 2003). Should temporal invariance not have been demonstrated then research which utilised GHQ-12 at two timepoints may not have been appropriate as retest effects may affect results or extraneous variables may have caused change over time.

The second part of this thesis used the GHQ-12 which had been established as appropriate in the previous chapters as a measure of general mental health in a longitudinal setting to identify different trajectories and explain what predisposes individuals to display various trajectories. This research represents a useful addition to the literature as it demonstrates the utility of using structural equation modelling frameworks to measure mental health over time. This analytic framework offers researchers opportunities to investigate mental health in a way that less advanced techniques could not.

The identification of stable and steadily deteriorating or increasing mental health trajectories facilitated investigation into the drivers of mental health changes, i.e. what causes individuals to exhibit either stable or changing mental health scores. This was further enhanced by the inclusion of time-varying covariates which facilitated investigation of the interaction between how changes in covariates drove change in trajectories of self-reported mental health. The results identified a number of covariates which were biological, social and psychological in nature which predisposed individuals to display various trajectories.

9.5.2 The Value of Longitudinal Mental Health Modelling

This thesis has demonstrated the importance of longitudinal modelling of mental health as this research allowed its trajectories to be investigated longitudinally. Previous research has adopted numerous research frameworks comparing GHQ-12 scores at different time points (Graetz, 1991; Mäkikangas, 2016). These research frameworks generally investigated fit at numerous time points separately and concluded that their dimensional representations remained stable due to good fit being demonstrated at all timepoints. These research frameworks did not facilitate the investigation of trajectories. They also did not facilitate the investigation of time-variant covariates as was done in this thesis where the change in time-variant covariates was specifically investigated.

Longitudinal modelling has facilitated the separation of variance in mental health scores from the variance associated with wording effects by using Ye's (2009) dimensional representation to retain the unidimensional factor of mental health while accounting for this aforementioned variance through the inclusion of a method factor.

Longitudinal modelling of the GHQ-12 has also facilitated the investigation of whether multidimensional representations of the GHQ-12 exhibited variance between covariates. Failure of multidimensional models investigated in this thesis (Greatz, 1991) to demonstrate such have lead researchers to question the utility of treating the GHQ-12 as multidimensional (Gao et al., 2004; Shevlin and Adamson, 2004) and was one of the main reasons why this thesis adopted Ye's (2009) unidimensional representation.

Finally, longitudinal modelling facilitated the extraction of latent classes which represented participants displaying different trajectories over time, which was integral to the aims of this thesis.

9.6- Clinical Implications

This research could be of use to clinicians as it provides an insight into the social, psychological and biological variables which predispose individuals to various mental health trajectories. In relation to the provision of mental health service provision, this research has identified that significant predictors of mental health trajectories were social in nature and consequently this research adds to the extensive research which suggests that interventions which address one's social circumstances may improve outcomes. An example of an existing organisation that adopts this approach is the Recovery College which was run by the Western Health and Social Care Trust, a branch of the UK National Health Service operating in Northern Ireland. This organisation adopts an educational approach in the treatment of individuals who experience poor mental health. This organisation offered courses which addressed social circumstances such as the ‘Money Matters’ and ‘Walking to Wellness’ courses in their Spring 2020 Prospectus. (Recovery College, 2020). These interventions align with the findings in this thesis that ‘financial situation’ and ‘satisfaction with leisure time’ were indicators of mental health trajectory and that improvements in how individuals perceive their financial situation and how satisfied with their leisure activities encouraged individuals to display improving GHQ-12 scores over time. In a research setting, leisure activities have been shown to help maintain mental health while experiencing adversity (Ponde & Santana, 2000). Jonsdottir et al. (2010) also stressed that leisure activities increased resilience and that individuals who engaged in leisure activities, especially physical exercise, were at reduced risk of developing mental health difficulties.

Within a more general public health care policy environment, previous research has investigated the relationship between physical and mental health, with research as early as the 1970’s (Vaillant, 1979) identifying the interdependency of mental health and physical health. The findings of this research have identified not only that physical health was a predictor of class membership, but that activities which aid in physical health such as leisure activities were predictors of mental health. These findings can be supplemented by extant research which suggested that leisure activities in so-called green spaces, were particularly effective at building resilience to mental health deterioration (Wood et al. 2017)

In relation to employment issues, clinicians could use the findings of this thesis to inform interventions which are caused by employment issues. The findings demonstrated that changes in job satisfaction were a major predictor of mental health trajectories exhibited by participants. This was consistent with research by Bogg and Cooper (1995), who identified that civil servants who reported poor levels of job satisfaction also reported poorer mental health. This research was cross-sectional in nature, however, this thesis has provided an insight into how changes in job satisfaction affect mental health over time. Rout and Rout (1994) have also identified that job satisfaction can deteriorate following the introduction of a change to contract and working conditions and found that mental health decreased concurrently with a deterioration of job satisfaction. Clinicians who are caring for individuals who experience a significant decrease in their job satisfaction would be at significant risk of relapse if their job satisfaction deteriorated and therefore increased monitoring of individuals who experience a deterioration in their job satisfaction may be beneficial given it was identified as such a strong predictor of deteriorating mental health. It may also be in employers and government’s interest to develop mental health initiatives designed to build resilience in employees as Harpman et al. (2003) demonstrated the negative impact that poor mental health could have on business and the economy in general.

More generally, clinicians could identify individuals who are at risk based on some of the covariates identified as significant predictors of the deteriorating class. The development of a register which utilises the findings of this thesis could represent a meaningful future research project. Clinicians may also be able to tailor interventions of individuals who are experiencing poor mental health by addressing some of the covariates which have been shown to predispose individuals to steadily increasing

GHQ-12 scores. An example of this could be the association of increased satisfaction with leisure time and membership of the recovery class. Clinicians could use the results of this research to justify the recommendation of increasing leisure time as a way of facilitating mental health improvement.

Finally, while ethnicity was largely found to exhibit no discernable relationship with class membership, UK citizens from an African background were found to exhibit an increased likelihood of consistently reporting poor mental health relative to individuals from a white British background. The research was pertinent as it had been historically suggested that ethnic minorities suffered poorer mental health than ethnic majorities based on the associated economic and social disadvantage that they endured (Kleiner, Tuckman & Lavell, 1960; Fischer, 1969; Kramer, Rosen & Willis, 1973; Cannon & Locke, 1977; Mirowsky & Ross, 1980). The research in this thesis has suggested that certain ethnic minorities have increased predisposition to trajectories of mental health, most notably that UK residents from an African background were most likely to be members of the class that indicated consistently poor mental health and Caribbean participants were more likely to be members of the recovery class. Pinto, Ashworth & Jones, (2008) identified high rates of schizophrenia in these cohorts, with incidence rates being as much as nine times higher than the UK population average. Schizophrenia would be likely to be indicative of poor self-reported mental health over a prolonged period of time and may have been responsible for the elevated GHQ-12 scores. Furthermore, Mantovani, Pizzolati & Edge (2016) identified a decreased likelihood that participants from this background would engage in mental health services. This unwillingness to engage in mental health services may be in part responsible for elevated GHQ-12 scores over time. Clinicians may wish to engage in the promotion of mental health services amongst this cohort of the UK population in order to address the consistently high GHQ-12 scores reported.

9.7- Research implications

The results of this thesis may have a number of implications for researchers in the field of mental health. In terms of researchers who use the GHQ-12 as a measure of mental health, this research established the utility of using the GHQ-12 within a structural equation modelling framework rather than the caseness approach using summed scores which have traditionally been used in a clinical setting and proposed by the measure's author (Goldberg, 1988). The establishment of Ye's (2009) unidimensional representation of the GHQ-12 as appropriate for representative UK data was meaningful as it retained the unidimensional nature of the measure while accounting for the variance caused by positively and negatively worded items which may have lead researchers like Graetz (1991) to conclude that the GHQ-12 was multidimensional. This thesis has established that this dimensional representation remained stable over time which is particularly pertinent to researchers who may wish to use the GHQ-12 in a longitudinal setting as it is important for researchers to ascertain the extent to which dimensional representations remain stable over time.

While the research has focused on the GHQ-12, other measures utilise a similar structure to that of the GHQ-12, such as the Rosenberg Self Esteem Scale (RSA). Like the GHQ-12 researchers have suggested that a multidimensional representation of the RSA may have been a product of the positively and negatively worded items included in the measure (Yang & Wang, 2002; Marsh, 1996; Wang et al., 2001). The findings of this thesis could inform appropriate dimensional representations for researchers using similar measures to the GHQ-12 in a wide range of research settings.

The findings of this thesis are pertinent to researchers who subscribe to the concept of a '*P-factor*' of general mental health (Caspi et al., 2014). This research proposed that mental health could be appropriately and more parsimoniously measured through a P-factor of general mental health, rather than being diagnostically driven. This research utilises a similar approach and conceptualises general mental health as a single latent variable which captures mental health within that latent factor. The analysis contained within this thesis provided support as to the utility of conceptualising mental health in this way in a research environment.

The statistical techniques which this thesis employed, i.e. MLR estimation and data imputation in Chapter 8 may be informative to researchers who wish to use UKHLS data in future research. Research has utilised a number of techniques when handling missing data, using this dataset. Sage (2015) utilised listwise deletion when investigating the effect of retraining programs on the wellbeing of unemployed people, whereas Griffith and Jones (2019) utilised Bayesian estimation techniques when investigating factor structures of the GHQ-12. The utilisation of data imputation techniques overcame the methodological constraints that regressing covariates with varying levels of missingness to the primary measure entailed.

This thesis is also informative to researchers who wish to use the UKHLS database for research into mental health. The findings in Chapter 8 attempted to explain the trajectories displayed by participants, however, the variables included were not exhaustive. Researchers who wish to expand upon the findings of this thesis could follow the analytic framework of this thesis but include other variables which they believe may explain the exhibition of mental health trajectories over time. Researchers may also wish to incorporate the new waves of UKHLS data which have been released since these analyses were conducted.

The findings in relation to personality traits demonstrated that individuals who reported higher levels of neuroticism were at highest risk of exhibiting consistently poor levels of mental health over time. This research, however, has added an extra dimension to mental health research. The findings of this thesis suggested that while neuroticism was the largest predictor of mental health trajectories over time, all aspects of personality have an impact on mental health trajectories. Research has shown that depending on the self-report measure being investigated, different aspects of personality were correlated with mental health to different extents (Hayes & Joseph, 2003). Ye (2009) investigated the same dimensional representation that was used in this thesis and found that only neuroticism and extraversion displayed statistically significant relationships with the general factor of mental health. The findings in this thesis have shown that in relation to a latent general mental health variable derived from the GHQ-12, all aspects of personality have an effect. Some researchers have suggested that a general measure of mental health could simply be a measure of neuroticism (Brandes et al., 2019), however, the findings of this thesis would suggest that clinicians should not focus on neurotic personality traits to the exclusion of all others and that while affecting mental health to a lesser degree, a wide range of personality traits are indicative of mental health trajectories.

9.8- Limitations

This research was not without limitations, and while specific analytic and methodological limitations are detailed in the relevant discussion section, general limitations are discussed below.

The research was conducted between waves 1 through 5 of the UKHLS database. This timeframe covered the economic depression of 2008, and some research has used UKHLS data in a similar timeframe to investigate the effect of the global economic downturn on wellbeing (Bayliss, Olsen & Walthery, 2017). The research conducted in this timeframe may have been adversely affected by global economic factors which were outside the control of those who collected the data and may not have been representative of mental health trajectories in more stable economic times.

The analyses contained within this thesis were limited due to the range of covariates that were included in the UKHLS database. Had, for example, variables which related to an individual's genetic predispositions been available then they would have been included as genetic predispositions have been shown to be a predictor of mental health (Vigod & Stewart, 2009). While efforts were taken to include a range of variables which corresponded to all components of the biopsychosocial model (Gatchel, 1969) the majority of the variables included were social in nature and did not give as full a picture of the predictors on mental health trajectories as would have been possible if other pertinent variables were available.

Some of the variables that were included were either simplistically measured through the use of the single item variables. Examples of this were the use of an unvalidated truncated measure of social cohesion used in analysis conducted in chapter 4 and a lack of desirable variables that ideally would have been included in chapter 8's covariate analysis. Chapter 8's covariate analysis utilised the biopsychosocial model as a framework for covariate analysis, however, the UKHLS database did not include variables that versions of the biopsychosocial model (Gatchel, 1996) commonly included, especially psychological variables such as coping techniques.

The class solution that was selected in chapter 5 was chosen for reasons of interpretability and parsimoniously. Four classes which represented high and low stable scores alongside a decreasing and increasing trajectory were extracted from the data. These trajectories were linear in nature to aid interpretation. While the reasons for class structure selection are given in the relevant chapter, a 5 class structure would have added an extra dimension to research as it would have provided an insight into what predisposed participants to consistently report GHQ-12 scores which were indicative of better mental health than the reference group. The imposition of linear trajectories may also have masked trajectories which were not linear in nature such as exponential or quadratic in nature.

9.9- Future research

As mentioned in the limitations section above, some important components of the biopsychosocial model (Gatchel, 1996) were not available in the UKHLS database and therefore, were not investigated. Future research using a wider range of covariates may be meritorious for a fuller understanding of variables which affect mental health trajectories.

During the course of the research of this thesis, a number of extra waves of the UKHLS database have been released. While the waves investigated, have provided a sufficient period of time from which to observe change over time, the inclusion of extra waves in any future research may provide additional scope to observe variations over time.

Finally due to the linear nature of the classes investigated future research may wish to investigate class structures which were not only linear in nature and may wish to investigate quadratic or exponential trajectories of mental health in the UK population.

9.10- Summary

This thesis has provided a useful insight into the utility of measuring mental health longitudinally within a structural equation modelling framework. The research has identified an appropriate dimensional representation of the GHQ-12 for the UK population and has demonstrated that this remains stable over time.

The research has identified that different participants exhibited different trajectories over time with four trajectories being extracted and has attempted to explain those trajectories through modelling a wide range of covariates which were social, biological or psychological in nature. It found that social, biological and psychological variables were statistically significant predictors of mental health trajectories over time and that variables such as financial situation and satisfaction with their jobs and leisure time were the largest predictors of mental health trajectories. It also found that certain ethnicities, such as participants from a Caribbean and African background, were statistically significantly associated with class memberships.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Andrich, D., & Van Schoubroeck, L. (1989). The General Health Questionnaire: a psychometric analysis using latent trait theory. *Psychological Medicine*, *19*(2), 469-485.

Asparouhouv, T., & Muthén, B. (2013). Appendices for auxiliary variables in mixture modeling: 3-step approaches using Mplus. *MPlus user's guide*.

Barrett, P. (2007). Structural equation modelling: adjudging model fit. *Personality and Individual Differences, 42,* 815–824.

Bayliss, D., Olsen, W., & Walthery, P. (2017). Well-being during recession in the UK. *Applied research in quality of life*, *12*(2), 369-387.

Bogg, J., & Cooper, C. (1995). Job satisfaction, mental health, and occupational stress among senior civil servants. *Human relations*, *48*(3), 327-341.

Brandes, C. M., Herzhoff, K., Smack, A. J., & Tackett, J. L. (2019). The p factor and the n factor: Associations between the general factors of psychopathology and neuroticism in children. *Clinical Psychological Science*, *7*(6), 1266-1284.

Brodman, K., Erdmann, A. J., & Wolff, H. G. (1949). *Cornell medical index health questionnaire: Manual*. Cornell University Medical College.

Cannon, M. S., & Locke, B. Z. (1977). Being black is detrimental to one's mental health: Myth or reality?. *Phylon (1960-), 38*(4), 408-428.

Carr, L. G., & Krause, N. (1978). Social status, psychiatric symptomatology, and response bias. *Journal of Health and Social Behavior*, 86-91.

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., ... & Moffitt, T. E. (2014). The p factor: one general psychopathology factor in the structure of psychiatric disorders?. *Clinical Psychological Science*, *2*(2), 119-137.

Clark, L. A., Watson, D., & Reynolds, S. (1995). Diagnosis and classification of psychopathology: Challenges to the current system and future directions. *Annual review of psychology*, *46*(1), 121-153.

Deary, I. J. (2020). *Intelligence: A very short introduction*. Oxford University Press.

Fischer, J. (1969). Negroes and whites and rates of mental illness: Reconsideration of a myth. *Psychiatry*, *32*(4), 428-446.

Gao, F., Luo, N., Thumboo, J., Fones, C., Li, S. C., & Cheung, Y. B. (2004). Does the 12-item General Health Questionnaire contain multiple factors and do we need them?. *Health and Quality of Life Outcomes*, *2*(1), 63.

Gatchel, R. J. (1996). Psychological disorders and chronic pain: cause-and-effect relationships.

Goldberg DP, Williams P: A User's Guide to the General Health Questionnaire. 1988, Windsor: nferNelson

Goodchild, M. E., & Duncan-Jones, P. (1985). Chronicity and the general health questionnaire. *The British Journal of Psychiatry*, *146*(1), 55-61.

Graetz, B. (1991). Multidimensional properties of the general health questionnaire. *Social psychiatry and psychiatric epidemiology*, *26*(3), 132-138.

Greenberger, E., Chen, C., Dmitrieva, J., & Farruggia, S. P. (2003). Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: Do they matter?. *Personality and individual differences*, *35*(6), 1241-1254.

Griffith, G., & Jones, K. (2019). Understanding the population structure of the GHQ-12: evidence for multidimensionality using Bayesian and Exploratory Structural Equation Modelling from a large-scale UK population survey. *bioRxiv*, 584169.

Hankins, M. (2008). The reliability of the twelve-item general health questionnaire (GHQ-12) under realistic assumptions. *BMC public health*, *8*(1), 1-7.

Harpman, T., Reichenheim, M., Oser, R., Thomas, E., Hamid, N., Jawsal, S., Ludermir, A. and Aidoo, M. (2003). How to do (or not to do). (20).. *Health Policy and Planning*, *18*(3), pp.344-349.

Hasin, D., & Kilcoyne, B. (2012). Comorbidity of psychiatric and substance use disorders in the United States: current issues and findings from the NESARC. *Current opinion in psychiatry*, *25*(3), 165.

Hayes, N., & Joseph, S. (2003). Big 5 correlates of three measures of subjective well-being. *Personality and Individual differences*, *34*(4), 723-727.

Huang, C., & Dong, N. (2012). Factor structures of the Rosenberg self-esteem scale. *European Journal of Psychological Assessment*.

Jensen, A. R. (1998). *The g factor: The science of mental ability* (Vol. 648). Westport, CT: Praeger.

Jonsdottir, I. H., Rödjer, L., Hadzibajramovic, E., Börjesson, M., & Ahlborg Jr, G. (2010). A prospective study of leisure-time physical activity and mental health in Swedish health care workers and social insurance officers. *Preventive medicine*, *51*(5), 373-377.

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of general psychiatry*, *62*(6), 617-627.

Keyes, C. L. (2002). The mental health continuum: From languishing to flourishing in life. *Journal of health and social behavior*, 207-222.

Kleiner, R. J., Tuckman, J., & Lavell, M. (1960). Mental disorder and status based on race. *Psychiatry*, *23*(3), 271-274.

Kramer, M., Rosen, B., & Willis, E. (1973). Definitions and distributions of mental disorders in a racist society. *Racism and mental health*, 353-459.

Kramer, M., Rosen, B., & Willis, E. (1973). Definitions and distributions of mental disorders in a racist society. *Racism and mental health*, 353-459.

Kupfer, D. J., Kuhl, E. A., & Regier, D. A. (2013). DSM-5—The future arrived. *Jama*, *309*(16), 1691-1692.

Mäkikangas, A., Feldt, T., Kinnunen, U., Tolvanen, A., Kinnunen, M. L., & Pulkkinen, L. (2006). The factor structure and factorial invariance of the 12-item General Health Questionnaire (GHQ-12) across time: evidence from two community-based samples. *Psychological assessment*, *18*(4), 444.

Mantovani, N., Pizzolati, M., & Edge, D. (2017). Exploring the relationship between stigma and help-seeking for mental illness in African-descended faith communities in the UK. *Health Expectations*, *20*(3), 373-384.

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors?. *Journal of personality and social psychology*, *70*(4), 810.

Mirowsky, J., & Ross, C. E. (1980). Minority status, ethnic culture, and distress: A comparison of Blacks, Whites, Mexicans, and Mexican Americans. *American Journal of Sociology*, *86*(3), 479-495.

Montazeri, A., Harirchi, A. M., Shariati, M., Garmaroudi, G., Ebadi, M., & Fateh, A. (2003). The 12-item General Health Questionnaire (GHQ-12): translation and validation study of the Iranian version. Health and quality of life outcomes, 1(1), 66.

Newman, D. L., Moffitt, T. E., Caspi, A., & Silva, P. A. (1998). Comorbid mental disorders: implications for treatment and sample selection. *Journal of abnormal psychology*, *107*(2), 305.

Perenboom, R., Oudshoorn, K., van Herten, L., Hoeymans, N., & Bijl, R. (2000). Lifeexpectancy in good mental health: establishing cut-offs for the MHI-5 and GHQ-12. *Leiden: TNO*.

Pinto, R., Ashworth, M., & Jones, R. (2008). Schizophrenia in black Caribbeans living in the UK: an exploration of underlying causes of the high incidence rate. British Journal of General Practice, 58(551), 429-434.

Ponde, M. P., & Santana, V. S. (2000). Participation in leisure activities: Is it a protective factor for women's mental health?. *Journal of Leisure Research*, *32*(4), 457-472.

Rosenberg, Morris. "Rosenberg self-esteem scale (RSE)." *Acceptance and commitment therapy. Measures package* 61.52 (1965): 18.

Rout, U., & Rout, J. K. (1994). Job satisfaction, mental health and job stress among general practitioners before and after the new contract—a comparative study. *Family Practice*, *11*(3), 300-306.

Sage, D. (2015). Do active labour market policies promote the well-being, health and social capital of the unemployed? Evidence from the UK. *Social Indicators Research*, *124*(2), 319-337.

Shevlin, M., & Adamson, G. (2005). Alternative factor models and factorial invariance of the GHQ-12: a large sample analysis using confirmatory factor analysis. *Psychological assessment*, *17*(2), 231.

Smith, A. B., Oluboyede, Y., West, R., Hewison, J., & House, A. O. (2013). The factor structure of the GHQ-12: the interaction between item phrasing, variance and levels of distress. *Quality of Life Research*, *22*(1), 145-152.

Spearman, C. (1961). " General Intelligence" Objectively Determined and Measured.

Tupes, E. C. (81). i Christal, RE (1961.). *Recurrent personality factors based on trait ratings*.

Vaillant, G. E. (1979). Natural history of male psychologic health: Effects of mental health on physical health. *New England Journal of Medicine*, *301*(23), 1249-1254.

Vigod, S. FN., & Stewart, D. E. (2009). Emergent research in the cause of mental illness in women across the lifespan. *Current Opinion in Psychiatry*, *22*(4), 396-400.

Wang, J., Siegal, H. A., Falck, R. S., & Carlson, R. G. (2001). Factorial structure of Rosenberg's Self-Esteem Scale among crack-cocaine drug users. *Structural Equation Modeling*, *8*(2), 275-286.

Westerhof, G. J., & Keyes, C. L. (2010). Mental illness and mental health: The two continua model across the lifespan. Journal of adult development, 17(2), 110-119.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain.

Wood, L., Hooper, P., Foster, S., & Bull, F. (2017). Public green spaces and positive mental health–investigating the relationship between access, quantity and types of parks and mental wellbeing. *Health & place*, *48*, 63-71.

Yang, Y., & Wang, D. (2002). Retest of the bidimensional model of Rosenberg self-esteem scale. *Chinese mental health journal*, (09).

Ye, S. (2009). Factor structure of the General Health Questionnaire (GHQ-12): The role of wording effects. *Personality and Individual Differences*, *46*(2), 197-201.

Appendix- Review of Fit Statistics Used Throughout this Thesis

The purpose of this appendix was to provide an overview of the statistical procedures used throughout this thesis. Listed below are the various fit statistics utilised throughout this thesis alongside details on characteristics that are pertinent to their use.

RMSEA (Root Square Mean Error of Approximation)

Equation of RMSEA as shown in Kenny (2005)

$$\frac{\sqrt{(\chi^2 - df)}}{\sqrt{[df(N - 1)]}}$$

X^2= Chi squared result
df= Degrees of Freedom
N= Participants

Root square mean error of approximation (RMSEA) is a chi-squared derived fit statistic that overcomes problems with small samples by analysing the difference between the hypothesised model and the population covariance. The model tends to be positively biased, i.e. it scores higher than other fit statistics (Kenny, 2015). It is currently the most commonly reported statistic of fit (Kenny, 2015). This measure was shown to be ineffective in models with low degrees of freedom (DF) to such an extent that Kenny, Kanistan and McCoach (2014) argue that it should not even be included in models with low DF. Brown (2015) argued that acceptable models have a RMSEA of <0.06.

TLI (Tucker Lewis Index)

Equation of TLI as shown in Kenny (2005)

$$\frac{\chi^2/df(\text{Null Model}) - \chi^2/df(\text{Proposed Model})}{\chi^2/df(\text{Null Model}) - 1}$$

X^2= Chi squared result
Df = Degrees of Freedom
N= Participants

The Tucker Lewis Index or TLI as it is also known, the non-normed fit index (NNFI) is an incremental index that compares a model's chi-squared values with that of a baseline or null model. Both the TLI and comparative fit index (CFI) have assigned cut off points of 0.9 (Awang, 2012). TLI has been shown to have a large standard error and not being scaled, it is not as easy to interpret as scaled indices. Bentler and Hu (1993) suggested that these type of analyses are not as vulnerable to sample size as many other fit indices.

CFI (Comparative Fit Index)

Equation for CFI as shown in Kenny (2005)

$$\frac{X^2\ \textbf{(Null Model) - } X^2\ \textbf{(Proposed Model)}}{X^2\ \textbf{(Null Model)}}$$

X^2= Chi-squared result

The Comparative Fit Index (CFI) was created by Bentler (1990). It operates by comparing the chi-squared values of a proposed model with a baseline model, as shown above. This index is superior to the TLI in terms of having a much smaller standard error, lower bias and being scaled; it is easier to interpret and less vulnerable to distortion by small sample sizes. Hu and Bentler (1999) also raised concerns that a value of 0.9 was too low a threshold to indicate a good fit. They suggested that a cut off of 0.95 would be more appropriate.

SRMR (Standardised Root Mean Square Residual)

The equation for SRMR as taken from Hu and Bentler (1999)

$$SRMR = \sqrt{\{2\sum_{i=1}^{P}\sum_{j=1}^{i}[(S_{ij} - \hat{o}_{ij})l(s_{ii}s_{jj})]^2\}lp(p+1)}$$

T_T = Statistic for Target Model

T_A = Statistic for Baseline Model

DF_r= Degrees of Freedom for Target Model

DF_a= Degrees of Freedom for Baseline Model

P= Number of Observed Variables

S_{ij}= Observed Covariances

$\hat{O}_{ij}$= Reproduced Covariances

S_{ii} and S_{jj} = Observed Standard Deviations

Standardised Root Mean Square Residual (SRMR) is an absolute measure of fit (Kenny, 2015). It measures the standardised difference between the observed correlation and the predicted correlation. It tends to produce positively biased results, and this effect is more significant in studies with a small sample and fewer degrees of freedom (Kenny, 2015). Concerning cut-off points, a value of less than 0.08 is sufficient to be considered a good fit (Hu and Bentler, 1999). This model does not have a penalty for complexity as many other models do.

Akaike Information Criterion (AIC)

The AIC is computed using the formula outlined in Akaike (1987), using the formula shown below. It is a comparative statistic and in the case of structural equation modelling compares the suggested model against an unrestricted model. It is important to note that the AIC is a measure of 'badness of fit' and as such better fitting models will score lower on this test. As can be seen from the formula, AIC imposes a penalty of 2 for each parameter estimated, however, this does not scale with sample size. While, researchers have questioned the accuracy of fit statistics, including the AIC, (Morgan, 2012), the same simulation study suggested that the AIC had the highest accuracy of identifying appropriate class structures. It must be noted that Morgan did stress that under the circumstances he tested, i.e. when classes were separated by a small degree, that all tests struggled.

$AIC = (- 2) \log + 2 (p)$

Where p equals the number of parameters.

Bayesian Information Criterion (BIC)

This technique is similar to the AIC and was developed based on it. This statistic increases the penalty to a model as sample size increases, unlike the aforementioned AIC. Some statisticians have suggested that this technique places too much emphasis on parsimony (Kenny, 2015), however, others argue that the increased likelihood of identifying fewer factors predisposes this measure to more accurate outcomes (Morgan, 2015). BIC techniques were cited as having a number of limitations, as cited in Giraud, 2015. The only relevant limitation, however, was that the sample size must be significantly larger than the number of parameters in the model. Considering the sample size in this analysis, this was not viewed as a major limitation. Furthermore, Morgan (2015) found that this statistic was particularly effective in larger samples.

BIC = (- 2) log + 2 (p) log (n)

In the above example, 'p' equals the number of free parameters and n equals the sample size

Sample Size Adjusted BIC (SSABIC)

This variation on the BIC, proposed by Sclove (1987) places a parameter based penalty on the model which is relative to sample size, however, this penalty is not as severe as in the BIC. This fit statistic was found to be particularly accurate relative to its competitors in 3 and 4 class solutions and in samples with a small sample size (Morgan, 2012).

SSBIC = (−2) log L + plog [(n + 2)/24]

In this formula, 'p' equals the number of parameters and 'n' the sample size.

Adjusted Lo Mendel Reuben (ALMR)

This technique is described as a nested test, and these are commonly used in SEM frameworks. They refer to testing two similar models which only vary in respect to parameterisation. The LMR and its successor, ALMR global measure of model fit where the likelihood ratios are estimated as shown in Lo, Mendel and Rubin (2001). This estimation had been necessary as previously, the fact that the likelihood ratios between models of different classes did not follow a chi-squared pattern had hampered analysis. The aforementioned estimation technique made comparison analysis between two models possible (Lo Mendel & Rubin, 2001). It is used in class analysis to determine the difference between two classes, for example, this test on a two-class solution tests the difference in fit between a two-class solution and one class. A P value of <0.05 would imply that the class being tested is significantly better fitting than the preceding class (Nylund et al., 2007). Following the development of the LMR, the authors developed ad hoc adjustments, which were designed to make the inferences that this test generated more accurate (Morgan, 2015).

Entropy

Entropy is a measure of how well the classes in a model are identified (Asparouhov and Muthen 2018). It uses the below formula, taken from Asparouhov and Muthen (2018), shown below. In this calculation, Entropy values close to one indicate clear delineation of classes (Celeux & Soromenho, 1996).

$$E = 1 + \frac{1}{N\log(k)}\left(\sum_{i=1}^{n}\sum_{k=1}^{k} P(C = k|U_i)\log(P(C = k|U_i))\right)$$

In this formula, C is the latent variable, K the number of classes, N the number of participants and U_i the vector of all the indicator variables.

References

Akaike, H. (1987). Factor analysis and AIC. In *Selected papers of hirotugu akaike* (pp. 371-386). Springer, New York, NY.

Asparouhouv, T., & Muthén, B. (2013). Appendices for auxiliary variables in mixture modeling: 3-step approaches using Mplus. *MPlus user's guide*.

Awang, Z. (2012). Structural equation modeling using AMOS graphic. Penerbit Universiti Teknologi MAR

Brown, S., Gray, D., & Roberts, J. (2015). The relative income hypothesis: A comparison of methods. *Economics Letters, 130*, 47-50.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of classification*, *13*(2), 195-212.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424–453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Kenny, D. (2019). *SEM: Fit (David A. Kenny)*. [online] Davidakenny.net. Available at: http://davidakenny.net/cm/fit.htm [Accessed 29 Aug. 2019].

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research, 44*, 486-507.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*(3), 767-778.

Morgan, G. B. (2015). Mixed mode latent class analysis: An examination of fit index performance for classification. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(1), 76-86.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling: A multidisciplinary Journal*, *14*(4), 535-569.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*(3), 333-343.

Tupes, E. C., & Christal, R. E. (1992). Recurrent personality factors based on trait ratings. *Journal of personality*, *60*(2), 225-251.