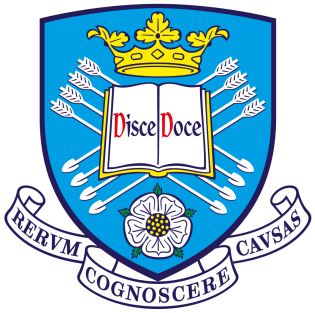

The University Of Sheffield.

DOCTORAL THESIS

A Quantitative Exploration of the Mechanisms Relating Obesity, Depression and Socioeconomic Position

Author:
Thomas BAYLEY

Supervisors:
Dr. Mark STRONG
Dr. Samantha CATON
Dr. Paul BISSELL

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Wellcome Trust Doctoral Training Centre in Public Health
Economics and Decision Science
School of Health and Related Research

15th April 2022

Abstract

A Quantitative Exploration of the Mechanisms Relating Obesity, Depression and Socioeconomic Position

Current research on the relationship between obesity and depression suggests that it is complicated and potentially complex. In particular, there are multiple mechanisms that might relate obesity and depression, which may interact with each other as well as being influenced by factors such as age, sex and socioeconomic position. However, there is currently a lack of longitudinal evidence exploring these proposed mechanisms with evidence in support of them relying on cross-sectional associations. Additionally, whilst the associations between socioeconomic position and both obesity and depression are well established, the way in which socioeconomic position influences, or is influenced by mechanisms that might relate the two conditions is unclear.

In this thesis, I will explore evidence for proposed mechanisms relating obesity, depression and socioeconomic position. This exploration is presented in two main stages. In the first stage, I review the literature on the relationship between obesity and depression to examine broadly what is currently known about mechanisms that relate the two conditions. In the second stage, I will analyse longitudinal data from the United Kingdom using a combination of statistical and mathematical modelling. The first of two models presented uses a Structural Equation Model to examine whether functional impairment, physical activity and diet mediate the relationship between obesity and depression bi-directionally, adjusting for the influence of age, sex and socioeconomic position. The second model then uses Agent-based simulation to examine the role of stigma as a mechanism that might produce relationships between obesity, depression and socioeconomic position.

Acknowledgements

I would like to thank my supervisors Dr. Mark Strong, Dr. Samantha Caton, and Dr Paul Bissell for all of their advice, guidance and support, both academically and pastorally. I feel very fortunate to have worked with such a great supervisory team.

To Mark, thank you for your unwavering commitment to helping me to develop both academically and personally over the last few years. I am very grateful for your prescience in knowing how to push my work on, your generosity with your time, and all of the efforts you have taken to support my well-being throughout a challenging period.

To Sam, thank you for all of your pragmatic guidance, feedback and support in both academic and personal matters that has helped me to always keep moving forward in this project. And to Paul, thank you for the useful discussions that greatly helped me to sculpt and develop my research whilst in its early stages.

I would also like to thank Dr. Alan Brennan for his support in helping me keep my project moving forward during the early stages of my PhD.

Thanks also to all my fellow PhD students at ScHaRR, current and former, for the friendships and community that have made this PhD all the more worthwhile. Finally, I would like the thank my mum, brother and sister for all their love and care; I look forward to celebrating with you all soon.

This work was supported by the Wellcome Trust [203956/Z/16/Z].

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

In the United Kingdom (UK), the vast majority of the disease burden comes from non-communicable diseases. Based on data from the global burden of disease study, in 2019, 89% of deaths in the UK were attributable to non-communicable diseases (NCDs, 30% cardiovascular diseases (CVDs), 31% cancers, 8% chronic respiratory diseases, 1% diabetes, 19% other NCDs) (Global Burden of Disease Collaborative Network., 2019). This share of the disease burden has also remained largely stable: in 1990, NCDs were also responsible for 89% of deaths in the UK, although the split between the causes was different (45.5% CVDs, 26.5% cancers, 6% chronic respiratory diseases, 1% diabetes, 10% other NCDs).

Evidence suggests that the prevalence of these diseases is largely affected by lifestyle oriented risk factors, primarily: smoking, alcohol consumption, poor diet and low physical activity levels. Outside of these primary risk factors, other intermediary outcomes are often investigated for their effect on NCDs. One such outcome is obesity, which occurs principally as a result of poor diet and low physical activity, although a wider range of environmental, social, psychological and biological factors can also influence its development (Vandenbroeck, Goossens and Clemens, 2007). Studies have shown that being obese raises an individual's risk of developing many of the aforementioned conditions such as cardiovascular disease (Lavie, Milani and Ventura, 2009), stroke (Suk et al., 2003), Type 2 diabetes (Abdullah et al., 2010), coronary heart disease (Logue et al., 2011) and various cancers (Bianchini, Kaaks and Vainio, 2002).

According to the World Health Organisation, for an individual to be obese, they must possess "abnormal or excessive body fat accumulation that may

impair health" (World Health Organization, 2021). Further to this, they classify individuals as overweight or obese based on their body mass index (BMI). Adults who have a BMI above 25 are considered overweight, and those with a BMI above 30 are considered obese (World Health Organization, 2021). For children, obesity and overweight are defined with reference to the WHO child growth standards median. Specifically, children under 5 are considered overweight if their weight for height is two standard deviations above the WHO growth standards median and obese if it is three standard deviations above the median. Similarly, children ages 5 to 19 are overweight if their weight-for-height is one standard deviation above the growth reference median, and obese if it is two standard deviations above (World Health Organization, 2021).

Data from the Health Survey for England suggests that recently the prevalence of obesity has almost doubled from 15% in 1993 to 27% in 2015. Additionally, it was found in the same study that in 2015, 58% of adult women and 68% of adult men are either overweight or obese and in children, over one in five of reception age are overweight or obese with this proportion increasing to over one in three by school year 6 (NHS Digital, 2017). This high prevalence is particularly troubling when considering evidence that shows many individuals who attempt weight-loss are unable to maintain a lower weight long-term (Wing and Phelan, 2005).

The increase in obesity prevalence, combined with its health consequences, has meant obesity is now a major contributor to ill health in the UK population. Alongside the health impacts, the management of obesity and its consequences creates a huge burden on the resources within the UK's health system. For example, in 2010 diabetes cost the UK an estimated £13.8bn (Kanavos and Aardweg, 2012) and NHS trusts have faced the need to extend obesity specific services such as bariatric ambulances and surgery. In particular, the NHS spent £6.1bn treating obesity-related ill-health in 2014-15 and this figure has been projected to rise to £9.7bn per year by 2050 (Holmes, 2021).

Alongside a growing burden of physical illness, there has also been greater recognition of the effects of poor mental health. Previous estimates on the impact of poor mental health found it contributed up to 22.8% of disability (Department of Health, 2011). This is greater than that of cardiovascular disease and cancer, which are the most common causes of death. Similarly to obesity, rates of poor mental health have risen over the last few decades.

Data from the Adult Psychiatric Morbidity Survey suggested that since the year 2000, the prevalence of common mental health disorders have steadily risen in women, whilst remaining stable in men (Mcmanus et al., 2016).

One of the most prominent common mental disorders in the United Kingdom is depression. In the 2014 Adult Psychiatric Morbidity Survey, around 3% of adult men and 4% of adult women had depression (Mcmanus et al., 2016). Whilst depression is far less widespread than is the case for other health issues such as obesity, depression still has a sizeable effect on the U.K population. Data from the global burden of disease study found that depression was responsible for nearly 3% of all disability-adjusted life years (DALYs) lost in 2019 in the U.K (Global Burden of Disease Collaborative Network., 2019).

Within this health burden, depression has been shown to have a wide range of health impacts. Studies of the general population have shown that depression has a negative impact on quality of life (Pyne et al., 1997), which may occur through impaired social functioning and reduced social support (Angermeyer et al., 2002). Additionally, it has been linked to shrinkage in the hippocampus and reduced memory function (Gorwood et al., 2008; Sheline, Gado and Kraemer, 2003).

Depression is also linked with poorer physical health, and particularly chronic health conditions. Individuals with chronic health conditions are significantly more likely to have depression than those who do not and those with comorbidity of depression and a chronic health condition invariably experience the worst health states (Moussavi et al., 2007). Depression can worsen the effects of existing chronic diseases through additional functional impairment and is associated with a 50% increase in medical costs when treating chronic illnesses (Katon, 2003). Depression is also associated with an increased incidence of physical chronic diseases, such as coronary heart disease (CHD) (Ferketich et al., 2000) and with increased general and cause-specific mortality (Mykletun et al., 2007; Mykletun et al., 2009).

As is the case with obesity, depression also has a significant economic cost. For example, in 2011, it was estimated that depression costs the U.K economy £7.5bn a year (Department of Health, 2011).

Physical health and mental health conditions do not evolve in complete isolation. Investigations into the relationship between physical and mental health have suggested that long-term physical health conditions in particular are

frequently associated with mental health problems (Naylor et al., 2012). A report into the association between long-term health conditions and mental health found that 30% of people living with a long-term health condition had a mental health problem, and 46% of people with a mental health problem had a long-term health condition (Naylor et al., 2012).

Within the relationship between mental and physical health, depression has been shown to be associated with multiple conditions for which obesity is a major risk factor. In particular, depression is between two and three times more common in those who have a cardiovascular disease and people living with Diabetes are between two and three times more likely to have depression than those in the general population (Naylor et al., 2012). These associations, as well the association between depression and chronic physical health issues generally presents considerable opportunity for comorbidities to occur between obesity and depression.

Such comorbidities between physical and mental health conditions have been shown to exacerbate the negative outcomes associated with ill-health. For example, patients with these comorbidities experience poorer clinical outcomes and disease prognosis, lower quality of life and poorer self-care and health behaviours (Naylor et al., 2012). These additional complications then increase the costs to the health system of these conditions through increased service use (Naylor et al., 2012).

However, despite knowledge of the strength of these relationships, and the profound impacts they can have, there is limited knowledge about the pathways through which physical and mental health conditions are related (Ohrnberger, Fichera and Sutton, 2017). The underlying mechanisms may be complex, involving a range of factors, and the associations may be bi-directional (Naylor et al., 2012).

Given the already complex challenges that the United Kingdom faces from dealing with obesity and depression in isolation, understanding the way in which they may relate is important. Supporting sustained weight-loss already presents a significant challenge, with the majority of individuals who engage in weight-loss eventually regaining any weight they lose (Hall and Kahan, 2018). However, a portion of the difficulties associated with obesity management may be attributable to the impacts of depression. Hence, resolving complications within obesity treatment that are due to depression, or preventing people with obesity from developing depression therefore might reduce the complexity of obesity management, allowing for more effective

weight management and weight loss. Additionally, understanding how obesity may confer risk for depression will allow interventions and treatments to be designed that prevent these challenging comorbidities developing.

Another significant challenge surrounding obesity and depression is the impact of health inequalities. Across many different health outcomes, more deprived individuals are more likely to suffer from these poor health outcomes than are less deprived individuals. This pattern is no different for obesity and depression. Within depression, Lorant et al. (2003) found that individuals with low socioeconomic position were at higher odds of developing a new depressive episode as well as being at higher odds of experiencing persisting depression. Similarly, evidence synthesised for the 2010 Marmot Review suggested that individuals in the most deprived quintile of society were twice as likely as those in the least deprived quintile to experience a common mental health disorder, which includes depression (Marmot, 2010).

In the case of obesity, both adults and children with low socioeconomic position are more likely to be overweight or obese than those with higher socioeconomic position. Data from the National Child Measurement Program suggests that obesity prevalence is roughly two times higher in the most deprived decile of children compared to the least deprived decile (Public Health England, 2017). In adults, data from the Health Survey for England found that there was a difference in obesity prevalence of 17% in women and 8% in men between the most deprived and least deprived areas, such that more deprived areas had a higher prevalence of obesity. This difference also appears to be widening over time (Holmes, 2021).

Reducing health inequalities, such as those seen in obesity and depression, is a key health priority in the UK (NHS, 2019). In order to facilitate the development of interventions and solutions to the impact of health inequalities, it is important to understand how socioeconomic position influences and is influenced by the development of poor health. As with the association between mental and physical health, these mechanisms of influence may be complex and so exploring the nature of these mechanisms allows more comprehensive solutions to be developed that minimise the chance of unintended consequences occurring.

1.2 Thesis Structure

1.2.1 Thesis aims

In light of the above, in this thesis I will conduct a quantitative exploration of the relationships between obesity, depression and socioeconomic position with two aims. The main aim of this thesis will be to explore the mechanisms that might relate obesity to depression, so that we may better understand how comorbidities of these two highly prevalent conditions might develop. A secondary aim will be to explore how socioeconomic position relates to these mechanisms.

1.2.2 Chapter Structure

The remainder of this thesis is presented in seven main chapters (2-8). In Chapter 2, I have conducted a broad review of the literature with the aim of getting a comprehensive understanding of the relationship between obesity and depression. Particularly, this chapter will focus on what is currently known about the strength of the relationship between the two conditions, as well as what is known about the potential mechanisms that relate them.

Following on from this literature review, in Chapters 3-7, I then build upon this literature by investigating some of the hypothesised mechanisms between obesity and depression that were highlighted in Chapter 2. In Chapters 3, 4 and 5, I present a Structural Equation Model (SEM) exploring whether physical function, physical activity and diet relate obesity to depression, and how socioeconomic position interacts with these mechanisms. Within this, Chapter 3 contains an exploratory analysis of the Whitehall II dataset, which will be used to provide context for the substantive Structural Equation Model analysis. Chapter 4 then discusses how missing data will be treated, before Chapter 5 presents my main Structural Equation Model of the hypothesised mechanisms.

In Chapters 6 and 7, I then present an Agent-based exploration of the role of stigma in generating the relationships between obesity and both depression and socioeconomic position. In Chapter 6, a simple initial model is presented, before a revised model that includes more detail on the proposed mechanisms is presented in Chapter 7.

Finally, in Chapter 8 the results of the two modelling studies are discussed within the context of what they can tell us about causality in the relationship between obesity and depression.

Chapter 2

The Association between Obesity and Depression

2.1 Introduction

An early review of the relationship between obesity and depression noted that studies overall had not found consistent evidence of a general association between obesity and depression (Friedman and Brownell, 1995). However, instead of following the conclusions of previous studies that suggested this indicated no association between obesity and psychological outcomes existed, they instead argued that the inconsistent evidence was indicative of an inconsistent phenomenon, in which some, but not all, obese individuals were at risk of psychological outcomes such as depression. Friedman and Brownell (1995) also suggested that the methodologies employed thus far in researching this association had not been adequate to exclude this possibility. From this, they suggested a new framework for future research aimed at examining which obese individuals are more likely to be depressed, and also what the causal mechanisms behind any uncovered associations may be.

In this chapter I have conducted a review of reviews, with the aim of exploring as widely as possible the research into the association between obesity and depression conducted since the meta-analysis by Friedman and Brownell (1995). The structure of the review will loosely follow the generations of research described by Friedman and Brownell in their review. Firstly, I will examine the statistical evidence by analysing meta-analyses and systematic reviews of data into the association between obesity and depression. Secondly, I will outline the range of risk factors explored in the literature, and in the final section of the review I will explore the range of hypothesised causal

mechanisms that have been highlighted in the reviews. In addition, evidence presented in the reviews supporting the effect of these risk factors and causal pathways will be examined.

2.2 Review Methods

2.2.1 Search Strategy

To explore the literature surrounding the association between obesity and depression, review papers published from 1995 onwards were searched for in the Medline, psychINFO and Web of Science databases. Searches were conducted by combing of a mixture of focused subject heading searches with keyword searches of obesity, depression and a number of similar terms. Searches for obesity related and depression related literatures were conducted separately, with the final search looking for papers in the overlap of these two search categories. To limit the amount of basic science literature retrieved, keyword searches were limited to retrieving papers only where the keyword appeared in the title of the paper. A total of 242 papers were retrieved (79 from Medline, 20 from psychINFO, 143 from Web of Science). After removing duplicates 208 reviews remained and were screened for relevance using the titles and abstracts.

Reviews were excluded primarily based on considerations of scope and population of interest. Reviews were sought to provide evidence for obesity-depression relations in the general population, and therefore reviews focusing on non-representative populations were excluded. The main populations that were excluded were those of expectant and new mothers, and those seeking treatment for weight loss, as evidence suggests that this population has different characteristics to the obese in the wider population (Fitzgibbon, Stolley and Kirschenbaum, 1993). Reviews that focused narrowly on one aspect of obesity-depression associations were also excluded. For example, papers that considered only biological pathways and physiological overlaps between the two conditions were excluded. Following this procedure, 21 articles were included for full text review.

2.2.2 Data Extraction

Data extraction was conducted in two broad sections. The first section was concerned with extracting what evidence the articles presented on the direct association between obesity and depression outcomes, and the second focused on the evidence presented about risk factors for comorbidity and causal mechanisms relating to the two conditions. For the first section, data was extracted from the reviews under the following main questions:

1. Whether a systematic review or analysis of data had been conducted
2. What type of evidence was presented
3. What this evidence suggested about the association between obesity and depression

For the second section, data was extracted under the following main themes:

1. What risk factors were examined/presented
2. What evidence was presented surrounding these factors
3. What causal pathways were examined/presented
4. What evidence was presented regarding these causal pathways
5. What conclusions were drawn

In addition to these main themes, reviews were also examined for comments on sources of heterogeneity and publication bias. All data was collated on a Microsoft excel spreadsheet.

2.3 Data on the Association between Obesity and Depression

Of the 21 papers that were included in the final review, 15 provided a comprehensive review of data investigating the association between obesity and depression. Of these 15 papers, 10 were meta-analyses and five provided a systematic data summary of epidemiological studies. The remaining six papers did not provide a comprehensive review of data investigating the association between obesity and depression in general. However, narrative summaries of the research findings from epidemiological research were presented in two of these papers, whilst the remaining four papers focused on a discussion of

the other variables relevant to the obesity-depression relationship, and exploration of the mechanisms behind the association.

2.3.1 Meta-analyses

Five meta-analyses included studies of longitudinal data only, three included studies of cross-sectional data only, and two included both studies of longitudinal and of cross-sectional data. In the seven studies of longitudinal data, four investigated both obesity as a predictor of future depression and depression as a predictor of obesity, one investigated depression as a predictor of obesity only and two studies examined obesity as a predictor of depression only. A summary of the results found in each of the studies is presented in Tables 2.1 and 2.2. However, the longitudinal data from Pereira-Miranda et al. (2017) has been omitted from the table since only one cohort study was included.

2.3.1.1 Longitudinal Data

Obesity as a Predictor of depression

All six meta-analyses which examined longitudinal data looking at obesity as a predictor of depression found significant evidence to support this association. The strongest effect estimate was found in Luppino et al. (2010) who found that the odds of depression were 1.55 times higher in people with obesity than those without. They also found that being overweight increased the risk of future depression (OR=1.27), however, this was only significant in adults over 20 years of age (Luppino et al., 2010). The more recent meta-analysis by Mannan et al. (2016a) found evidence of a similar but weaker association between obesity and future depression. They found that individuals with obesity had an 18% increased risk of being depressed compared those who were not obese, but found no significant risk difference, and no increased risk of depression in those who were only overweight. The same authors conducted an equivalent analysis for adolescent populations, in which they found adolescents with obesity were 40% more likely to become depressed than adolescents who were not obese, with a risk difference of 1% between the groups (Mannan et al., 2016b). Rooke and Thorsteinsson (2008) found similar results in their meta-analysis, with obesity having a small but significant correlation with future depression (r = 0.05). The papers by Jung et al. (2017) and Pereira-Miranda et al. (2017) included both longitudinal and cross-sectional data.

In the Meta-analysis by Jung et al. (2017), six longitudinal studies were included that investigated obesity as a predictor of future depression. In this subset of studies there was a non-significant pooled odds ratio of 1.13 for obesity as a predictor of depression. However, when stratifying by gender, the odds in women increased to 1.26 which was significant at the 5% level. A similar trend was found for overweight as a predictor of depression, with pooled results in the cohort studies producing a small non-significant odds ratio of 1.02 which rose to 1.16 in women when stratified by gender. This odds ratio was also significant at the 5% level. Neither obesity nor overweight in men was significantly associated with depression in the subgroup of cohort studies in general. The meta-analysis by Pereira-Miranda et al. (2017) included only one cohort study, which showed an increased risk of depression in men with obesity, but not in women. However, given this is the result of a single study, it presents much weaker evidence than that of the aforementioned papers.

TABLE 2.1: Summary of Longitudinal results from Meta-analyses

Author (Year)	Data Reviewed	Obesity/Overweight as Predictor of Depression	Depression as Predictor of Obesity/Overweight
Blaine (2008)	23 samples from 16 included studies		18/23 samples provided evidence of depression leading to weight gain. OR = 1.47, CI = (1.16,1.85)
Jung et al. (2017)	Extracted data from 76 studies. For both overweight to depression, and obesity to depression, 6 cohort studies were analysed	Obesity: OR = 1.33, CI = (0.96-1.34); Overweight: OR = 1.02, CI = (0.84-1.23)	
Luppino et al. (2010)	Longitudinal data from 15 studies	Obesity: OR = 1.55 (1.22,1.98); Overweight: OR = 1.27, (1.07,1.51)	Obesity: OR = 1.58 (1.33,1.87); Overweight: not predictive.

TABLE 2.1: Summary of Longitudinal results from Meta-analyses (continued)

Author (Year)	Data Reviewed	Obesity/Overweight as Predictor of Depression	Depression as Predictor of Obesity/Overweight
Mannan et al. (2016a)	21 articles, 19 included in meta-analysis of longitudinal data	Obese: RR = 1.18 (1.04,1.39) RD = 0.01 (-0.01,0.05); Overweight: RR and RD non-significant.	Obese: RR = 1.37 (1.17, 1.48) RD = 0.02 (0.01,0.03) Overweight: RR = 1.17 90.77,1.77) RD non-significant.
Mannan et al. (2016b)	13 longitudinal studies, 7 examined obesity to depression, 6 depression to obesity	Obese adolescents at increased risk for depression: RR = 1.4, CI = (1.16,1.70), RD: 0.01, CI (0.00,0.02).	Depressed adolescents at increased risk for obesity: RR = 1.70, (1.40-2.07) RD: 0.04, (0.01, 0.06)
Rooke and Thorsteinsson (2008)	2 Meta-analyses, one for each direction of association. 26 studies for depression as a predictor of weight gain. 13 studies for weight gain as a predictor of depression.	Small but significant weighted average correlation between obesity/overweight and future depression (r=0.05)	Small but statistically significant weighted average correlation between depression and future weight gain (r=0.07)

Depression as a Predictor of Obesity

Five meta-analyses of longitudinal data provided evidence about depression as a predictor of obesity. The review by Luppino et al. (2010) found that the odds of future obesity were 1.58 times higher in the depressed than the non-depressed, but that depression was not significantly predictive of overweight. Similarly, the review of adult populations by Mannan et al. (2016a) found that depressed adults were 37% more likely than non-depressed adults to develop obesity, with a risk difference of 2% between the two groups. In their review of adolescent samples, the risk for future obesity in the depressed was even greater: depressed adolescents were 70% more likely to become obese than non-depressed adolescents, with a risk difference of 4% (Mannan et al., 2016b).

Rooke and Thorsteinsson (2008) found that depression had a small but statistically significant correlation with future weight gain (r=0.08) that was stronger when considering only females (r=0.12). However, they also noted that this effect diminished in size with longer follow-up periods, suggesting that depression may confer a propensity for more short-term weight gain, which may then begin to reverse as the depression lifted.

Blaine (2008) focused specifically on depression as a predictor of obesity, rather than a bi-directional association. In their systematic review and meta-analysis, 16 of 23 papers found statistically significant evidence that depression was associated with future obesity. In addition to this, they conducted a meta-analysis that found that the odds of weight gain in the depressed was 1.47 times the odds of weight gain in the non-depressed. When controlling for initial BMI this odds ratio dropped to 1.23 but was still significant at the 5% level.

2.3.1.2 Cross-sectional Data

Five studies included a meta-analysis of cross-sectional data within their review. All of these studies provided evidence that supported the hypothesised association between obesity and depression. In de Wit et al. (2010), analysis of community based studies found that the odds of depressive symptoms were 1.18 times higher in people with obesity than in those without. This relationship became even stronger when outlier studies were removed from the analysis. Abou Abbas et al. (2015) found a similar association in a review of 8 studies of middle-eastern populations. For this sample there was a significant association between obesity and depression (OR = 1.27).

Author (Year)	Data Reviewed	Cross-sectional Association
Abou Abbas et al. (2015)	8 studies (5 cross-sectional and 3 case-control)	OR = 1.27, (1.11,1.44)
de Wit et al. (2010)	17 cross-sectional studies	Pooled OR = 1.18, (1.01,1.37)
Jung et al. (2017)	Extracted data from 76 studies. For both overweight to depression and obesity to depression, 29 cross-sectional studies were analysed.	Obese: OR = 1.18 (1.11,1.1.26) Overweight: OR = 0.98 (1.01 excluding outlier)
Pereira-Miranda et al. (2017)	9 studies, 8 cross-sectional and 1 cohort	PR = 1.29 (obese vs non-obese); 1.32 (obese vs normal BMI)
Quek et al. (2017)	Pooled data from 18 studies	Obese: OR = 1.34 (1.1-1.64) for depression. SMD = 0.23 (0.025-0.44) for depressive symptoms Overweight: non-significant

TABLE 2.2: Summary of Meta-analysis results for cross-sectional data

The review by Jung et al. (2017) included 29 cross-sectional studies in the analysis of the association of depression with obesity and overweight. In this subset of studies, they found that obesity was associated with depression in both men and women. The pooled odds ratio for the studies was 1.18, and remained significant when stratified by gender. However, in the case of overweight being associated with depression, no statistically significant evidence of an association was found in the pooled estimate. Finally, Quek et al. (2017) meta-analysis of 18 studies found that adolescents and children with obesity were at significantly greater risk of depression, with a pooled odds ratio from the studies of 1.34.

2.3.2 Data Summaries

Five papers provided systematic summaries of data in studies investigating the relationship between obesity and depression (Table 2.3). The first of these examined 24 studies, 4 longitudinal and 20 cross-sectional, to investigate obesity as a cause of depression (Atlantis and Baker, 2008). Overall they

found weak evidence to suggest that obesity was linked to future depression.

The four longitudinal studies presented all provided consistent evidence that obesity was associated with future depression. However, the cross-sectional studies provided much less consistent results. Studies conducted within the United States were much more consistent in finding an association between obesity and depression than those conducted outside the United States. Within the studies from the U.S the association was found consistently only in women, with many studies finding either no association or at times a negative association between obesity and depression in men. The majority of studies conducted outside the U.S found no significant association between obesity and depression, and those that did produced conflicting results. In addition, all but three of the cross-sectional studies included in the review lacked over half of the authors' quality indicators, indicating that their results were susceptible to bias.

The study by Faith et al. (2011) found stronger evidence of an association between obesity and depression. They reviewed 25 longitudinal studies, 10 of which investigated obesity as a predictor of future depression and 15 that investigated the reverse direction. They found consistent evidence that obesity leads to future depression, with 80% of the reviewed studies finding a significant association between obesity and future depression. However, evidence for depression leading to future obesity was less consistent with only 53% of studies finding a significant association.

Author (year)	Data Reviewed	Cross-sectional Association	Obesity/Overweight as Predictor Depression	Depression as Predictor of Obesity/Overweight
Atlantis and Baker (2008)	24 studies (4 longitudinal, 20 cross-sectional)	Supported association in women but not in men.	Results from prospective studies supported hypothesis that obesity increased risk of future depression.	
Faith et al. (2011)	Data from 25 longitudinal studies.		8/10 studies found significant evidence of that obesity was associated with future depression onset or higher levels of depression	8/15 studies found that depression was a significant predictor of future obesity/weight gain.
Liem et al. (2008)	21 cross-sectional studies and 11 longitudinal	16 cross-sectional studies found positive association.		9 longitudinal studies reported significant association.
Korczak et al. (2013)	16 studies, 9 of which included a depression specific measure.		Obese adolescent females at more risk than non-obese peers.	Depressed adolescent females and potentially males more at risk.
Mühlig et al. (2016)	24 'High-quality' studies.	5/8 studies found a significant association.	7/10 cross sectional and 3/8 longitudinal studies found significant effect.	2/2 cross-sectional and 3/9 longitudinal studies found significant effect.

TABLE 2.3: Summary of Results from systematic data summaries

Liem et al. (2008) analysed 21 cross-sectional studies and 11 longitudinal studies to examine the association between adolescent and childhood depression and future overweight. They found consistent evidence that adolescent depressive symptoms increased the risk of future depression. In their review 16 cross-sectional and 9 longitudinal studies reported a significant positive association between depression and future overweight. When the analysis was restricted to only the highest quality studies, similar findings were produced, with results from four longitudinal studies suggesting depressive symptoms in adolescence or childhood were associated with a 1.9 to 3.5-fold increase in the risk of subsequent overweight. High quality cross-sectional studies also suggested that depressive symptoms in girls aged 8-15 was associated with overweight.

The review by Korczak et al. (2013) analysed evidence that suggested adolescent obese females were more at risk of future depression than non-obese peers, and that depressed adolescent females were at risk of future weight gain. Results also suggested the possibility of depression leading to future obesity in males.

Mühlig et al. (2016) examined data from 24 high quality studies, in which they found mixed evidence for a bidirectional association between obesity and depression in adolescence and childhood. Evidence from cross sectional data was stronger than that from longitudinal studies. In the cross-sectional data five out of eight studies found a significant positive association between obesity and depression, most of which looked at associations in early adolescence rather than late adolescence. Seven out of ten studies that examined obesity as a predictor of depression and both of the studies examining depression as a predictor of obesity found a significant positive association. In the case of longitudinal data, only three out of eight found evidence for an influence of BMI on future depression, and three out of nine studies found evidence supporting depression's influence on BMI. BMI's influence on depression was also only found longitudinally in females.

2.3.3 Conceptual Papers

The remaining six papers did not provide a systematic analysis of data on the association of obesity and depression. For the purpose of this review I have labelled these papers 'conceptual papers', due to their focus on discussing mechanisms linking the two conditions and other variables associated with

the relationship. Despite this focus, two papers provided narrative summaries of the research on the general association between obesity and depression (Faith, Matz and Jorge, 2002; Markowitz, Friedman and Arent, 2008). In both papers the summary was broken down into the findings from cross-sectional, longitudinal and intervention studies. Similar to other reviews, both papers concluded that cross-sectional data provided weak but potentially inconclusive evidence of an association between obesity and depression. In terms of longitudinal studies, both papers suggested that data supported both causal pathways between obesity and depression. However, Markowitz, Friedman and Arent (2008) noted that at the time of writing, the studies linking depression to future obesity were less consistent in their findings. The intervention studies presented also supported the notion that obesity and depression were related, in so far as those who received weight loss treatment also often experienced improvements in their mood. However, it was also noted that this relationship had not been shown to be mediated by the actual weight loss itself, and instead other mechanisms were posited to mediate the improvements in mood such as treatment participation and realistic goal setting.

2.3.4 Heterogeneity and Publication Bias

In their early review, Friedman and Brownell (1995) cited heterogeneity within study methodology as a potential source of the inconsistent findings into association between obesity and depression. Within the meta-analyses and systematic data summaries 13 papers investigated whether heterogeneity was present in either the results of studies or in methodologies. Of these reviews, 12 found evidence of heterogeneity when considering either effects or methodology (Atlantis and Baker, 2008; de Wit et al., 2010; Faith et al., 2011; Jung et al., 2017; Korczak et al., 2013; Luppino et al., 2010; Mannan et al., 2016b; Mannan et al., 2016a; Mühlig et al., 2016; Pereira-Miranda et al., 2017; Quek et al., 2017; Rooke and Thorsteinsson, 2008). Of particular note was that in two studies methodological heterogeneity was cited as the reason that a quantitative data analysis was not undertaken (Atlantis and Baker, 2008; Faith et al., 2011).

The impact of publication bias on reported associations was also considered. Six reviews examined evidence of publication bias, two of which provided significant evidence that publication bias may be affecting the reported results (Mannan et al., 2016b; Mannan et al., 2016a). Both of these papers suggested that possible reasons for this bias were clinical and methodological

Author (Year)	Data Reviewed	Findings Presented
(Fabricatore and Wadden, 2004)	Brief review of studies in general and clinical populations.	Obesity not associated with depression in general but certain obese individuals are at risk.
(Faith, Matz and Jorge, 2002)	Summary of findings from cross-sectional, longitudinal and intervention studies.	Longitudinal evidence supports the possibility of both directions of association. Intervention studies suggest that weight loss intervention may also confer secondary benefits to mood. Results from cross-sectional studies do not suggest a strong or simple association.
(Markowitz, Friedman and Arent, 2008)	Summary of findings from cross-sectional, longitudinal and intervention studies.	Evidence from cross-sectional studies suggestive but not conclusive of an association. Longitudinal research more supportive of notion that obesity can raise risk of future depression. Research suggests adolescent depression may confer obesity risk in later life but less so for adult depression. Weight loss intervention involvement often incurs mood benefits alongside.
(Preiss, Brennan and Clarke, 2013)	No review of data on the association in general.	
(Stunkard, Faith and Allison, 2003)	No review of data on the association in general	
(Hoare et al., 2014)	No review of data on the association in general	

TABLE 2.4: Summary of results from conceptual papers

heterogeneity, and also a bias towards publication of results that find positive associations. One review found no significant evidence for the existence of publication bias in studies examining obesity as a predictor of depression, but found a trend towards significance for publication bias in studies examining the opposite temporal association (Luppino et al., 2010). The remaining three reviews found no evidence of publication bias (Abou Abbas et al., 2015; Pereira-Miranda et al., 2017; Quek et al., 2017).

2.3.5 Discussion

Overall the evidence presented by these reviews suggests that there is an association between obesity and depression. In all analyses of longitudinal data, evidence was found to support both obesity as a predictor of depression, and depression as a predictor of obesity. However, in line with the findings of Friedman and Brownell, the evidence concerning the strength and direction of association have been inconsistent, with many studies providing conflicting results.

Similar to the conclusion of Friedman and Brownell, heterogeneity within the studies analysed could account for some of this inconsistency. In the included reviews, evidence of heterogeneity between the studies was consistently found, which could be down to both heterogeneities within the population and methodological variability.

This hypothesis is perhaps supported by evidence suggesting that the obese population is highly heterogeneous (Green et al., 2015). Within this study, the authors found that the obese population fit into 6 distinct clusters which were broadly described as: "heavy drinking males, young healthy females; the affluent and healthy elderly; the physically sick but happy elderly; the unhappy and anxious middle aged, and a cluster with the poorest health." The differences in the characteristics between these groups perhaps makes it unsurprising that the association of obesity and depression differs across studies, as samples may be unlikely to have similar populations with respect to these groupings.

Depression is also a highly heterogeneous condition. One place this is apparent is the inclusion of many different symptoms across different depression assessment scales (Fried, 2017). In addition, studies have also shown that even when using a single form of depression assessment symptom profiles of the depressed can be hugely heterogeneous (Fried and Nesse, 2015), and

it is possible to identify aetiologically distinct subgroups of depressed individuals within the population (Chen et al., 2000). It stands to reason that this heterogeneity might also be responsible for inconsistent findings across studies, as different clusters of depression symptoms may have different relationships with body weight, and other obesity risk factors and outcomes.

The data here reinforce the importance of looking at which obese individuals are at risk of developing depression, and who out of the depressed population is at risk of developing obesity. This question, along with the potential mechanisms that link the two conditions is explored in the next section.

2.4 Risk Factors in the Relationship between Obesity and Depression

Following on from the discussion in the previous section, I will now summarise the research findings surrounding the potential risk factors that may influence the relationship between obesity and depression. A risk factor is defined as "*any attribute, characteristic or exposure of an individual that increases the likelihood of developing a disease or incurring injury*" (Organisation of Economic Co-operation and Development, 2019). When speaking of risk factors here, I am referring to risk factors for comorbid obesity and depression. In this way, risk factors presented here may be characteristics that increase the risk of depression in the obese, characteristics that increase the risk of obesity in the depressed, and characteristics which way raise risk for the development of each condition concurrently.

Risk factors discussed in this section that raise the risk of comorbidity in someone with either of the two conditions will be those that can be considered one the following: proxy risk factors, overlapping risk factors, independent risk factors or moderators (Kraemer et al., 2001). The section discussing potential causal pathways will then be reserved for mediating variables.

2.4.1 Severity of Obesity

A widely posited and investigated moderator of the relationship between obesity and depression is severity of obesity. Four of the conceptual papers presented evidence to suggest that individuals with more severe obesity

were at greater risk of depression (Fabricatore and Wadden, 2004; Markowitz, Friedman and Arent, 2008; Preiss, Brennan and Clarke, 2013; Stunkard, Faith and Allison, 2003).

In the systematic review by Preiss, Brennan and Clarke (2013) eight out of nine studies that investigated severity of obesity as a variable associated with the relationship between obesity and depression found a significant association. Of these, four found that severity of obesity was positively associated with depressive symptoms and four found a positive association with diagnosis of depression. Additionally, Fabricatore and Wadden (2004) suggested, with reference to a study on obesity treatment, that extreme obesity was a risk factor for depression which could be mediated by the increased likelihood of health complications and stigmatization.

Two reviews referred to data from studies conducted in the United States which suggested severe obesity may confer greater risk of depression. Stunkard, Faith and Allison (2003) referred to NHANES data in which high rates of depression were found for both adolescent boys and girls within the most obese 5-percent. Markowitz, Friedman and Arent (2008) referred to data from NHANES-III which suggested that those with severe obesity of BMI over 40 were significantly more likely to have Major Depressive Disorder than those in the 30-34.9 BMI range.

Data from the meta-analyses included here also support the idea that more severe obesity confers a greater risk of comorbid depression. In the cross-sectional evidence analysed by Jung et al. (2017), studies that investigated individuals with severe obesity (BMI greater than 40) had a greater pooled odds ratio than was found for studies that investigated just obesity (BMI greater than 30). This was greater still than the association the authors found between overweight and depression, which was significant in female samples but not in males. This led the authors to suggest that obesity-depression associations followed a dose-response relationship.

Other reviews also found that the strength of the association was greater in higher BMI cut-offs. Pereira-Miranda et al. (2017) calculated prevalence ratios comparing indiviudals with obesity to normal weight, overweight to normal weight and obesity to non-obese. Within these results, the greatest effect size was found in the comparison of individuals with obesity to those with normal weight, followed by those with obesity vs non-obese and finally those with overweight vs normal weight. The grading of these comparisons

is indicative of a dose response relationship, one which potentially becomes stronger as individuals become more obese.

Similar results were found in the meta-analyses of longitudinal data. In the meta-analysis of adult samples conducted by Mannan et al. (2016a), the authors found significant evidence of a bi-directional association with depression for obesity but not for overweight. The analysis by Luppino et al. (2010) found that the effect size for future depression in the obese was also greater than that found in the overweight, when both groups were compared with individuals of normal weight. However, the meta-analysis by Rooke and Thorsteinsson (2008) did not concur with the results regarding the difference in relationship between obese and overweight. Instead they found that there was no moderation effect by the weight cut-off and suggested that both overweight and obesity conferred the same risk of future depression.

Taken as a whole, these reviews have presented consistent evidence to suggest that severity of obesity moderates the relationship between obesity and depression.

2.4.2 Depression Characteristics

From this sample, four reviews provided evidence that depression characteristics could moderate the association between obesity and depression (Faith, Matz and Jorge, 2002; Liem et al., 2008; Quek et al., 2017; Stunkard, Faith and Allison, 2003). The conceptual review by Stunkard, Faith and Allison (2003) suggested that those with more severe depression may be more likely to experience an increase in body weight. They noted that studies on the effect of major depression had found associations with increased future BMI, whilst studies examining the relationship between obesity and subclinical depression had less consistently found an association. The reviews by Faith, Matz and Jorge (2002), Quek et al. (2017) and Liem et al. (2008) proposed that the individuals who exhibit a specific type of depression known as atypical depression (AD), could especially be at risk of developing obesity.

The exact definition of atypical depression has varied over the years, but nowadays, the definition is taken to be that of the fifth edition of the American Psychiatric Associations Diagnostic and Statistics Manual, which specified atypical depression as a depressive disorder with specific clinically relevant features (American Psychological Association, 2013). In particular, the depressive disorder must occur along with the following features: significant

mood reactivity (mood brightness in response to actual or potential positive events) and two or more of the following symptoms: significant weight gain, increase in appetite, hypersomnia, leaden paralysis, and a long-standing pattern of interpersonal rejection sensitivity that results in significant social or occupational impairment (American Psychological Association, 2013; Łojko and Rybakowski, 2017). The disorder must also not meet the specific criteria for the melancholic or catatonic depression subtypes. The articles here make note of the specific weight and appetite related symptoms of this subtype of depression to suggest that those with atypical depression may be at greater risk of future obesity than those with other subtypes.

Overall the evidence provided here is not conclusive that either severity of depression, or whether individuals with particular subtypes of depression are more at risk of developing obesity than others, but it does provide some suggestion that this may be the case.

2.4.3 Sex

One of the most consistently posited moderators of the relationship between obesity and depression is sex. Five out of the six conceptual reviews presented evidence surrounding the effect of sex on the association between obesity and depression. Markowitz, Friedman and Arent (2008), Stunkard, Faith and Allison (2003), Fabricatore and Wadden (2004) and Faith, Matz and Jorge (2002) all presented data from the studies by Istvan, Zavela and Weidner (1992) and Carpenter et al. (2000) to suggest that female sex was a risk factor for comorbid obesity and depression. In addition, Faith, Matz and Jorge (2002) also presented data from the Midtown Manhattan Study (Moore, Stunkard and Srole, 1997), putting forward that female sex may also influence the effect of socioeconomic position on obesity-depression relations.

Preiss, Brennan and Clarke (2013) investigated 20 cross-sectional studies that examined the effect of sex on the relationship between obesity and depression, of which 12 found a significant association. In all 12 of these studies, female sex conferred greater risk of comorbid obesity and depression. There were no systematic differences between studies that did and did not find an association based on assessment of depression or the population origin. In addition to the cross-sectional data, 2 longitudinal studies were analysed that examined sex moderation of obesity-depression relations. Neither of these found a general moderation effect by sex, but one study did find and age-sex interaction, whereby middle-aged women were most at risk.

Sex was also investigated as a potential moderator of the relationship between obesity and depression in the reviews that conducted a systematic review of data on the general association between obesity and depression.

Analyses of cross-sectional data were consistent in finding a gender difference in the association of obesity and depression. Four out of five meta-analyses of cross-sectional data found a statistically significant difference between the association in men and women, such that the association between obesity and depression was greater in women than in men (de Wit et al., 2010; Jung et al., 2017; Pereira-Miranda et al., 2017; Quek et al., 2017). In the study that didn't find a significant difference there was still an observed trend towards this finding (Abou Abbas et al., 2015). Cross-sectional data presented in the systematic data summaries was consistent with this. All three reviews of this data found evidence for a stronger positive association between obesity and depression in females than in males.

Despite the consistent evidence of a sex moderation effect found in the analyses of cross-sectional data, the evidence presented in the analyses of longitudinal data found less evidence of such effects. In the case of obesity as a risk factor for future depression, the systematic reviews by Mühlig et al. (2016) and Korczak et al. (2013) both found evidence that obese females were at higher risk of depression than males prospectively. However, this finding was inconsistent with all meta-analyses of longitudinal data, none of which found significant evidence for a general difference between males and females (Blaine, 2008; Luppino et al., 2010; Mannan et al., 2016b; Rooke and Thorsteinsson, 2008).

Although no overall sex moderation was found, evidence was presented within the meta-analyses of longitudinal data of interactions between sex and other potential moderators. Three reviews presented evidence to suggest that sex interacts with age to produce a moderating effect. Across two reviews, Mannan et al. (2016a) and Mannan et al. (2016b) found that young and middle aged females with obesity were at increased risk for future depression when compared to older females, and adolescents, particularly when exposure was over a longer period of time.

When considering overweight as a predictor of depression Jung et al. (2017) noted that overweight was predictive of future depression in females, but protective of depression in males. In addition to this finding, they found opposite gender differences for underweight as a predictor of depression: males being underweight was predictive of future depression, whereas in

females it was protective. This suggests that the previously hypothesised moderators of gender and severity of obesity interact. This may also support the idea of gender dependent, ideal body sizes, where in men the optimal size places one in the overweight category, and in females the optimal size is one that is underweight, with deviation from this optimal sizes increasing one's likelihood for depression.

When considering depression as a predictor of obesity, gender differences in the relationship were more prominent, but still not significant in general, with only one of the five studies finding an overall significant moderation of the association by gender (Rooke and Thorsteinsson, 2008). However, there was evidence for an age-gender interaction. The meta-analysis by Blaine (2008) found that depressed adolescent girls were at particularly high risk of future obesity in comparison to other groups. Contrasting evidence was found in the reviews by Mannan et al. (2016a) and Mannan et al. (2016b), which suggested that depressed young and middle aged adult females were at elevated risk for developing obesity. The conflicting results presented in these reviews suggest that conclusions about the age-gender interaction in this direction of causality are less certain.

Some reviewers have suggested that obese females are more at risk of depression than obese males (Fabricatore and Wadden, 2004; Friedman and Brownell, 1995; Stunkard, Faith and Allison, 2003). However, this hypothesis was made based on data from only a few studies. When considering a wider pool of data such as that presented by Preiss, Brennan and Clarke (2013), and that of the meta-analyses and data-summaries presented in this review, it would appear this gender difference is not so clear cut. In particular, sex differences were only found in a small minority of analyses of longitudinal data and in some but not all analyses of cross-sectional data. Despite this, evidence does suggest that females may be at greater risk of comorbid obesity and depression, but perhaps only under certain conditions.

Evidence presented here suggests that sex may interact with other moderating variables such as age, socioeconomic position and severity of obesity. As such, detection of sex moderation in cross-sectional studies may be influenced by these other moderating variables. It might also be the case that women are more vulnerable to some, but not all of the mechanisms that related obesity and depression. For example, women may be more vulnerable to stigma and body image concerns than men, whereas functional limitations may affect the mental health of both sexes equally. As such, a particular

gender difference observed in a study could depend on the particular mechanisms responsible for the comorbidities observed in that population.

2.4.4 Socioeconomic Position

Overall the evidence presented in these reviews regarding the effect of socioeconomic position on the relationship between obesity and depression has been mixed. Out of the six conceptual papers, five suggested that socioeconomic position was associated with the relationship between obesity and depression (Faith, Matz and Jorge, 2002; Hoare et al., 2014; Markowitz, Friedman and Arent, 2008; Preiss, Brennan and Clarke, 2013; Stunkard, Faith and Allison, 2003) with the remaining article making no mention of socioeconomic position as a moderator of the relationship between obesity and depression.

The evidence presented about the nature of this association was inconsistent. Due to the established negative associations between socioeconomic position and depression and obesity separately, the obvious hypothesis is that comorbidity would also follow the same pattern, with the most deprived experiencing the highest rates of comorbid obesity and depression (Everson et al., 2002). This was suggested in the reviews by Hoare et al. (2014) and Markowitz, Friedman and Arent (2008), but was not backed up by any data to support the hypothesis.

Despite the logical deductions this hypothesis is based upon, three out of six reviews suggested that it is in fact high socioeconomic position puts obese individuals at higher risk of depression, particularly when socioeconomic position was measured by education status. Markowitz, Friedman and Arent (2008) cited two studies, one in treatment-seeking African-American women, and another in a nationwide sample in the U.S that both found having a greater level of education increased the likelihood of comorbid obesity and depression. Similarly, both Stunkard, Faith and Allison (2003) and Faith, Matz and Jorge (2002) used results from the Midtown Manhattan study (Moore, Stunkard and Srole, 1997) to suggest that high SES was a risk factor for depression in women with obesity but that there was no relationship with SES in men.

The most comprehensive consideration of the effect of socioeconomic position provided in these reviews was given by Preiss, Brennan and Clarke (2013). In their review they included seven studies that investigated the effect

of education status on the relationship between obesity and depression, and three that investigated the effect of income status. The results for the effect of education were more ambiguous than suggested by early reviews, with only four of the seven studies finding a significant association. Within these, the direction of association was also inconsistent, with one study finding that having only high school education put individuals at more risk than any other level of education, and another finding that the association between education and the obesity-depression was only observed in some samples.

The studies investigating the effect of income on the relationship between obesity and depression were more consistent. Two out of the three studies included found a significant association between both income and the obesity-depression relationship. In particular, both lower income and financial problems were associated with higher prevalence of comorbid obesity and depressive symptoms. The one study that found no significant association used major depressive disorder as its measure of depression, suggesting that low income may be associated with increased depressive symptoms, but not with the relationship between obesity and major depressive disorder.

Outside of the conceptual reviews, the effects of socioeconomic factors were only considered or examined in three papers, all of which either presented similar evidence and conclusions to those found in the conceptual reviews or null findings regarding an association. de Wit et al. (2010) also suggested that individuals of high SES are more at risk of comorbid obesity and depression, although these conclusions were based on the same papers presented in the reviews by Stunkard, Faith and Allison (2003)and Markowitz, Friedman and Arent (2008). Liem et al. (2008) noted that none of the included studies in their review that controlled for socioeconomic position found a significant effect. Faith et al. (2011) made reference to only a single paper to suggest that education may moderate the association between obesity and depression.

Overall these results suggest that there may be an association between socioeconomic position and comorbid obesity and depression, but that the association may depend on which indicators of SEP the relationship is being measured with respect to. Further research is therefore needed to unpick the relationship between the wider determinants of health and comorbid obesity and depression outcomes. It remains unclear whether socioeconomic position interacts with other proposed moderators such as gender, and there is also the possibility that the effects of particular causal mechanisms differ with respect to socioeconomic factors.

2.4.5 Ethnicity and Geographical Region of Study

Five reviews directly examined the role of ethnicity (Faith et al., 2011; Faith, Matz and Jorge, 2002; Hoare et al., 2014; Mühlig et al., 2016; Preiss, Brennan and Clarke, 2013), and six looked at the effect of country or region of study (Atlantis and Baker, 2008; de Wit et al., 2010; Jung et al., 2017; Mannan et al., 2016b; Mannan et al., 2016a; Quek et al., 2017) on the obesity-depression relationship.

For ethnicity, the evidence presented in studies was not sufficient to conclude whether general ethnic differences in obesity-depression relations existed. However, there was some evidence provided to suggest that ethnic differences may occur in the effects of some specific mediators or moderators of the relationship between obesity and depression.

Preiss, Brennan and Clarke (2013) included eight U.S based studies in their review which investigated the association of ethnicity of obesity-depression relations. Of these studies, only two found that ethnicity had a significant effect. From this they concluded that there was not consistent evidence supporting ethnic variation in obesity-depression associations. However, they did find early evidence that certain ethnic identities may be associated with the relationship between obesity and depression within certain populations, with one study that examined the association of ethnic identity with obesity-depression relations finding a significant effect.

Hoare et al. (2014) suggested that there was some evidence to suggest that there may be ethnic differences in the relationship between obesogenic risk factors and depression, but that these differences may themselves be moderated by socioeconomic differences that are commonly observed between ethnic groups. Mühlig et al. (2016) put forward the idea that there may be ethnic differences in the mediators and moderators or the relationship between obesity and depression, and that the effect of ethnic background in BMI may also play a role. However, due to the small number of studies in different regions, they were unable to conclude whether ethnic variation was responsible for any differences between study results in their sample. It is also worth noting that this suggestion was based on studies conducted in different regions, rather than a direct observation of different ethnic groups, mean the support for ethnic differences overall is weak.

Faith, Matz and Jorge (2002) also presented evidence from two studies that

found conflicting evidence for a difference in the association of BMI and depression between Caucasian and African American populations. In Faith et al. (2011) only one of the five included studies that conducted a formal analysis for moderation by ethnicity found a significant result, leading them to suggest that further investigation was needed to understand any effect that ethnicity has on this association.

Similar to ethnicity, the reviews did not provide consistent evidence that obesity-depression associations varied depending on the country or region in which the study was population was sampled. Atlantis and Baker (2008) found that, in the studies they analysed, only those conducted in populations in the United States consistently found an association between obesity and depression. However, this difference was not found in the more recent meta-analyses by Mannan et al. (2016b) and Mannan et al. (2016a) who compared the U.S with the rest of the world but found no significant difference between these subgroups in either adults or adolescents.

The subgroup analysis based on country of study conducted by de Wit et al. (2010) did not find a significant difference between the associations found in North America, Europe, and elsewhere in the world. Similarly, Jung et al. (2017) performed a subgroup analysis based on the continent of study of the samples in their included studies. There were no significant differences between continents for obese or overweight outcomes but the authors did find a significant difference in underweight outcomes in cross-sectional studies at the 5% significance level. Finally, Quek et al. (2017) found that non-western obese adolescents and children were at greater risk of depression and severe depressive symptoms than those from western populations.

Overall, these results do not provide consistent evidence that ethnicity or geographical region of study are systematically associated with the relationship between obesity and depression. However, ethnicity may play a role in the association between obesity and depression through a differential effect on the other moderators and mediators of the relationship. For example, ethnic identities may overlap with body size ideals, leaving certain individuals more prone to stigmatisation and discrimination, and with it lower self-esteem and depression. Again though, it is important to highlight that this suggestion was largely speculative when made in the reviews, and so firm conclusions can not be made.

2.4.6 Adverse Childhood Experiences

Although it has been considered in only two reviews, there is evidence to suggest that adverse childhood experiences could moderate the relationship between obesity and depression. Stunkard, Faith and Allison (2003) suggested that although adverse experiences had not been studied in direct relation to comorbid obesity and depression, adverse childhood experiences may promote the occurrence of both conditions, which may lead to greater levels of comorbid obesity and depression than is found in those who do not experience such adverse events. Preiss, Brennan and Clarke (2013) also included one study which found that childhood abuse increased the risk of comorbid obesity and depression in adulthood. However, since this was the only study included in their review they did not deem it sufficient evidence to draw conclusions.

2.4.7 Gene-by-environment Interactions and Genetic Predispositions

The effect of genetic predispositions on comorbid obesity and depression are hypothesised but not well evidenced in the reviews. Faith, Matz and Jorge (2002) and Stunkard, Faith and Allison (2003) presented similar evidence from studies of the ObD7S 1875 gene, which found that variations in the gene's expression were associated with varying levels of depression. Due to this gene's proximity to gene loci previously found as relevant to weight regulation, it was hypothesised that variation at this gene could be relevant to both weight management and depression. Additionally, both authors proposed that a genetic clustering of Syndrome X, which expresses symptoms such as obesity, insulin resistance and obesity related conditions, may also give credibility to genetic predispositions of other obesity related conditions such as depression (Faith, Matz and Jorge, 2002; Stunkard, Faith and Allison, 2003). However, the evidence presented here is still largely speculative, and as such is not conclusive that any genetic effects are present.

2.4.8 Other Risk Factors

Alcohol consumption, smoking status, marital status, family depression, were all examined by Preiss, Brennan and Clarke (2013) to see whether these variables were associated with the relationship between obesity and depression. For marital status, all included studies produced a null result, which was

considered sufficient evidence to dismiss marital status as potentially associated with obesity-depression relations. In the cases of alcohol and smoking, evidence was not consistent enough to conclude on whether there is an effect from smoking status or alcohol consumption on the relationship between obesity and depression. The one study that looked at the influence of family depression on obesity-depression relations found a significant result, but this was deemed insufficient to draw robust conclusions.

2.4.9 Summary

In this section I have outlined a range of non-mediating risk factors which may influence the association between obesity and depression. Consistent findings about the effects of these risk factors was lacking. Only in the case of severity of obesity was there consistent and substantial evidence to suggest that more severely obese individuals are at higher risk of developing depression. For other risk factors such as socioeconomic position, sex and ethnicity, findings presented were largely inconsistent. However, there was evidence for interactions between these risk factors, which may account for some of the inconsistencies found in different studies. The effect of these risk factors may also depend on the existence of the underlying mechanisms that are explored in the next section.

2.5 Potential Causal Pathways between Obesity and Depression

In addition to searching for which individuals in the population are more at risk for comorbid obesity and depression, research has also investigated the mechanisms by which people with obesity may become depressed and vice versa. Six reviews focused on a discussion of the findings from epidemiological studies, using them to explore possible mediating pathways between obesity and depression (Fabricatore and Wadden, 2004; Faith, Matz and Jorge, 2002; Hoare et al., 2014; Markowitz, Friedman and Arent, 2008; Preiss, Brennan and Clarke, 2013; Stunkard, Faith and Allison, 2003). In this section I will explore the arguments covered and conclusions reached surrounding the various mechanisms presented in these reviews. Where pertinent I will also include evidence presented in the 'non-conceptual' reviews. Whilst the mechanisms here are described separately, it is important to note

that they are not entirely disjoint, as many have overlapping features or the potential to interact with one another.

2.5.1 Pathways from Obesity to Depression

2.5.1.1 Stigma, Body-image Dissatisfaction and Self-esteem

One of the most commonly hypothesised and well supported mediating links from obesity to depression is through stigma and body image dissatisfaction. All 6 of the conceptual reviews provided evidence that the stigma in the form of weight based teasing and discrimination mediated the link between obesity and depression.

As part of a bidirectional model Markowitz, Friedman and Arent (2008) presented evidence for a route from obesity to depression through stigma and body image dissatisfaction under what they called the 'health concern' pathway. Within this, they summarised evidence that showed constant stigma can directly affect negative mood, as well as having a negative impact on self-concept. The authors also included studies of obese treatment-seeking populations that found higher degrees of obesity were associated with increased bodily dissatisfaction, which in turn was associated with increased depression via a lowering of self-esteem. Experiences of discrimination were found to increase bodily dissatisfaction, although bodily dissatisfaction was also suggested to have an effect independently of previous weight based teasing.

Preiss, Brennan and Clarke (2013) presented two studies that both indicated stigma was significantly associated with the obesity-depression relationship. In samples of treatment seeking obese, these studies found that increased stigma was positively associated with increased depressive symptoms, but not with depressive disorder, with one study also finding that stigma explained most of the participants' depressed mood. They also presented five studies examining the effect of body image and weight and shape concerns on the relationship between obesity and depression. All five of these studies found significant associations such that individuals who were more concerned with their weight or body shape, or were less satisfied with their bodies, were more likely to experience comorbid obesity and depression.

In addition to the evidence presented on stigma, two studies in the review by Preiss, Brennan and Clarke (2013) investigated the effect of interpersonal relations on the obesity-depression relationship. The first of these studies reported that distressing interpersonal interactions had a significant effect on

obesity-depression relations and the second found that decreased social interaction increased the likelihood of meeting criteria for depressive disorder.

Fabricatore and Wadden (2004) suggested that the widespread stigma and discrimination that is experienced by individuals with obesity from a young age is a chronic stressor that could lead to a deterioration in the mental health of obese subjects. They also, note however, that at the time of writing there had been little empirical investigation of this theory. Stunkard, Faith and Allison (2003) and Faith, Matz and Jorge (2002) presented results from three studies that suggested weight based teasing led to increased dissatisfaction with bodily appearance in the obese which in turn led to depression. Hoare et al. (2014) suggested that greater levels of body image dissatisfaction in women leaves them more susceptible to weight-related mental health issues. However, this was not backed up by any references to evidence.

Discussion presented outside of the conceptual reviews also pointed to the possibility of stigma and body image concerns providing a causal pathway between obesity and depression. In all but 2 reviews, stigma and body image concerns were outlined as potential causal pathways relating obesity to depression.

Overall, the studies included in this review have presented consistent evidence to suggest that stigmatization in the obese, along with body image concerns may facilitate the onset of depression.

2.5.1.2 Diet

The relationship between diet, obesity and depression is complex, with diet and dietary practices potentially mediating relationships from both obesity to depression, and depression to obesity. In this part I will focus on the evidence presented in the reviews for dietary focused mechanisms from obesity to depression. Mechanisms dealing with the opposite pathway will be considered in a later section.

Alongside the widespread effects of stigma and body image concerns, it is also suggested in the review by Markowitz, Friedman and Arent (2008) that repeated dieting may provide a pathway from obesity through to depression. They note that while some studies have suggested there is not a relationship between repeated dieting, weight cycling and depression, more recent studies have provided contrary evidence. They suggest that this mechanism may work in two ways. Firstly, repeated unsuccessful attempts to diet

may lead obese individuals to begin a negative internal dialogue with themselves, labelling themselves as a failure and thus contributing to feelings of low self-esteem. Pereira-Miranda et al. (2017) also used a number of studies to suggest that initial attempts to diet may also be encouraged by experiences of stigmatization, and that failed dieting attempts may also lead on to the occurrence of binge-eating episodes, which have been extensively connected to depression by evidence presented in these reviews.

Secondly, those obese individuals who use food as an emotional regulator may also find the experience of dieting difficult and stressful, particularly those who have an eating disorder such as binge-eating disorder. They also provide evidence that the experience of dieting can have a negative effect on mood, in particular through increased irritability.

The review by Preiss, Brennan and Clarke (2013) included four studies that investigated the effect of a range of eating behaviours on obesity-depression relationships, including one on dieting, two on dietary restraint and one that covered a range of different eating behaviours. Significant associations were found in all but one of the studies investigating the effect of dietary restraint, which the authors noted was carried out in a non-representative population. Within these studies it was found that dieting was associated with the obesity-depression relationship at all weight levels. Furthermore, in obese treatment-seeking individuals, those who had the highest levels of depression had significantly higher levels of dietary restraint, and among women in this population, purging was found to be associated with greater levels of depressive symptoms.

In their review of the relationship between obesity risk factors and depression, Hoare et al. (2014) examined three studies that examined the cross-sectional association between measures of dietary quality and depression. All three of these found associations between poorer diet quality and increased depression which the authors suggested could potentially be due to the effects of nutrient deficiency and a reduced immune system, both of which can underpin mental illness. However, since no studies in the review examined the link between diet and depression longitudinally, these results are not sufficient to conclude on temporal precedence.

Overall there is some evidence to suggest that pathways from obesity to depression via diet and dietary practices exist. Although some of the evidence presented in these reviews doesn't allow a precise direction of causality to be concluded, the evidence here does give some suggestion that the eating

practices of obese individuals, either by becoming disordered or restrictive, could facilitate the onset of depression in those individuals.

2.5.1.3 Disordered and Binge Eating

The role that disordered eating and particularly binge-eating play in the relationship between obesity and depression is complex. Overall, it is unclear how binge eating disorder interacts with obesity and depression to effect the association between the two conditions. In particular, evidence presented in this review suggests that binge-eating and other disordered eating may mediate the association between obesity and depression bi-directionally. Here I will outline the cross-sectional evidence provided in these reviews before exploring how the authors have suggested that disordered eating may raise the risk of depression in the obese.

Four out of the six conceptual reviews suggested that binge-eating behaviours and binge-eating disorder may be cross-sectionally associated with obesity and depression. Preiss, Brennan and Clarke (2013) included eight studies that investigated the effect of binge-eating on the relationship between obesity and depression, of which five found that it was significantly associated. Associations were found mostly in studies of non-surgical treatment seeking obese individuals and those three that did not find any such association all examined surgical treatment seeking populations. As a result, the authors concluded that these findings may not be applicable to general populations.

Fabricatore and Wadden (2004) presented data that shows the prevalence of binge-eating disorder is much higher in people with obesity than in general population based samples. Additionally, they note that binge-eaters report higher depression rates than in the rest of population. Similarly, Faith, Matz and Jorge (2002) noted that binge-eating disorder is one of the most reliable discriminators of psychological outcomes in people with obesity to suggest that the condition may be a risk factor for depression in the obese. In particular, they make reference to a study of individuals with obesity in which the prevalence of major depression in people with obesity that suffer from with binge-eating disorder was three times higher than the prevalence in those without binge-eating disorder. In the same study, dysthymia prevalence was two times higher in the binge-eating disorder group.

The evidence presented here does not allow conclusions to be drawn on the temporal associations between binge-eating and depression and obesity.

However, from this evidence it is possible to hypothesise a variety of ways in which binge-eating may interact with obesity and depression. Obesity may serve as a proxy risk factor for the association of binge-eating disorder and depression. As such, the development of binge-eating and other dysregulated eating may temporally coincide with the development of obesity, followed by an effect on the mental health of individuals due to the disordered eating.

Evidence presented in this review also allows for the possibility of obesity temporally preceding disordered eating before leading to depression. Stunkard, Faith and Allison (2003) provided evidence that disordered eating may increase the risk for depression in the obese. They note that the symptoms of binge-eating disorder such as impaired control over eating, and distress about binges are strongly associated with depression. Similarly, those suffering from night eating syndrome may also suffer consequences in their mental health due to their disturbed eating practices. Markowitz, Friedman and Arent (2008) provide evidence from two studies of overweight and obese children that found self-esteem was lower in those with BED than those without. They also provided evidence from 3 other studies across which it was found binge-eating predicted future depression, as well as higher depressive and anxiety symptoms, lower self-esteem and bodily self-esteem.

Discussions within the non-conceptual reviews also covered the potential for binge-eating to mediate the link from obesity through to depression. With reference to studies on the association between obesity, binge-eating and depression, Atlantis and Baker (2008), Pereira-Miranda et al. (2017) and Luppino et al. (2010) all put forward that disturbed eating patterns and eating disorders developed as a consequence of obesity could increase one's risk for developing depression.

From the above, it stands to reason that if the dietary practices of an obese individual develop into a more disordered structure, that individual could be at higher risk for depression.

2.5.1.4 Physical Activity

Longitudinal data presented in the review by Hoare et al. (2014) suggests that physical activity could be a pathway in which obesity leads to the onset of depression. In their review they included six studies that investigated the association between physical activity and depression longitudinally, and

six which examined this association cross-sectionally. All six of the longitudinal studies found evidence that physical activity was associated with future levels of depression, although the studies did not all use the same measures of physical activity. Overall, these studies found that increased physical activity was associated with reduced depressive symptoms and a lower risk of future depression.

Similar evidence was found in the cross-sectional studies. In addition, three of these studies examined the association of sedentary activity with depression in adolescents, all of which found that higher amounts of screen use, through things such as television and social networking was associated with more depressed mood and depressive symptoms. In discussing these results, Hoare et al. (2014) also presented results from three studies to suggest the mechanisms behind these observed longitudinal associations. These studies suggested that physical activity may have a protective effect against depression through improvements in self-efficacy, resilience and discipline.

The reviews by Stunkard, Faith and Allison (2003) and Markowitz, Friedman and Arent (2008) also offered evidence that higher rates of activity may help in the treatment of depression. Stunkard, Faith and Allison (2003) made reference to a single study which found that physical activity had been a somewhat successful means of treating depression. This was also suggested in Markowitz, Friedman and Arent (2008)), who made reference to six meta-analyses that consistently found moderate to large effects for the reduction of depressive symptoms with exercise. They also presented evidence that has suggested exercise can be as effective as traditional treatments of depression such as pharmacological treatments, and may exert this effect partly through reducing inflammation and also through an effect on the HPA axis and autonomic nervous system.

2.5.1.5 Physical Health Concerns

Another potential mechanism linking obesity though to depression is through its effect on physical health and physical function. Four out of the six conceptual reviews posited that physical health effects could mediate the link from obesity to depression.

Markowitz, Friedman and Arent (2008) suggest that physical health in people with obesity affects their risk of depression in two ways: through functional impairment and through self-rated health. The authors highlight studies that

have shown that people with obesity face greater levels of functional impairment through increased physical disability, reduced ability to exercise and reduced capacity to carry out their activities of daily living. These examples of reduced function are themselves associated with increased rates of depression and depressive symptoms. People with obesity are also at higher risk of developing a variety of chronic diseases. Studies have suggested that complications associated with these chronic diseases may also negatively impact physical functioning and in this way have a depressive effect on individuals with obesity.

In terms of self-rated health, Markowitz, Friedman and Arent (2008) suggested that the poor self-rated health associated with obesity leads individuals with obesity who believe their health to be poor to also have other depressive beliefs about themselves via various cognitive processes. In particular, they may feel unable to adequately exercise or eat a nutritious diet, and sentiments from popular media surrounding the obesity crisis may reinforce these negative beliefs that they hold about themselves.

Preiss, Brennan and Clarke (2013) examined seven studies that examined the association between various measures of physical health, including pain, fatigue, physical functioning, somatic symptoms and general physical health, and the relationship between obesity and depression. All seven of these studies found that worse physical health was associated with greater levels of comorbid obesity and depression, which the authors concluded provided consistent evidence that physical health was associated with obesity-depression relations.

The remaining two conceptual reviews only provided sparse evidence. Introducing their study of risk factors, Hoare et al. (2014) cited one study that showed mobility issues and medical problems led to emotional problems in obese children. When highlighting that people with the most severe obesity are at the greater risk for depression than are those with moderate obesity, Fabricatore and Wadden (2004) suggested that one potential explanation for this was the increase in medical complications and impairments in health related quality of life.

Overall there is consistent evidence to suggest that poor physical health, and concern over physical health experienced by people with obesity could facilitate the onset of depression. However, little of the evidence provided came from longitudinal studies, and so again it is not possible to be certain about the direction of association. In addition, many of the studies included in

the reviews focused on more general mental well-being measures which included depression and as such it is possible that physical health concerns have an impact on other aspects of mental well-being more so than depression.

2.5.2 Pathways from Depression to Obesity

2.5.2.1 Stress, HPA Axis Dysregulation and Inflammation

Three conceptual reviews provided evidence of shared biological pathways between obesity and depression. The role these biological pathways play, however, is not entirely straightforward, as evidence suggests that these pathways may be bidirectional between the two conditions.

Evidence presented surrounding the biological links between obesity and depression has largely followed the same pattern across the reviews. In general, authors have referred to associations found separately between obesity, depression and a variety of biological disturbances to suggest that these common disturbances may increase one's risk of developing a comorbidity, following the onset of one of them (Stunkard, Faith and Allison, 2003; Faith, Matz and Jorge, 2002; Markowitz, Friedman and Arent, 2008). In terms of inflammation, authors have noted that both depression and obesity are associated with inflammatory states in the body, and used this to hypothesise that inflammation may mediate the association between two conditions (Markowitz, Friedman and Arent, 2008; Pereira-Miranda et al., 2017; Luppino et al., 2010).

Stunkard, Faith and Allison (2003) suggested that physiological stress responses in depressed individuals may put them at increased risk of weight gain. They explain that HPA axis activation is associated with stress, as well as increased levels of cortisol. Elevated levels of cortisol are then believed to give rise to greater depositing of fat around the abdomen. They cite three studies that have found significant associations between HPA axis activation and measures of obesity, particularly abdominal fat.

Markowitz, Friedman and Arent (2008) provide similar evidence that links both HPA axis activation and elevated cortisol to increased body weight and visceral fat. They combine conclusions from five studies to suggest that elevated reactivity to stress manifests itself through dysregulation of the HPA axis and elevated cortisol, which in turn promotes weight gain of a chiefly visceral nature. They also cite an additional study which found that

stress can influence eating behaviours and suggests that these effects may be worsened in depressed individuals due to the already heightened activity of the stress system. Faith, Matz and Jorge (2002) used the same studies as Stunkard, Faith and Allison (2003) to suggest the role of HPA axis activation in linking depression to obesity. However, they also provided evidence that HPA axis activation and cortisol levels did not always link obesity to psychopathology. In particular, they suggested that changes to cortisol levels were not different between people with obesity and BED and those who with obesity but no binge-eating disorder, within a population undergoing a weight-loss intervention.

To support claims of an association with inflammation, Markowitz, Friedman and Arent (2008) present studies that have shown associations between depression and inflammatory cytokines. In particular, they present evidence from three articles which suggested that depression is associated with increased inflammation, which is further exacerbated in the presence of obesity. In addition, the authors present the conclusions of one study which found that administration of cytokines can have a beneficial effect on both mood and eating behaviours.

Discussion presented outside the conceptual reviews provides similar evidence supporting the route from depression to obesity via HPA axis dysregulation and cortisol levels. In addition, some authors have also posited that the same mechanisms could raise the risk of depression in the obese. Pereira-Miranda et al. (2017) combined conclusions from four studies to suggest HPA dysregulation and inflammatory states as a consequence of obesity could lead to depression via degradation of neurons and neuronal structure. Similarly, Luppino et al. (2010) combines the results from four articles that together suggest obesity may influence HPA axis activation, which is in turn associated with depression. The authors also cite four articles showing that obesity and depression are both associated with increased inflammation, suggesting that this may mediate the association.

Alongside the suggestion of links between obesity and depression via HPA axis dysregulation and inflammation, a few studies also presented evidence to suggest other shared pathologies of the two conditions may exist. Mühlig et al. (2016) lists three additional areas in which studies have found pathological links between obesity and depression. These were in Vitamin D deficiency, dopamine responses, and neuroendocrine links via the hormone

Leptin. Quek et al. (2017) also provided evidence that Leptin may be involved in the association, and additionally they cite two articles that show that levels of Neuropeptide Y are associated with both stress and appetite, which taken with depression's known association with stress provides potential for this being a mediating link from depression to obesity.

Overall, the reviews here have presented a wide range of evidence to suggest that shared pathologies between obesity and depression may mediate the link between the two conditions. In addition, it would appear that this mechanism may operate bi-directionally, such that shared pathologies are able to raise the risk of depression in the obese, and raise the risk of obesity in the depressed. However, evidence showing temporal associations all the way through from depression to obesity, or vice versa, via pathology was not presented in these reviews. As such, from this it is not possible to conclude with certainty that one condition causes the other through shared pathology.

2.5.2.2 Physical Activity Levels

Low levels of physical activity are a major risk factor for the development of obesity, and as such it seems reasonable to hypothesise that the effect of depression on the development of obesity might occur, in part, through an effect on physical activity levels. This hypothesis is hinted at in the reviews by Stunkard, Faith and Allison (2003) and Faith, Matz and Jorge (2002) who noted that reduced physical activity commonly occurs in depressed people. However, little evidence of a direct association between depression and future physical activity was presented in these reviews.

Two of the six conceptual reviews included summaries of cross-sectional evidence on the association between obesity, depression and physical activity. Preiss, Brennan and Clarke (2013) examined five studies that cross-sectionally investigated the effect of physical activity on obesity-depression relations, three of which found significant effects. Two of these found that major depressive disorder led to lower levels of physical activity, with the third showing that exercise was associated with obesity-depression relations cross-sectionally.

In the paper by Hoare et al. (2014), six studies were included that cross-sectionally examined the association between physical activity measures and depression. As was noted in the previous section, these provided consistent evidence that depression was associated with lower levels of physical activity and greater levels of screen based sedentary activity. However, none of

the included longitudinal studies presented evidence on the effect of depression on levels of physical activity, only the reverse temporal association was considered. Therefore, whilst this evidence certainly supports the possibility that physical activity provides a mediating pathway from obesity to depression, it does not provide sufficient evidence to conclude as much.

Overall, these reviews provide some evidence that levels of physical activity may mediate the link between obesity and depression. However, all of the studies presented analysed cross-sectional evidence; evidence supporting a longitudinal association between depression and future levels of physical activity is lacking and as such it is not possible to conclude with certainty that depression effects future levels of physical activity.

2.5.2.3 Diet

Poor diet is another of the main risk factors for obesity, and so similar to physical activity it is logical to expect that if depression does cause future obesity, this could be through an effect on diet and dietary practices. The most commonly hypothesised way in which depression may affect diet is through the occurrence of binge-eating, and other disordered eating. Whilst this has been discussed in a previous section, the focus there was on disordered eating mediating the association between obesity and future depression. In this section we will focus on the evidence that supports mediation in the reverse temporal association.

Markowitz, Friedman and Arent (2008) present evidence from three studies that suggest instances of emotional distress and low mood can lead to episodes of binge-eating, particularly in individuals with binge-eating disorder. From this it is hypothesised that the use of food as an emotional regulator would be highly likely to result in subsequent overweight in the depressed, particularly in those whose depression is chronic. Stunkard, Faith and Allison (2003) suggest that poor diet may confer obesity risk in the depressed based on observations of symptoms in DSM-IV criteria for depression. They note that over-eating (and under-eating) forms part of this criteria and hence that depressive symptoms may confer direct risk of obesity. With respect to binge-eating disorder, the authors also presented findings from a study which found that the percentage of obese binge-eaters with a history of depression was 54% in comparison to just 14% in non-binge eating obese (Yanovski, 1993). As was noted in the previous section, Hoare et al. (2014)

examined three studies that found an association between poor diet and depression. As an alternative to their suggestion of poor diet affecting mood, they also hypothesised that this could be due to the symptoms of depression such as reduced motivation leading to unhealthy diet choices.

Three non-conceptual reviews also provided evidence that depression could lead to obesity via an effect on diet. Mannan et al. (2016a) referenced two studies that suggest the association between stress and common mental health disorders could lead to changes in adiposity via binge-eating behaviour. In their meta-analysis of adolescent samples they also suggest that depression may affect adolescent dietary practices such as through binge-eating, and favouring carbohydrate rich food (Mannan et al., 2016a). Liem et al. (2008) refer to data from two longitudinal studies that found depression was a risk factor for the onset of binge-eating behaviours. From this they hypothesise that this could be due to the use of binge-eating to relieve depressive symptoms.

In summary, there is consistent evidence to suggest that binge-eating may mediate the association between depression and obesity. Cross-sectional evidence also supports the possibility that depression may be generally associated with poorer diet quality. However, longitudinal evidence of this association was not presented in these reviews, making it difficult to conclude whether dietary practices and habits other than binge-eating mediate the association between depression and future obesity.

2.5.2.4 The Effect of Depression on Diet and Physical Activity

Despite the lack of clear evidence presented surrounding an effect of depression on physical activity and diet, reviews have hypothesised how aspects of depression could have an impact on diet and physical activity.

Focused around weight loss maintenance, Markowitz, Friedman and Arent (2008) present three ways in which depression may influence an individual's following of a healthy diet and exercise routine: 'Adherence', 'Negative Thoughts', and 'Social Support'.

In support of their adherence hypothesis, they provide evidence from three articles that show depression predicts attrition from weight loss programs, and can also worsen the actual and perceived effects of physical health symptoms individuals experience. From this they suggest that depressed individuals may find it harder to adhere to healthy diet and exercise regimes. Two other reviewers also hypothesise that depression may contribute to obesity

through its effect on adherence to healthy routines. Hoare et al. (2014) cites one study which suggests depressive symptoms such as low motivation may lead to individuals making unhealthy food choices. Stunkard, Faith and Allison (2003) hypothesise that stress may disrupt healthy diet and exercise habits, although they present no evidence to directly support this.

The arguments Markowitz, Friedman and Arent (2008) present in support of the effect of 'negative thoughts' are based around self-efficacy theory, which hypothesises that beliefs that an individual hold about their abilities to achieve certain outcomes can directly affect those outcomes (Bandura, 1977). They presented results from three articles that found depression in obese women was associated with lower weight-loss self-efficacy, suggesting that, in line with the theory, this could lead to reduced weight-loss (Markowitz, Friedman and Arent, 2008). Additionally, they presented evidence from one article suggesting that predicted weight loss was positively associated with actual weight loss in a sample of women, and posited that pessimistic beliefs held by depressed individuals may therefore negatively affect weight loss-outcomes.

The review by Preiss, Brennan and Clarke (2013) also included one study which showed that perceived physical activity self-efficacy was associated with the relationship between obesity and depression. However, the cross-sectional nature of this study, and the lack of specific detail given by the reviewers prevents us from concluding about the temporal nature of the observed effects.

Markowitz, Friedman and Arent (2008) suggest that social support may affect weight-loss attempts by worsening adherence to healthy lifestyle, and also by reducing the likelihood of weight loss maintenance. To support these claims, they reference four articles which together found that both weight-loss success and weight-loss maintenance outcomes improved when individuals had support from others. Additionally, they cite two studies that found depression can cause strain and erosion of family support, and that this may therefore increase the likelihood that depressed individuals achieve poorer weight-loss and maintenance.

Preiss, Brennan and Clarke (2013) included two studies that investigated the association of interpersonal relationships with obesity-depression relations, both of which found significant evidence of an effect. The first study found that distressing interpersonal relations were significantly associated with the relationship between obesity and depression, and the second found

that in people with obesity decreased social activity increased the likelihood of meeting criteria for depressive disorder. The results suggest that mechanisms involving social withdrawal may operate bi-directionally, and potentially in conjunction with stigmatization.

2.5.2.5 Psychological Characteristics

Preiss, Brennan and Clarke (2013) examined four studies that considered the effects of a range of psychological characteristics on the obesity-depression relationship. All four found significant associations between psychological characteristics and obesity-depression relations in treatment-seeking obese and overweight populations. Significant associations were found for: Self-esteem, hostility and maladaptive schemas. Significant gender specific characteristics were also found in the included studies. For women, anger, sadness and excitement were significantly associated with obesity-depression relations, whereas boredom, anxiety, loneliness and poor conflict resolution were not. In men low levels of interpersonal effectiveness, poor conflict resolution and loneliness had significant associations whereas boredom, anxiety, sadness, anger and anxiety did not.

Whilst the evidence from Preiss, Brennan and Clarke (2013) does not provide longitudinal evidence that psychological variables mediate either direction of association from obesity to depression, when taken in conjunction with evidence presented in other reviews, the evidence presented is consistent with the idea that the experience of certain moods and other psychological states may be influential in the relationship between obesity and depression. Furthermore, these psychological states may vary depending on a individual's sex.

2.5.3 Discussion

This section of my review has focused on two main aims. The primary aim of the review was to outline the range of mechanisms that might link obesity and depression, and the secondary aim was to examine the evidence presented in support of these hypothesised mechanisms. With regards to the first aim, the reviews here have presented a wide range of potential causal mechanisms between the two conditions, including pathways through physical health, social experiences and physiology. This is line with comments

made in previous reviews, in which it was posited that there existed multiple obesity-depression associations in the population rather than a single general association (Faith, Matz and Jorge, 2002).

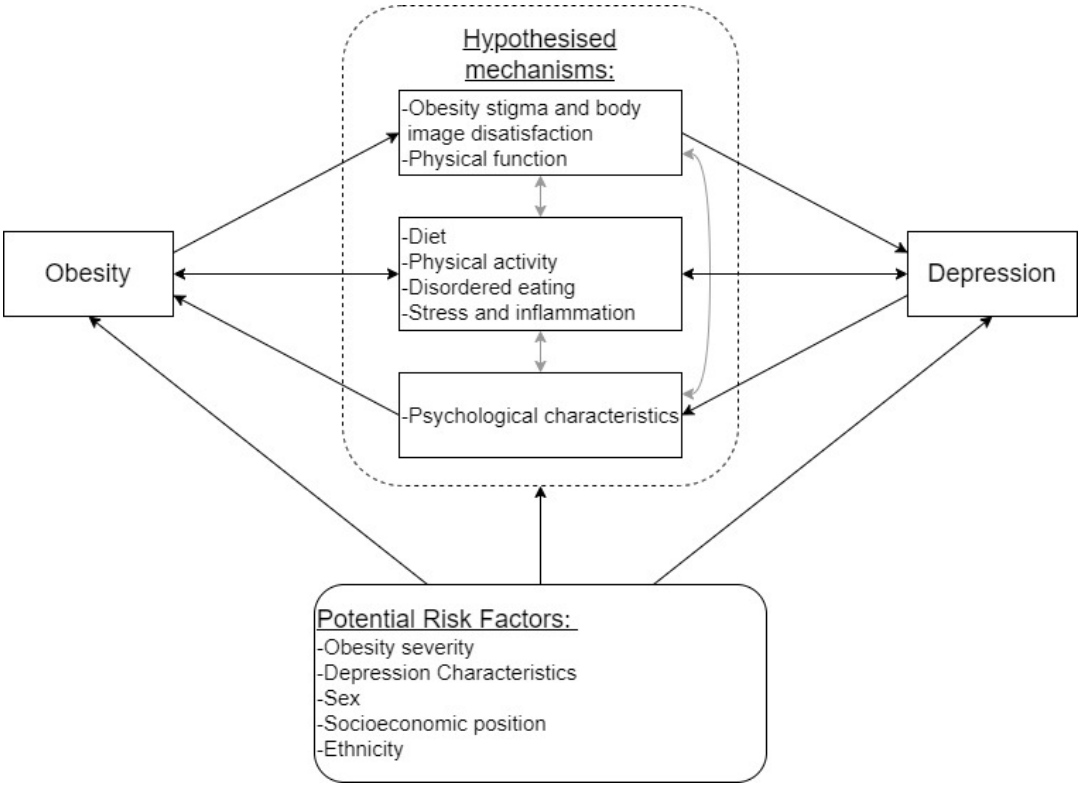

FIGURE 2.1: Conceptual diagram summarising the hypothesised mechanisms that may exist between obesity and depression, along with potential risk factors. Grey arrows represent potential for mechanisms to interact with one another.

The wide ranging influences on the relationship between obesity and depression can be seen in Figure 2.1, which presents a summary of the mechanisms hypothesised to relate obesity with depression, along with potential risk factors that may influence the relationship either individually or together. Within the diagram mechanisms are split based on suggestions from the reviewed literature about their direction of influence. For example, single headed arrows connecting obesity to depression via the top box of hypothesised mechanisms suggests that these mechanisms are implicated in the relationship between obesity and future depression. Mechanisms that are connected to obesity and depression via black double headed arrows are then suggested to operate bidirectionally.

The diagram also presents the potential for mechanisms to interact with each other, as depicted by the grey bidirectional arrows. In this case, bidirectional arrows do not necessarily imply a bidirectional relationship exists between

the groups of mechanisms, but instead that relationships may be present in general. Finally, risk factors were included in the diagram if they were highlighted as potentially having a impact on the relationship between obesity and depression either individually or in conjunction with other risk factors. Risk factors may influence the individual conditions themselves or have an impact on the proposed mechanisms.

Evidence presented in this review suggests that socio-demographic risk factors such as age, gender and socioeconomic position do not have straightforward influences on the relationship between obesity and depression, and may have interactive effects. It is perhaps unsurprising that there are not simple and general effects of these factors on the relationship between obesity and depression. Different mediating factors that link the two conditions may interact with these socio-demographic factors in different ways. For example, studies have suggested that sedentary activity and unhealthy eating are more stigmatised in high socioeconomic groups (Pampel, Krueger and Denney, 2010), which may explain why some reviewers found data suggesting people with obesity of high socioeconomic position were at greater risk of depression. Conversely, physical activity rates have been shown to be negatively associated with high socioeconomic postion (Giles-Corti and Donovan, 2002), particularly in women (Ford et al., 1991).

These mediating variables have complex relationships with socio-demographic factors when considered in isolation, and so it is foreseeable that the relationship between conditions associated by a combination of these factors has a complex relationship with socio-demographic factors. Mechanisms presented in the reviews may also interact. For example, dieting and social stigma may be related to the likelihood of binge-eating episodes (Puhl, Moss-Racusin and Schwartz, 2007), and may also be stressors that influence the association through stress's effect on dysregulation of the HPA-axis and adherence to healthy lifestyle (Tomiyama, 2014).

Alongside the main aim of outlining potential mechanisms, a secondary aim of this review was to examine what empirical evidence was presented in the reviews to support the proposed mechanisms. Although most of the mechanisms proposed by reviewers were based on observations and evidence from empirical studies, a large number of these were cross-sectional. Of the few longitudinal studies that were used, many did not examine the mechanism from one condition to another, and have largely focused on temporal association rather than attempting to infer causation. Therefore, based on the

evidence in this review there is still a need for studies analysing these mechanisms in a framework in which it is possible to infer whether they are true causal pathways or not.

2.6 Conclusions and Next Steps

In this chapter, I have conducted a broad review of the literature surrounding the relationship between obesity and depression, covering both data on the association between the two conditions, and proposed pathways that might link them. Overall, the reviewed data suggested that obesity and depression are related such that increased obesity is associated with increased risk for depression, but that this relationship may be heterogeneous. Relatedly, a wide range of explanatory mechanisms that might link the two have been proposed, alongside a plethora of potential risk factors. These risk factors and mechanisms may interact to contribute to the heterogeneity found in the relationship between obesity and depression.

Despite the reviews proposing a wide range of explanatory mechanisms and risk factors for developing comorbidities in obesity and depression, there was little longitudinal evidence presented to support their existence. As a result, many of the proposed mechanisms have been presented somewhat speculatively by the authors of the included reviews. Similarly, the precise effects of socio-demographic risk factors such as socioeconomic position, age and sex are not entirely clear based on the current literature.

In light of the above evidence gaps for the remainder of this thesis, I will attempt to bridge this gap by analysing longitudinal observational data from the United Kingdom in order to achieve two main aims. The first is to explore whether there is empirical support for mechanisms relating obesity and depression that have been proposed here, and the second is to explore how socio-economic status influences, or is influenced by, these mechanisms.

In order to achieve these aims, in Chapters 3, 4 and 5 I will present a Structural Equation Model that aims to explore the role of Physical function, Diet and Physical activity as mediators in the relationship between obesity and depression and how socioeconomic position influences these mediating pathways. Then, in Chapters 6 and 7, I present an exploration of the role of stigma in generating the relationships between obesity and both depression and socioeconomic position using an Agent-based Model.

Chapter 3

Exploring the Whitehall II Data Set

3.1 Introduction

As was noted in Chapter 2, the current literature examining the relationship between obesity and depression suggests that it is complicated and potentially involves multiple interacting mechanisms and risk factors. Within the proposed mechanisms, pathways that operate via individuals' physical health and health behaviours may be important. In terms of the impact of physical health, evidence presented by the conceptual reviews suggested that poor physical function may be a mechanism through which obesity leads to an increased risk of depression. Four out of six of the conceptual reviews presented evidence that consistently suggested poor physical health and concern about one's physical health could facilitate the onset of depression in individuals with obesity.

When looking at the impact of health behaviours, depression and obesity may be related via diet and physical activity bi-directionally. Evidence presented in the reviews found that physical activity was associated with depression both cross-sectionally and longitudinally such that increased physical activity was associated with reduced future depressive symptoms and lower risk of future depression. As such, low physical activity in people with obesity may confer risk for future depression.

Similarly, depression may impact individuals ability to engage in healthy exercise routines, leading to an increased risk of developing depression. These issues of depression impacting adherence and maintenance of a healthy lifestyle may also extend to dietary habits, suggesting this is another pathway through which depression may increase individuals' risk of developing obesity. Furthermore, binge-eating behaviours in people with obesity may put these individuals at higher risk of developing depression.

Despite these suggestions it was also noted that longitudinal evidence supporting the existence of these mechanisms was generally lacking. Furthermore, where longitudinal evidence does exist, it has not investigated the mechanism in its entirety, from one condition through to the other, thus limiting the strength of conclusions one can draw about whether these factors form a mechanism that relates obesity and depression. Similarly, it is not known how socioeconomic position might interact with these mechanisms. As such, in the next three chapters I will analyse longitudinal data from a cohort of civil servants (the Whitehall II data set) using a Structural Equation Model to examine the following hypotheses:

- **Hypothesis 1:** *Physical Function mediates the association between obesity and future depression.*
- **Hypothesis 2:** *Diet mediates the association between obesity and depression bidirectionally.*
- **Hypothesis 3:** *Physical activity level mediates the association between obesity and depression bidirectionally.*
- **Hypothesis 4:** *Socioeconomic position (SEP) affects the relationship between obesity and depression via multiple effects on obesity, depression and their mediating variables*

The analysis of these hypotheses is conducted in three main stages each presented in their own chapter. In this chapter I will conduct an exploratory analysis of the Whitehall II data set. This is followed in Chapter 4 by a description of how missing data will be handled within the main Structural Equation Model that is presented in Chapter 5. A more detailed description of the Structural Equation Model used is given in Chapter 5, but to provide some context for Chapters 3 and 4 I will briefly outline the Structural Eqaution Modelling methodology here.

Structural Equation Modelling (SEM), is a general statistical modelling technique that includes a combination of other multivariate data analysis methods such as regression, factor analysis, and path analysis (Hox and Bechger, 1998). Modelling using structural equations typically involves describing a path diagram between the constructs of interest, and then estimating regression or path coefficients, assuming the data follows the given structure (Hox and Bechger, 1998; Little and Card, 2013). In this way Structural Equation Models can be used to check the consistency of a model with available data and estimate causal effects between constructs (Anderson and Gerbing, 1988;

Bollen and Pearl, 2013). They can also be applied to both longitudinal and cross-sectional data, making it a versatile tool for many kinds of data analysis.

Structural Equation Models can make use of both latent variables and observed variables to estimate the relationships between constructs. In the more simple case, constructs can be represented by a single observed variable from the data, with regression paths between the observed variables estimated by the Structural Equation Model (Usami, Murayama and Hamaker, 2019). An example of this is the models using the path analysis methodology, upon which much of the SEM framework has been built (Wright, 1960; Petraitis, Dunham and Niewiarowski, 1996). However, constructing a Structural Equation Model in this way will mean that some unwanted variance from sources such as measurement error will impact the estimation of structural relationships between constructs (Little and Card, 2013). At best, this will introduce some bias into the estimates of the size of relationships between constructs, and at worst could affect the substantive conclusions of the study (Little and Card, 2013).

To tackle this problem, we can use latent variables within our Structural Equation Model. Specifically, constructs can be represented by a latent variable which is measured using multiple observed variables within the data, referred to as indicators. This allows variance in the measured data that is not related to the underlying construct to be filtered out (Little and Card, 2013). As a result, the latent variable captures the shared variance of the indicators which in theory better captures the variance of the underlying construct than any one indicator can alone (Little and Card, 2013).

3.1.1 Chapter Aims

The exploratory analysis presented in this chapter has two main aims. The first aim of this chapter is to give a general description of the variables that will be used as indicators in the Structural Equation Model that is presented in Chapter 5, and the second aim is to provide insight into the distributions of important variables within the Whitehall II data set as well as how they relate to one another. Fulfilling the second aim will provide necessary context in which the model presented in Chapter 5 can be interpreted.

To achieve these aims the remainder of this chapter is broken down into five

main sections. In the first of these sections (Section 3.2), I will give a general description of the Whitehall II data set, followed by a description of the sample that have been used for this analysis. In Section 3.3, I then describe the measures that have been considered for use as indicators in the model for each construct. In particular, I describe precisely how the data was measured using the Whitehall II questionnaire, as well as how derived measures have been created. This section will also contain descriptions of the distributions of these variables that are structured around satisfying the two aims of this chapter.

Overall the description of the distributions will focus on two main aspects. Firstly, I will be looking for departures from Normality, as this may impact the analysis of of the SEM in Chapter 5. Secondly, the distributions of the indicators provides a broad picture of the pattern of important characteristics, such as obesity and depression, within the Whitehall II sample. This provides important context in which the model results can be interpreted. For example, if there is little variation in key variables such as those measuring socioeconomic position, then the relationships between socioeconomic position and other variables might be affected by this lack of variation.

Sections 3.4 and 3.5 will then give an overview of the relationships that are being investigated within the model. Specifically, in Section 3.4, I examine the cross-sectional correlations between measures on different constructs, which is then built upon in Section 3.5 where I analyse some simple longitudinal linear regression models. Similar to the examination of the distributions of measured variables, this analysis aims to provide a crude snapshot of the relationships in our model, which as mentioned previously will provide some context in which to interpret the results in Chapter 5. Finally, in Section 3.6, the findings from our exploratory analysis are discussed in light of the literature in Chapter 2, and possible implications for the model in Chapter 5 are eluded to.

3.2 The Whitehall II Dataset

The Whitehall II study is a cohort of 10,308 participants who were recruited from the British Civil Service in 1985. In the first wave of data collection, the participants were aged between 35 and 55; 3314 of them were female and 6895 were male (Marmot and Brunner, 2005). So far data has been collected on the cohort in 12 phases, or waves. Going forward I will refer to these

phases as waves. Table 3.1 gives a summary of the data collected in each wave. In each of the waves, except for wave 10, data was requested from all remaining eligible participants, which included all those who had not died or withdrawn from the study. In addition to the full follow-ups a pilot of new measures was carried out at wave 10.

Within each wave of full data collection, participants were sent a questionnaire to fill in and return that covers five main topic areas: family history; health; health behaviours; personality, mental health and well-being; and psychosocial and socioeconomic. In addition to the self-report questionnaire participants also took part in a clinical evaluation in waves 1,3,5,7,9,11 and 12. The clinical evaluation collected data on the following topics: anthropometry, blood analysis, cardiovascular measures, cognitive function, cortisol and physical functioning.

Whilst these topic areas remain consistent throughout the study, the exact measures included within each questionnaire and clinical evaluation differed. When designing a structural equation model it is preferable that the measures used at different time points remain the same as this allows us to describe a simpler model that is easier to analyse and interpret. If the indicators for a construct change over the course of a study, then the underlying meaning of that construct might also change. Failing to account for the effect of changing measures could lead to a changed construct definition halfway through a study that is unaccounted for, which could affect both the interpretability of the model, and the accuracy of the study's findings. As a result, in order to account for this potential change, we would need to add in additional parameters to the model so that the comparability of the different construct definitions could be quantified. This naturally adds additional complication into the analysis procedure.

As a result of these concerns, and the changing measures found in the Whitehall II data set, our analysis of the Whitehall II data set has been limited to variables in waves 5, 7 and 9, as this will allow us to use the same measures over time for our constructs. Going forward, I will refer to this subset of the data as 'our sample of the Whitehall data'.

Wave	Dates	Type	Participation	Response Rate (Alive)	Response Rate (Eligible)
1	1985-1988	Questionnaire and Clinic	10,308	-	-
2	1989-1990	Questionnaire	8,132	79.3%	79.3%
3	1991-1994	Questionnaire and Clinic	8,815	86.6%	86.6%
4	1995-1996	Questionnaire	8,628	85.3%	92.4%
5	1997-1999	Questionnaire and Clinic	7,870	78.7%	86.3%
6	2001	Questionnaire	7,335	74.4%	82.5%
7	2002-2004	Questionnaire and Clinic	6,967	71.6%	82.2%
8	2006	Questionnaire	7,173	75.2%	87.2%
9	2007-2009	Questionnaire and Clinic	6,761	72.3%	84.5%
10(**)	2011	Questionnaire and Clinic	277	n/a	n/a
11	2012-2013	Questionnaire and Clinic	6,318	70.9%	84.1%
12	2015-2016	Questionnaire and Clinic	5632	66.6%	80.2%
13	2019-2020	In progress	-	-	-

(**) - Wave 10 was a pilot study used to test new measures

TABLE 3.1: Summary of data collection for the Whitehall II study. *Reproduced on 26/02/2021 from https://www.ucl.ac.uk/epidemiology-health-care/research/epidemiology-and-public-health/research/whitehall-ii/data-collection*

3.2.1 The Whitehall II Sample Characteristics

Before embarking on an exploration of the constructs included in our model, first I examined the characteristics of the sample. The review of the literature presented in Chapter 2 showed that the relationship between obesity and depression may vary within different population groups. In particular, the relationship between obesity and depression was often found to be stronger in women than in men, and in younger women in particular. Some authors also suggested that ethnicity may interact with mechanisms that might relate obesity and depression such as body size ideals, however, there was little evidence presented in the reviews supporting this. As such, understanding

the distribution of these characteristics provides a useful and necessary context in which to interpret any results that arise from our analyses. With this in mind, our first step in exploring the Whitehall II dataset is to explore the distribution of these characteristics within the sample.

	Wave 5 (N = 7,870)	Wave 7 (N = 6,967)	Wave 9 (N = 6,761)
Age			
Mean ± sd	55.95 ± 6.04	61.24 ± 6.01	66.01 ± 5.98
Sex			
Male	5,473 (70)	4,893 (70)	4,759 (70)
Female	2,397 (30)	2,074 (30)	2,002 (30)
Ethnicity			
White	7,186 (91)	6,393 (92)	6,218 (92)
Non-white	675 (9)	561 (8)	527 (8)

sd - Standard Deviation

TABLE 3.2: Sample characteristics in waves five, seven and nine

Overall, the sample from the Whitehall II data set that we are analysing is comprised of older adults, 70% of whom are males. In all our analysis waves over 90% of the sample identify their ethnicity as white.

3.3 Examining the Distributions of Construct Indicators

In this section we will examine the distributions of variables that are intended to be used as construct indicators in the model. The variables analysed in this section were chosen based on two factors: their conceptual suitability for being an indicator of the construct of interest and their availability within the dataset. Conceptual suitablity was judged based on the description of the variable given within the Whitehall documentation. Data availablity was important as the SEM methods employed in Chapter 5 require there to be consistent measures used for the constructs over time. Similarly, the measures used for different constructs needed to be available at the same timepoints as each other. As such I only selected measures to be analysed if they satisfied these availability conditions.

In this way, variables were chosen if I considered them a part of the best available representation of the underlying construct that could be used within the structural equation model. For example, both BMI and waist circumference

standard ways of measuring body weight and body size respectively, and are often used to make inferences about an individual's level of obesity. Hence, these variables provide a good basis on which to describe a latent variable that represents obesity. Similarly, both variables are available within waves 3, 5, 7, 9 and 11. The statistical properties that groups of indicators need to form a 'good' construct will be examined during Chapter 5 and so the final indicators to be used in the model are subject to change from those examined here.

3.3.1 Indicators of Obesity

3.3.1.1 BMI

One of the primary indicators of obesity contained in the Whitehall II dataset is BMI. In the data, BMI is derived from measures of height and weight that were collected as part of the clinical evaluations in waves 1, 3, 5, 7, 9, 11 and 12. As such they provide a value for each participant's BMI that is not subject to self-report bias. Table 3.3 gives summarises the mean (with standard deviation) and median (with interquartile range) of the BMI distribution in wave 5, 7 and 9. For both females and males, average BMI in the sample increases minimally over time. In wave 5, the averages for BMI sit just above the overweight cut-off of 25 and they continue to rise further above this value as time progresses. This means that, in both sexes, at least half the sample are either overweight or obese in every wave of our data. In general, there is little difference between the two sexes in terms of these statistics, with women having a fractionally higher average BMI, as well as a larger amount of variation in the sample.

Figure 3.1 shows the observed trends in BMI from wave 1 through to 11 without separating the sample by sex. In early waves, the interquartile range spans only the cut-off between normal weight and overweight, whereas from wave 7 onwards, this range spans both the cut-off between normal weight and overweight and the cut-off between overweight and obesity. This suggests that alongside the slight increase in average BMI over time in the sample, there is also a slight increase in variability of BMI in the sample.

This trend of increasing overweight and obesity in the sample can also be seen when analysing the number of people that fall within each BMI category within each wave. The BMI categories used are displayed in Table 3.4. From wave 1 through to wave 11 the proportion of the cohort who have a normal

	Males	Females
BMI - Wave 5 (1997-1999)		
Median (IQR)	25.68 (23.75, 27.84)	25.54 (23.06, 29.11)
Mean $\pm$ sd	26.04 $\pm$ 3.48	26.43 $\pm$ 4.96
BMI - Wave 7 (2002-2004)		
Median (IQR)	26.20 (24.00, 28.50)	26.10 (23.50, 30.00)
Mean $\pm$ sd	26.56 $\pm$ 3.83	27.21 $\pm$ 5.47
BMI - Wave 9 (2007-2009)		
Median (IQR)	26.15 (23.99, 28.62)	26.42 (23.32, 30.48)
Mean $\pm$ sd	26.60 $\pm$ 3.95	27.32 $\pm$ 5.60

IQR - Interquartile Range; sd - Standard Deviation

TABLE 3.3: Descriptive Statistics for BMI distribution of time by Sex

BMI Classification	BMI range
Underweight	< 18.5
Normal weight	≥ 18.5 and < 25
Overweight	≥ 25 and < 30
Class I Obesity	≥ 30 and < 35
Class II Obesity	≥ 35 and < 40
Class III Obesity	≥ 40

TABLE 3.4: Table of BMI classifations

BMI drops from 59.4% to 36.4%, whereas the proportion of the sample who are overweight and class I obese increases from 34.3% and 4.6% to 46.5% and 13.2% respectively. The proportion of individuals with class II and class III obesity in the cohort also increases over the study period. This data is visualised in Figure 3.2. Despite the seemingly small trend for increased BMI over time shown in Figure 3.1 it is clear that when split into BMI categories, in later waves a greater proportions of the sample are found in the overweight and obese categories.

Having looked at the general BMI trends in the data, it only remains to examine whether there are any departures from Normality in the BMI distributions. From inspection of Figure 3.3 it is clear that overall, all three distributions follow a bell-curve similar to that of the Normal distribution. However, there is also a notable right skew in the distribution, as the right hand tail of the distribution is larger than the left. This indicates that there are more individuals with higher BMIs than would be the case if the data followed a Normal distribution more closely.

	Males	Females
Wave 1 (1985-1988)		
Underweight	80 (1.16%)	69 (2.02%)
Normal Weight	4,090 (59.42%)	2,015 (59.07%)
Overweight	2,361 (34.30%)	953 (27.94%)
Class I Obesity	318 (4.62%)	272 (7.97%)
Class II Obesity	29 (0.42%)	79 (2.32%)
Class III Obesity	5 (0.07%)	23 (0.67%)
Wave 3 (1991-1994)		
Underweight	35 (0.63%)	45 (1.81%)
Normal Weight	2,907 (51.99%)	1,228 (49.46%)
Overweight	2,250 (40.24%)	818 (32.94%)
Class I Obesity	358 (6.40%)	273 (10.99%)
Class II Obesity	35 (0.63%)	81 (3.26%)
Class III Obesity	6 (0.11%)	38 (1.53%)
Wave 5 (1997-1999)		
Underweight	17 (0.43%)	27 (1.60%)
Normal Weight	1,630 (40.79%)	727 (43.15%)
Overweight	1,872 (46.85%)	596 (35.37%)
Class I Obesity	409 (10.24%)	230 (13.65%)
Class II Obesity	59 (1.48%)	75 (4.45%)
Class III Obesity	9 (0.23%)	30 (1.78%)
Wave 7 (2002-2004)		
Underweight	21 (0.46%)	30 (1.59%)
Normal Weight	1,631 (35.70%)	713 (37.91%)
Overweight	2,189 (47.91%)	668 (35.51%)
Class I Obesity	594 (13.00%)	307 (16.32%)
Class II Obesity	113 (2.47%)	98 (5.21%)
Class III Obesity	21 (0.46%)	65 (3.46%)
Wave 9 (2007-2009)		
Underweight	25 (0.57%)	36 (2.04%)
Normal Weight	1,600 (36.17%)	659 (37.25%)
Overweight	2,077 (46.95%)	583 (32.96%)
Class I Obesity	571 (12.91%)	308 (17.41%)
Class II Obesity	125 (2.83%)	136 (7.69%)
Class III Obesity	26 (0.59%)	47 (2.66%)
Wave 11 (2012-2013)		
Underweight	31 (0.77%)	43 (2.73%)
Normal Weight	1,469 (36.37%)	584 (37.06%)
Overweight	1,876 (46.45%)	517 (32.80%)
Class I Obesity	531 (13.15%)	286 (18.15%)
Class II Obesity	111 (2.75%)	104 (6.60%)
Class III Obesity	21 (0.52%)	42 (2.66%)

TABLE 3.5: Table of BMI category distribution in males and females. Weight classes described in Table 3.4

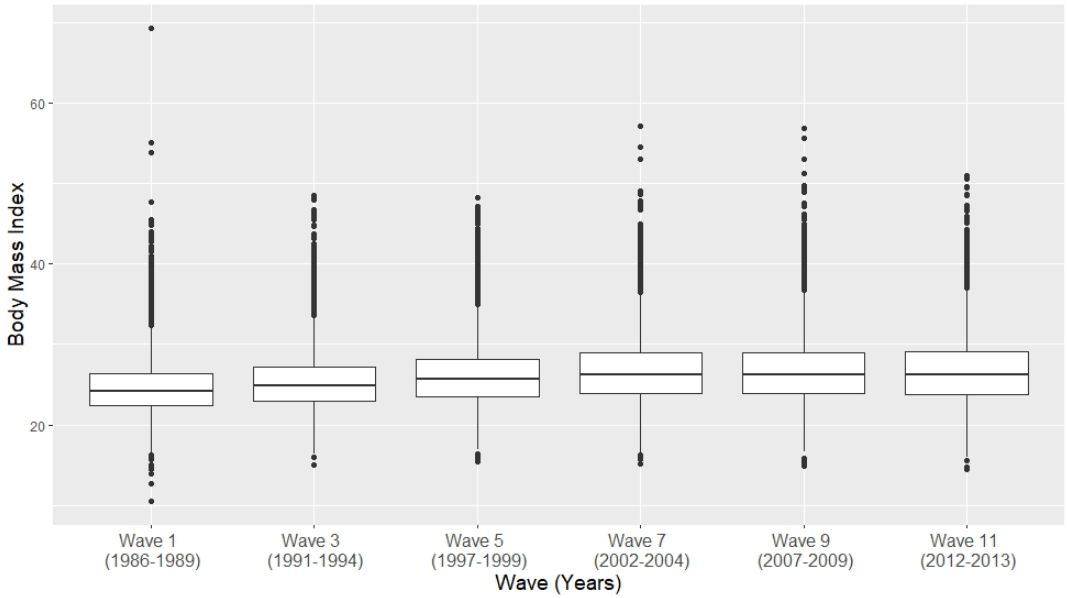

FIGURE 3.1: Box-plot of BMI distribution across waves of Whitehall Data

This departure from Normality can be seen more clearly in the Q-Q plot in Figure 3.4. Below a BMI of about 27, which is close to the mean BMI of the sample, the BMI distributions closely matches a normal distribution, as indicated by the lines in the Q-Q plot being straight in this portion of the plot. However, above a BMI of 27, the lines in the Q-Q plot begin to curve upwards slightly, indicating that quantiles in this part of the BMI distribution are at higher values than would be the case if the data were normally distributed.

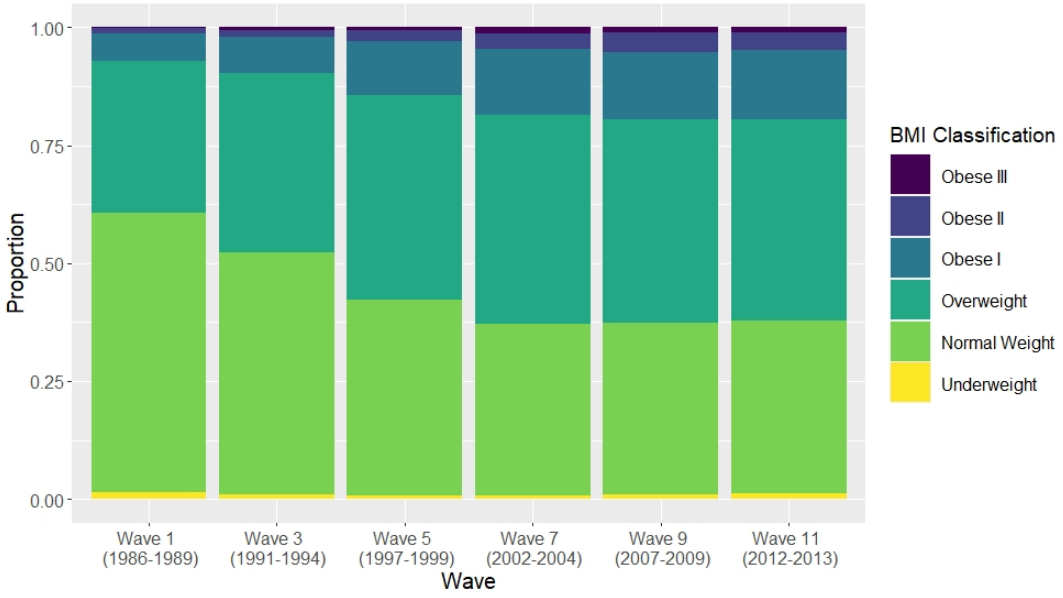

FIGURE 3.2: Distribution of BMI categories across each wave

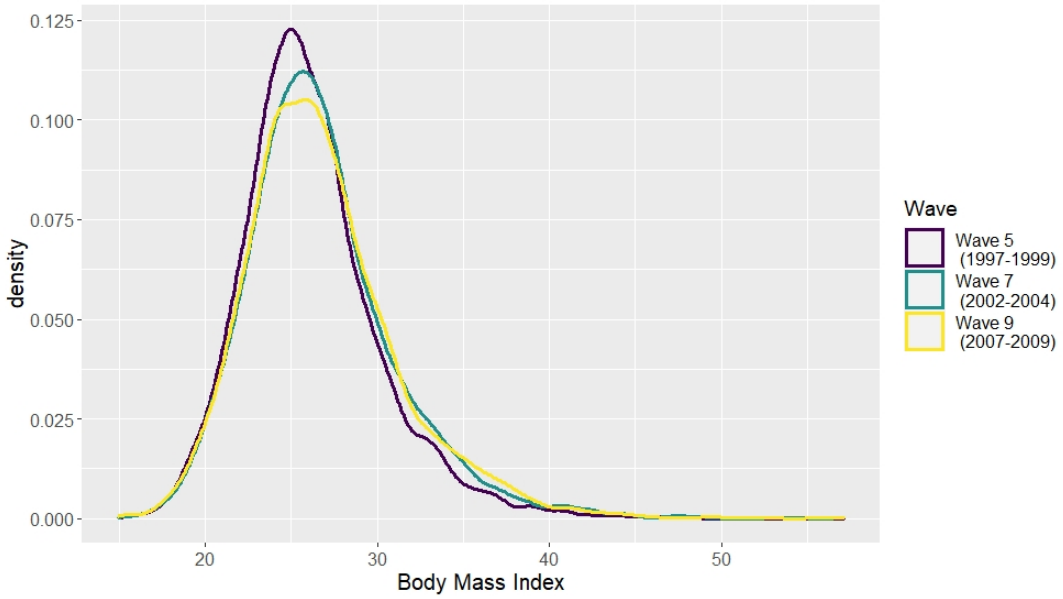

FIGURE 3.3: Distribution of BMI in waves 5, 7 and 9

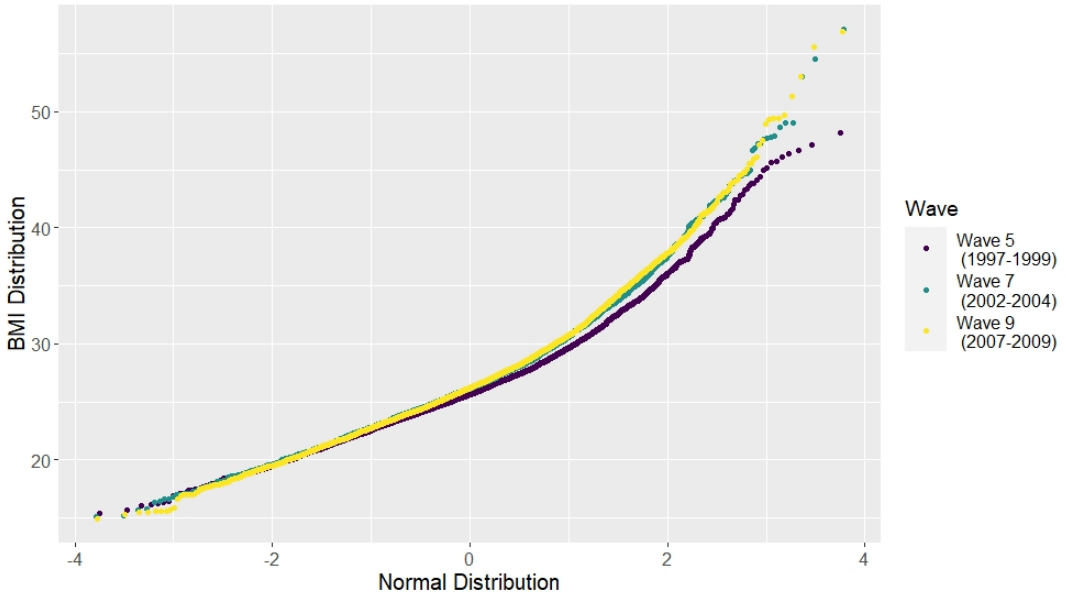

FIGURE 3.4: Q-Q plots comparing the BMI distributions in waves 5, 7 and 9 to the normal distribution

3.3.1.2 Waist Circumference

Waist circumference was collected from participants during the clinical evaluation in waves 3, 5, 7, 9 and 11. Exact measurements, in centimetres, of the participants' waist circumferences are available in the data, however, for the purpose of examining how the distribution of waist circumference developed across the waves, here we have categorised the participants' waist circumferences base on the associated risk to health. Waist circumference was classified into three risk categories: low, medium or high based on NHS recommendations (NHS Digital, 2017). In recent years these categories have been renamed to desirable, high and very high respectively. The exact cut-offs used for males and females are displayed in Table 3.6.

Risk Group	Cut-off (Males), cm	Cut-off (Females), cm
Low	≤ 94	≤ 80
Medium	≥ 94 and ≤ 102	≥ 80 and ≤ 88
High	> 102	> 88

TABLE 3.6: Cut-off values for Waist Circumference Risk groups

Figure 3.5 shows the proportion of individuals within each wave's sample whose waist circumferences are considered low, medium or high risk. Exact counts, split by sex, of the number of individuals in each risk group in each wave are also shown in Table 3.7. Percentages of the sample found in each risk group within each wave for each sex are also displayed.

Similar to the trend found in BMI across waves of data collection, the proportions of the sample with a medium or high-risk waist circumference increases over time. This could indicate that for those individuals whose BMI has increased, this is due to an increase in abdominal adiposity. If so, this is of concern since excess body fat around the abdomen has been suggested to be particularly harmful for cardiovascular health (Emery et al., 1993; Broom, 2006) and is a predictive of the developing metabolically unhealthy obesity (Hwang et al., 2015).

Also of note is that females were more likely to have high-risk waist circumferences compared to men. Additionally, females were less likely to have either a medium or low risk waist circumference compared to men. This suggests that, whilst BMI differences between the sexes are largely non-existent,

	Males	Females
Wave 3 (1991-1994)		
low	4,333 (78.43%)	1,712 (69.34%)
medium	799 (14.46%)	413 (16.73%)
high	393 (7.11%)	344 (13.93%)
Wave 5 (1997-1999)		
low	2,189 (61.44%)	851 (53.96%)
medium	845 (23.72%)	319 (20.23%)
high	529 (14.85%)	407 (25.81%)
Wave 7 (2002-2004)		
low	2,352 (51.41%)	810 (42.97%)
medium	1,261 (27.56%)	432 (22.92%)
high	962 (21.03%)	643 (34.11%)
Wave 9 (2007-2009)		
low	2,106 (47.57%)	660 (37.33%)
medium	1,244 (28.10%)	422 (23.87%)
high	1,077 (24.33%)	686 (38.80%)
Wave 11 (2012-2013)		
low	1,654 (40.88%)	493 (31.34%)
medium	1,135 (28.05%)	370 (23.52%)
high	1,257 (31.07%)	710 (45.14%)

TABLE 3.7: Table of Risk to Health due to waist circumference across waves

there are sex-differences in the distribution of abdominal obesity. As a result, females in the sample may be at a greater risk for health issues that are associated with abdominal obesity than men.

As with BMI, having outlined the general trends in waist circumference, all that remains is to assess departures from Normality in the distribution of waist circumference. Inspecting the distributions of waist circumference in Figure 3.6 we can see again that the shape of the distributions follow a bell shaped curve similar to that of the Normal distribution. However, the Q-Q plot in Figure 3.7 show that the waist circumference distributions do deviate somewhat from Normality. This can be seen by the fact that the lines in the Q-Q plot are not straight and instead curve gently upwards as one moves along the quantiles of the distribution. Again this indicates that the data is slightly right-skewed, which given the right-skew of BMI is unsurprising.

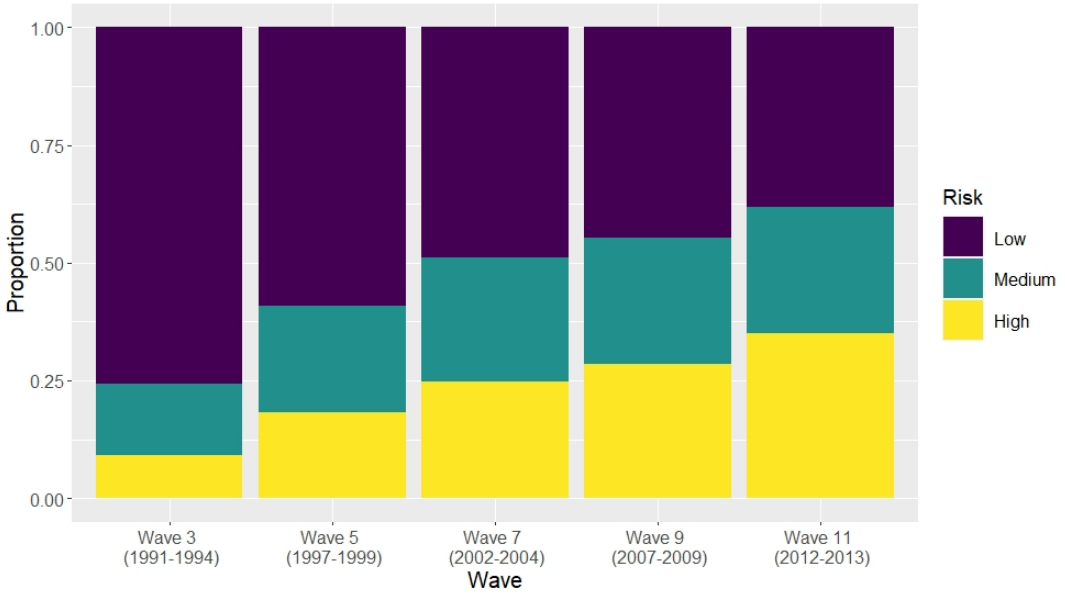

FIGURE 3.5: Risk profile of sample waist circumferences across each wave

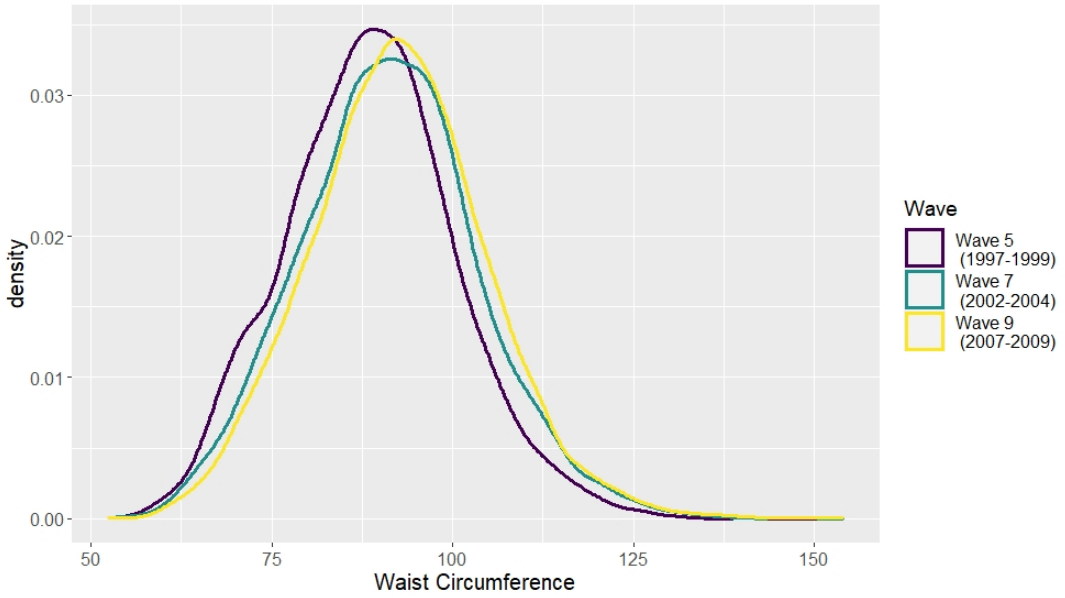

FIGURE 3.6: Distribution of waist circumference in waves 5, 7 and 9

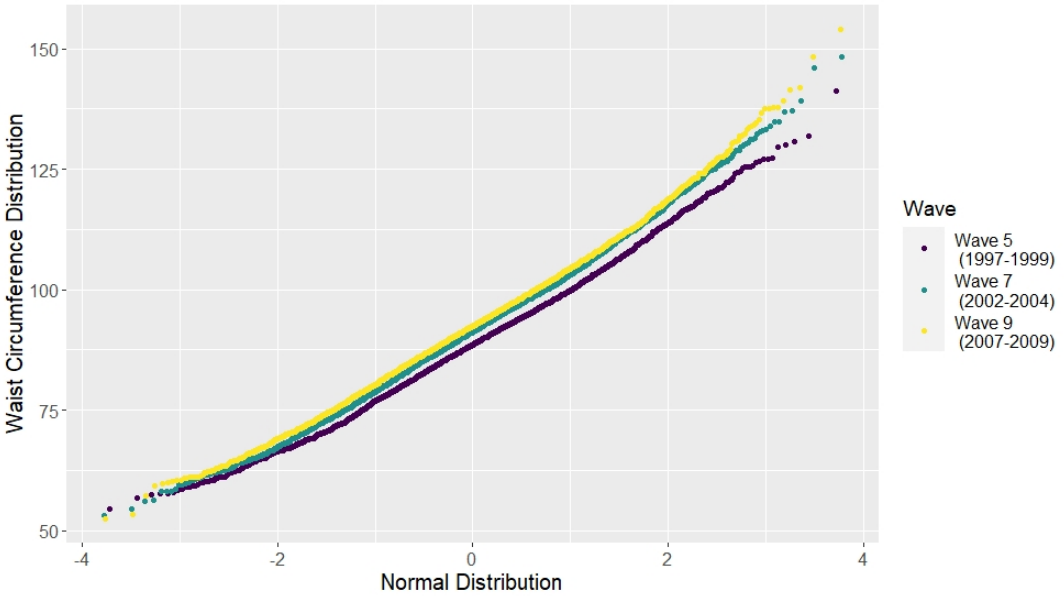

FIGURE 3.7: Q-Q plots comparing the waist circumference distributions in waves 5, 7 and 9 to the normal distribution

3.3.2 Indicators of Depression

3.3.2.1 The 30-item General Health Questionnaire (GHQ-30)

The GHQ-30 is a general mental health screening instrument which aims to detect diagnosable psychiatric disorders. Originally introduced in a 60-item form by Goldberg (1972), the questionnaire now comes in multiple formats: The full 60-item questionnaire (GHQ-60), a 30-item questionnaire (GHQ-30), a 28-item questionnaire (GHQ-28) (Goldberg and Hillier, 1979) and a 12-item short form version (GHQ-12). These different forms of the GHQ have been created to fill specific needs within mental health screening. As such, they do not all provide identical information about an individual's mental health. For example, the GHQ-30 only provides a single sum score to indicate general risk of an individual developing mental health issues, whereas the GHQ-28 provides information using four distinct sub-scales: somatic symptoms, anxiety and insomnia, social dysfunction and severe depression (Goldberg and Hillier, 1979).

Within the Whitehall II data set, the GHQ-30 is administered to the participants as part of the main questionnaire. Within this questionnaire, participants are asked to record how they been "over the past few weeks" on each of the items to give an overview of their recent mental health. Since the GHQ-30 has been used, there is no prescribed depression-based sub-scale that we can make use of as an indicator of depression. This means that I will have to define my own, or else simply use the GHQ-30 sum score as the indicator. Since the total sum score is intended to provide an indicator of general risk of mental health issues, using the sum score is undesirable, as our estimate of the relationship between obesity and depression may be confounded by variation in the GHQ scores that comes from items that are not related to depression.

Given that the GHQ-30 does not have a specific depression score or subscale, an indicator of depression was created using individual items from the GHQ-30. The subscale was created similarly to the GHQ-30 based measure used in a previous analysis of the Whitehall II data by Stansfeld et al. (2003). In Stansfeld et al. (2003), the authors summed the likert scores of four items from the GHQ-30 to form their subscale. The items were included based a factor analysis and comparison with items from the depression subscale in the GHQ-28. For my sub-scale, I chose items to be included in the based on the factor analysis of the GHQ-30 performed in Huppert et al. (1989). In 10

samples containing 600 British adults each, Huppert et al. (1989) identified a 'depression' factor within the data consisting of five items from the GHQ-30, which are displayed in Table 3.8. As in Stansfeld et al. (2003), to create the GHQ-based depression score that will be used in all our analysis of the Whitehall II dataset, I have summed the Likert scores from each of the items before subtracting the minimum possible score that could be obtained. This subtraction is made so that the minimum possible depression score that can be obtained is zero.

TABLE 3.8: Items from GHQ included in our measure of depression

Item code	Item Content
ghq-24	'Been thinking yourself as a worthless person'
ghq-25	'Felt that life is entirely hopeless'
ghq-26	'Been feeling hopeful about your own future'
ghq-29	'Felt that life isn't worth living'
ghq-30	'Found at times you couldn't do anything because your nerves were too bad'

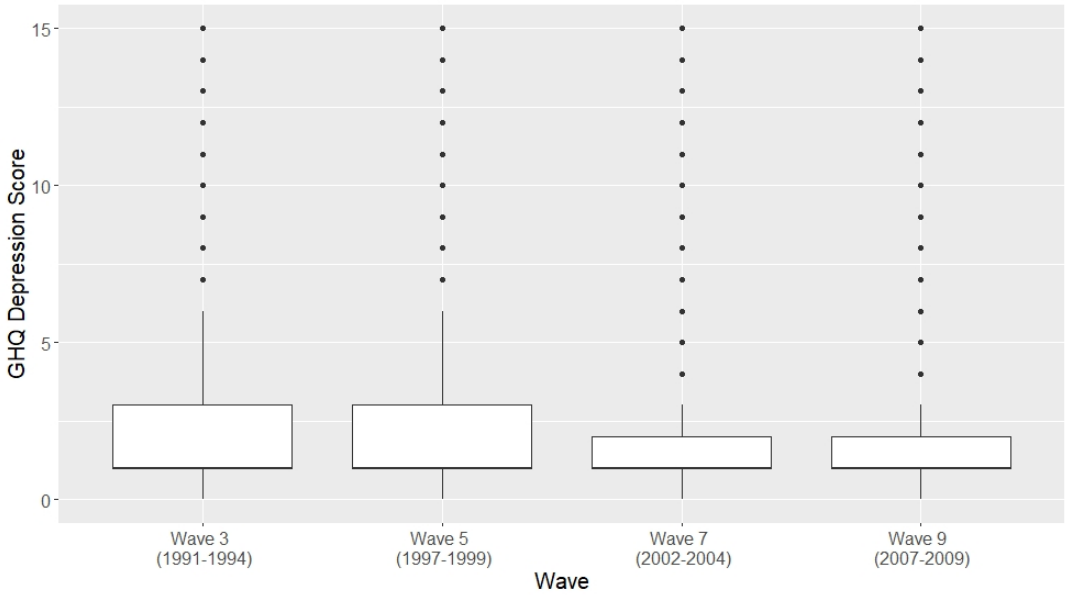

FIGURE 3.8: Box-plots of the GHQ based depression score within waves 3, 5, 7 and 9

Figure 3.8 summarises the distributions of the GHQ-based depression score within waves, 3, 5, 7 and 9. Overall, it is largely the same in each of the waves, with the majority of the mass being concentrated over very low scores, with a long tail to higher scores. This indicates that most individuals only scored a few items above the lowest option, with very few responding with a 3 or

4 on any of the items. Taking the interpretation of scores from Goldberg and Hillier (1979) i.e. that only a 3 or 4 on an item suggests the presence of a symptom, this suggests that there may be very few individuals in the sample with any depressive symptoms whatsoever. If true this may make it difficult to explore the relationships of interest in our substantive model.

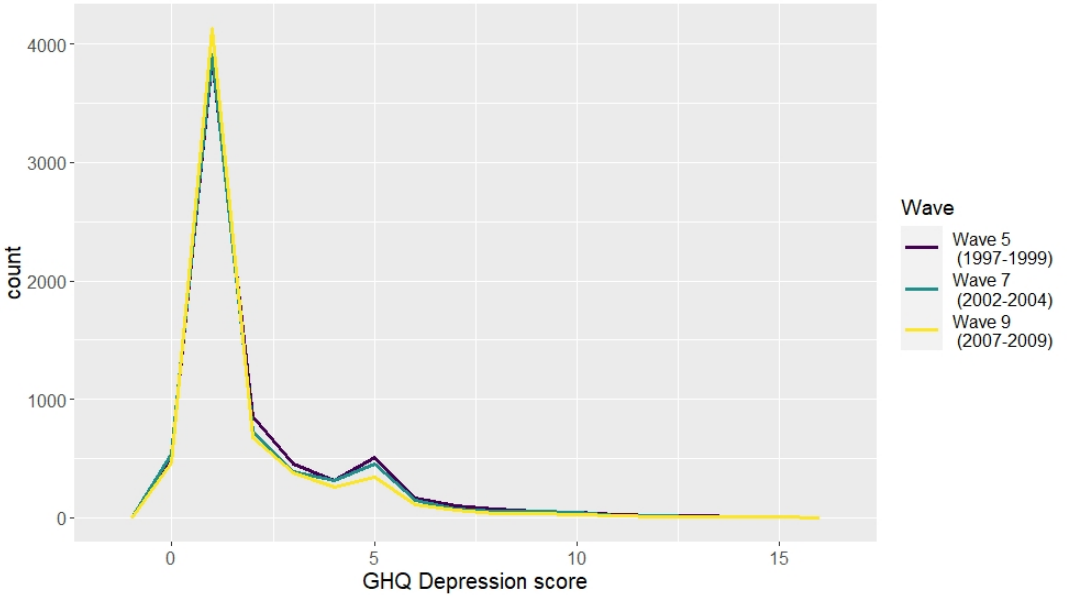

FIGURE 3.9: Frequency polygons for GHQ based depression score in waves 5, 7 and 9

The GHQ-based depression score that is derived here can only take integer values between zero and fifteen. As such, it is not a continuous variable and so cannot be Normally distributed. Even if we decide to treat the discrete measure as continuous, the right skew of the distribution and the focus of its mass around very low values further violate any assumptions of Normality that we might make in our Structural Equation Model. It is important also to note that the apparent smoothness of the distribution portrayed in Figure 3.9 is only due to the size of the bins used to calculate the frequency polygon. Moving to smaller bins would have resulted in a more jagged graph with multiple peaks, highlighting the effect of the variable's discrete structure.

3.3.2.2 SF-36 Mental Health Score

A second mental health measure collected in Whitehall-II is the SF-36 emotional well-being sub-scale. This sub-scale of the SF-36 is comprised of five items, shown in Table 3.9. Participants indicate the amount of time in the past four weeks they have been in each of the states listed in the items. For each question there are six options: "All of the time", "Most of the time",

"A good bit of time", "Some of the time", "A little of the time" and "None of the time". Item responses are coded from 1-6 with 1 representing "All of the time" and 6 representing "None of the time". To calculate the sub-scale's overall score, the item scores for items three and five are reverse coded, before all the individual scores are combined using the formula in 3.1, where i is an index of the five items.

Item	Item Content
1	'Have you been a very nervous person'
2	'Have you felt so down in the dumps that nothing could cheer you up'
3	'Have you felt calm and peaceful'
4	'Have you felt downhearted and low'
5	'Have you been a happy person'

TABLE 3.9: Items from SF-36 used in the emotional well-being score

$$S = \left(\frac{\sum_i s_i - 5}{25} \right) \times 100 \tag{3.1}$$

This process gives a score that ranges from 0-100, with higher scores interpreted as better emotional well-being (Ware and Sherbourne, 1992). As such to make this a measure of depressed affect, we will reverse the scale by taking $S_{SfMh} = 100 - S$. It is worth noting that, as with the GHQ, this measure is designed to be a general measure of mental health, rather than being specifically designed to detect depression. Therefore, there is likely to be some confounding due to other mental health issues that are picked up by this screening instrument.

Figure 3.10 shows the distribution of our SF-36 emotional well-being score in waves 3, 5, 7 and 9. Unsurprisingly, the distributions follow a very similar pattern to the GHQ-based measure, with most of the mass being concentrated on low scores of the measure. However, the median of the distribution sits slightly higher in the range of possible values than was the case with the GHQ-based measure. Relatedly, the mass of the distribution is more spread out across the range of possible values than was the case with the GHQ depression score. Overall, this again points to the majority of the Whitehall II sample having little to no mental health difficulties.

In terms of departures from normality, the SF-36 emotional well-being score is again non-continuous and so cannot be Normally distributed. As with the

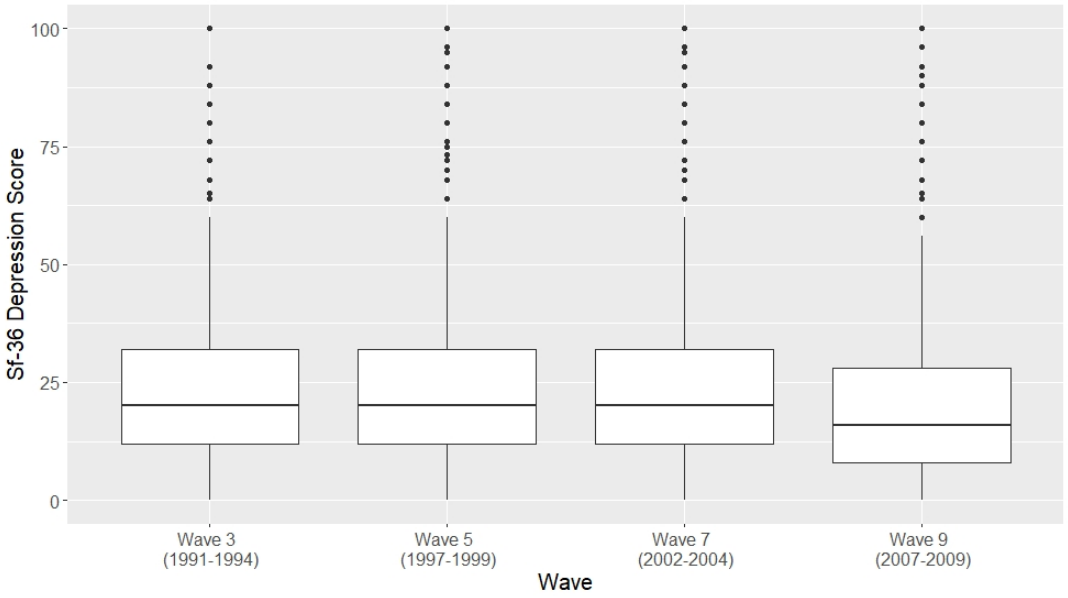

FIGURE 3.10: Box-plots of the SF-36-based depression score within waves 3, 5, 7 and 9

GHQ-based depression score, even if we are to treat this discrete measure as continuous, there are several departures from Normality. Again there is a right skew and additionally there are multiple peaks in the distribution. These peaks are likely a result of the underlying discrete nature of the data. Therefore, similar to the GHQ-based depression score this violation of Normality will need to be accounted for in our analysis of the substantive model. Furthermore, the apparent smoothness in the frequency polygons in Figure 3.11 is an artefact of the aggregation of scores into bins. The size of the bins was chosen to make the overall trends in the data more clear. Choosing a smaller size for the bins would highlight the discontinuity of the data and give a very jagged polygon.

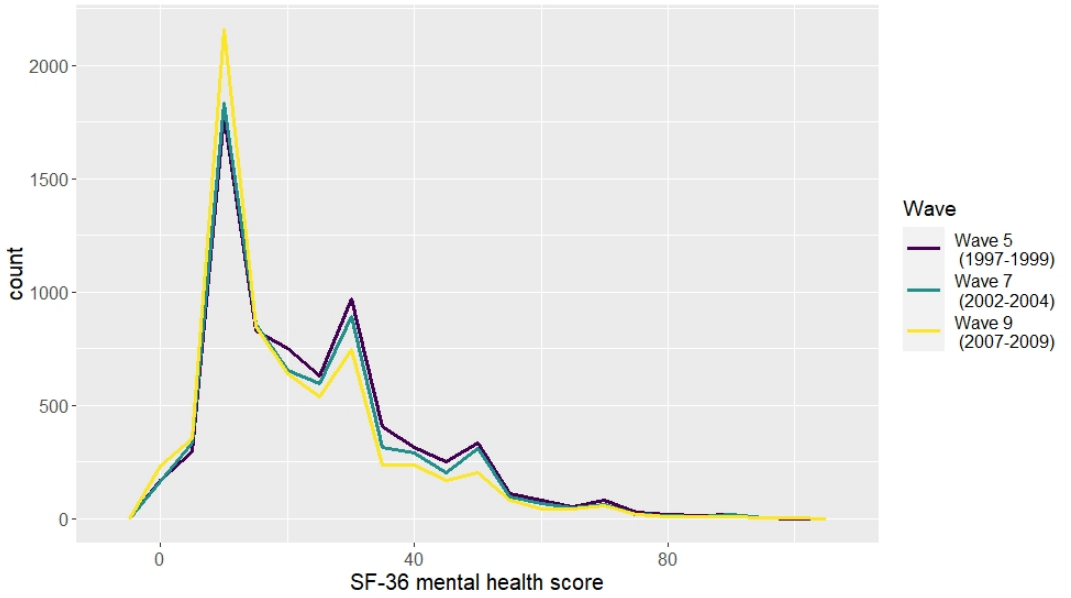

FIGURE 3.11: Frequency polygon for SF-36 mental health score in waves 5, 7 and 9

3.3.3 Indicators of Physical Function from the SF-36

The SF-36 is a measure of health-related quality of life (HRQoL) that was developed as part of the Medical outcomes study (Tarlov et al., 1989). The measure does not provide an overall score of HRQoL but instead provides a summary of eight different components of health: physical functioning, bodily pain, role limitations due to physical problems, role limitations due to emotional problems, emotional well-being, social functioning, energy and fatigue, and general health perceptions (Ware and Sherbourne, 1992). As well as the eight component scores there are also two summary scores associated with the SF-36: the Physical Component Score (PCS) and the Mental Component Score (MCS).

The SF-36 questionnaire has been included as part of the main Whitehall II questionnaire since wave 3. To investigate the distribution of physical function within the Whitehall sample I focused on three of the component scores from the SF-36: the physical function score, the role limitation due to physical problems score, and the bodily pain score. All sub-scale scores included in the Whitehall II dataset were calculated as described in the SF-36 User guide (Ware et al., 1993).

3.3.3.1 SF-36 Physical Function Score

The physical function score is designed to give an indication of how limited an individual is in doing various physical activities (Ware and Sherbourne, 1992). The score is calculated using 10 items from the SF-36 questionnaire, which ask participants about their ability to perform a wide range of physical activities, including vigorous activity, moderate activity, and a range of other activities associated with daily living, such as carrying groceries, walking up stairs and bathing oneself. Items are scored from one to three, with one representing being 'limited a lot' on that activity, and three representing not being limited at all.

If a participant had missing data on no more than half the items that made up the physical function subscale, the Whitehall II researchers replaced these missing values with the mean of the participant's other responses on the sub-scale's items. This process is known as mean imputation and is discussed further in Chapter 4 when I outline the my strategy for dealing with missing data in my own analysis of the Whitehall II data set. Note that this mean value may not have a value that corresponds to the categorical responses available for each item.

The sub-scale score is then calculated from a rescaled sum-score of the individual item responses as described in Ware et al. (1993). Having a high score on the SF-36 physical function sub-scale indicates that the participant is able to perform all types of physical activity without limitations due to health. Having a low score instead indicates that the participant is 'limited a lot' in their ability to perform all physical activities due to their physical health.

Figures 3.12 give frequency polygons of SF-36 physical function score in waves 5, 7 and 9. All three distributions are skewed towards higher values on the scale, indicating that a large proportion of the sample has little or no limitations to their physical function due to health concerns. Unsurprisingly, in wave 9 the peak of the distribution has shifted slightly to the left, indicating increased prevalence of limitations. These are in part likely to be caused by the ageing of the sample between these waves of data collection. Figure 3.13 shows this pattern via a series of box-plots. In Wave 5 the interquartile range for the data covers the very top end of the range of possible values, however, in waves 7 and 9 the median physical function scores and the interquartile range are shifted down. In wave 9 the lower quartile decreases further down the range.

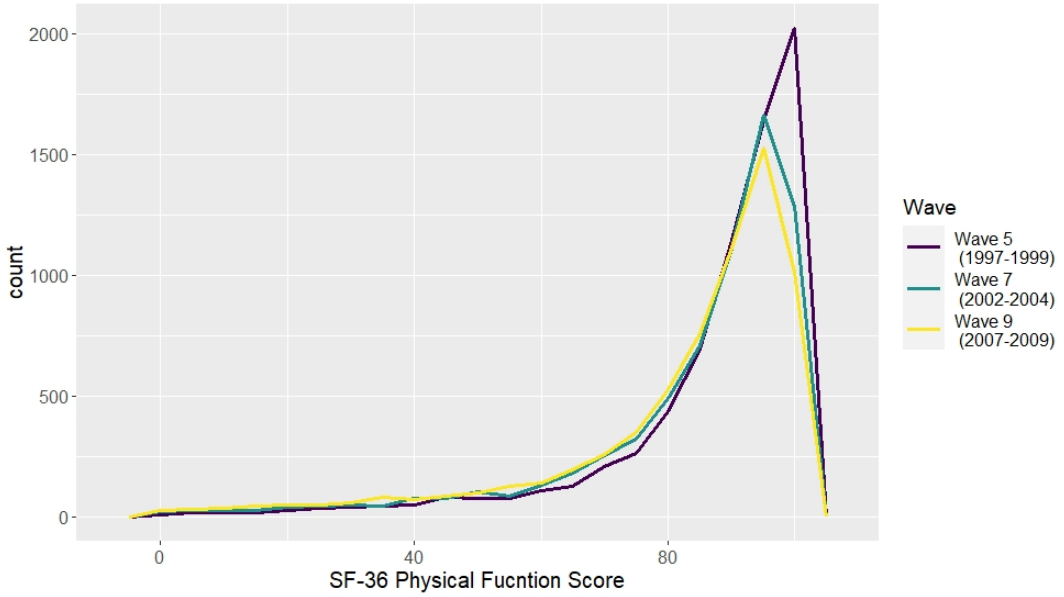

FIGURE 3.12: Distribution of SF-36 Physical Function scores within Waves 3 to 9

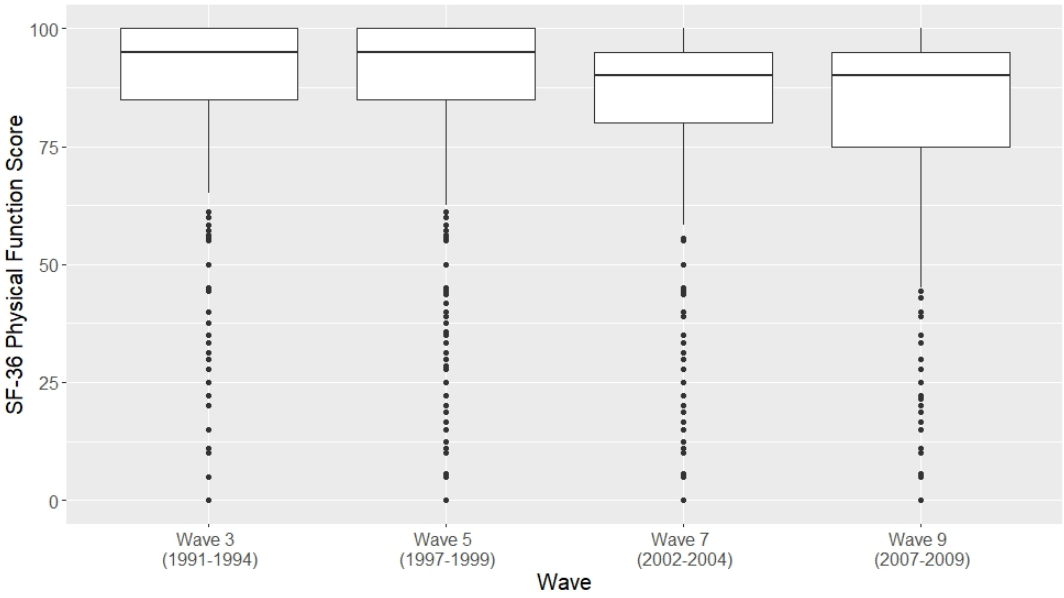

FIGURE 3.13: Boxplots of SF-36 Physical Function scores within waves five, seven and nine

3.3.3.2 Role Limitations due to Physical Problems Score

Similar to the physical function score, the role limitations due to physical problems sub-scale gives an indication of to what extent a participant's physical health causes them problems with work or other daily activities. It is calculated using four binary questions in which answering yes indicates that the participant experiences that kind of limitation due to their physical health. The score is given by the average number of 'no'-responses the participant gave, multiplied by 100 to set the scale between 0 and 100. Similarly to the physical function score, the Whitehall II researchers replaced missing item responses with the mean number of 'no'-responses in the participant's observed responses if no more than 50% of the responses on the subscale were missing.

A high role limitations score indicates that an individual has 'little to no' limitations to their daily activities due to physical health issues, whereas a low score represents multiple such limitations. Generally speaking, lower values on the subscale are taken to indicate more severe limitations due to physical problems. Aside from the cases where mean imputation was conducted, five different Role limitation scores are possible: 0, 25, 50, 75 and 100, which are equivalent to zero, one, two, three, or four 'no'-responses on the individual items respectively. When mean imputed values are present different scores between 0 and 100 are possible.

Figure 3.14 shows the distribution of SF-36 role limitation scores, with values from participants with mean imputed data removed. From waves 3 through to 9 there is a similar pattern within the distribution to that of the physical function score. In wave 3, the overwhelming majority of participants have no limitations to their role due to physical health issues, and small proportions of individuals make up each of the remaining categories. However, as time progresses, a greater proportion of the sample experience role limitations within all severity categories.

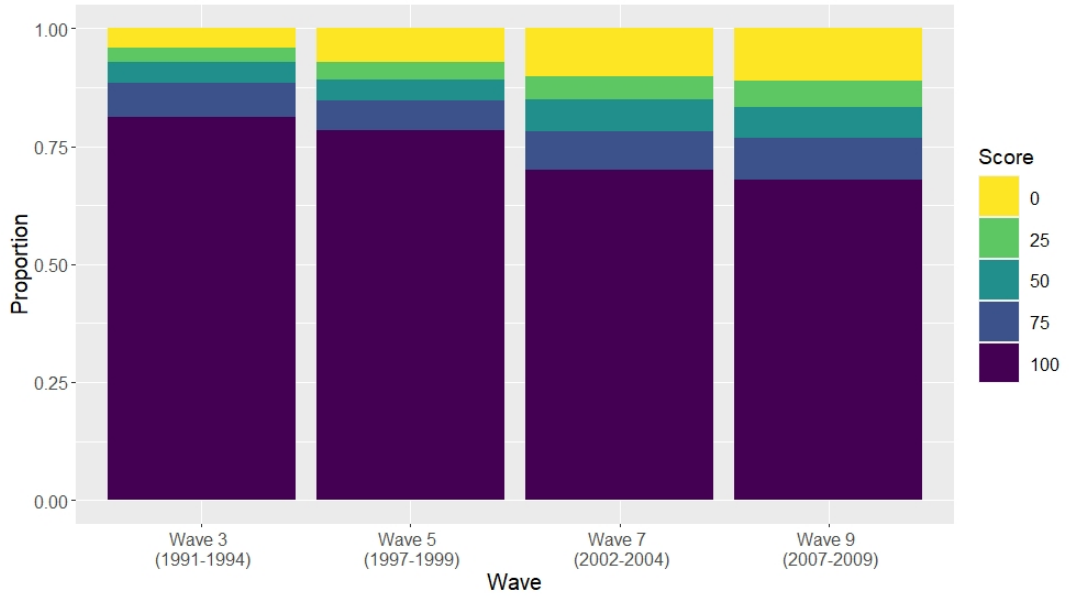

FIGURE 3.14: Distribution of SF-36 role limitation scores within waves 3 to 9

3.3.3.3 Bodily Pain Score

The SF-36 bodily pain score summarises two aspects of bodily pain that a participant might experience: extent of bodily pain, and how much this bodily pain interferes with normal work. For the first item on the extent of bodily pain, individuals were asked 'how much bodily pain have you had in the past four weeks' and were able to give one of six responses: 'none', 'very mild', 'mild', 'moderate', 'severe' or 'very severe'. For the second item on interference with normal work, participants were asked 'how much did pain interfere with your normal work in the past four weeks' and were able to give one of five responses: 'not at all', 'A little bit', 'moderately', 'quite a bit' or 'extremely'.

The bodily pain score is then calculated as suggested in the SF-36 user manual (Ware et al., 1993). Firstly, the item responses are precoded with a numerical value that corresponds to their severity. In particular, the first item is coded with integers from one to six, with one representing no bodily pain, and six representing severe bodily pain. Similarly, the second item is coded from one to five, with one representing no effect on work from pain, and five representing extreme effect. The item scores are then recoded as per Table 6.3 from the SF-36 user manual (Ware et al., 1993). After recoding, missing responses were mean imputed by the Whitehall II researchers in the case that only one

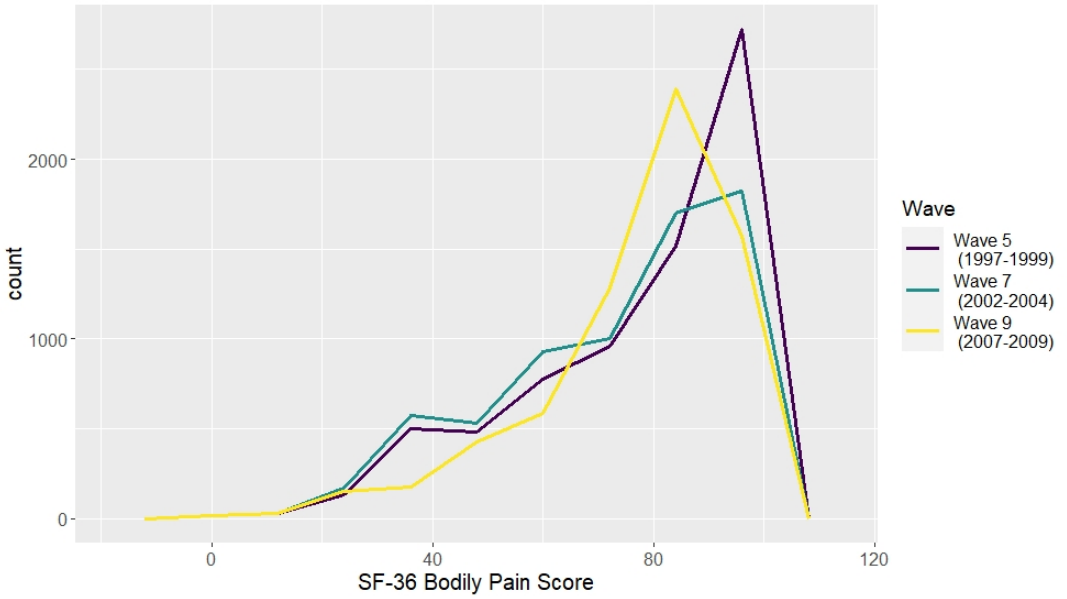

FIGURE 3.15: Frequency polygon of SF-36 bodily pain score within waves 3 to 9

item of the two is missing. Mean imputing values in this way effectively duplicates the single present score in place of the missing value. After imputing missing values, the scores on the two items are summed together and then rescaled so that the score runs from 0 to 100.

Figure 3.15 shows frequency polygons of SF-36 bodily pain scores in waves 5, 7 and 9. Similar to the other SF-36 scores, the distribution is heavily skewed towards higher scores, with lower bodily pain scores becoming increasingly less likely. There is also the same trend of more frequent low scores being found at later time points. This is again likely associated with the ageing of the sample participants.

As was the case with the depression measures, all three measures of physical function are not continuous, and have skewed distributions. As a result they are not Normally distributed.

3.3.4 Indicators of Physical Activity

In waves 5, 7 and 9 a subsection of the Whitehall II questionnaire is devoted to measuring participants' physical activity. Within this questionnaire are 20 items that assess the amount of time that participants spend doing the following activities: walking, sports (cycling, football (soccer), golf, swimming, and two open-ended questions on other sport), gardening (weeding/hoeing, mowing the lawn, and one open-ended question on other gardening),

housework (carrying heavy shopping, cooking, hanging out washing and two open-ended questions on other housework), DIY (painting/decorating, washing the car and one question on other DIY), and two further questions on other physical activity. These questions constitute a modified version of the Minnesota leisure-time physical activity questionnaire (Taylor et al., 1978; Sabia et al., 2012).

Taken on their own, individual items from this questionnaire are unlikely to provide a comprehensive picture of the participants' physical activity levels. As such, choosing only a handful of indicators to represent the physical activity construct in our model is likely to result in significant bias in the model. Conversely, using all of these items as individual indicators for our structural equation model will create an unwieldy model with either a difficult to interpret or multidimensional physical activity construct. As a result, three summary measures of physical activity volume have been derived from the raw item scores, and represent the participants' volumes of mild, moderate and physical activity.

3.3.4.1 Deriving the Measures of Mild, Moderate and Vigorous Physical Activity

An estimate of mild, moderate and vigorous activity was derived in the following way. First, I have estimated the number of hours per week that each participant spends performing each activity separately. For each activity (except for walking and cycling), participants were asked how many hours of each physical activity they partook in during the last four weeks. Respondents could choose one of seven possible responses: no hours, half an hour, one to one and a half hours, two to three hours, four to five hours, six to ten hours or 11 or more hours. For walking and cycling participants were asked how many minutes they spent walking and cycling on weekdays and weekends. Hence for walking and cycling, weekly hours were calculated using the equation 3.2

$$x_{week} = (5x_{weekday} + 2x_{weekend})/60 \tag{3.2}$$

where x_{week} denotes the weekly hours performed, $x_{weekday}$ denotes the minutes done on a weekday, and $x_{weekend}$ denotes the minutes done on a weekend, with $x \in \{walking, cycling\}$.

Response Category	Recoded Value
No hours	0
Half an hour	0.5
1-1.5 hours	1.25
2-3 hours	2.5
4-5 hours	4.5
6-10 hours	8
11 or more hours	11

TABLE 3.10: Possible responses categories and recoded estimate of duration for all activities in the Whitehall II Physical activity questionnaire except Walking and Cycling

For all the other activities, the estimate of weekly hours was calculated as follows. For each activity, the participants' responses were recoded so that they took a single value, rather than a range. In particular, I recoded the values to take the midpoint of their respective response category, except for responses in the final open-ended category which I recoded to 11. The full list of categories and their recoded values is shown in Table 3.10. Recoded response categories were then divided by four to give an estimate of weekly hours of activity.

Once I estimated the weekly hours performing an activity, this value was then multiplied by the metabolic equivalents (METs) associated with performing that activity to give an estimate of the volume of physical activity done per week in that activity. MET values for each activity were taken from the compendium of physical activity provided in (Ainsworth et al., 2011). All items were assigned a MET score based on a single entry from the PA compendium, except 'other sports' and cooking. The single entry was chosen to represent the activity if its description was deemed to most closely match the activity from the Whitehall questionnaire. 'Other sports' was assigned a MET value based on the average of the MET scores from all entries within the sports category on the PA compendium, excluding golf and football. Cooking was assigned a value based on the average of all the cooking items in the compendium. A detailed description of the MET values that were used for each activity is included in Table 3.11.

Once an estimate of physical activity volume has been derived for each activity separately, the activities were classified as mild, moderate and vigorous based on their associated MET values. Specifically, activities were classified as: mild if they had a MET value less than three; moderate if they had a

Activity	METs	Ainsworth (2011) (Code) and description
Walking	2.8	(17151) Walking 2mph, level, slow pace, firm surface
Cycling	7.5	(01015) Bicycling, general
Football	7.0	(15610) Soccer, casual, general
Golf	4.8	(15255) Golf, general
Swimming	6.0	(18310) Swimming, leisurely, not lap swimming, general
Other Sports	6.3	Average of Sports excluding golf and football
Weeding/Hoeing	5	(08241) Weeding, cultivating garden, using a hoe, moderate to vigorous effort
Lawn mowing	5.5	(08095) Mowing lawn, general
Other Gardening	3.8	(08245) Gardening, general, moderate effort
Carrying heavy shopping	2.5	(05056) putting away groceries (e.g carrying groceries, shopping without a grocery cart), carrying packages
Cooking	2.625	Average of cooking items (05049 - 05052)
Hanging Washing	2.0	(05090) Laundry, fold/hang clothes, implied standing, light effort
Other Housework	3.5	(05026) Multiple household tasks at once, moderate effort
Car Wash	2	(06225) Washing and Waxing car
Painting	4.5	(06165) painting, (Taylor Code 630)
Other DIY	4.5	(06127) home repair, general, moderate effort

TABLE 3.11: Table of Activities in Whitehall II questionnaire and their associated intensities

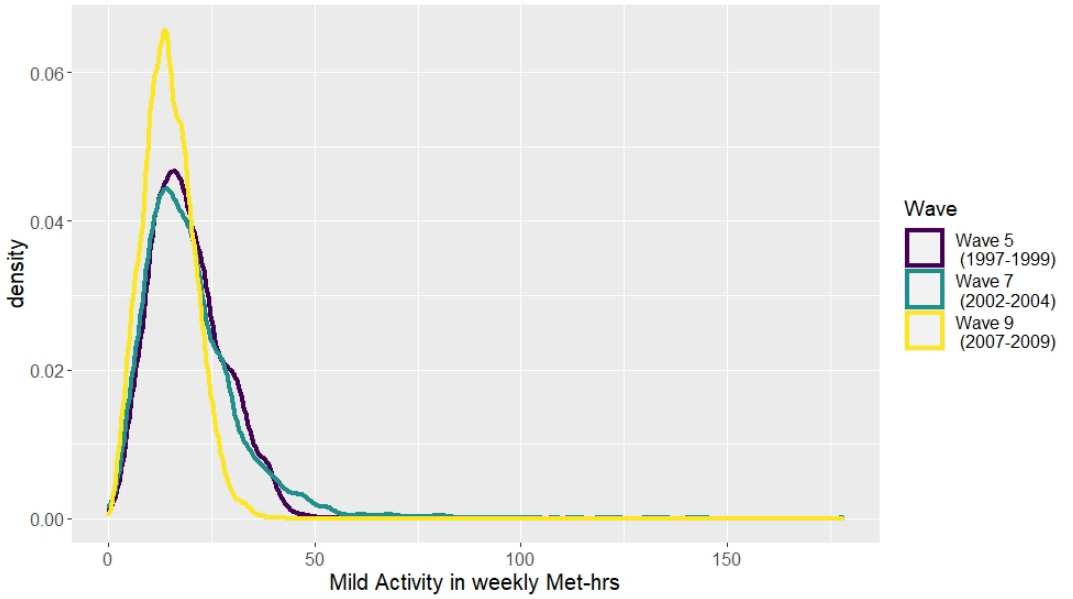

FIGURE 3.16: Distribution of mild activity volume in waves 5, 7 and 9

MET value greater than or equal to three but less than six; and vigorous if they had a MET value greater than or equal to six. Once activities had been classified, estimates of mild, moderate and vigorous activity were derived by summing together the physical activity volumes for activities within the same category. In this way, mild activity is the total volume of activity from walking, washing the car, carrying shopping, cooking and hanging washing. Moderate activity is the total volume of activity from golf, weeding, painting, mowing the lawn, general gardening, general housework and general DIY. Vigorous activity is then the total volume of exercise from football (soccer), swimming, cycling, and general sport.

3.3.4.2 Volume of Mild Activity

Figure 3.16 shows the distribution of mild activity volume in waves 5, 7 and 9. The distributions are largely similar in shape, however, in wave 9 there is a greater density concentrated around lower values of mild activity volume compared to waves 5 and 7. This is perhaps due to natural declines in physical activity associated with older age. Another difference between the distributions is that there is a much longer tail in the distribution of mild activity volume in wave 7. This is caused by a handful of individuals doing large amounts of walking. Besides these outliers, the distributions appear to mostly follow the bell shape of the Normal distribution. However, in both wave 5 and wave 7 the is a slight right skew.

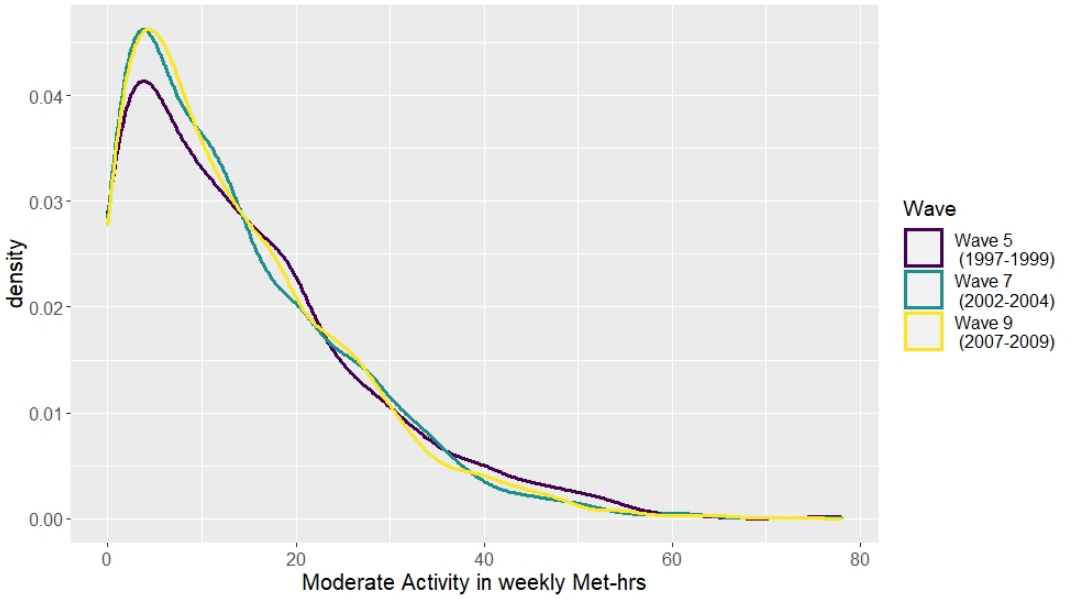

FIGURE 3.17: Distribution of moderate activity volume in waves 5, 7 and 9

3.3.4.3 Volume of Moderate Activity

Figure 3.17 shows the distributions of moderate physical activity in waves 5, 7 and 9. The distributions are all very similar, and are more skewed than is the case for mild activity volume, suggesting that higher volumes of moderate activity are increasingly less common in the sample. Unlike the distributions for mild activity, volume, the distributions for moderate activity are not normally distributed.

3.3.4.4 Volume of Vigorous Activity

Figure 3.18 shows the distributions of vigorous activity volume in waves 5, 7 and 9. The distributions are severely skewed, and as such are non-Normal. All three distributions have the majority of their mass concentrated on low values of vigorous activity volume which suggests that the vast majority of the sample does next to no vigorous physical activity.

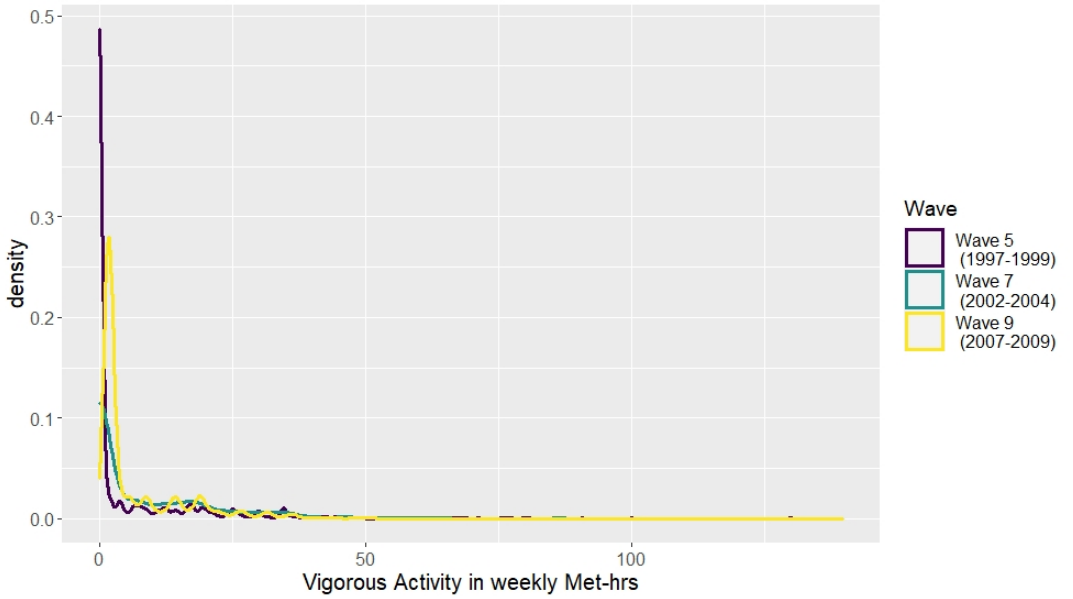

FIGURE 3.18: Distribution of vigorous activity volume in waves 5, 7 and 9

3.3.5 Indicators of Diet

The vast majority of the diet data in Whitehall II is collected in the form of a food frequency questionnaire (FFQ). This questionnaire asks about how often the participants eat food from a number of different groups including bread, dairy, meat, fish, fruit, vegetables, pasta, sweets and snacks as well as questions on drink consumption and the use of salt, fats and sauces. Using only a single item of food within a group as an indicator of the consumption of all food in that group within the model under utilises the vast amounts of data collected within the questionnaire. Furthermore, single foods are unlikely to be representative of a participant's consumption of foods in the same group, as consumption of foods of the same group is likely to vary within an individual's responses. In light of this, I have estimated the daily consumption frequency for each food group in the respondent's diet.

For each food, participants are asked to select from one of nine different options to indicate how many medium sized portions of the food they consumed, on average, over the last month. The options and their corresponding response categories are listed in Table 3.12. Even though the question asks for a frequency of medium sized portions that were consumed, it important to note that the derived measure I have calculated is focused purely on consumption frequency, rather than being interpreted as a number of portions. This is because there is often no objective description of what constitutes a

medium portion of a food, and hence what participants considered a portion size may vary considerably across the sample.

Response Category	Consumption Frequency
1	Never/ less than once a month
2	1-3 per month
3	Once per week
4	2-4 per week
5	5-6 per week
6	Once per day
7	2-3 per day
8	4-5 per day
9	6+ per day

TABLE 3.12: Whitehall II FFQ question response categories and meanings

In order to estimate the daily consumption frequency for a food, I have converted response categories that use either a monthly or weekly time scale into an equivalent daily frequency. To do this, I have taken the length of a month as 30 days so that monthly frequencies can be divided by 30 to obtain daily frequencies. Table 3.13 shows the response categories and their associated consumption frequency when converted to a daily time-scale.

Response Category	Consumption Frequency
1	Never - less than 1/30 a day
2	1/30 - 1/10 per day
3	1/7 per day
4	2/7 - 4/7 per day
5	5/7 - 6/7 per day
6	Once per day
7	2-3 per day
8	4-5 per day
9	6+ per day

TABLE 3.13: Whitehall II FFQ response categories and rescaled meanings

Having created these categories, a single reference point in each category is needed so that daily consumption frequency may be summed across different foods within the same food group to estimate the daily consumption of the entire food group. For response categories one to eight I have chosen the midpoint of the daily range associated with the response category, and for category nine I have chosen the reference point of 6 per day. The daily

estimate for each response category is shown in Table 3.14. I then calculated daily estimates for a food group by adding up an individual's estimated daily consumption for each food in that group. For example, the estimated daily consumption frequency of vegetables consumed was calculated by summing the individual responses to consumption of different vegetable items.

Response Category	Consumption Frequency
1	2/60 a day
2	2/30 per day
3	1/7 per day
4	3/7 per day
5	11/14 per day
6	Once per day
7	2.5 per day
8	4.5 per day
9	6 per day

TABLE 3.14: Whitehall II FFQ response categories and corresponding daily consumption frequency estimate

3.3.5.1 Daily Consumption Frequency of Vegetables

The first food group that I estimated the daily consumption frequency for was vegetables. The Whitehall II FFQ questionnaire contains 24 questions on the consumption of vegetables and legumes. The items covered the following foods: baked beans, other beans, broccoli, cabbage, carrots, cauliflower, coleslaw, dried lentils (and similar), garlic, green salad, leeks, marrows and courgettes, mushrooms, onions, parsnips (and similar), nuts, peas, soya meat, spinach, spring greens, sweet peppers, soy (such as tofu), tomatoes and vegetable soup. All of these items, except for nuts were used to estimate the daily consumption frequency of vegetables. Nuts were excluded due to their high fat content, which may have resulted in confounding with the measure of fat intake that is introduced later in this section.

The distributions of vegetable consumption frequency in waves 5, 7 and 9 are displayed in Figure 3.19. Inspection of these distributions suggests that the overall pattern of vegetable consumption does not change much across the three waves. Furthermore, it also suggests that on average, individuals are consuming vegetables around four times per day with the majority of participants eating vegetables between zero and ten times per day. Besides the long tail to the right the distribution also appears to be approximately normally distributed.

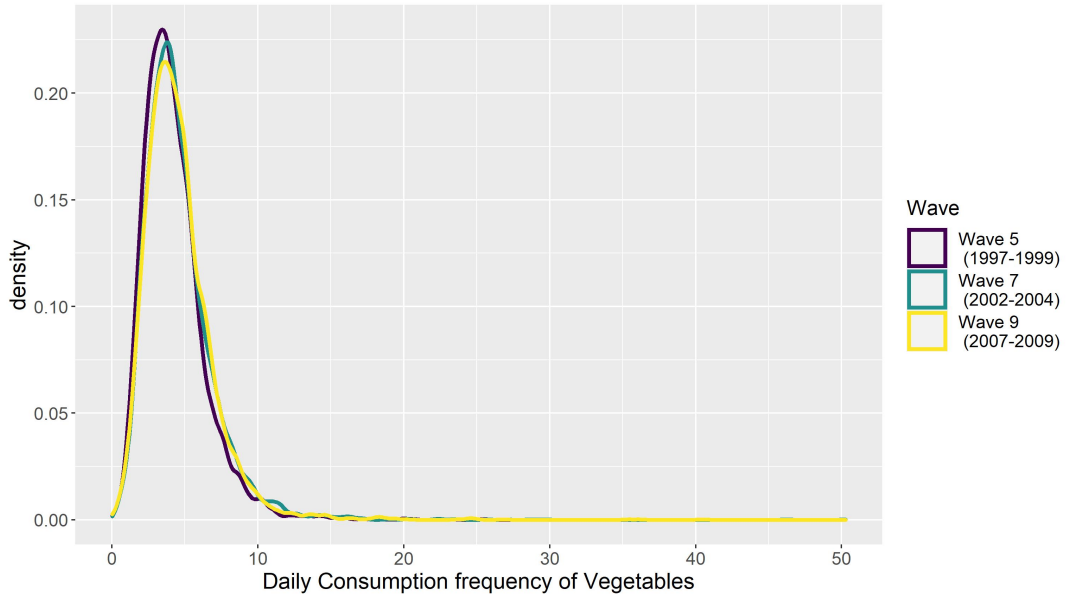

FIGURE 3.19: The distribution of daily vegetable consumption frequency in waves 5, 7 and 9

3.3.5.2 Daily Consumption Frequency of Fruit

The Whitehall II FFQ questionnaire contains 11 questions on the consumption frequency of fruits. The items covered the following fruits: apples, bananas, grapefruit, grapes, oranges, melon, peaches, pears, strawberries, tinned fruit and dried fruit. All of the above items were used to create the estimate of daily consumption frequency of fruit. Figure 3.20 shows the distributions of estimated daily consumption frequency of fruit in waves 5, 7 and 9. The majority of respondents eat fruit between zero and five times per day, with there being a small tail in the distribution out to higher numbers of daily portions. The mean consumption frequency for fruit is approximately two per day. Again there is a long tail out to the right of the distribution, however, from observing the overall shape of the distributions, departures from normality do not appear severe.

3.3.5.3 An Alternative Indicator of Fruit and Veg Intake

Outside of Whitehall II's food frequency questionnaire, data is collected within the main questionnaire on the frequency with which participants consume fruit and veg. Participants are asked how often they consume fresh fruit or veg, and are given eight options to choose from: "Seldom", "Less than once a month", "One to Three times a month", "Once or Twice a week", "Three to four times a week", "Five to six times a week", "Daily", Two or more times a day". Figure 3.21 shows how the responses to this question are distributed

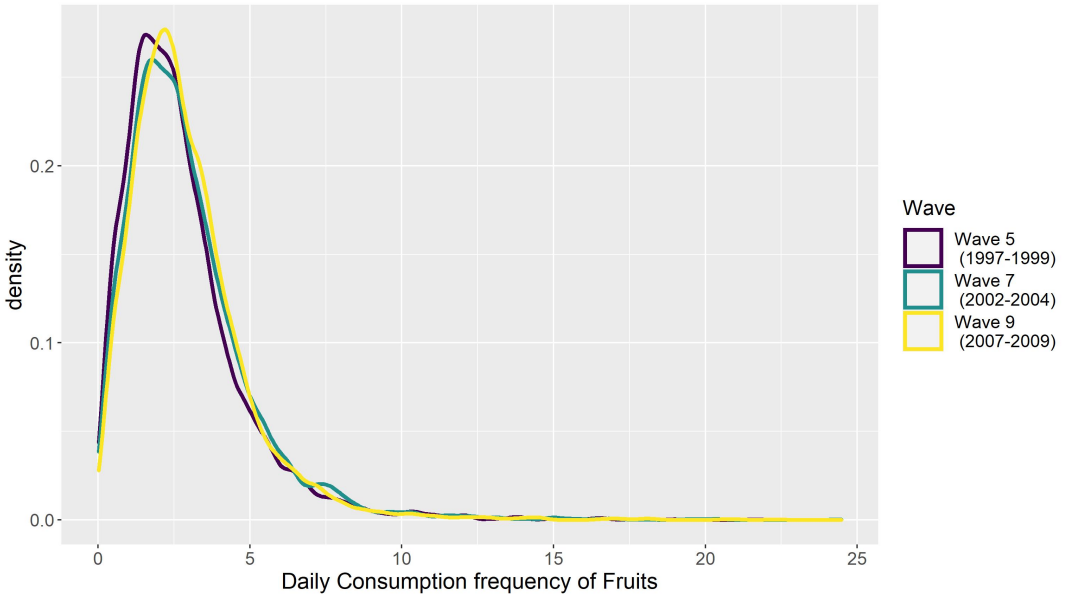

FIGURE 3.20: The distribution of daily fruit consumption frequency in waves 5, 7 and 9

across the different options in waves 3, 5, 7 and 9. In all waves of the sample more than half of the participants report eating only one piece of fresh fruit or vegetables at most each day. As such, this would indicate that in all waves, the majority of participants are not meeting recommended intake for fruit and vegetables.

This data provides a contrary picture to the data provided in our estimates of daily consumption based on the FFQ data, where it appeared that around half of the sample were consuming more than four vegetables per day and more than two fruits per day. From this we know that at least one of these measures is providing an inaccurate reading of participants' fruit and vegetable intake.

Estimates of daily consumption derived from the FFQ may be over estimating the dietary consumption of fruit and vegetables. Health Survey for England data suggests that English adults on average consumed approximately three and a half to four portions of fruit and vegetables per day in the years that this Whitehall data was collected (NHS Digital, 2020a). The FFQ data however suggests that the sample members on average consume fruit and vegetables around six times per day. Individuals may be influenced by the social desirability to follow a healthy diet and over report their fruit and vegetable consumption. It is, however, also possible that our sample of civil servants is systematically more likely to engage in healthy eating habits than is the general population. In the United Kingdom, higher SEP is associated

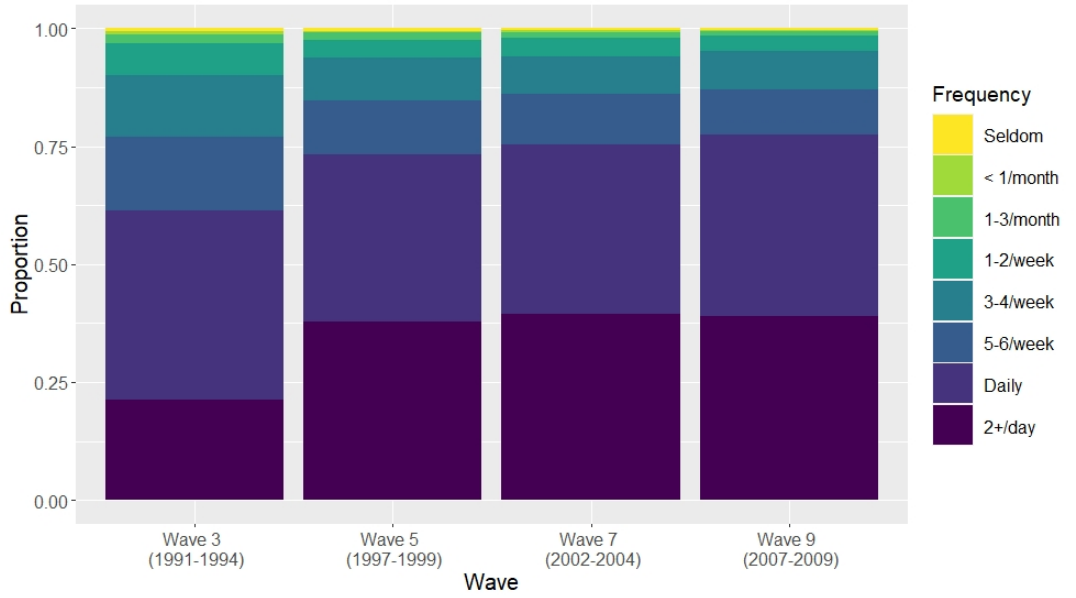

FIGURE 3.21: The distribution of fruit and vegetable consumption frequency within waves 3, 5, 7 and 9

with increased consumption of fruit and vegetables (Maguire and Monsivais, 2015). On average, one would expect that a sample of civil servants from London would have a higher average SEP than a general population sample and so this may explain some of the increase in reported fruit and vegetable intake.

Estimates from the single item alternative measure may be underestimating the amount of fruit and vegetables consumed; the proportion of the sample who report eating 2 or more pieces of fruit a day is similar to the proportion of the sample in the health survey for England who report eating five portions a day (NHS Digital, 2020a)). Recalling average behaviour is a cognitively expensive task and so participants may base their answer on an inaccurate heuristic recollection of their true diet. For example, participants may only recall fruit and veg eaten at meal times, or over the last few days, both of which might underestimate true consumption (Naska, Lagiou and Lagiou, 2017).

3.3.5.4 Daily Consumption Frequency of Sugary Foods

Another food group that I have derived an indicator for the consumption of is sugary food. In the FFQ there are 14 items within the sweets and snack group: biscuits, buns and pastries, cakes, chocolate, cocoa and hot chocolate, crackers, crisps, tarts (and similar), ice cream, jam (and similar), milk

puddings, sponges, sugar in hot drinks, and sweets. Of these, all items except for crisps and crackers were included in our derived measure of sugar consumption frequency. Note that this indicator does not describe total consumption of sugar, since other foods which contain sugar from other groups of diet questions in the FFQ are not included. For example, fruits contain high amounts of sugar but are not included in this measure.

Figure 3.22 shows the distribution of daily sugar consumption frequency in wave 5. Participants most commonly have a reported sugar consumption frequency of between 0.5 and 1.5, with higher consumption frequencies becoming increasingly less common in the sample. This would indicate that it is most common for participants to only consume a small number of sugary foods per day on average. However, there are many individuals who consume higher amounts of sugary snacks: over half of the sample have a reported sugar consumption frequency greater than 2.5 per day. Figure 3.23 shows the distribution of sugar consumption frequency in waves 5, 7 and 9. All three distributions are very similar which is perhaps an indication of persistent sugar consumption habits over time. Departures from Normality are more severe than is the case for fruit and vegetable intake with the main mass of the density having a less symmetrical shape. There is also a right skew in the distributions providing a further departure from Normality.

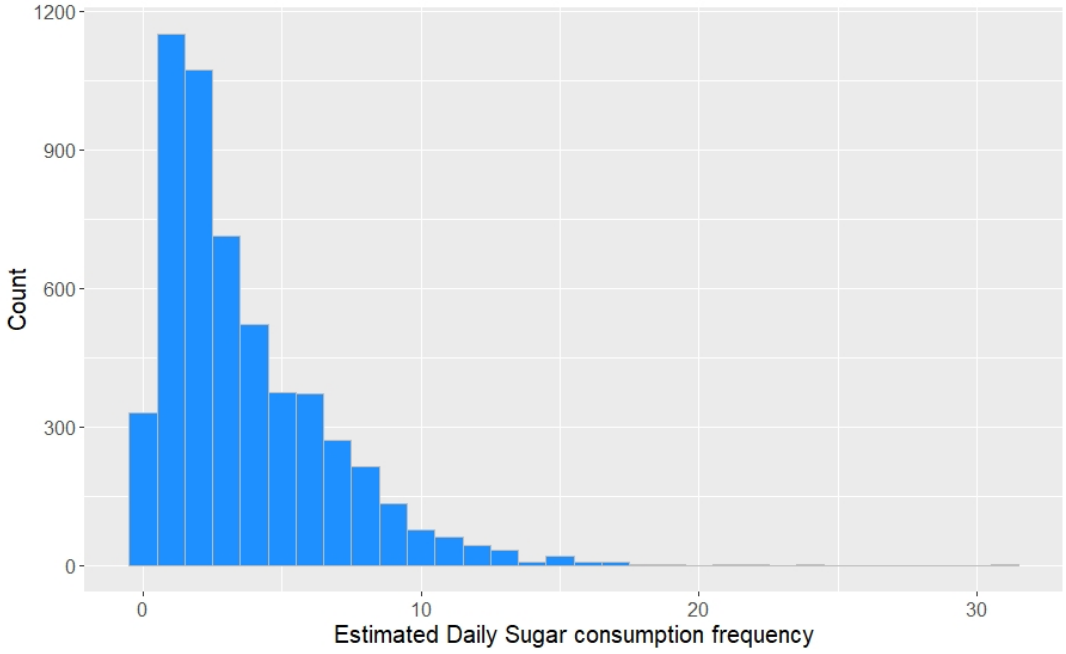

FIGURE 3.22: The distribution of daily sugar consumption frequency in wave 5

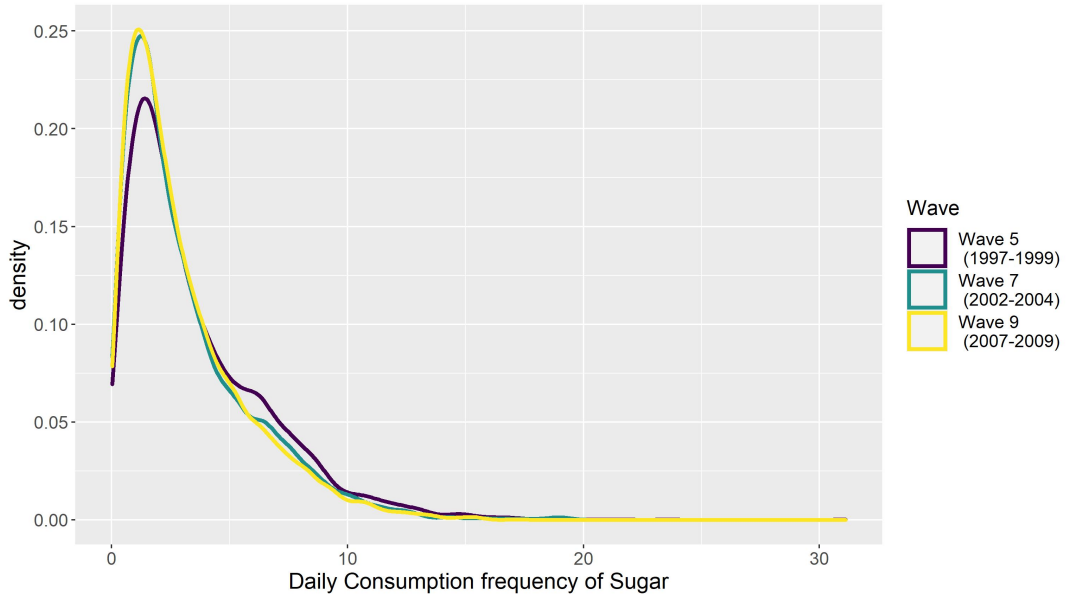

FIGURE 3.23: The distribution of daily sugar consumption frequency in waves 5, 7 and 9

3.3.5.5 Daily Consumption Frequency of Unhealthy Fat

A measure of unhealthy fat consumption was derived from four items from the fat section of the FFQ questionnaire. These items asked how many times on average the participants ate: butter, hard margarine, visible fat on food, or fried food. These items were chosen due to the fact that they contain high amounts of saturated and trans fat, which have been suggested to be more harmful to health than other types of fat, and than fat intake overall (Liu et al., 2017a).

Figure 3.24 shows the distribution of daily consumption frequency for unhealthy fat in waves 5, 7 and 9. The vast majority of participants in the sample have a low reported daily consumption frequency with higher consumption frequency becoming decreasingly common except for a spike in density just above a consumption frequency of 2.5. It is clear to see that none of the distributions for unhealthy fat consumption follow a normal distribution closely: there are multiple peaks in the distribution and the distribution reflects a high frequency of low values of consumption.

These multiple peaks are likely an artefact of how the measure was calculated. On inspection of Figure 3.24, it is apparent that the peaks of the distribution cluster around the consumption frequency categories from Table 3.14. For example there is peak just above 2.5 per day and another smaller bump just above 4.5 per day. As was noted before, the bulk of responses

on fat intake suggest participants are infrequently consuming unhealthy fat. As a result, we can expect that the majority of participants responses indicated low consumption frequencies on the individual items. If a participant has a higher consumption frequency on a single item, but low consumption frequency on the remaining items, then their overall score will sit just above the consumption frequency of the high consumption item, as this value will contribute most heavily to their overall score. This pattern of responses, in which participants have either reported low consumption on all items, or high consumption on a single item and low consumption on the rest, is likely responsible for the observed shape of the distribution.

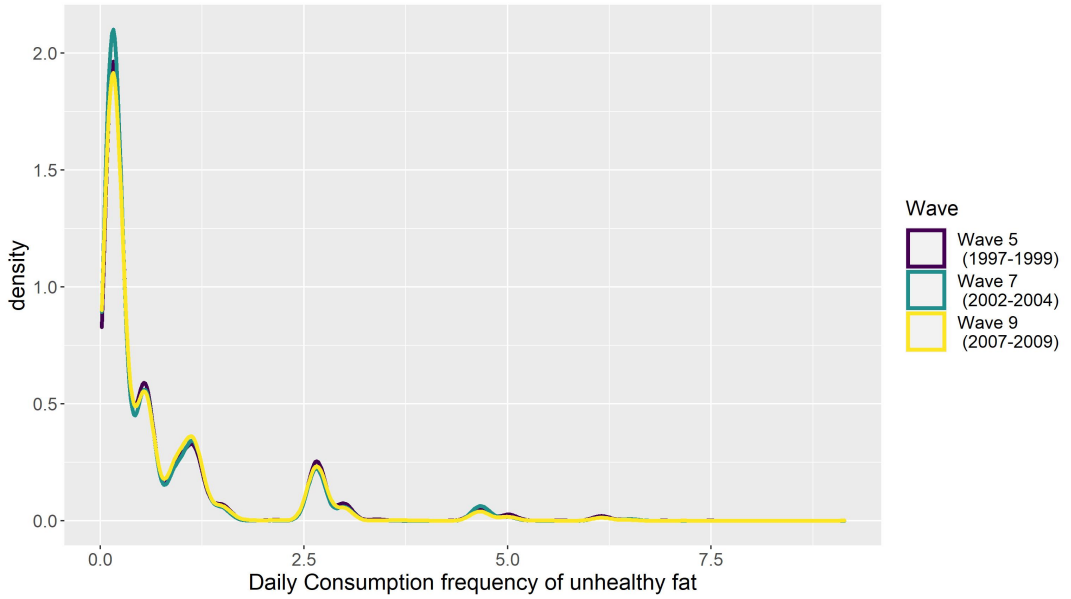

FIGURE 3.24: The distribution of daily unhealthy fat consumption frequency in waves 5, 7 and 9

3.3.5.6 Daily Consumption Frequency of Processed Meat

The last indicator that has been derived from the FFQ is an indicator of processed meat consumption frequency. This measure has been derived from five items that ask how frequently participants ate the following food stuffs: bacon, corned beef or luncheon meat, ham, sausages, and savoury pies. Figure 3.25 shows the distribution of consumption frequency for processed meat. Similar to the distributions of sugar consumption and fat consumption, low consumption frequency is the most common response in the sample, with higher consumption frequency becoming decreasingly likely. Again the shape of the distributions is consistent across the waves of data suggesting that the patten of dietary intake has not changed within the sample over time. Lastly,

the distributions are again not normally distributed although the departures from normality are less severe than is the case for sugar intake.

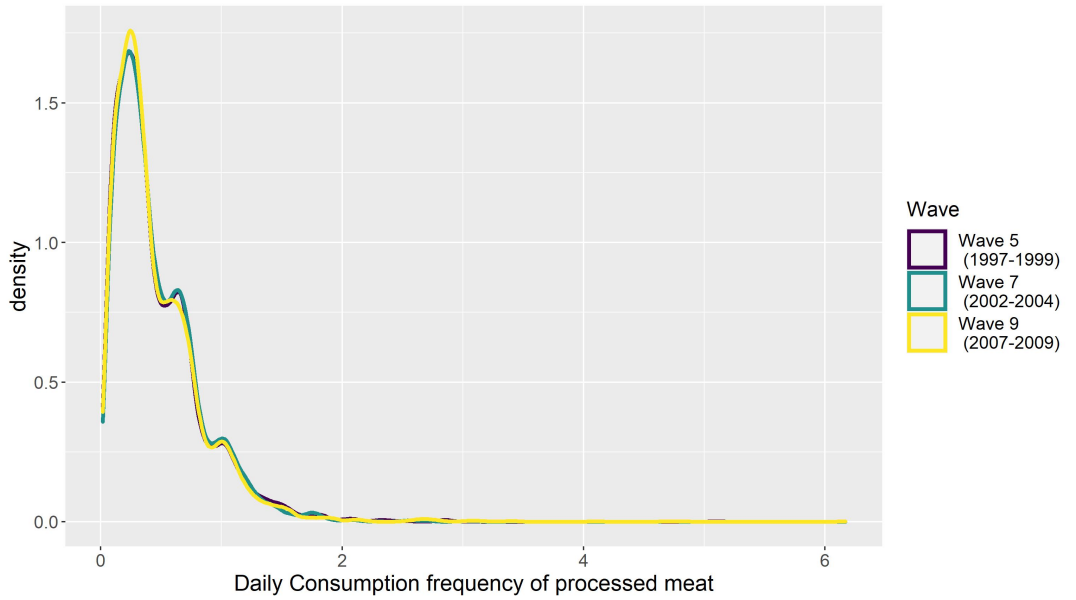

FIGURE 3.25: The distribution of daily processed meat consumption frequency in waves 5, 7 and 9

3.3.6 Indicators of Socioeconomic Position

The Whitehall II data set contains a range of measures associated with socioeconomic position that cover areas such as education status, occupational status, family problems and perceived socioeconomic position. In our model I have used three measures as indicators of socioeconomic position: two measures of family problems and one measure of perceived socioeconomic position. I did not include measures of occupational status for this analysis and the final model because they had very high proportions of missing data. I also did not examine educational status as it does not change over time, and therefore would be uninformative for a time-dependent measure of SEP.

3.3.6.1 Frequency of Problems with Money

The first measure of socioeconomic position I have examined measures the frequency with which participants experience problems with money. Specifically, participants were asked how often they did not have enough money to afford the kind of food and clothing they feel their family should have. Respondents could give one of five different answers: 'always', 'often', 'sometimes', 'seldom' or 'never'. Table 3.15 shows the number of individuals that responded in each category in waves 5, 7 and 9. Figure 3.26 then shows the

	Wave 5 (1997-1999)	Wave 7 (2002-2004)	Wave 9 (2007-2009)
Problem Frequency *N* **(%)**			
Always	72 (1.02%)	66 (0.98%)	54 (0.82%)
Often	159 (2.25%)	111 (1.65%)	63 (0.96%)
Sometimes	711 (10.05%)	426 (6.31%)	364 (5.54%)
Seldom	1,752 (24.77%)	1,185 (17.57%)	960 (14.61%)
Never	4,380 (61.92%)	4,958 (73.50%)	5,131 (78.07%)

TABLE 3.15: Distribution of responses to question on frequency of problems with money

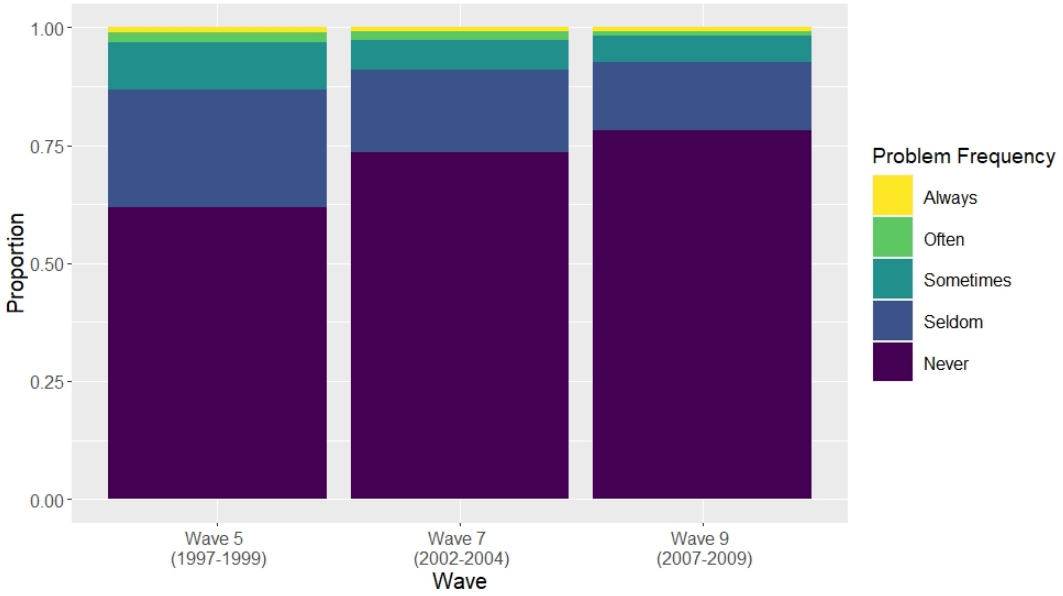

FIGURE 3.26: Distribution of Money Problems Responses

proportion of participants that responded in each category. Of note is that the majority of participants report that they never have money problems when providing food and clothing for their family and that this majority increases over time. In addition, more frequent money problems are increasingly uncommon over time.

3.3.6.2 Frequency of Problems Paying Bills

The second indicator of socioeconomic position measures another aspect of financial problems that might occur in participants' family lives. Participants were asked how much difficulty they have paying bills. Again respondents could give one of five options: very great, great, some, slight or very little. Table 3.16 shows the number of participants who responded in each of these

	Wave 5 (1997-1999)	Wave 7 (2002-2004)	Wave 9 (2007-2009)
Problem Frequency *N* **(%)**			
Very great	32 (0.45%)	12 (0.18%)	17 (0.26%)
Great	71 (1.00%)	59 (0.88%)	39 (0.60%)
Some	535 (7.57%)	425 (6.32%)	351 (5.36%)
Slight	900 (12.74%)	785 (11.68%)	655 (10.00%)
Very little	5,528 (78.23%)	5,442 (80.95%)	5,489 (83.79%)

TABLE 3.16: Distribution of responses to question on frequency of problems paying bills

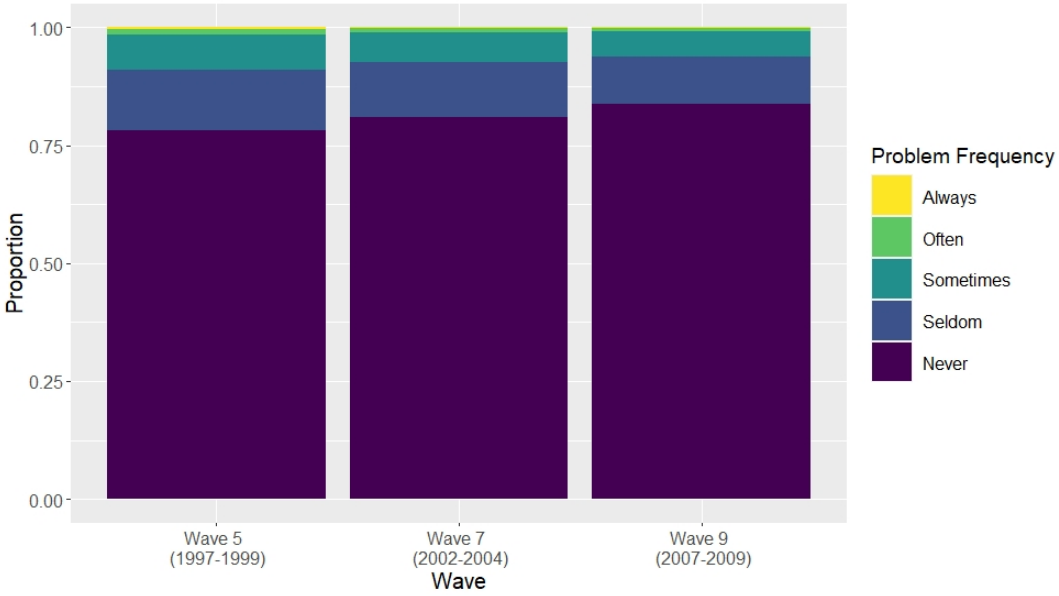

FIGURE 3.27: Distribution of Bills problems responses

categories in waves 5, 7 and 9 and 3.27 displays the proportion of participants who responded in each category, again in waves 5, 7 and 9. The pattern of responses is very similar to the pattern seen in the frequency of money problems. The vast majority of respondents indicate that participants have very few problems paying bills and as with money problems, more severe issues are increasingly infrequent over time. The pattern of responses here is also more skewed towards very little problems than is the case with the money problems variable.

3.3.6.3 Perceived Position on the Societal Ladder

The final indicator of socioeconomic position gives an indication of each respondent's perceived position within society. Participants were asked to

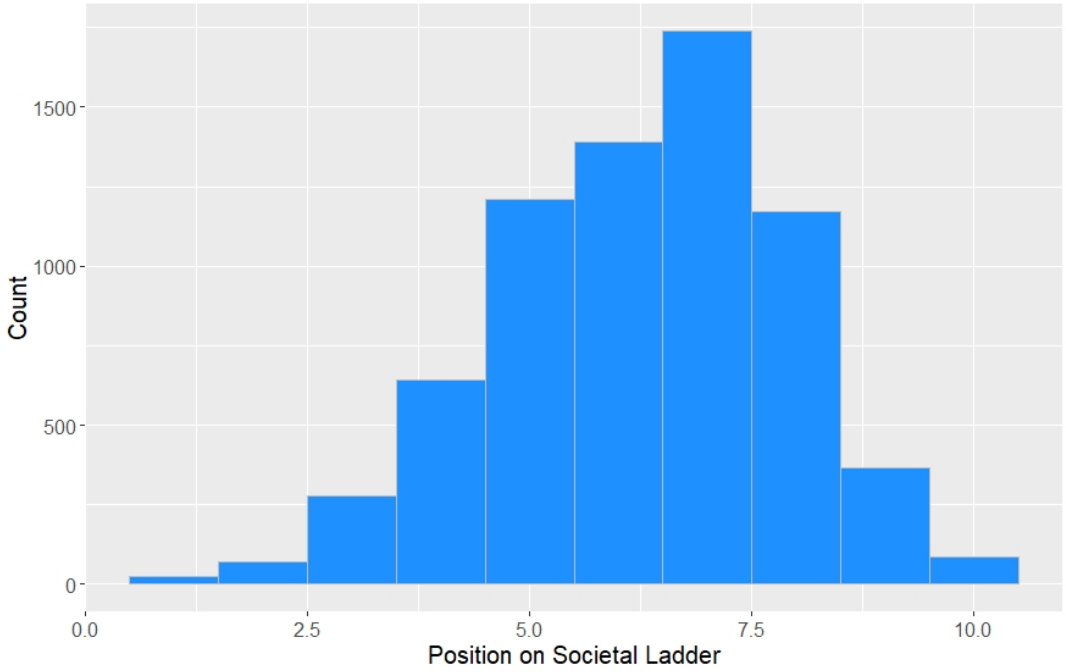

FIGURE 3.28: Distribution of perceived social ladder position in wave 5 (1997-1999)

mark on a ladder where they felt they were in society, with the top of the ladder indicating those best off in society, and the bottom of the ladder representing those who are worst off. Responses range from 0.5 to 10.5, with 0.5 representing a participant placing themselves below the bottom rung (on the ground) and 10.5 representing an individual placing themselves above the top rung. In later waves individuals were instructed to only place themselves on an exact rung, reducing the number of different response possibilities, and reducing the range of values to integers between 1 and 10.

Figure 3.28 shows the distribution of perceived societal position in wave 5. The majority of respondents consider themselves to be within the top half of society. Participants rarely viewed themselves as being at the very top of society, but instead were more likely to put themselves between halfway and four fifths of the way up the ladder. Very few individuals consider themselves in the bottom of society, with fewer and fewer participants placing themselves on lower rungs of the societal ladder.

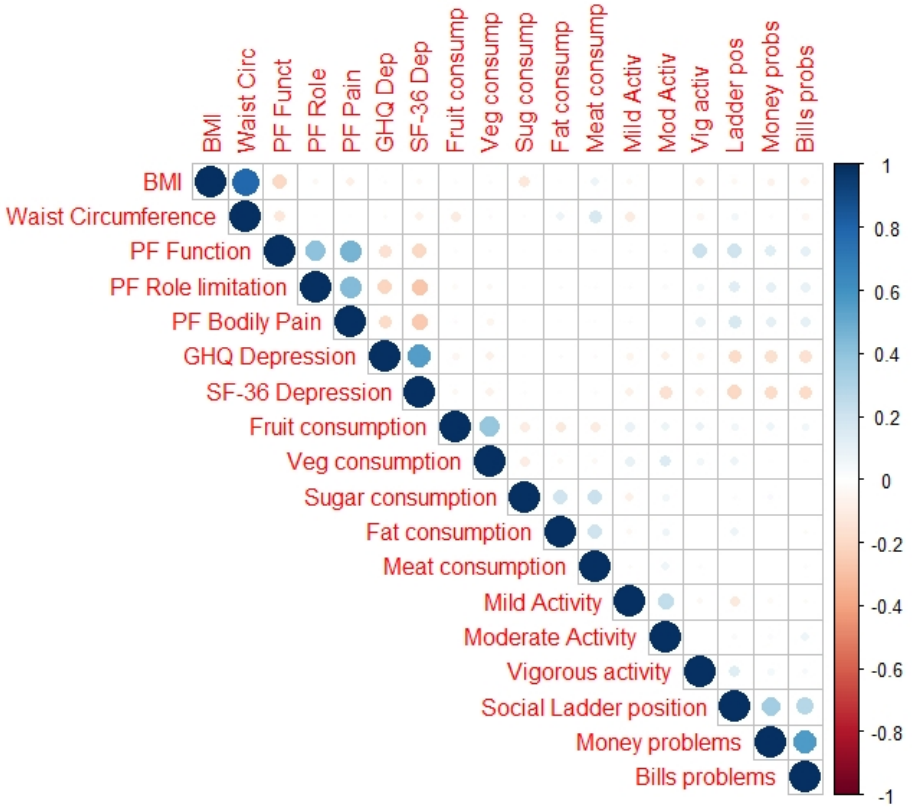

FIGURE 3.29: Correlation plot for measures in wave 5 (1997-1999)

3.4 Correlations Between Indicators of Different Constructs

In this section I will analyse how indicators of the different constructs examined in the previous section correlate with one another. This will serve to give some initial insight into the relationships that might exist between the different constructs being investigated. When discussing the correlations between variables I have followed the cut-off descriptions suggested by Cohen (1988). As such, correlations less than 0.1 in magnitude are described as negligible, between 0.1 and 0.3 in magnitude as small, between 0.3 and 0.5 as medium and above 0.5 as large. In calculating the correlations, missing data was dealt with using pair-wise deletion, which calculates each correlation using all cases for which data was observed on both of the variables being correlated.

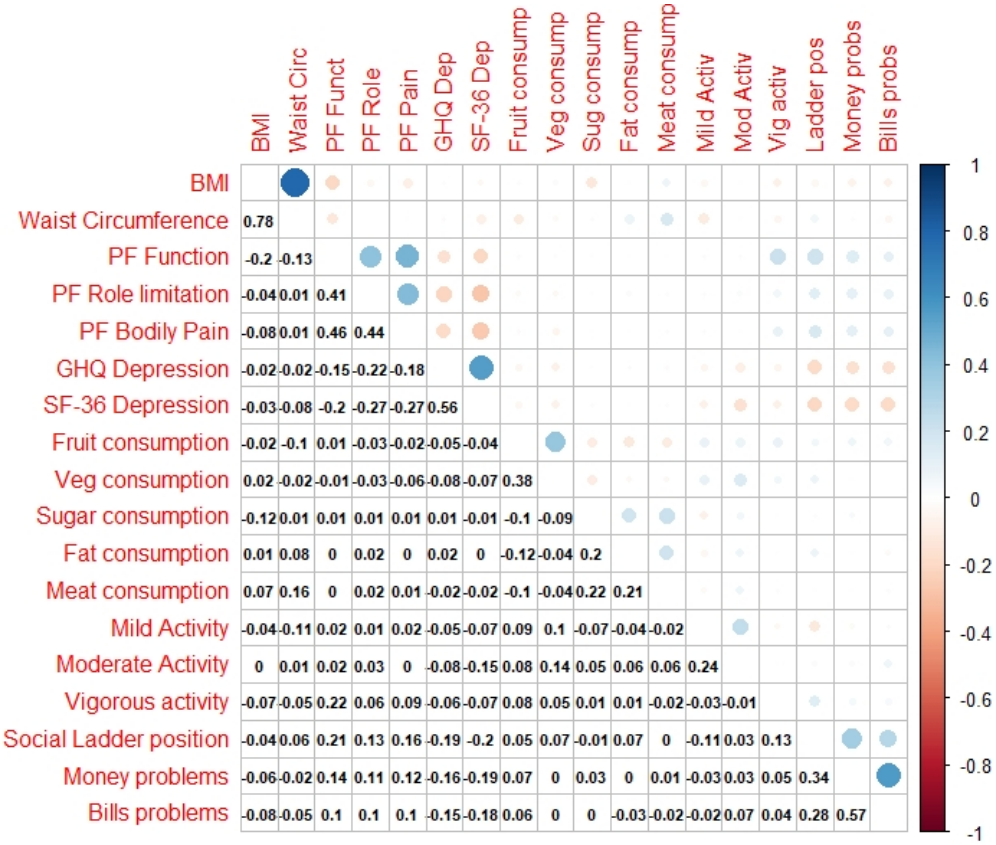

FIGURE 3.30: Correlation plot for measures in wave 5 (1997-1999)

3.4.1 Cross-sectional Correlations between Measures

When selecting measures to be indicators of the same construct, it is important that the measures share some common source of variance, as it is this common source of variance that we are explaining using a latent factor. Having a common source of variance due to an underlying latent construct implies that there will be a non-zero correlation between such measures when we do not account for this latent part of our model. Hence, if measures do not correlate with one another, then they cannot be measuring the same underlying latent construct. This is specifically the case when the relationship between construct and measures is specified as "construct *causes* measure".

With the above in mind, Figure 3.29 suggests that two of the constructs included in the model are not well indicated by their measures: diet quality and exercise. In terms of diet, whilst fruit and vegetable consumption correlate with one another, sugar consumption has only a small correlation with vegetable consumption, and only correlates negligibly with fruit consumption, although this correlation is only just below the cut-off value 0.1 for being considered a 'small' correlation.

Similarly, processed meat consumption and fat consumption only have small correlations with fruit consumption and negligible correlations with vegetable consumption. Overall, this suggests that using all of these measures as indicators of a single underlying diet construct is not appropriate, as they are not all measuring the same underlying construct. As such, only some of these measures will be carried forward as measures of the diet construct in the SEM model presented in Chapter 5.

The three physical activity measures also do not correlate non-negligibly with each other. Whilst volume of mild and moderate exercise have a small correlation between them, volume of vigorous activity has only a negligible correlation with both mild and moderate activity. This measure will be excluded from the formation of the physical activity construct. Besides these two constructs, the remaining four constructs: obesity, physical function, depression and socioeconomic position have good internal consistency giving some initial face validity to the measures being used.

3.4.1.1 Correlates of Obesity

When looking at correlations between measures on different constructs, there are very few non-negligible correlations and even when a non-negligible correlation exists it is at most small. In the obesity measures, BMI only has a non-negligible correlation with two measures that are indicators of other constructs: SF-36 Physical function score and daily sugar consumption frequency. Physical function had a small and negative correlation with BMI, indicating that better physical function was weakly associated with lower BMI. However, sugar consumption also had a small negative correlation with BMI, suggesting that higher daily sugar consumption was weakly associated with lower BMI. Waist circumference was only non-negligibly correlated with SF-36 Physical function score, daily meat consumption frequency and fruit consumption frequency, although the correlation with fruit consumption was on the borderline for being negligible. Both of these correlations were in the expected directions. Waist circumference had a small negative correlation with Physical function, and a small positive correlation with daily meat consumption frequency.

Also of note is that both BMI and Waist circumference have negligible correlation with both of the depression measures. Figure 3.31 shows a scatter plot of GHQ based depression scores against BMI, with the blue LOWESS line providing a localised estimate of the relationship between the two variables. From the LOWESS line it appears that as well as the non-negligible correlation between the two variables, the relationship between BMI and GHQ-based depression score is linear.

When the sample was partitioned based on sex, Frequency of money problems, and ethnicity separately, the correlation between BMI and GHQ-based depression score was still negligible. However, the LOWESS lines depicted in Figures 3.32 and 3.33 show some non-linearities in the relationship between BMI and GHQ-based depression score. When split by sex, a non-linear relationship between the variables appears at higher GHQ scores, with the association becoming increasingly positive in Males, and increasingly negative in females. However, in this range of GHQ scores, there is considerably more uncertainty in the local LOWESS estimate (indicated by the darker grey band), due to there being less data in this region of the scatter plot. As a result it is possible that the apparent non-linear behaviour is down to sampling variation, and that with more data, the relationship would be linear across the entire range.

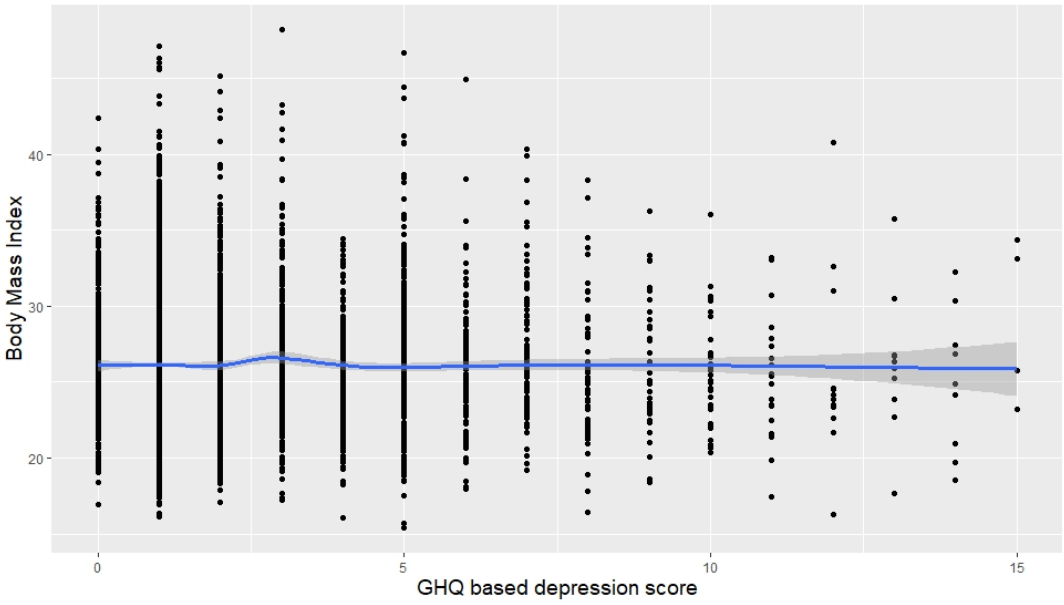

FIGURE 3.31: Scatter plot of GHQ based depression score against BMI in wave 5

The relationship in non-whites also displays some non-linear behaviours, with an oscillation in the relationship when GHQ scores are between 1 and 4. This oscillation is also found in the other scatter-plots, though to a lower extent. An important thing to note is that the size of the oscillation relates to the sample size of the group in that groups with a smaller sample size have a larger oscillation. As a result, this may be an artefact of the way the measure was derived that is eventually smoothed out by larger sample sizes. Above a GHQ score of 4, there appears to be a negative association between GHQ-based depression and BMI in non-Whites although again there is considerable uncertainty in the local estimates in this GHQ score range.

3.4.1.2 Correlates of Depression measures

We have already seen that depression is negligibly correlated with BMI and waist circumference, even when the relationship is stratified by sample characteristics such as sex, socioeconomic position and work status. The depression measures, however, unlike BMI and waist circumference, do correlate non-negligibly with some measures on the other constructs. In particular, both depression measures have small, negative correlations with all measures of physical Function and all measures of socioeconomic position. This translates to the expected relationships that poorer physical function and lower socioeconomic position are associated with higher levels of depression (Hoebel et al., 2017; Lorant et al., 2003; Stegenga et al., 2012; Russo et al.,

2007).

Within the physical function measures, bodily pain had the strongest correlation with the depression measures, followed by role limitation and then lastly physical function score. This suggests that the experience of role limitation due to physical health problems and the experience of pain may contribute additional risk for depression over and above the physical function problems they are associated with.

Within the socioeconomic position measures, social ladder position had the strongest correlation with both depression measures, followed by money problems and finally problems with bills. This is interesting as it perhaps suggests that one's perceived place in society may be more important than the experience of more objective adverse socioeconomic circumstances in the relationship between depression and SEP. Previous research has found that subjective socioeconomic position has an impact on ill-health independently of objective socioeconomic position (Doshi et al., 2016; Hoebel et al., 2017). In the case of depression, it has also been shown that subjective socioeconomic position mediates the relationship between objective socioeconomic position and depressive symptoms (Hoebel et al., 2017). In this way the larger correlation seen between the social ladder position and depression could be indicative of a cumulative effect of the perceived socioeconomic position itself and the objective socioeconomic position that may influence it. However, the difference in correlation is small and so could also be explained by sample variation.

The SF-36 based measure of depression also has a small negative correlation with average volume of moderate activity. All other correlations with the depression measures are negligible.

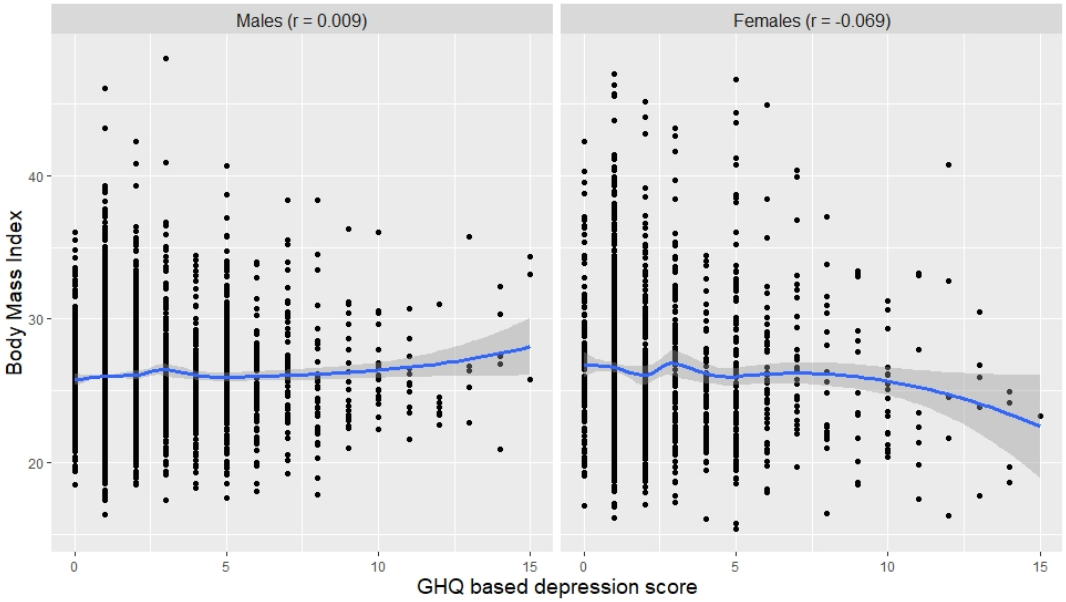

FIGURE 3.32: Scatter plot of GHQ based depression against BMI in wave 5 by sex

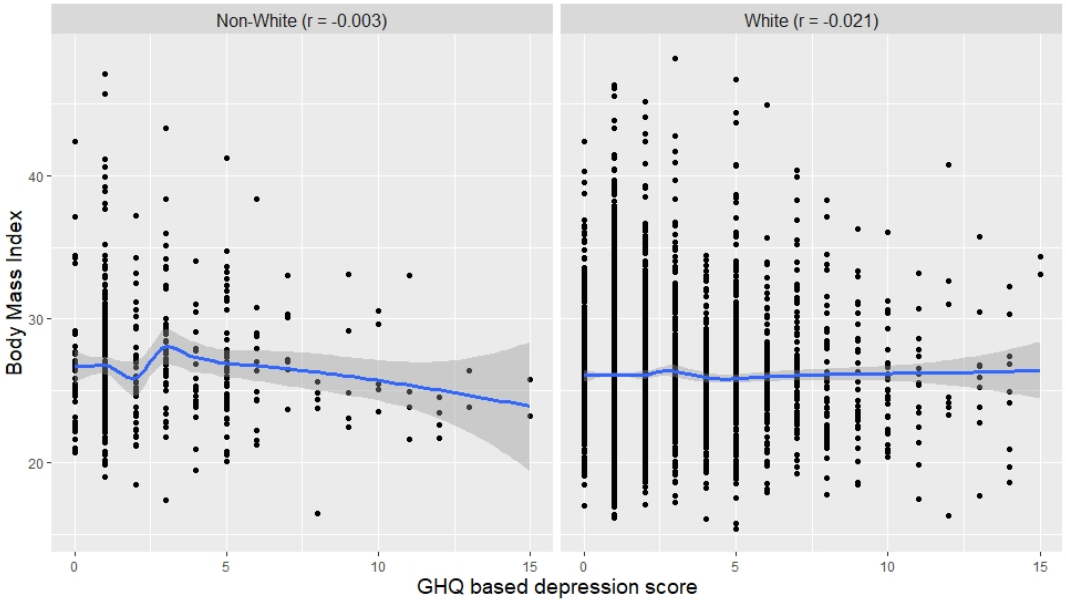

FIGURE 3.33: Scatter plot of GHQ based depression against BMI in wave 5 by ethnicity

3.4.1.3 Other Correlates

The majority of correlations between other constructs' measures were negligible. The only exceptions to this were the small positive correlations between indicators of physical function and socioeconomic position and a small correlation between physical function and vigorous physical activity. In the case of physical function and socioeconomic position, all three measures of physical function had small positive correlations with all three measures of socioeconomic position. Interestingly, within the physical function measures, only the physical function score correlated with any of the physical activity measures: it had a small positive correlation with vigorous activity. Both bodily pain and role limitation score correlated negligibly with all of the physical activity measures.

3.5 Exploratory Linear Models

Having explored cross-sectional correlations between the measures in the previous section, in this section I will present the results from some prospective exploratory linear models that aim to give a snapshot of what we might expect to see in the final Structural Equation Model. In particular, I will look at prospective predictors of obesity (Model 1) and depression (Model 2) separately. For each of these outcomes, three models will be presented. The first will be an unadjusted model (Model a), in which one outcome is regressed onto the other at the previous time point. The second model will then adjust for other covariates at the previous time-point (Model b), and the third will then estimate model 2 for males and females separately (Model c). For Model 2, a fourth model was also analysed to examine how strongly current depression was associated with future depression (Model d).

For this analysis, outcomes from wave 7 were regressed onto covariates at wave 5. Since the focus of this analysis is not to explore the stability of relationships over time, predictors of outcomes at wave 9 have not been analysed. This will be more easily explored in the Structural Equation Model presented in 5.

In all models I have chosen a single measure to represent each construct, as simple linear regressions are not capable of including a more complex factor structure. Obesity has been indicated by body weight (BMI), depression by GHQ-based depression score, physical function by SF-36 physical function score, diet by sweets consumption, physical activity by mild activity volume

and socioeconomic position by social ladder position. In all models, missing data was handled using list-wise deletion, whereby cases are excluded from the analysis if they have a missing value on any of the included variables.

3.5.1 Model 1 - Regressing BMI on Depression

The unadjusted association in Model 1a indicates that GHQ based depression score is a poor predictor of future BMI. The beta coefficient is not significantly different from zero and the R^2 values indicate that our model is not explaining any of the variation in BMI scores that is found in the sample. Once the other covariates were added into the model in Model 1b, a much higher proportion of the variation in BMI scores was explained. This is likely due to the inclusion of previous BMI in the model, which alongside age, was the only significant predictor of BMI at wave 7 in the model. All other beta coefficients in this model were either non-significant, very small in magnitude, or both. When model 1c was run separately in males and females, there were no substantive differences between the model coefficients. The small differences in statistical significance that were observed, such as in the age coefficient, are likely due to the different sample sizes available in the two groups. Full results for the models looking at prospective predictors of BMI can be found in Table 3.17.

TABLE 3.17: Model coefficients for linear regression models of BMI at wave 7 regressed on covariates at wave 5

	Model 1a	Model 1b	Model 1c (Males)	Model 1c (Females)
Coefficients β (s.e):				
Intercept	26.66 (0.08)***	1.59 (0.47)***	1.84 (0.53)***	1.43 (0.98)
Depression	0.01 (0.03)	0.03 (0.01)	0.02 (0.02)	0.04 (0.03)
Previous BMI		1.01 (0.01)***	1.00 (0.01)***	1.02 (0.01)***
Sweets Consumption		0.00 (0.01)	−0.00 (0.01)	0.01 (0.03)
Physical Function Score		−0.00 (0.00)	−0.00 (0.00)	0.00 (0.00)
Mild Activity METS		0.01 (0.00)	0.01 (0.00)	0.01 (0.01)
Social ladder position		−0.04 (0.02)	−0.02 (0.02)	−0.08 (0.04)
Age		−0.02 (0.01)***	−0.02 (0.01)***	−0.02 (0.01)*
Female Sex		0.06 (0.07)		
Goodness of Fit:				
R^2	0.00	0.86	0.85	0.87
Adjusted R^2	-0.00	0.86	0.85	0.87
N	5840	2859	2025	834
RMSE	4.28	1.57	1.40	1.91

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$; RMSE - Root Mean Squared Error

3.5.2 Model 2 - Regressing Depression on BMI

Similar to the unadjusted model in the previous subsection, regressing GHQ-based depression score on previous BMI (Model 2a) provided a model with very poor explanatory power. Again the beta coefficient was almost zero, and was non-significant at all alpha levels. The R^2 values also indicate that none of the variance that exists in GHQ scores is explained by previous BMI. When adding in the remaining covariates in Model 2b, a similar picture arises to that of the full model of BMI predictors. Previous GHQ depression score is a significant predictor of GHQ depression score at the next time point, and has the only beta coefficient greater than 0.1. Both Physical Function Score and Social ladder position are significant predictors of GHQ depression score at the 1% level, however, their associated beta coefficients are close to zero, indicating only a small effect. Overall model 2b explained just less than a quarter of the variation that was observed in GHQ-depression scores, meaning that large amounts of variation has been left unexplained.

	Model 2a	Model 2b	Model 2c - Males	Model 2c - Females
Coefficients β (s.e):				
Intercept	2.13 (0.20)***	2.68 (0.52)***	2.50 (0.63)***	3.45 (0.94)***
Previous BMI	−0.01 (0.01)	−0.01 (0.01)	−0.00 (0.01)	−0.01 (0.01)
Depression		0.43 (0.02)***	0.46 (0.02)***	0.38 (0.03)***
Sweets Consumption		0.01 (0.01)	0.01 (0.01)	−0.03 (0.03)
Physical Function Score		−0.01 (0.00)**	−0.01 (0.00)**	−0.01 (0.00)
Mild Activity METS		−0.01 (0.00)	−0.00 (0.00)	−0.01 (0.01)
Social ladder position		−0.07 (0.02)**	−0.05 (0.02)*	−0.11 (0.04)**
Age		−0.01 (0.01)	−0.01 (0.01)	−0.00 (0.01)
Female Sex		0.08 (0.08)		
Goodness of Fit:				
R^2	0.00	0.24	0.25	0.23
Adjusted R^2	-0.00	0.24	0.25	0.22
N	5019	2942	2077	865
RMSE	2.07	1.74	1.68	1.86

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$; RMSE - Root Mean Squared Error

TABLE 3.18: Model coefficients for linear regression models of GHQ depression score at wave 7 regressed on covariates at wave 5

Similar to the models of BMI, the majority of the explained variation is likely explained by previous GHQ-depression score. This can be seen clearly in Table 3.19 which displays the model co-efficients and goodness of fit when GHQ-based depression in wave 7 is regressed only on GHQ-based depression at wave 5 (Model 2d). In this model, the explained variance is almost identical to the explained variance of the model with covariates from the other constructs. This means that these additional covariates are responsible for almost none of the explained variance in model 2b.

	Model 2d Both Sexes	Model 2d - Males	Model 2d - Females
Co-efficients β (s.e))			
Intercept	0.96 (0.03)***	0.89 (0.04)***	1.11 (0.06)***
Previous Depression	0.49 (0.01)***	0.52 (0.01)***	0.45 (0.02)***
Goodness of Fit			
R^2	0.25	0.26	0.23
Adjusted R^2	0.25	0.26	0.23
N	6085	4335	1750

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

TABLE 3.19: Model coefficients for linear regression models of GHQ depression score at wave 7 regressed on GHQ depression at wave 5

Analysing Model 2 in males and females separately has suggested there may be some difference between males and females in the predictors of depression. Social ladder position had a larger beta coefficient in females than in males, and physical function had a significant beta coefficient in males but not in females. However, it is again possible that the difference in physical function significance is due to the different sample sizes in the groups. However, differing sample size should not influence the estimates of the β coefficients, so the differing effect of social ladder position is more likely to reflect a true difference between the groups, such that lower perceived social ladder position is more detrimental to future depression in females than is the case in males. The full results of the regression models looking at prospective predictors of GHQ-based depression score can be found in Table 3.18.

3.6 Discussion

In this chapter, I conducted an exploratory analysis of the Whitehall II data set, with two main aims. The first aim was to provide descriptions of the

variables in the Whitehall II data set that could be used as measures for the constructs in my SEM model presented in Chapter 5. The relevant variables have been described at length in the preceding sections and so there remains nothing more to add here. The second aim was to analyse the distributions of these variables and their relationships with one another, in order to provide some context for the SEM model presented in Chapter 5.

3.6.1 The Distributions of Construct Indicators

In terms of the distributions of variables, an important thing to note first is the number of non-normally distributed variables present. Many of the variables are categorical in nature, and even if we treat these categorical measures as continuous many would still be heavily skewed. For example, if we were to treat SF-36 subscale scores as continuous then they would still have a pronounced left skew. This presents a challenge for the final Structural Equation Model, as standard maximum likelihood estimators used in SEM rely on the data following a Multivariate Normal distribution.

Since this is not the case, I will need to make use of robust procedures for estimating the final model in order to avoid introducing bias into the parameter estimates. Failing to account for non-Normality would also lead to poor quantification of statistical significance, since standard errors (and hence p-values) associated with parameters in the model would be calculated based on a false assumption of Normality and so may not represent the true standard error.

Obesity and overweight are common in the Whitehall II sample, with more than half the sample being either overweight or obese in waves 5, 7 and 9. However, the Whitehall II sample appears to be less obese than the general population. Average BMI is slightly lower than is the case for the general UK population. Data from the Health Survey for England suggests that between the year 2000 and 2009, average BMI in England rose from 26.7 to 27.3 (NHS Digital, 2020b) whereas over the same time-scale, mean BMI in the Whitehall II cohort rose from 26 to 26.6. The proportion of obesity in the Whitehall II sample is lower than that found in the general population. In adults aged over 45, obesity levels are over 30% in the general population, whereas in the Whitehall sample they are less than 25%.

High scores on the depression measures was uncommon in the Whitehall II sample, suggesting that there is a low prevalence of diagnosable depression

in this population. However, comparing the observed distribution to wider data is difficult due the variety of ways in which depression symptoms are measured. The 2014 Adult Psychiatric Morbidity Survey suggested the prevalence of depression was 3.3% (Mcmanus et al., 2016). The Kings Fund Mental Health Review estimated that in 2007 the prevalence of depression in the over 55 age group was between 2.9% 4.2% (McCrone et al., 2008). Within our Whitehall II sample, in wave 9 (collected between 2007 and 2009) 4.8% of GHQ-based depression scores were above 5 and 3.1% responses were above 6. To obtain a score of 6 or more a participant would have to respond to at least one of the depression items in Table 5.1 with a 3 or 4 (1 or 2 for ghq-26). Responding in this way to one of the items suggests that the participant has been feeling this way more often than usual (less often than usual for ghq-26). This might suggest that despite the low scores on the GHQ-based depression measure, the overall prevalence of depression is similar to what one might expect in the population, albeit maybe slightly lower. Poor physical function and low SEP were also uncommon in the Whitehall II sample.

The diet measures derived from the FFQ gave some indication that the Whitehall II sample may have a more healthy diet than the general population on average. Average fruit and vegetable consumption was higher in the sample than is the case in the general population (NHS Digital, 2020a), however, it is possible that this increase is an artefact of the method used to derive the estimates. It has been suggested that intake estimated based on summing items together from an FFQ can lead to overestimation of consumption (Thompson and Subar, 2013; Kristal et al., 2000). Since the diet measures here were derived by summing together multiple items in this way, it is possible they are subject to the same bias.

Nearly all the sample did little or no vigorous physical activity, however, participation in moderate and mild physical activity was higher. Given the age of the sample this is perhaps unsurprising. Older age is associated with reduced functional capacity such as lower muscle strength and endurance. As such, the overall ability to do vigorous activity declines making it more likely that individuals will meet their activity needs via less intensive means (McPhee et al., 2016; Milanović et al., 2013).

3.6.2 Relationships between Indicators of Different Constructs

When looking at the cross-sectional relationships between constructs, there were very few non-negligible relationships found, and those that were non-negligible were at most small. Even correlations between constructs that are widely accepted to be related had largely negligible relationships in this sample. For example, within the diet and physical activity measures, only sugar consumption correlated non-negligibly with BMI, and only processed meat consumption and mild activity volume correlated non-negligibly with waist circumference. All other diet and physical activity measures have negligible correlation with both obesity construct measures.

The prospective linear regression models also found few significant predictors of future BMI and depression. For BMI, the strongest predictor of future BMI in the models was BMI at the previous time point, as we would expect. Similarly, the strongest predictor of GHQ-based depression score was previous GHQ-based depression score. In general, other predictors in the model had little predictive power.

Whilst there has been historic scepticism about the role of diet and exercise in causing obesity (Lincoln, 1972) nowadays it is widely accepted that they both play a role. In its simplest form, obesity is described as a result of energy imbalance, whereby people gain weight and hence become obese when they consume more energy than they expend (World Health Organization, 2021). This energy imbalance may be more important than the specific quantities of different foods that are consumed. High consumption of food does not necessarily correlate positively with an individual consuming excess energy as the high consumption may be compensated by equivalently high energy expenditure. Similarly, high consumption in a single food group may be compensated by lower consumption in other food groups as well as high energy expenditure. In this way a participant with high consumption of one of the food group measures analysed here may not be living a lifestyle in which energy is imbalanced overall. If this was the case then this might explain the negligible associations found between measures of obesity and diet.

Further to this point, studies investigating the effect of sugar on obesity have given mixed conclusions as to whether consumption is associated with obesity

independently of energy balance. The review by Ruxton, Gardner and McNulty (2009) suggested that there was not enough credible evidence for an association between sugar consumption and BMI. They also cited methodological concerns that gave them scepticism over the association between sugar-sweetened beverage intake and obesity. Trumbo and Rivers (2014) similarly suggested that when adjusting for energy balance, evidence for an effect from sugar-sweetened beverages on obesity was inconsistent. Contrary to the results presented by Ruxton, Gardner and McNulty (2009) and Trumbo and Rivers (2014), Malik, Schulze and Hu (2006) and Malik et al. (2013) presented meta-analyses that suggested intake of sugar sweetened beverages was associated with increased body weight. Lastly, Morenga, Mallard and Mann (2013) suggested that, although sugar consumption was correlated positively with body weight, it was mediated by an effect on overall energy consumption.

Studies investigating the effect of fat and processed meat on obesity have been more consistent in suggesting that high fat intake and high processed meat intake are associated with increased body weight. Hooper et al. (2012) found that diets low in fat were associated with reduced body weight, and also found that in RCTs investigating the effect of diet in non-weight-loss populations lower weight gain was observed in the lower fat arms of the trials. Bray and Popkin (1998) concluded similarly from a review of 28 clinical trials that dietary fat contributed to obesity. Tobias et al. (2015) also suggested that low fat diets could be effective for weight-loss when compared to usual dieting, but were less effective than carbohydrate reduction diets. In terms of processed meat intake, Rouhani et al. (2014) found that groups who consumed higher amounts of red and processed meat had higher BMI and waist circumference. Taken as a whole, whilst the above literature does suggest that certain food groups are associated with increased risk of obesity, it is not entirely clear whether these effects are due to an impact on energy balance, or whether consumption has an impact on obesity independently.

Studies investigating the effect of fruit and vegetable intake on body weight have given mixed results (Kaiser et al., 2014; Schwingshackl et al., 2015), and it has been suggested that the effect of increased fruit and vegetable consumption on body weight is limited unless overall energy consumption is reduced (Kaiser et al., 2014). For example, increased fruit and vegetable consumption may be beneficial in producing short-term weight loss, but this weight loss might not be maintained after longer periods unless the increase

in fruit and vegetable intake is accompanied by an overall energy deficit as well (Tapsell et al., 2014; Tanumihardjo et al., 2009; Whigham et al., 2012). It may also be the case that energy-deficit diets are more important for weight loss than those that focus purely on increasing fruit and veg intake (Tanumihardjo et al., 2009). Given the aforementioned literature, participants who are consuming fruit and vegetables more frequently in the Whitehall II sample might not be consuming diets that are less calorific overall, hence attenuating the relationship between this measure and body weight in the sample.

Another possible explanation for the negligible correlations between diet and obesity measures is that biases and measurement error within the estimates of consumption, obtained via the FFQ, are attenuating estimates of the association. Obtaining accurate data on diet is a notoriously difficult problem, and methods based on self-report such as the FFQ are known to be susceptible to both person-specific bias and general bias from measurement error (Kipnis et al., 2002). For example, individuals may respond in a way that they feel is socially desirable, leading to under-reporting of foods considered 'bad' and an over-reporting of foods considered 'good' (Hebert et al., 1995). Studies have also shown that FFQ estimates derived from the summation of long lists of food items, such as I have done with the Whitehall II data, leads to overestimation of consumption frequencies (Thompson and Subar, 2013; Kristal et al., 2000). Overall, studies have found that using FFQ data results in biases that attenuate the estimate of the diet-disease relationship unless suitable adjustments can be made with using a more objective reference measure (Kipnis et al., 2001; Kipnis et al., 2002; Kristal et al., 2000). As no such reference measure is available in the Whitehall II data set, it is reasonable to expect that the lack of correlation is at least, in part, influenced by this attenuation effect.

Whilst FFQ data is generally recognised as being better than other diet measures, such as 24-hr dietary recalls, for assessing long-term trends (Thompson and Subar, 2013) there is still the possibility that diet data collected is not representative of overall diet patterns that participants engaged in since the last wave. The FFQ questionnaire only collects data on the most recent month, whereas the time between data collection is several years. As such, estimates of associations assume that participants diet patterns remain stable over the time between waves. However, participants' dietary patterns may change for various reasons between the study waves. Seasonal changes may effect the responses that individuals give to the FFQ (Thompson and Subar, 2013)

and individuals who are overweight or obese may also have changed their diets as an attempt at losing weight. A recent meta-analysis found that 42% of adults from general populations reported trying to lose weight in the past year, with a further 23% reporting trying to maintain their weight (Santos et al., 2017). Given this, as well as the prevalence of obesity and overweight in our sample, it is reasonable to expect that many participants in the Whitehall II study will also have changed their diet at some point between the waves in which diet data is collected.

Of the obesity measures, only waist circumference correlated non-negligibly with a measure of physical activity, and only with mild activity volume. At first glance, this is a surprising finding as physical activity is widely recognised as an important factor in the development of obesity (Paolicelli, 2016; Fox and Hillsdon, 2007; Wareham, 2007) and studies have shown that individuals with obesity perform a lower volume of physical activity than do people of healthy weight. However, physical activity can be performed in a variety of settings and is multifaceted in that duration, intensity, and type of activity are likely to be important to energy expenditure. As such, the measures derived to represent physical activity in this analysis may not be sufficiently representative of individuals overall physical activity and by extension may not be a good proxy of the calories expended during exercise. The forms of physical activity included in my derived measures may not describe all the possible ways in which participants keep themselves active; it is possible that participants are engaging in other forms of activity outside of these means. Additionally, individuals activity habits may fluctuate in between the different waves of data collection. This missing data on other forms of physical activity and variation between the data collection points may be confounding the association between our physical activity and obesity measures.

Both BMI and waist circumference were also negligibly correlated with all three measures of socioeconomic position. At first glance this is again surprising. Obesity has an established relationship with socioeconomic position in which lower socioeconomic position is associated with increased prevalence of obesity. Previous studies using the Whitehall II dataset have also found a relationship between obesity and SEP (Brunner et al., 2001). However, that study measured SEP using a seven level measure of civil service employment grade. This particular measure was not available to me, necessitating the selection of alternative measures.

This difference in the measures used may explain the lack of association found between socioeconomic position and both BMI and waist circumference. The measures used in my analysis relate to a different component of socioeconomic position than does employment grade. The relationship between obesity and socioeconomic position is recognised as being complex, and so it is also possible that the exact relationship differs when different components of socioeconomic position are investigated. Studies of the obesity-socioeconomic position relationship in highly developed countries that use income-based measures of SEP have provided weaker evidence of a relationship than have education or occupation (McLaren, 2007). Svedberg et al. (2016) showed that measures of socioeconomic position often do not correlate well with one another, and so different measures of SEP may produce different associations with obesity.

At this stage in the analysis I have not found any significant evidence of a relationship between diet and depression, nor between physical activity and depression. Whilst the literature reviewed in Chapter 2 suggested that there was some evidence that both diet and physical activity were related to depression, the results were not conclusive. Despite the evidence of diet quality affecting depression presented in Hoare et al. (2014) across the reviews greater emphasis was put on the affect of patterns of consumption and eating behaviours more so than the total consumption itself. In the case of diet influencing depression, it was suggested that eating behaviours such as dietary restraint, repeated dieting, disordered eating and the use of food as an emotional regulator could lead increased risk of depression in the obese (Markowitz, Friedman and Arent, 2008; Preiss, Brennan and Clarke, 2013). Similarly, for depression leading to obesity, binge eating behaviours were consistently associated with depression, albeit only cross-sectionally. The measures I have analysed here are not capable of assessing whether individuals are binge eating, engaging in dietary restriction or how participants view their eating habits and instead focus reported consumption. In light of this, the lack of association we have seen between our diet measures and depression might suggest that if there is an association between diet and depression, the way individuals consume food, and their perceptions of their dietary behaviours may be more important than the consumption itself.

Much of the literature presented in Chapter 2 suggested that females might be most at risk of comorbid obesity and depression, particularly younger females. As such, our sample population might not include many individuals

who are high risk for comorbid obesity and depression. In addition, there are relatively small numbers of non-white participants, meaning that the relationship between obesity and depression in this sample will be dominated by the association found within the white participants.

Our exploratory analysis also suggests that the only mechanism though which socioeconomic position operates within the obesity-depression system, is through relationships with depression and physical function.

3.7 Conclusions and Next Steps

Overall, this exploratory analysis has found little evidence for an association between obesity and depression. Furthermore, of the proposed mediators of diet, physical activity and physical function, physical function is the only construct with indicators that are associated with both obesity and depression. Going forward into the next phase of the modelling this might indicate that of these mediators, only physical function is implicated in the relationship between obesity and depression. It would also appear that any influence that socioeconomic position has on this system might only come through relationships with depression and physical function.

Despite these indications, it is of course important to wait until the analysis of my substantive model is complete before making firm conclusions. This analysis has only used crude analysis methods such as cross-sectional correlations and simple linear regressions which are not capable of including the potentially complex structure of the system I am investigating. In addition, missing data has only been dealt with in an ad-hoc way. For the cross-sectional correlations, pair-wise deletion was used, and for the prospective linear models list-wise deletion was used.

The inadequacies of these missing data techniques will be addressed in the next chapter, where I will also describe how missing data will be dealt with more comprehensively in the substantive SEM model that I analyse in Chapter 5.

Chapter 4

Treatment of Missing Data within Whitehall II

4.1 Introduction

In the previous chapter I explored the Whitehall II data descriptively, to give some initial insights into the sample characteristics and potential relationships that might exist within the data. This exploration was based only on the data that was observed; cases with missing observations in a given variable were excluded from any analysis in which that variable was used. Whilst this is acceptable practice when doing exploratory analysis, in order to avoid introducing bias into the results of our main model, it will be necessary to take a more considered approach to dealing with any missing data.

In this chapter, I describe the development of a principled approach to dealing with missing data that I will use in my analysis of the Whitehall II data set presented in the next chapter. The model development will be described across three main sections. Firstly, I will describe briefly the issues posed by missing data and how different types of missing data are conceptualised. Secondly, I will attempt to quantify the extent of missing data within the Whitehall II data set, as well as exploring the pattern of missingness within the data. Specifically, this will focus on examining whether systematic differences exist in the data between participants with missing data and those without. Thirdly, I will briefly review some of the available methods for dealing with missing data before discussing in detail two of the main methods: Multiple Imputation (MI) and Full Information Maximum Likelihood (FIML). These two methods are discussed within the context of analysing the Whitehall II data set and the particular challenges that presents. This is followed by a justification of my use of Robust FIML (RFIML), supplemented

by nearest neighbour imputation, to deal with missing data in the analysis presented in Chapter 5.

4.1.1 The Issue of Missing Data

When conducting epidemiological research, that some participants will have missing data is almost inevitable. Particularly in questionnaire based longitudinal studies, participants may drop out from a study as it progresses, or may simply not answer specific items for both deliberate and accidental reasons. Additionally, data may be missing due to recording and data input errors.

Whatever the cause of missing data, its presence can lead to problems during data analysis. Many statistical methods, such as regression, require complete data in order to produce the parameter estimates for the specified model. Useful quantities such as means and standard deviations cannot be calculated when missing entries are included.

4.1.2 Types of Missingness

Before describing methods of dealing with missing data, first we need to describe the different kinds of missingness that might occur in a dataset. To do this, missing data problems are usually considered to belong to one of three categories: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). These categorisations were introduced by Rubin (1976), and they each describe a different assumption about the probability that data is missing.

4.1.2.1 Missing Completely at Random

Data are considered missing completely at random (MCAR) if the probability that data is missing doesn't depend on either the observed data or unobserved data.

If we suppose that R is a matrix that represents the locations of missing data points in our dataset Y such that $R_{ij} = 0$ if the observation for participant i on variable j is missing, and $R_{ij} = 1$ if it is observed, then the three categorisations relate to different claims made about the distribution of R, $P(R)$. Specifically, if we decompose Y into the parts that are observed Y_{obs} and missing Y_{mis} such that $Y = (Y_{obs}, Y_{mis})$ then the data being missing completely at

random is equivalent to the following equality holding true:

$$P(R \mid Y) = P(R).$$

In this way the distribution of the missing data is unchanged irrespective of whether we are conditioning on the data or not. For example, individual questions that are missed out accidentally by a participant might be considered MCAR (assuming all individuals are equally error prone), as might coding errors made by researchers.

4.1.2.2 Missing at Random

Data are considered missing at random (MAR) if the probability that data is missing in the dataset depends only on the observed data, but not on the missing data. Using the same notation as above the data being MAR is equivalent to the following equality holding true:

$$P(R \mid Y) = P(R \mid Y_{obs}).$$

Under this condition, some participants may be more likely than others to have a particular response missing, but it is possible to account for this difference using other pieces of observed data. An example of this might be in longitudinal cohort studies where missing data due to a participant dropping out of the study is predicted perfectly by data on the participant that is available at previous time points.

4.1.2.3 Missing not at Random

Finally, data are considered missing not at random (MNAR) if the probability that data are missing depends on both observed and unobserved data. For example, if individuals with high income are less likely to report their income than individuals with low income then the data will be MNAR. Reconsidering the example of data being missing due to a participant dropping out of a cohort study, if the probability of a participant dropping out depends on the data that would only have been available after the participant dropped out, then this data would be MNAR.

Again, in the above notation, data are MNAR if the following equality holds:

$$P(R \mid Y) = P(R \mid Y_{obs}, Y_{mis})$$

4.1.3 Dealing with Missingness

MCAR data is generally the easiest to deal with, as ad-hoc methods such as those presented in 4.3 don't introduce bias under this condition (Schafer and Graham, 2002). MAR is also straightforward to deal with. More thorough methods of dealing with missing data assume that the data are at worst MAR, and should produce unbiased results under this condition (Schafer and Graham, 2002; Enders and Bandalos, 2001). MNAR is much trickier to deal with, as it requires explicit modelling of the missing data mechanism, which may be unknown. That being said, methods for dealing with MNAR are being developed (Galimard et al., 2018).

It is not possible to determine for certain whether data are MNAR or not, since to do so would rely on us knowing the value of the missing value. In some studies, MNAR can be assessed by following up on some of the individuals who have missing data to recover both the missing observations and the reasons why it was missing. However, since this is not an option in this analysis, instead we are limited to making assumptions about the missingness in our studies. The methods for dealing with missing data are usually based on the assumption that the data are MAR and hence, when using these methods, it is important to ensure that this assumption is plausible. Hence, in the next section I will explore the missing data patterns in the Whitehall II dataset with the aim, in part, of improving the plausibility of the MAR assumption.

4.2 Missing Data within the Whitehall II Dataset

Before describing the methods that are available to us for dealing with missing data, first I will explore the patterns of missing data that are present in our variables of interest from the Whitehall II data set. First, I have explored the extent of missing data for construct indicators that were explored in Chapter 3. In general, the less data that is missing the better, as higher rates of missing data are more likely to prevent our missing data methods from recovering accurate parameter estimates for the model. As such, it is important to understand how much missing data there is so that any biases created by this factor can be considered in the analysis.

Second, I have investigated whether there are systematic differences between the characteristics of participants with missing data and participants without. Most methods of dealing with missing data assume that the data are at most

MAR i.e. that the data are not MNAR. However, it is often impossible to know definitively whether data are MNAR or not, as this would require knowing the missing value and the specific cause for it not originally being recorded, the first of which is necessarily not knowable and the second of which is also difficult to know in a secondary data analysis setting.

Despite this, analysing the patterns of missing data can be helpful for discounting whether the data are MCAR. For example, if participants and non-participants differ systematically on variables that were collected in an earlier wave, then this suggest that missing data due to non-participation is at least MAR. As a result, understanding whether these systematic differences exist or not will help motivate my choice of missing data method, and may also highlight variables that can be used to improve its performance.

4.2.1 Quantifying the Extent of Missing Data

A sensible first step into exploring the patterns of missing data that are present in our sample of Whitehall II variables, is quantifying the "degree of missingness" that exists in the data. To do this, I have calculated the proportion of data that is missing for the overall data set, for each variable separately and for each participant separately.

The overall percentage of missing data across the entire collection of construct indicators was 37.4%. Figure 4.1 shows the distribution of missing data proportions for all the 51 variables that may be used as construct indicators. From this, we can see that very few variables have less than 10% missing data, with the vast majority having between 10 and 40 percent missing data. Particularly concerning is the high frequency of variables that have over 30% missing data as above this rate even the more sophisticated methods of dealing with missing data can experience issues such as bias in the parameter estimates and in some cases estimated models failing to convergence on a solution (Savalei and Falk, 2014; Jia, 2016).

Table 4.1 categorises the construct indicator variables depending on how much missing data the variable has. This data is also depicted in Figure 4.2, where variables are ranked from highest proportion of missing data to lowest. The worst performing constructs in terms of missing data are physical activity and by diet. In the case of diet, all of its indicators have at least 30% missing data, whereas for physical activity all but two indicators (mild activity volume at waves seven and nine) have more than 40% missing data.

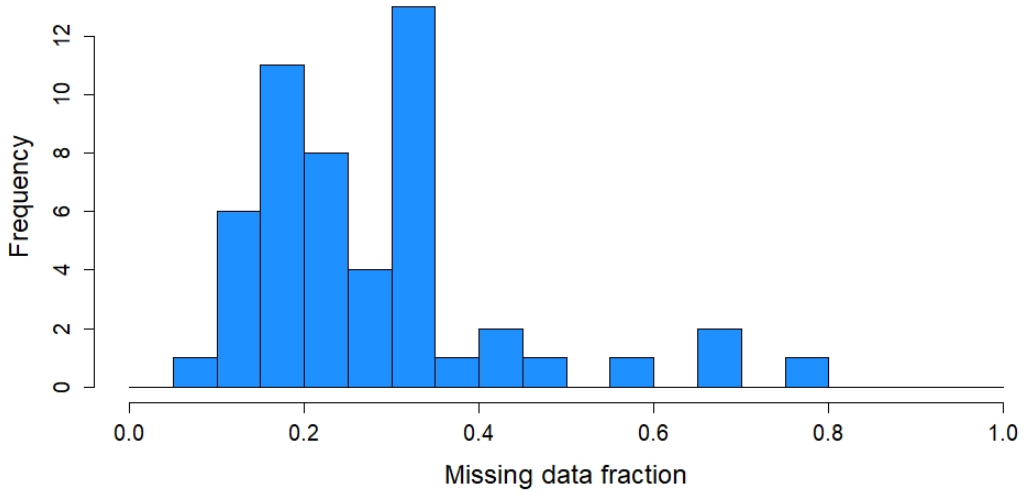

FIGURE 4.1: Histogram of the fraction of missing data within our sample of the Whitehall II data on variables to be used as construct indicators

Missing data rates in the moderate and vigorous activity in waves five and seven are particularly high (between 50% and 80%).

These high missing data rates are likely a result of the fact that both the diet and physical activity measures are derived from a combination of individual items. As such, when one or more of these individual items is missing, the composite score is missing as well. Since different individuals will likely have missing data on different items this translates into much higher missing data rates in the composite score than might be found on the items separately.

Obesity variables also have moderate to high missing data in all waves, with both indicator variables containing at least 20% missing data. Depression and physical function indicators from the GHQ and SF-36 generally had better rates of missing data, all being less than 20%, except in wave nine where GHQ-based depression, SF-36 physical function and SF-36 role limitation had between 20 and 30 percent missing data. Socioeconomic position variables similarly had less than 20% missing data until wave nine where the missing data rate rose to between 20 and 30%.

	Wave 5	Wave 7	Wave 9
Missing 40% +	Mild activity volume, moderate activity volume, vigorous activity volume	Moderate activity volume, vigorous activity volume	Moderate activity volume, vigorous activity volume
Missing 30 - 40%	BMI, Waist circumference, fruit consumption, vegetable consumption, sweets consumption, meat consumption	fruit consumption, vegetable consumption, sweets consumption, meat consumption	fruit consumption, vegetable consumption, sweets consumption, meat consumption
Missing 20 - 30%		BMI, waist circumference, mild activity volume	BMI, waist circumference, GHQ-depression, SF-36 Physical Function, SF-36 Role limitation, mild activity volume, social ladder position, problems with money, problems with bills
Missing 0 - 20%	GHQ-depression, SF-36 depression, SF-36 Physical Function, SF-36 role limitation, SF-36 bodily pain, social ladder position, problems with money, problems with bills	GHQ-depression, SF-36 depression, SF-36 Physical Function, SF-36 role limitation, SF-36 bodily pain, social ladder position, problems with money, problems with bills	SF-36 depression, SF-36 bodily pain

TABLE 4.1: Table of missing data categorisations for construct indicators used in the model. Colour coding shows which construct the indicator belongs to: Obesity, Depression, Physical activity, Diet, Physical function, Socioeconomic position
.

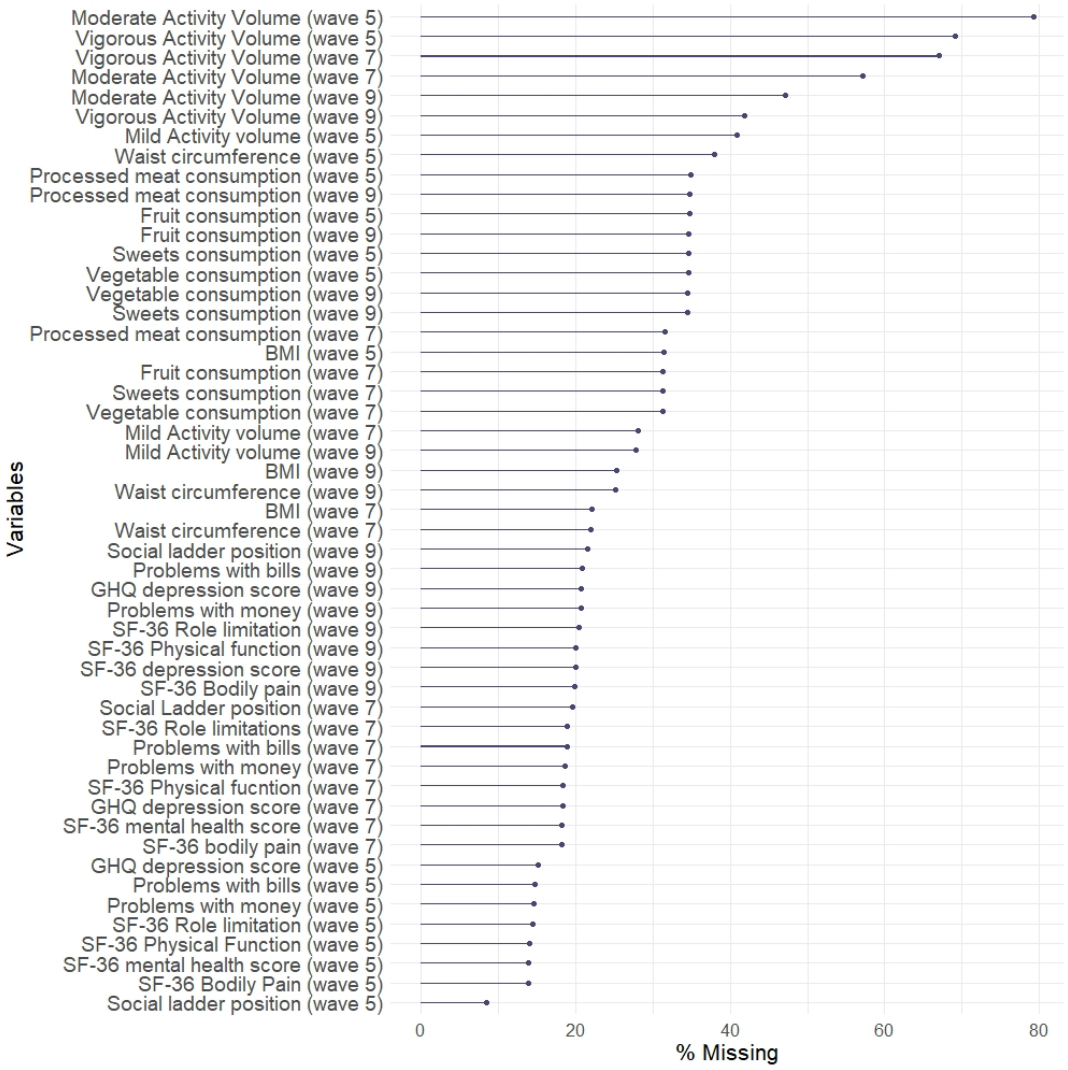

FIGURE 4.2: Plot of missing data percentage in construct indicators within our sample of the Whitehall II data

Figure 4.3 then shows the distribution of the missing data fraction within individuals. The most frequent rate of missingess within a participants data is just below 10%, and higher rates of missing data become rapidly less frequent. The extended tail of the distribution, however, shows that there are still some participants with a very high proportion of missing data. There is also a moderate frequency of participants who have less than 5% missing data and overall, the majority of participants have less than 20% of their data missing.

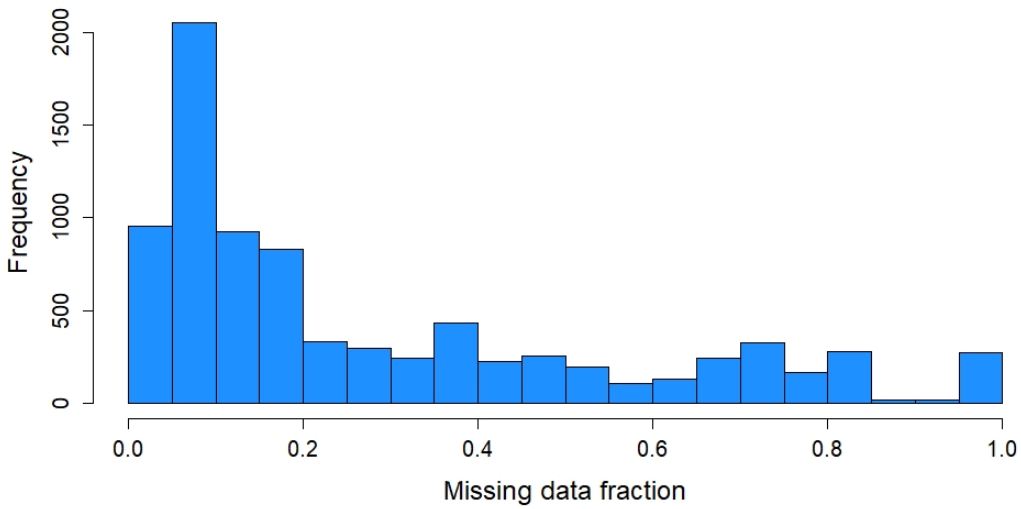

FIGURE 4.3: Histogram of the fraction of missing data within members of our sample of the Whitehall II data

4.2.2 Causes of Missing Data

Having looked at the overall missing data rate in our variables of interest, we can now turn to look at the potential causes and predictors of missing data in the Whitehall II study. This is useful for two main reasons. Firstly, it allows us to investigate whether we believe the observed data to be representative of the entire sample, and secondly it allows us to find variables outside of our analysis model that might provide useful additional information that will inform my missing data treatment. This is particularly helpful for supporting the assumption that data are missing at random. If we are able to find variables that are associated with missingness in the data, then these variables can be added into the study as predictors of missing data, hence making the assumption that data are MAR more plausible.

4.2.2.1 Non-participation

One of the most prominent causes of missing data within the Whitehall II data set is non-participation. Figure 4.4 shows the proportion of the sample that participated in odd numbered waves of the data collection. There is a clear decrease in the number of participants over time, to the extent that in wave 11, nearly 40% of the original sample did not participate.

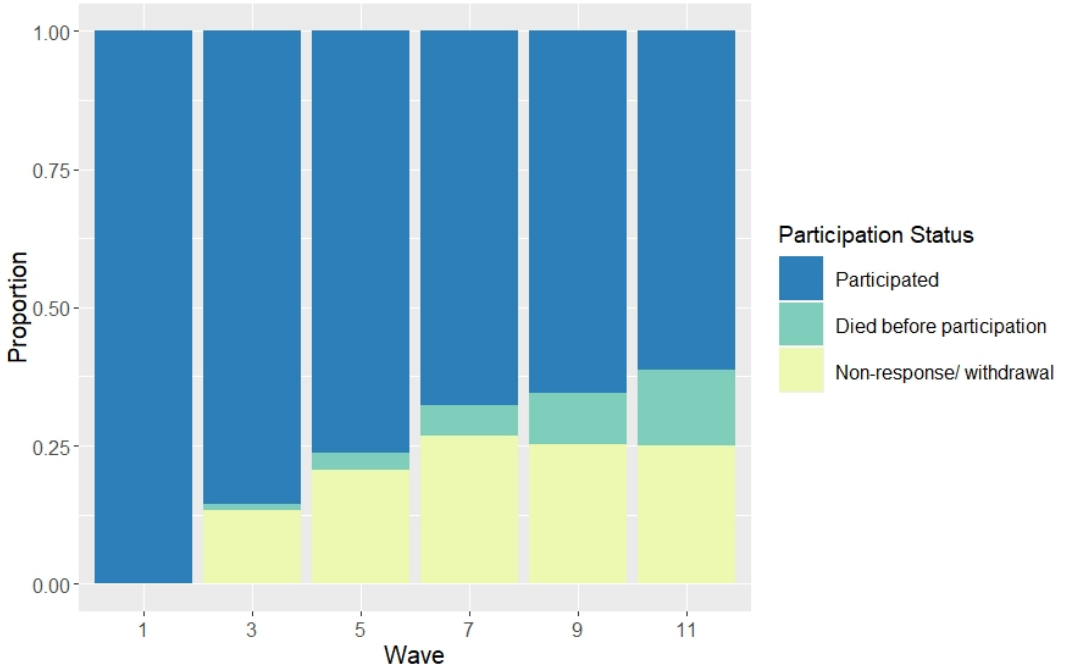

FIGURE 4.4: Plot showing proportion of sample in each participation status by wave

Non-participation can lead to two main missing data patterns. Firstly, participants may have missing data on one wave due to a failure to fill in the questionnaire, but then re-participate in a future wave. Secondly, participants may drop out permanently, due to death or other reasons. Given that non-participation is responsible for such a large proportion of missing data in the sample, understanding which variables might predict non-participation is central to determining whether data in the study are MAR or not. As such, here I will address two main questions. Firstly, does the sample of participants with recorded data for our model differ substantively from those who have no recorded data due to non-participation? And relatedly, are there any predictors of non-participation that could be useful for our missing data treatment?

To answer these questions I will explore whether there are differences between the characteristics of those who participated in at least one of waves five, seven or nine and those who didn't. Examining differences over every possible variable from the Whitehall II data set is impractical, hence I will explore how the distributions of key variables and sample characteristics differ between the two groups. The key variables included are a selection of the variables that will be used as indicators to measure the constructs included in the model in Chapter 5, however, this time taken from wave three of the

Whitehall II data. The only exception to this is the variable for vigorous activity, which here is taken from a single question on the number of hours of vigorous activity participants per week on average. This is because the individual items for physical activity were not available at wave three and so the measures outlined in Chapter 3 could not be calculated. The sample characteristics analysed cover participants' general health, marital status, age, sex, ethnicity and employment grade.

Table 4.2 gives descriptive statistics for some of the construct indicators included in our model, but at wave three in the model. Specifically it includes descriptive statistics for: BMI, SF-36 physical function score, GHQ-based depression score, daily vegetable consumption frequency, weekly hours of vigorous activity and the extent of money problems. There are few observable differences between the distributions of the indicators in the two groups and those differences that exist are either very small, or likely explained by the very different sample sizes in the two groups. The non-participant group have slightly higher average BMI, slightly lower average physical function, slightly higher average depression score, slightly higher average vegetable consumption and slightly lower average vigorous activity. There was a slight trend towards increased extent of money problems in the non-participant group as well. However, these differences do not provide strong evidence of there being a substantial difference between the distributions of the indicators in the two groups. In the case of money problems extent, the proportions of group members found in each category are almost identical in the two groups, and for the other five indicators, both the averages and measures of spread are almost identical. Differences of this size could be due to random sample variation, and are perhaps too small regardless to be considered a meaningful difference.

Whilst the measures of location and dispersion shown in Table 4.2 don't suggest that there are major differences between the group of participants and non-participants for our model, there may be differences in the skewness and kurtosis of the distributions between the two groups. Distributions with similar measures of location and spread can have substantially different skewness and kurtosis. Skewness provides a measure of symmetry for the distributions. A skew of zero indicates that the distribution is symmetrical in shape, whilst a positive (negative) skew means that the right (left) tail is heavier than than the left (right) tail. This can be seen clearly by analysing the formula used to calculate skew, given in equation 4.1. Suppose that

	Participation Status:	
	No participation (N = 2,028)	Participation (N = 8,280)
BMI		
Minimum	15.03	16.00
Median (IQR)	25.07 (23.18, 27.67)	24.83 (22.89, 27.13)
Mean ± sd	25.67 ± 3.89	25.29 ± 3.72
Maximum	45.91	48.48
SF-36 Physical Function		
Minimum	0.00	0.00
Median (IQR)	90.00 (80.00, 100.00)	95.00 (85.00, 100.00)
Mean ± sd	84.89 ± 19.62	89.58 ± 14.86
Maximum	100.00	100.00
Depression (GHQ)		
Minimum	0.00	0.00
Median (IQR)	1.00 (1.00, 3.00)	1.00 (1.00, 3.00)
Mean ± sd	2.23 ± 2.26	2.05 ± 2.05
Maximum	15.00	15.00
Vegetable Consumption		
Minimum	0.02	0.38
Median (IQR)	3.38 (2.39, 4.93)	3.42 (2.47, 4.61)
Mean ± sd	3.89 ± 2.33	3.74 ± 1.91
Maximum	26.44	32.12
Vigorous Activity hours		
Minimum	0.00	0.00
Median (IQR)	0.00 (0.00, 0.50)	0.00 (0.00, 1.00)
Mean ± sd	0.55 ± 1.27	0.77 ± 1.53
Maximum	11.00	21.00
Money Problems extent		
Very great	1 (0)	35 (0)
Great	18 (3)	124 (2)
Some	90 (13)	826 (11)
Slight	89 (13)	874 (11)
Very little	243 (36)	2,986 (39)
None	232 (34)	2,791 (37)

TABLE 4.2: Descriptive statistics for selected construct indicators in wave three, for model participants and non-participants

our variable X has more data points that are above the mean than are below it, such that the variable X has a heavier tail to the right than to the left. In this case, when calculating the skew of the sample data, the higher frequency

	Participation Status:	
	No participation (N = 2,028)	Participation (N = 8,280)
Body Mass Index		
Skewness	0.9859623	1.1536919
Kurtosis	5.302225	6.039716
SF-36 Physical Function		
Skewness	-1.885543	-2.560578
Kurtosis	6.483909	10.982748
Depression (GHQ)		
Skewness	1.830228	2.048323
Kurtosis	7.127394	8.353271
Vegetable Consumption		
Skewness	2.930387	2.278350
Kurtosis	20.89176	17.66556
Vigorous Activity		
Skewness	3.890947	4.165098
Kurtosis	23.39547	32.45250

TABLE 4.3: Table showing skewness and kurtosis of distributions of BMI, physical function, depression, vegetable consumption and vigorous activity in wave three, separated by participation in the model

of data points that sit above the mean will mean that the positive terms included in the expectation sum will 'outweigh' the negative terms, resulting in a positive skew value. Similarly, when the left tail is heavier than the right tail, the negative terms will 'outweigh' the positive terms in the expectation sum, giving a negative skew value.

$$\text{Skew} = \frac{E[(X-\mu)^3]}{\sigma^3} \tag{4.1}$$

As an example, in Table 4.3, depression has a positive skew in both groups. This is because the mean depression Score is very close to the minimum possible score of zero on the scale, meaning that there is a very short left tail. The right tail, however, is much longer, and stretches all the way to the maximum value on the scale of 15. Since the right tail is much longer than the left tail (see Figure 4.5) the data are positively skewed.

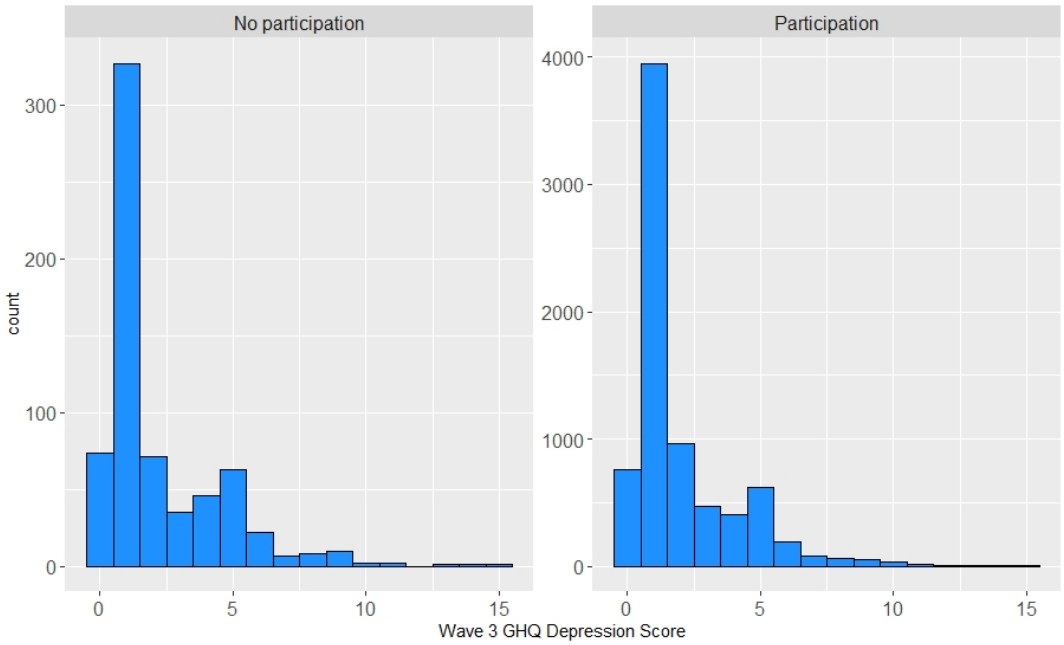

FIGURE 4.5: Histograms of GHQ based depression score at wave three in those who participated in at least one of waves five, seven or nine vs those who didn't

Kurtosis describes how extreme the tails of the distribution are (Westfall, 2014). When more data is found in the tails of the distribution, the kurtosis of the distribution becomes larger, and when less data is found in the tails of the distribution, the kurtosis becomes smaller. This can be seen more clearly when examining the formula for kurtosis given in equation 4.2. The expression $(X - \mu)/\sigma$ gives the standardised data of the variable X and so the Kurtosis represents the average of the standardised data when raised to the power of four. When the data are less than one standard deviation σ away from the mean value μ the value of $(X - \mu)/\sigma$ gets smaller when it is raised to the power of four, whereas for data that are more than one standard deviation away from the mean $(X - \mu)/\sigma$ gets larger when raised to the power of four. As a result when a variable has heavy tails in which lots of the data are more than one standard deviation away from the mean, the value of 4.2 and hence the kurtosis is high. Conversely, if all the data is clustered around the mean with very little data further than one standard deviation form the mean, the value of 4.2 will be lower.

The quartic nature of the right hand size of equation 4.2 also means that data that are further from the mean contribute more to the kurtosis of the variable and data that are closer to the mean, contribute less. A final point to note is that, unlike skew, kurtosis is always positive, due to the even exponent in the

formula.

$$\text{Kurtosis} = E\left[\left(\frac{(X-\mu)}{\sigma}\right)^4\right] \tag{4.2}$$

As an example, the distributions of body mass index are much less kurtotic than the distributions of vigorous activity. This can be easily seen by comparing BMI densities depicted in Figure 4.6 with the histograms of wave three vigorous activity hours shown in Figures 4.7 and 4.8. Both BMI densities have very little mass in the tails their distributions, whereas the distributions of vigorous activity hours have long right tails with outliers that are many standard deviations away from the mean. This is why the kurtosis is higher for these distributions.

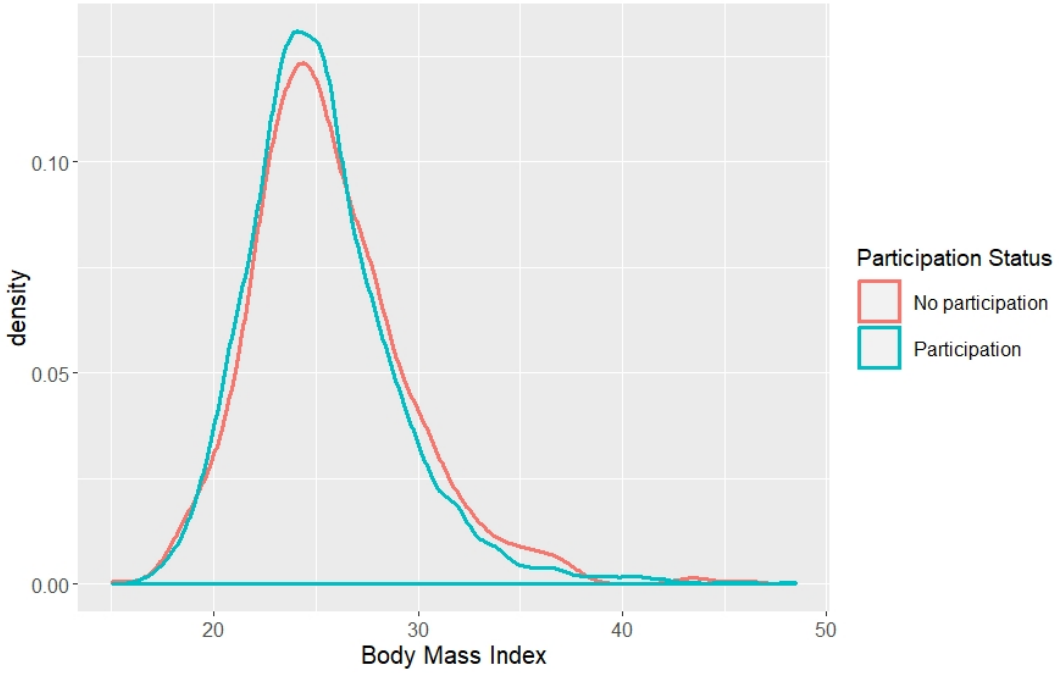

FIGURE 4.6: BMI densities in wave 3 for those who participated in one of waves five, seven or nine (blue) and those who did not (red)

Overall, there were very few differences between the skew statistics of the same indicators in the two participation groups. Whilst the magnitude of the skew differed between groups, the direction of the skew was the same for all the indicators. Kurtosis values were also different in magnitude between the two groups but this is not likely to be of any qualitative significance. In order to assess whether the different sizes of skew and kurtosis were problematic, I have looked at the histograms in the different groups for vigorous activity;

a variable with one of the largest differences between the two groups in both skew and kurtosis. It is obvious from comparison of the two histograms that the distributions are very similar in shape. The difference between the statistics obtained for these distributions is therefore likely to be due to the presence of particularly extreme outliers that are present in the participation group increasing the values of skew and kurtosis in these groups. Very severe outliers have a stronger effect on skew and kurtosis than more moderate outliers.

	Participation Status:	
	No Participation (N = 2,028)	Participation (N = 8,280)
Age		
Minimum	34.67	34.06
Median (IQR)	45.95 (40.27, 51.53)	44.03 (39.55, 50.03)
Mean (sd)	45.77 $\pm$ 6.18	44.75 $\pm$ 6.00
Maximum	56.15	56.11
Sex		
Male (%)	1,158 (57)	5,737 (69)
Female (%)	870 (43)	2,543 (31)
Ethnicity		
Non-white (%)	311 (16)	724 (9)
White (%)	1,647 (84)	7,534 (91)
Education Level		
Secondary	870 (59)	3,181 (52)
University/ polytechnic	399 (27)	2,107 (34)
Nursing	6 (0)	36 (1)
College	117 (8)	526 (9)
Other	88 (6)	305 (5)
Civil Service Grade		
Administrative	377 (19)	2,651 (32)
Professional/ Executive	861 (42)	4,082 (49)
Clerical/ Support	790 (39)	1,547 (19)

TABLE 4.4: Descriptive statistics for sample characteristics (age, sex, ethnicity, education level and Civil Service grade) in participants and non-participants for the model.

There were more pronounced differences in the sample characteristics of participants when compared to non-participants. There were higher proportions of females and people from non-white ethnic backgrounds in the non-participation group than was the case in the participation group. Additionally, there was a higher proportion of clerical support staff, and a lower proportion of administrative staff in the non-participant group compared to

the participation group. Finally, There was a lower proportion of university/ polytechnic educated individuals in the non-participant group than was the case in the participation group which was compensated with a higher proportion of participants only having secondary level education in the non-participant group. The distributions of age were mostly similar, except for the non-participant group having a higher average age by one to two years.

So far we have looked exclusively at the difference between individuals who haven't participated at all, and those who participated in at least one of waves five, seven or nine. In general, differences between these groups were minimal, except for some differences in sample characteristics. However, combining those who participated in some but not all waves of the study with those who participated in all waves may mask some differences between individuals with missing data and those without it. To investigate this possibility, the above comparisons were repeated, but this time between individuals who participated in all of waves five, seven and nine, referred to as the full participation group, and those who missed at least one wave of the study, referred to as the missed participation group.

Overall, the comparison showed a similar picture, with there being minimal differences between the distributions of construct indicators at wave three, but greater differences between the distributions of sample characteristics such as sex, ethnicity, education level and employment grade. In addition to the variables examined in the previous comparison, I also examined the distributions of other potential correlates of missingness in each of the participation groups. In particular participants' marital status, general health, previous diagnoses of depression and the presence of long-standing illnesses were analysed. Of these additional variables, only the distribution of long-standing illness did not differ substantially between the two groups; there was only a slight increase in the presence of long-standing illness in the missed participation group. Counts and proportions for these variables in the two groups can be found in Table 4.5.

Differences in marital status, general health, and depression history were more pronounced between the two groups, although the difference in depression history was still only slight. In the missed participation group, there was also a slightly higher proportion of participants who had ever been told they had depression, and there was an increase in the number of individuals with average health which was compensated by a decrease in the proportion of individuals with very good health. This provides some suggestion that

the missed participation group has poorer health and slightly higher depression on average, when compared to the full participation group. There was also a lower proportion of married participants which was compensated by higher proportions of single and divorced participants as well as a marginally higher proportion of widowed participants.

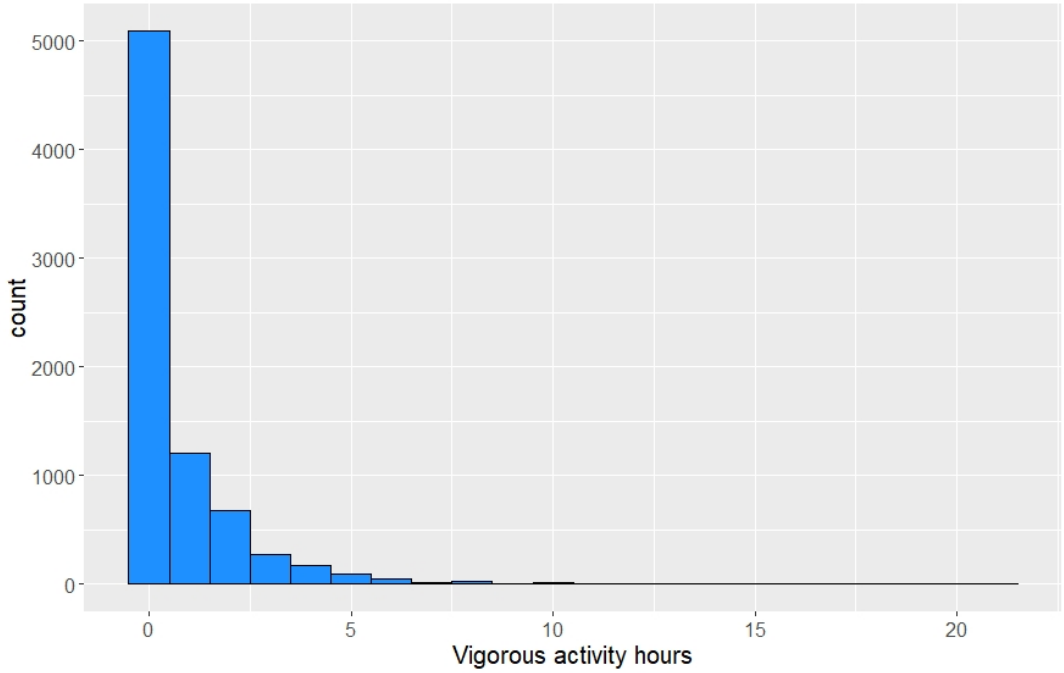

FIGURE 4.7: Histogram of vigorous activity hours in wave three for participants in the model

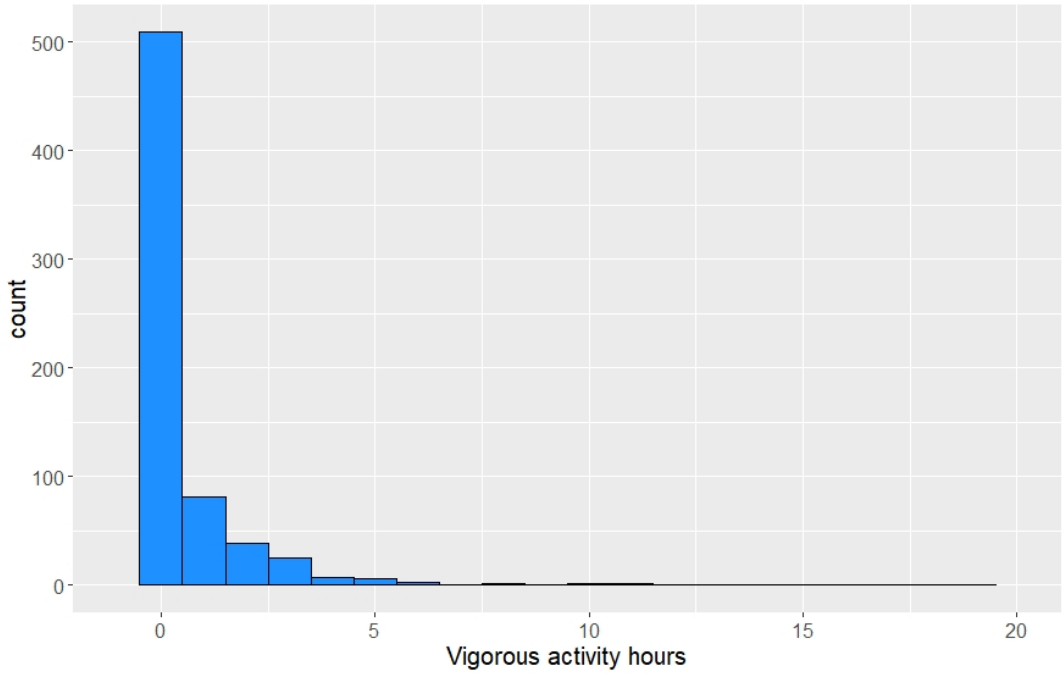

FIGURE 4.8: Histogram of vigorous activity hours in wave three for non-participants in the model

	Participation Status:	
	Full Participation (N = 6,113)	Missed Participation (N = 4,195)
Marital Status		
Married/cohabiting	4,670 (77)	2,938 (70)
Single	928 (15)	762 (18)
Divorced	430 (7)	403 (10)
Widowed	71 (1)	68 (2)
Ever told they had depression		
Yes	660 (12)	322 (15)
No	5,051 (88)	1,800 (85)
General health last year		
Very Good	1,983 (34)	649 (26)
Good	2,514 (43)	1,060 (43)
Average	1,084 (19)	606 (25)
Poor	216 (4)	117 (5)
Very poor	53 (1)	28 (1)
Presence of longstanding illness		
Yes	1,381 (31)	1,046 (33)
No	3,120 (69)	2,107 (67)

TABLE 4.5: Distributions of potential missing data predictors across different participation groups

4.2.3 Plausibility of MAR in Whitehall II

In the previous section, we have seen how a major contributor to missing data within the Whitehall II dataset is non-participation. The exploration of variables that might be associated with non-participation has highlighted a number of variables that might be predictive of non-participation. Specifically, characteristics such as sex, ethnicity, education level, employment status and marital status might be associated with non-participation as might health variables surrounding depression diagnosis and general health. Using these variables to supplement out missing data treatment will therefore improve the plausibility of the MAR assumption.

Outside of non-participation, another prominent missing data pattern is the high missing data rates found in the diet and physical activity indicators. This pattern emerges due to the combining of multiple items to form a summary measure. If any of the individual items are missing, then the derived score is also missing. As a result, this data may be missing via a different mechanism than non-participation.

As an example of the patterns of missing data in the diet variables, Figure 4.9 shows the different missing data patterns within the FFQ items that are included in the sugar consumption frequency indicator in wave 5. Within this it is clear that the main source of missing data is still non-participation, likely due to individuals not participating in this wave, but participating in either wave 7 or 9. Assuming the variables associated with non-participation overall are still associated with non-participation in individual waves, the MAR assumption should still be plausible for this portion of the missing diet data. The other missing data in the variables is perhaps mostly MCAR, since many of the remaining missing values are scattered evenly across variables.

The missing data patterns within the physical activity data displayed in Figure 4.10 paint a different picture, one in which there is a greater risk of data being MNAR. Specifically, there is a lot more data that is missing for other reasons than non-participation, and within these missing data patterns, some variables appear to be more likely missing than others. For example, the second 'other housework' variable and the 'other gardening' variable are missing in more missing data pattens than is the case for other variables.

This may indicate that these have been left blank for a reason that depends on their own value. For example, the participants may have left these items blank because they did not do any exercise of this type. This would make the

data MNAR. However, without following up on these participants, this is a strong assumption to make. Furthermore, similar variables such as 'cooking', 'hanging washing', 'lawn mowing' and 'weeding/hoeing' also feature in the dataset and have much lower missing data rates. Hence we may be able to leverage information provided by these similar variables to make the MAR assumption plausible.

The Spearman rank correlation between 'carrying washing' and 'other housework' in the sample is 0.26, and rank correlation between weeding and 'other gardening' is 0.65, hence suggesting that observed values of the similar variables are associated with observed values of our potentially problematic variables.

We can also examine whether these other variables might be predictive of the missingess found in 'other gardening' and 'other housework' using logistic regressions. Specifically, regressing the missingness indicator for 'other gardening' onto the values for weeding suggested that being one response category higher on the weeding variable is associated with a decrease in the odds of data being missing on 'other gardening' of 0.93 on average. Similarly regressing the missingness indicator of 'other housework' onto responses for cooking suggested being one response category higher on the cooking variable is associated a decrease in the odds of data being missing on the 'other housework' variable of 0.89 on average.

As such, the observed values of these similar variables are somewhat predictive of the missingness found in the 'other gardening' and 'other housework' variables. Therefore, using the observed data on other physical activity items provides a way for us to recover missing data in way that leverages the relationships between items, and does not require the making of strong untestable assumptions. For example, we do not need to make the strong assumption that missing data should be replaced by zeros. Overall, the use of these additional variables again makes the MAR assumption plausible.

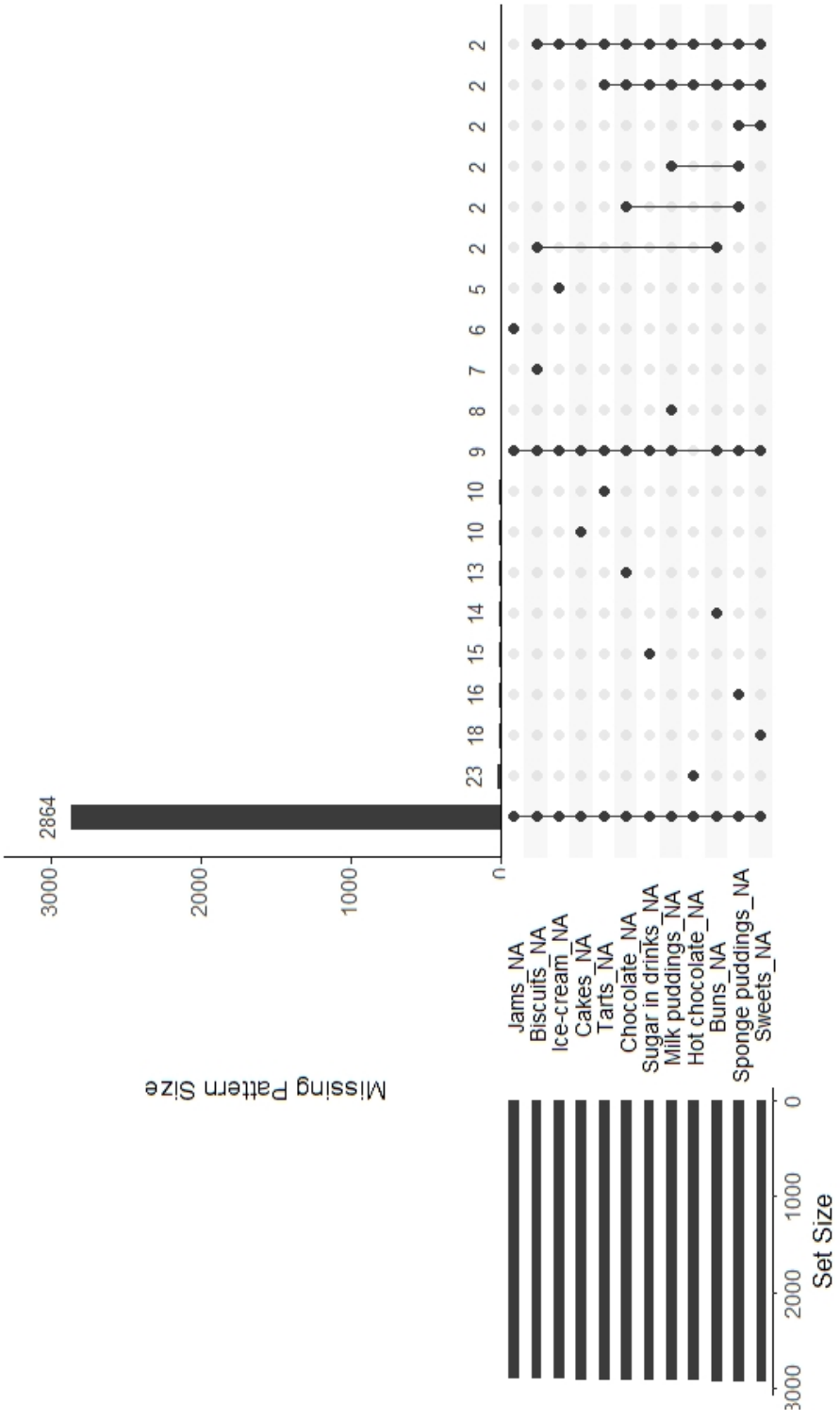

FIGURE 4.9: Missing data patterns for sugar consumption items in wave 5

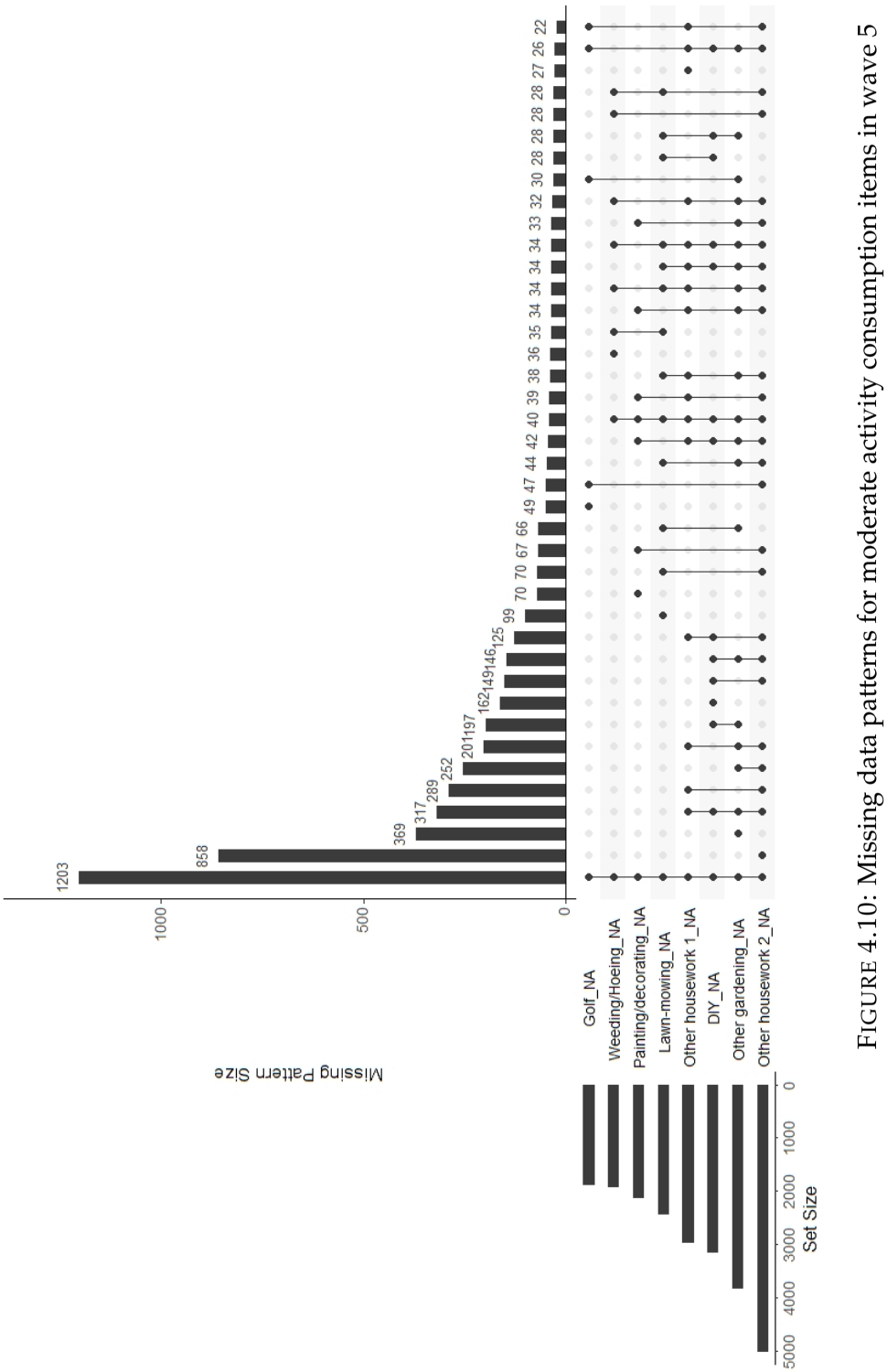

FIGURE 4.10: Missing data patterns for moderate activity consumption items in wave 5

4.3 Methods of Dealing with Missing Data

Having described some of the issues that can arise due to missing data, the different categorisations of missing data problems, and the particular missing data patterns that can arise in the Whitehall II dataset, in this section I will briefly review some of the available methods for dealing with missing data. A focus will be placed on describing the two main methods of dealing with missing data: Multiple Imputation (MI) and Full-information Maximum Likelihood (FIML). Within the missing data literature these methods are referred to as principled since they combine statistical assumptions with information from the observed data in order to deal with the missing data, instead of relying on quick, convenient fixes (Dong and Peng, 2013). Following this I will provide a brief comparison of the performance of MI and FIML, which will lead onto the final section where I outline the approach to dealing with missing data that I have taken in my analysis of the Whitehall II dataset.

4.3.1 Simple 'Ad-hoc' Methods for Dealing with Missing Data

As was mentioned in section 4.1, estimates of parameters in statistical models often require complete data for the estimate to be calculated. Historically, the issue of missing data has been dealt with using simple 'ad-hoc' fixes, which provide a quick and convenient solution to the missing data issue. More complete reviews of 'ad-hoc' missing data methods have already been conducted (see e.g. van Buuren (2012) and Schafer and Graham (2002)) and so here I will highlight some of the most well known methods for the purpose of discussing their limitations as tools for dealing with missing data.

4.3.1.1 Listwise Deletion

One of the most popular 'ad-hoc' methods for dealing with missing data is list-wise deletion, whereby cases with any missing data are removed from the analysis. However, list-wise deletion suffers from a number of limitations. For example, this approach can lead to problems when attempting to add new variables into a model, as this can affect the usable sample and hence make comparing the models difficult (van Buuren, 2012).

Removing cases from the analysis also comes at a cost to the power of any statistical analysis performed. In general, more cases gives the analysis more power to detect smaller observed effects, and so by deleting some cases we may miss the effect we are interested in observing. Lastly, individuals with

missing data may be systematically different in important aspects to those without missing data, and as such removing them may introduce bias into the analysis (Schafer and Graham, 2002)).

Listwise deletion is, however, a very simple method to implement, and so as a result can at times be useful when this simplicity is worth the cost of the potential bias. For example, in my analysis in Chapter 3, the simple regressions were run using listwise deletion. Since the aim was to explore some initial trends in the data, exact parameter estimates were not necessary and so taking the time to use a more considered missing data approach would not have added much to the analysis for the cost of additional time and complexity.

4.3.1.2 Pairwise Deletion

Another 'ad-hoc' method of dealing with missing data that operates in a similar vain to listwise deletion is pairwise deletion. Instead of deleting all cases that have missing data on any of the variables in the study, pairwise deletion instead calculates parameters with only the cases deleted that have missing data on variables necessary to calculate that parameter. For example, if one wanted to calculate correlations between three variables X_1, X_2 and X_3, then the calculation of the correlation between X_1 and X_2 could include cases with missing data on X_3 so long as the data was observed for both X_1 and X_2. This is the method of missing data treatment that was used to calculate the correlations analysed in Chapter 3.

Despite this method of dealing with missing data being useful for exploratory analysis where more principled methods of missing data treatment might make the analysis too cumbersome, it has limitations that make it unsuitable for an analysis of our substantive model. As with listwise deletion, unless the data are MCAR, their is a risk that the parameters derived using pairwise deletion are biased, and it is not possible to determine the size of this bias (Schafer and Graham, 2002). In addition, pairwise deletion also creates the possiblility of estimating impossible parameters such as correlations outside the range [-1,1] (Schafer and Graham, 2002).

Furthermore, studies comparing simple 'ad-hoc' methods of dealing with missing data with more principled methods invariably conclude that the simple methods suffer from severe limitations and are outperformed by more principled approaches (Ali et al., 2011; Allison, 2003; Olinsky, Chen and Harlow, 2003). As a result of these concerns, the use of principled methods of dealing with missing data is advised, even when the data are MCAR.

4.3.2 Single Imputation

As was mentioned before, estimating quantities of interest such as means, variances and regression coefficients requires that we have no missing data included in the calculation. When some data are missing however, instead of removing cases with missing data from the analysis we can generate a set of complete data by replacing any missing values with a plausible estimate of what its value might be. Quantities of interest can then be calculated as normal with these replacement values included.

The process above describes the general procedure behind imputation. When only a single plausible value is chosen for each missing data entry this is known as single imputation. Single imputation has the advantage over listwise and casewise deletion in that it doesn't reduce the power of one's analyses, since all the cases in the dataset are retained. When multiple studies are planning to run analyses on the same dataset, single imputation also has the benefit of creating a dataset that will be the same across all the studies, aiding comparability of results.

Despite these benefits, single imputation does has its drawbacks. The precise weaknesses of a single imputation method depends upon the method used to impute the plausible value, but include biased estimates, increased false positive rates and underestimation of variances (Schafer and Graham, 2002). For example, one way in which missing values are imputed is by substituting in the mean value of the observed data. This method will force the mean of the complete data to be the same as that of the observed data, which may not truly be the case, and also underestimates the variance since more data is now clustered around the mean.

There are many methods by which one can generate replacement plausible values. These methods have been extensively reviewed elsewhere, (see Schafer and Graham (2002) and van Buuren (2012)) and so here I will focus on outlining the methods that I have used in my analysis of the Whitehall II data.

4.3.2.1 Hot Deck and K-Nearest Neighbour Imputation

Hot deck imputation is a method of imputation in which missing values are replaced with observed values from a 'similar' unit (Andridge and Little, 2010). When data are only missing on a single variable Y, a 'deck' of variables $X_1, ..., X_n$ is selected on which cases in the sample can be compared. Then for each case i with a missing value on Y, a donor case j is chosen from a case

that matches closely on the X variables. The missing value y_i is then set as the value of y_j.

The method of matching may create a donor pool from which a value for y_i is chosen at random, know as random hot-deck imputation (Andridge and Little, 2010). In other cases a distance metric, $d(i, j)$, such as the Malhalanobis distance may be used to find the single donor that matches most closely to the recipient. This second method of hot-deck imputation is known as nearest-neighbour imputation (Beretta and Santaniello, 2016).

The nearest neighbour method of imputation can be extended to incorporate data from the k-nearest neighbours to a unit in the sample, in a process known as k-nearest neighbour imputation. In this method, values from the k-nearest neighbours are aggregated to derive the donor value. For example, for continuous data the mean of the k-nearest neighbours' values may be used and for catergorical data, a value could be selected based on the mode (Beretta and Santaniello, 2016; Kowarik and Templ, 2016).

When data are missing on multiple variables within a data set, the process of hot-deck imputation is more complicated. This is because the 'deck' of variables used to compare units in the sample needs to consist of only complete data, and so there may not to be a single deck of variables that is suitable for imputing every missing value on a variable. However, methods for dealing with multivariate missing data are available, and can deal with both monotone and general missing data patterns (Andridge and Little, 2010).

Hot-deck imputation is a non-parametric method of imputation, as there is no prescribed model from which the missing data are assumed to come. Since there is no prescribed model used to predict the missing values, hot-deck imputation may be less sensitive to model misspecification than methods that take a parametric approach, such as regression imputation (Andridge and Little, 2010). However, it is important to note that the results of the imputation still rely on choosing an adequate set of variables over which donor cases can be chosen. If poor choices of matching variables are used the resulting imputations may be of low quality.

Another advantage of this method is that only plausible values can be imputed, since values are drawn from another unit in the sample. This gives a level of face validity to the imputations that may not be present when imputing from a predictive distribution. Additionally, if variables used within the

hot-deck for matching are associated with both non-response and the variable being imputed, then non-response bias can be reduced (Andridge and Little, 2010).

Despite the benefits of using hot-deck imputation there are also downsides. One such downside is that in some cases it is not possible to find a donor from which a suitable value can be borrowed. If a unit with missing data is substantively different on the observed variables to other units in the sample then there may not be an appropriate value for it to receive. This can particularly be an issue when the sample size of the data is small (Andridge and Little, 2010).

4.3.3 Multiple Imputation

The ideas outlined above can be extended upon to create a more robust treatment of missing data that better incorporates uncertainty over what value should imputed into the data. Rather than using a single plausible value, Multiple Imputation deals with missing data by creating multiple complete datasets each with their own set of plausible replacement values. Each of these datasets are then analysed simultaneously using complete data methods (van Buuren, 2012) and the quantities of interest from each of these analyses are combined using rules known as Rubin's rules (Rubin, 1987).

Rubin's rules assume that repeated estimates of a parameter θ produced by the analysis of each imputed dataset follow a Normal distribution. When this does not hold for a given parameter, that parameter can be transformed before Rubin's rules can be applied. The overall estimate for the parameter θ is then given by:

$$\bar{\theta} = \frac{1}{m} \sum_{i=1}^{m} \theta_i,$$

where m is the total number of imputed datasets, and θ_i is the estimated parameter value in dataset i.

The standard error of $\bar{\theta}$ is then calculated by combining uncertainty from within each imputation and across imputations. Denoting by σ_{θ_i} the standard error estimate for theta in dataset i, the within imputation variance is given by

$$V_W = \frac{1}{m} \sum_{i=1}^{m} \sigma_{\theta_i},$$

and the between imputation variance is given by

$$V_B = \frac{1}{m-1}\sum_{i=1}^{m}(\theta_i - \bar{\theta})^2.$$

The total variance of θ is then estimated as

$$V_T = V_W + \left(1 + \frac{1}{m}\right) V_B$$

and the standard error estimate $\sigma_{\bar{\theta}}$ is then equal to $\sqrt{V_T}$. The above estimates can then be used for significance testing via the Wald test, as described in (Marshall et al., 2009).

4.3.3.1 Joint Modelling and Full Conditional Specification

Broadly, there are two main methods used to generate the plausible values to be imputed: Joint Modelling (JM), and Full Conditional Specification (FCS) (van Buuren, 2012). Under the Joint Modelling method, the data is assumed to be described by a multivariate distribution, usually (but not necessarily) the Multivariate-Normal distribution. Missing data values are then imputed from the relevant marginal distribution for the missing data pattern associated with that case, conditional on the observed data. For example, if some data Y follow a Multivariate-Normal distribution with parameters $\theta = (\mu, \Sigma)$ and a case in the data has missing data on the first two variables in the dataset Y_1 and Y_2 but is observed on all other variables, then the missing data will be drawn from a bivariate Normal distribution described by $P(Y_1, Y_2 \mid Y_3, ..., Y_n, \theta)$ (van Buuren, 2012).

In practice the underlying parameters θ of the distribution are unknown (since some data are missing) and in general it can be hard to estimate them from observed data (van Buuren, 2012). As a result an algorithmic approach is used to create the imputed datasets.

Suppose that we wish to generate M imputed datasets and that we have data $Y = (Y_{obs}, Y_{mis})$. Let $\dot{Y}^t_{mis}$ denote the imputed values for Y_{mis} in data set t and $\dot{\theta}^t$ the estimated parameters of the multivariate Normal distribution at step t of the algorithm. Then the algorithm for generating the imputations is as follows:

1. Specify a plausible set of starting values for $\theta^0 = (\mu^0, \Sigma^0)$.

2. For $t = 1, ..., M$ imputed values are generated by alternating between two steps
   (a) $\dot{Y}^t_{mis}$ are predicted using $P(Y_{mis} \mid Y_{obs}, \dot{\theta}^{t-1})$
   (b) $\dot{\theta}^t$ is predicted using $P(\theta \mid Y_{obs}, \dot{Y}^t_{mis})$
3. Once $\dot{Y}^t_{mis}$ have been imputed for $t = 1, ..., M$ the algorithm stops.

In other words, $\theta^0 = (\mu^0, \Sigma^0)$ is used to impute missing values for data set one, which is in turn used to generate parameter estimates for θ, which in turn are used to generate imputations for the second dataset and so on until all M datasets have been imputed. Additional information on the process of imputation via Joint Modelling, can be found in van Buuren (2012) and the references therein.

In comparison to the Joint Modelling approach, Full Conditional Specification does not describe the joint distribution of the data. Instead, variables are imputed on a variable-by-variable basis in which the multivariate distribution of the data is specified using a collection of conditional densities (van Buuren et al., 2006). Specifically, a missing data model is described for each variable with missing data in the data set. The missing data model for a variable with missing data is specified conditional on other variables in the dataset as well as the missing data pattern. Algorithms such as the MICE (Multiple Imputation by Chained Equations) algorithm (van Buuren and Groothuis-Oudshoorn, 2011) then generate imputations by iterating over these conditionally specified models a fixed number of times.

4.3.4 Full Information Maximum Likelihood

The other main method of dealing with missing data is Full-Information Maximum Likelihood (FIML). In FIML, missing data is not imputed, but instead the parameters of the substantive model are estimated directly using all available information in the dataset. The procedure is an extension of standard Maximum-likelihood (ML) methods, which work as follows. For each case in the data set the log-likelihood of each case's observed data is obtained and then these are summed together to give an overall log-likelihood for the sample data. Maximising this log-likelihood then provides the parameter estimates that are interpreted to be most likely to have generated the observed data.

The main difference between standard ML and FIML is that in the former, the log-likelihood includes observations on all variables for each case, whereas in FIML, the log-likelihood for one case may use different variables to the log-likelihood for another due to the two cases having a different missing data pattern. The usual assumption made in FIML, is that the data are Multivariate-Normally distributed. Then, the log-likelihood for case i is given in equation 4.3, from Enders (2001):

$$\log(L_i) = K_i - \frac{1}{2}\log(|\Sigma_i|) - \frac{1}{2}(x_i - \mu_i)^T \Sigma_i^{-1}(x_i - \mu_i), \tag{4.3}$$

where x_i gives the vector of complete data for case i, μ_i gives the mean estimates from the entire sample, and K_i is a constant that depends on the number of complete data points in case i. In equation 4.3 both μ_i and Σ_i are based only on the variables that are observed for case i (Enders, 2001). The overall log-likelihood of the data is then calculated by summing each of the case-wise log-likelihoods, since each of the cases are assumed to be independent. Hence the log-likelihood of the data is given by equation 4.4. To find the parametrisation that best fits the observed data an algorithm then iterates over different combinations of μ and Σ until it finds the combination of these parameters that maximises the value of equation 4.4.

$$\log(L(\mu, \Sigma)) = \sum_{i}^{n} \log(L_i) \tag{4.4}$$

Our analysis in Chapter 3 demonstrated at many of the variables we wish to use in our model are not Normally distributed. If the data in the model were Multivariate-Normal, then the marginal distribution of each variable would be a univariate Normal. Since this is not the case for the marginal distributions of our variables, we know that the assumption of Multivariate-Normality does not hold, and so estimating parameters and making inferences based on this assumption could be problematic. However, violations of the Multivariate-Normality assumption can be adjusted for using Robust FIML procedures (RFIML), such as those given in Yuan and Bentler (2000). These methods still generate parameter estimates for the model using equations 4.3 and 4.4, however, the standard error estimates are modified to compensate for the non-Normality of the data.

4.3.5 Performance of MI and FIML

In general, both methods perform well when dealing with missing data that is Multivariate-Normal and at most MAR. Schafer and Graham (2002) demonstrated that, in line with their theoretical properties, both MI and FIML performed well under both MCAR and MAR conditions when data was missing on a bivariate normal distribution. Even in the authors test of a small sample size and high missing data scenario, parameter estimates produced by both methods had acceptable levels of bias. Bias in the confidence intervals was also able to be remedied in a high missing data circumstance by increasing the sample size. Enders and Bandalos (2001) showed that FIML outperformed listwise deletion, pairwise deletion and single imputation that used a similar response matching condition when conducting a Confirmatory Factor Analysis with Normally distributed indicators.

Studies investigating the use of MI with non-Normal data, however, have had more mixed results, and have often served to highlight the limitations of the methods. Results of MI using Joint Modelling have been particularly poor, suggesting that issues can arise when the underlying distributional assumptions don't hold. Leite and Beretvas (2010) found that using Multivariate-Normal based MI (MI-MVN) could lead to bias when imputing likert data, especially when the data were more than 10% missing. Statistical tests based on MI in these conditions were also underpowered. Lee and Carlin (2012) also found that when missing data rates were high, Multivariate-Normal based MI can become unreliable and introduce bias. In addition, missing data on exposures of interest were also found to limit the benefits of MI.

An extensive examination of MI-MVN's robustness to non-Normality was included in Jia (2016), which found that MI-MVN was mostly robust to non-Normality, except under certain conditions when data are severely non-Normal. For non-Normal continuous data in SEM, MI-MVN was not unacceptably biased, however, when data were severely non-normal, confidence interval coverage was considered inadequate when the data were either missing in the tail of the distribution or the sample size was small. MI-MVN on non-Normal categorical data also performed well in all but the most extreme conditions: missing data in the tail of severely asymmetric distributions.

Results from studies using MI with FCS have been mixed when examining non-Normal missing data. Despite MI with FCS's theoretical weaknesses,

van Buuren et al. (2006) showed that it produced essentially unbiased estimates in both linear and logistic regression, even when the implied joint distribution didn't exist. In a recent thesis examining the performance of MI and Robust FIML (RFIML) in SEM with non-Normal data, MICE with predictive mean matching (PMM) performed well for non-Normal continuous data in under all conditions that were investigated, including high missing data (30%), severe non-Normality, and small sample size (n=300) (Jia, 2016). Additionally, it outperformed the other methods investigated (RFIML, MI-MVN and two other implementations of the MICE algorithm) across the conditions investigated. However, performance of PMM was not investigated for categorical data, and performance of the other MICE methods was generally poor.

RFIML methods have been suggested to suffer from theoretical limitations that limit their use to MCAR data. For example, the methods presented in Yuan and Bentler (2000) were described as applicable to one of two assumptions: Normal data which was MAR, or non-Normal data that was MCAR. Despite this, there are promising results for the performance of RFIML when data are both non-Normal and MAR. Jia (2016) found that RFIML performed well for both non-Normal continuous and categorical data with only a few exceptions. Specifically, for non-Normal continuous data RFIML was always in the acceptable range for bias, and confidence interval coverage was only inadequate when data were MAR in the tail of the distribution. For categorical data, problems only arose when data were dichotomous or three-category, and severely asymmetric. Savalei and Falk (2014) similarly found that standard RFIML could perform badly when the missing data rate was above 30% and the missing data was concentrated in the heavy tail of the distribution. Performance was adequate outside of this condition, however.

4.4 The Whitehall II Missing Data Strategy

4.4.1 RFIML vs MI in the Analysis of Whitehall II

Despite some of the performance limitations highlighted in the above literature, both MI and RFIML are likely to be suitable methods for dealing with missing data that are found in the Whitehall II study. Here we will discuss some of the pros and cons of each method in the context of the particular missing data issues that arise in the Whitehall II data set. In particular we will discuss the issues of: high missing data rates, potential violations of the

MAR assumption, and the problem of drop out due to death. General comparisons between the two methods will also be made. Finally, I will outline the method I have chosen to deal with missing data in my analysis of the Whitehall II data set.

4.4.1.1 Missing Data Rate

As was noted in section 4.2, some variables in the Whitehall data set have a missing data rate that is greater than 30%. This could be particularly problematic if the data are missing in the tail of distributions that are heavily kurtotic, as under these conditions, as both MI and RFIML can perform badly (Jia, 2016; Savalei and Falk, 2014). Performance issues may be ameliorated by the large sample size in the Whitehall II data set. In Jia (2016), bias and confidence interval coverage improved within each missing data condition as sample size increased, and Schafer and Graham (2002) found that both MI and FIML performed well even when over 70% of the data was missing so long as the sample size was not very small. Whilst this second study did not include MI with FCS or RFIML techniques, it seems reasonable to expect that both would experience similar benefits from a large sample size. However, without evidence from simulation studies, this is of course not certainly the case.

4.4.1.2 Violations of MAR

One of the main issues that can worsen the performance of both MI and RFIML is violation of the MAR assumption. It is very difficult to assess whether data are MNAR because, fundamentally, the missing data value is unknown. However, it is possible to strengthen the likelihood of the MAR assumption holding, by including auxiliary variables into the treatment of missing data. An auxiliary variable is not of substantive interest in the model, but is included in case it is a significant predictor of missingness. If any of the missingness predicted by the auxiliary variable was uniquely predicted by that variable, then failing to include it would lead to data that is MNAR.

Adding auxiliary variables into an imputation model is both intuitive and straightforward. One simply adds the auxiliary variable into the list of variables that are to be imputed, along with a description of which variables it is a predictor of in the imputation model (van Buuren, 2012). Historically, this simplicity gave it an advantage over RFIML, where adding auxiliary variables was much more complicated. However, in the context of SEM, modern

software such as MPlus can now add in auxiliary variables just as easily as can be done in MI (Asparouhov and Muth, 2008) with results that are as good as those that can be achieved with MI (Graham, 2003). This is the case even when the auxiliary variable itself has a high proportion of missing data which is MNAR (Enders, 2008).

This perhaps lends ML based methods an advantage over MI with FCS, as adding auxiliary variables directly into the model estimation rather than during an imputation step avoids issues that can arise due to in-congeniality between the imputation model and the substantive model. Simply put, the analysis model and imputation model are congenial if there is a joint model whose conditionals include both analysis model and the imputation model (van Buuren, 2012). This condition effectively ensures that the two models do not imply that substantially different relationships exist within the data. Imputing with an in-congenial model would be problematic since it could artificially alter the relationships that we are investigating in the substantive model. For example, this can arise in the situation where the substantive model contains many non-linear relationships, with a typical MI FCS imputation model usually having only linear relationships between the variables. Since we are investigating a complex latent structure for our data, using a linear MI technique such as FCS unchanged has the potential to influence the results unwontedly unless alterations are made to the imputation model (Bartlett et al., 2015).

4.4.1.3 Dealing with Censorship due to Death

The Whitehall II study has been collecting data from its sample for over 30 years. Given that the initial sample members were between 35 and 55 years of age, it is of no surprise that some of these individuals have died during the study. Between waves five and nine 648 participants died and across the other waves a total of 766 participants died. These dead individuals present a particularly complex missing data issue, especially when attempting to do Multiple imputation. Participants who have died during the study have useful data in the waves in which they are alive, however, including them in the sample to be imputed on results in improper values being imputed for them on variables that were collected in waves after they had died. These improper values should not be used in the final data analysis, as if we include them any findings and conclusions would be based on the analysis of data

that is impossible to attain in the real world, thus invalidating them. However, simply re-censoring the individuals who have died after imputation, may reduce the overall reliability of the method, as the improper imputed values may have unduly influenced the imputed values of the living. For example, Ning et al. (2013) found that imputing values for those who died in a study of cardiovascular health, resulted in an increased assignment of difficulties in activities of daily living to those who had missing values but were not deceased.

Despite these issues, there are methods suggested in the literature to deal with data that is missing due to death. Ning et al. (2013) suggested that imputing using a series of cross-sectional imputation models, one for each wave which includes only the participants who were alive in that wave was preferable to imputing on the whole sample and then re-censoring. Harel et al. (2007) suggested a two stage imputation method in which first one imputes a time to death for all participants, and then uses this information to improve imputations on the other variables. Whilst these methods may be suitable and preferable in some analysis cases, in the context of our Whitehall analysis they have a number of drawbacks. Firstly, the method described by Harel et al. (2007) relies upon on having a variable in the dataset that gives an accurate time and date of death. No such variable exists in the Whitehall II dataset and so this method is not useful in this situation. The methods described in Ning et al. (2013) are more feasible in terms of practicality, however, creating many different imputation models for each wave of data is cumbersome, and the cross-sectional nature of each imputation model means that we will be losing the predictive power of longitudinal relationships that exist between variables in different waves. The cross-sectional imputation strategy could therefore affect the parameters of longitudinal relationships in the data. Lastly, the cross-sectional imputation strategy may also affect the plausibility of the MAR assumption. Data collected at previous waves may well be predictive of missing data in future waves, particularly when it comes to drop out. For example, it is easy to imagine a case in which an individual with poor health in one wave drops out of a study in the next waves.

RFIML has a significant advantage over MI approaches in this regard. Since no data are imputed, we completely avoid the problem of creating improper values entirely. Parameters are estimated based only the available data and as such any inference done is based only the cohort who are alive at any time

point, rather than a hypothetically immortal one (Wen, Terrera and Seaman, 2018).

4.4.1.4 Other Issues

Another advantage of RFIML over MI in general is that the model parameters do not depend on any imputed values. In practice, this means that if a different researcher were to re-run the analysis on the same data, using the same code, they would obtain the same parameter estimate and results. With an imputation based method, the same parameter estimates would only be obtained if exactly the same imputed datasets were analysed. If the data was re-imputed, even using the same method, parameter values would almost certainly vary. Whilst this is perhaps unlikely to affect the substantive conclusions that one draws from the analysis, it is still a source of additional uncertainty.

Using RFIML also avoids a number of practical issues that arise when using Multiple imputation. Firstly, issues may arise when attempting to impute derived variables in a dataset, as failure to specify the derived variable's relationship with its component variables could generate imputed data that are inconsistent. Creating imputations using FCS also requires extensive post-imputation diagnostics to make sure that the imputed data that have been generated are reasonable, and also to ensure that the algorithm for generating imputations has converged (van Buuren, 2012).

4.4.2 The Overall Strategy

Keeping in mind the discussion above, I chose to use RFIML as my method for dealing with missing data when Analysing the Whitehall II dataset. Performance of RFIML was largely similar to MI approaches in the literature, and overall it is a practically simpler technique to implement in my research context. Firstly, it does not require the construction of a complex imputation model that would be required by our dataset. Instead, MPlus is able to implement the missing data strategy alongside the analysis using robust Maximum Likelihood estimation similar to that which is described in Yuan and Bentler (2000).

The only additional work required is to specify the auxiliary variables that are needed to make the MAR assumption more plausible, which are then handled alongside the analysis procedure by MPlus using the saturated model

method outlined in Graham (2003). The advantage that using MI has, in general, over RFIML in terms of adding auxiliary variables to improve the missing data model and make the MAR assumption more reliable is undermined in our setting due to the modern auxiliary variable capabilities of ML methods, as well as the fact that sequential imputation would be required to avoid potentially damaging effects of improper imputations of variables for individuals who have died. It is worth noting, that despite its demonstrated improvements over standard RFIML procedures, I have not used the two-stage RFIML approach outlined in Savalei and Falk (2014), because at the time of performing the analysis, this method was not available within MPlus.

Two kinds of auxiliary variables will be included in the model. The first kind is those variables that are correlated with missingness in the data. Specifically, I will be including the variables examined earlier in this chapter whose distributions differed substantially between the participants with missing data, and those without. This criteria leads to the selection of the following variables to be included: ethnicity, marital status, education level, employment grade, depression history and general health. The second kind of auxiliary variable to be included in the model is those variables that are highly correlated with measures in the substantive model (Collins, Schafer and Kam, 2001). To include variables of this nature, I have chosen to include measures of the model constructs taken in earlier waves of the Whitehall dataset, where available. This means including the following variables in the model as auxiliary variables: BMI, waist circumference, daily fruit consumption frequency, daily vegetable consumption frequency, GHQ-based depression score, SF-36 mental health score, SF-36 physical function score, SF-36 physical role limitation score, SF-36 bodily pain score, total hours of vigorous activity, problems paying bills and problems with housing. All of these measures are taken from wave three except for the SF-36 variables which are available at wave four.

On top of the overall strategy of using RFIML, I will also be using single imputation to recover some of the missing data within the physical activity and diet data. In particular, if participants had responded to 10 or more of the physical activity items in a particular wave, then missing values were imputed for their responses to the remaining physical activity items in that wave. Similarly, if a participant responded to more than half of the FFQ items in a given wave, the remaining values for their FFQ data for that wave were

imputed. The imputations were carried out using kNN imputation. In this approach, missing values are imputed by borrowing an observation from another participant who is one of k neighbours that closely matches that of the participant with the missing entry. Closeness was assessed based on the Gower distance between participants using the kNN function with 10 nearest neighbours from the R package VIM (Kowarik and Templ, 2016).

Using 10 neighbours was chosen as a compromise between improved prediction of the imputed values and preserving the underlying data structure. When imputing Likert data, as we are here, it has been suggested that using the square root of complete cases in the data to set the number of neighbours obtains optimal imputed values (Jonsson and Wohlin, 2004). Given the large sample size of the Whitehall II data, the number of complete cases for the physical activity and diet questionnaires ranged between 1252 and 3334, suggesting that between 35 and 57 nearest neighbours could be used. However, previous investigations of the kNN imputation method have shown that using more than one neighbour can result in substantial distortions of the underlying data structure, even when as few as five nearest neighbours are used (Beretta and Santaniello, 2016). Hence, 10 neighbours were used to provide a compromise between these two recommendations. This value of k was used for all imputations to provide consistency in the way missing data was imputed.

4.5 Conclusions and Next Steps

In this chapter, I have developed and described the principled approach to dealing with missing data that will be used in the Structural Equation Modelling analysis presented in Chapter 5. Overall, the development showed that there are significant missing data challenges involved with this analysis, including high rates of missing data, challenges in supporting the MAR assumption and censorship resulting from participants in the cohort dying during the study period. However, the approach of using Robust FIML supplemented with auxiliary variables and single imputation in cases of very high missing data rates in derived summary measures provides a pragmatic way to minimise the impact of missing data on our analysis without making the analysis impracticably cumbersome.

Having outlined this missing data strategy, in the next chapter I will build

upon the exploratory analysis in Chapter 3 by examining diet, physical function and physical activity as potential mechanisms between obesity and depression using a structural equation model.

Chapter 5

A Structural Equation Model of the Relationship Between Obesity and Depression

5.1 Introduction

In Chapter 3, I conducted an exploratory analysis of the Whitehall II dataset with two main aims. The first aim was to explore the statistical properties of the distributions of variables in the Whitehall II dataset that might be used in a SEM model. Overall the analysis demonstrated that there were frequent deviations from Normality in the distributions of available measures, due to skewness in the data and also due to some measures being categorical in nature. The findings around this aim fed into the work presented in Chapter 4, where I analysed the patterns of missing data within the Whitehall II data set, before describing the missing data procedures that I will be using to handle this missing data in the Structural Equation Model presented in this chapter.

The second aim of Chapter 3 was to examine how the measures that might be used in the model relate to one another, particularly in terms of their correlations, in order to give some preliminary insight into what relationships exist within the model. Overall, that exploratory analysis found little evidence to support a direct relationship between obesity and depression and that, of the proposed mediators (diet, physical activity level and physical function) physical function might be the only one to have a mediating influence between the two variables.

In this chapter I will expand upon the analyses presented in Chapter 3, in

order to more thoroughly investigate whether there is support for the hypotheses presented below. Note that, due to changes in the measures used to represent the diet construct, the precise specification of the diet hypothesis has changed since Chapter 3. In particular, the diet construct now focuses on energy intake rather than consumption frequency. More detail on these changes will be given on in section 5.3.2.2.

This chapter is broken up into four further main sections. In section 5.2, I will briefly outline the specific type of Structural Equation Model that I will be using, followed by an aside on the concept of determining goodness of fit in SEM. In section 5.3, I will then present the analysis of the measurement model, before the structural model is presented in section 5.4. The results of these section are then discussed in section 5.5 with reference to the exploratory analysis presented in Chapter 3, and the literature presented in Chapter 2.

Study Hypotheses:

- **Hypothesis 1:** *Physical function mediates the association between obesity and future depression.*
- **Hypothesis 2:** *Dietary energy intake mediates the association between obesity and depression bidirectionally.*
- **Hypothesis 3:** *Physical activity level mediates the association between obesity and depression bidirectionally.*
- **Hypothesis 4:** *Socioeconomic position (SEP) affects the relationship between obesity and depression via multiple effects on obesity, depression and their mediating variables*

5.2 The Cross Lagged Panel Model

To investigate the study hypotheses, I used a latent-variable multiple-group Cross-Lagged Panel Model (Usami, Murayama and Hamaker, 2019; Selig and Preacher, 2009; Little and Card, 2013; Cole and Maxwell, 2003). The equations in 5.1 describe a simple version of this model structure within a single group:

$$\begin{aligned} X_2 &= \beta_1 X_1 + \beta_2 Y_1 + \xi_{X_2} \\ Y_2 &= \beta_3 X_1 + \beta_4 Y_1 + \xi_{Y_2}. \end{aligned} \tag{5.1}$$

In this simple example, the variables X and Y are two different constructs measured at two time points. Auto-regressive effects, whereby a construct at one time point is associated with the same construct at a later time point, are estimated by the β_1 and β_4 terms, and the cross-lagged effects, whereby a construct at one time point is associated with a different construct at the next time point are estimated by β_2 and β_3. Lastly, ξ_{X_2} and ξ_{Y_2} are error terms. The model can be extended to include more constructs and more time points, as has been done in this chapter for my analysis (Little and Card, 2013; Zyphur et al., 2020).

The auto-regressive effects describe how individuals' standings on a construct at one time point effect individuals' standings on the same construct at a later time-point (Mulder and Hamaker, 2021). For example, if individuals with high values on the construct X at time 1 still have high values on X at time 2, and similarly for those with low values, then their will be a strong positive auto-regressive effect β_1. However, if individuals' standings on X at time 1 have little to no bearing on their standings on X at time 2, then the auto-regressive effect β_1 will be close to zero (Selig and Little, 2012).

The cross-lagged effects are interpreted similarly, with large β values representing individuals' standings on a construct at one time point being closely related to their standings on a different construct at the next time point, and small β values meaning that the standings are not closely related (Selig and Preacher, 2009; Selig and Little, 2012; Mulder and Hamaker, 2021). Of note is that the inclusion of auto-regressive effects means that cross-lagged effects have controlled for the impact of prior correlations between the constructs at an earlier time-point, preventing any bias that may have been introduced due to their omission (Selig and Little, 2012).

The above interpretation of parameters within the cross-lagged panel model is important for the type of conclusions we are able to make about our hypotheses. In particular, the model is evaluating the effect of differences between people, rather than the differences within people (Selig and Little, 2012; Mulder and Hamaker, 2021). For example, the model analysed here will be able to tell us whether having a higher body weight is associated with having a subsequent lower level of physical function, and similarly whether having lower physical function is associated with having a higher level of subsequent depression. However, what the model is not able to say, is how much an individual's personal change in body weight subsequently impacts their physical

function and how much in turn this affects their level of depression (Mulder and Hamaker, 2021).

5.2.1 Model Structure

The equations in 5.1 represent what is known as the structural model within the Cross-Lagged Panel Model. This part of the model describes the hypothesised relationships between the constructs of interest. Within the structural model, X and Y can be single manifest variables, or latent variables measured by multiple indicators. Generally, latent variables are used to represent constucts that cannot be directly observed themselves, and so must have their values inferred from direct observations of other variables (Little and Card, 2013). As an example, in psychology, personality traits might be inferred from observations of behaviour (Bollen, 2002). However, latent variables need not only be used for constructs that cannot be directly observed. Even when a construct is more directly observable, such as obesity, latent variables can be used as a means of separating out reliable variance in a measure from variance due to error (Little and Card, 2013; Bollen, 2002). For example, whilst BMI is a validated and trusted measure of body-weight that can be used to make inferences about obesity, it is not a perfect representation of whether one has excess adiposity; some individuals may have a high BMI due to a high muscle mass rather than excess adipose tissue.

The estimated size of relationships between constructs may be influenced by this measurement error (Lomax, 1986). Hence, we may prefer to try and remove these measurement variances so that only 'true' variability in the underlying construct is represented in the parameters of the structural model (Little and Card, 2013). As a result of these concerns, where possible in the this model, latent variables have been used to represent the constructs.

When there are latent variables in the model, the SEM also includes a measurement model, which specifies how the manifest variables reported in the dataset relate to the underlying latent construct (Little and Card, 2013). This measurement model must be analysed prior to the fitting of the structural model, as a poorly specified measurement model may lead to erroneous and misleading conclusions being made about the structural model.

In the model presented in this chapter, there are six main constructs (obesity, depression, physical function, physical activity, dietary energy intake and socioeconomic position) measured at three time points. Auto-regressive effects

analogous to those in 5.1 were included in the model, however cross-lagged paths were only included if they corresponded to a path in the hypothesised relationships. For example, physical function at time 2 was regressed on obesity at time 1 as this path represents part of hypothesis 1, however, obesity at time 2 was not regressed on physical function at time 1, as this path is not relevant to any of the hypotheses. Figure 5.1 shows a summary of the initial model structure that was used to examine the hypotheses. In the diagram different coloured paths are included to represent the different hypotheses, and 'general' paths are included that represent important paths that do not relate to any one hypothesis specifically. These paths are the direct cross-lagged effects between obesity and depression, and auto-regressive effects for each construct over time.

The constructs in the model are represented by a combination of manifest and latent variables, as no one type of variable was universally most appropriate. Hence, our model includes both a measurement model and a structural model. Details of how the measurement model was developed and analysed are given in section 5.3. Details on how the structural model was analysed are then given in section 5.4.

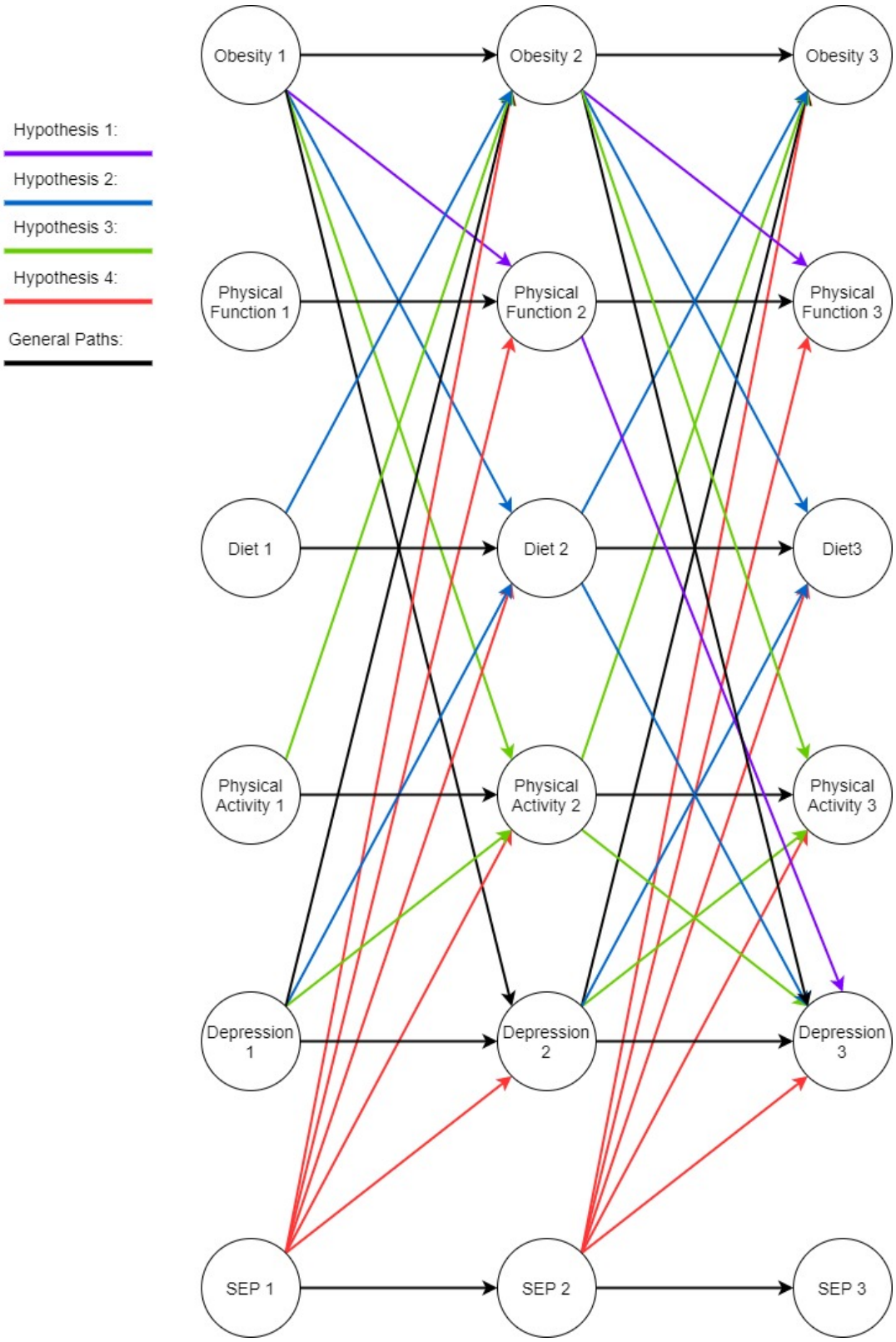

FIGURE 5.1: Initial Structural Model used to represent hypothesised relationships. *Note:* The structural model also includes covariances between the constructs at time 1, as well as covariances between the constructs' residual variances at time 2 and 3, however, these associations have been omitted from the diagram to avoid the figure becoming cluttered.

5.2.2 Model Fit

When estimating any Structural Equation Model, alongside the parameter estimates, it is important to understand additionally how well the model fits the data. When we estimate a Structural Equation Model we generate statistics that are implied by the model which can then be compared to equivalent statistics for the observed data (Little and Card, 2013). The principal statistic that is generated is the model's implied variance-covariance matrix Σ which can then be compared to the observed variance-covariance matrix for the data, S. Depending on the model, other statistics may also be compared, such as the model's implied mean values and the mean values of the observed data (Hox and Bechger, 1998; Little and Card, 2013). The closer the statistics reproduced by the model are to those that summarise the data, the better the model fits.

Deciding how to objectively assess the difference between the model and the data is not straightforward and as a result multiple fit indices have been developed to capture different aspects of how and why a model might fit poorly. Using multiple fit indices in this way provides a more comprehensive assessment of model fit. Rather than describing all the possible model fit indices that can be used, here I will outline the general principles behind the different fit indices, as well as some of the fit indices that I have used in my analysis.

5.2.2.1 The Model Chi-Square

One of the principal measures for goodness of fit within Structural Equation Modelling is the model Chi-Square value. The model Chi-square provides a way to test statistically whether the model fits the data against a null hypothesis that the observed variance-covariance matrix S does not differ from the model implied matrix Σ (Little and Card, 2013). A significant p-value from this Chi-square test for goodness of fit means there is evidence to suggest that the model does not fit the data, and conversely a non-significant p-value means there is not sufficient evidence to suggest the model does not fit the data (Cheung and Rensvold, 2002).

Whilst the Chi-square test has the advantage of providing a way to define statistically whether a model fits or not, it also has a major disadvantage. The Chi-square test of model fit is sensitive to sample size, and as a result

models that are estimated with larger amounts of data become very likely to fail the Chi-square goodness of fit test (Perry et al., 2015).

Despite this drawback, the Chi-square can still provide a useful way of comparing nested models. A model m_0, with degrees of freedom df_0, is nested inside another model m_1, with degrees of freedom df_1, where $df_0 \geq df_1$, if the estimation of m_1 is capable of reproducing the variance-covariance structure and mean structure of the model m_0 (Bentler and Satorra, 2010). In general a more narrow type of nesting is considered called *parameter nesting* which occurs when one the model m_0 can be created by fixing the value a parameter that is free in m_1 (Bentler and Satorra, 2010). When speaking further of nesting I will implicitly be referring to parameter nesting, since this is the form of nesting used in this analysis.

The fit of these nested models can be compared in the following way. Suppose that m_0 and m_1 have model Chi-square values χ_0^2 and χ_1^2 respectively. Then, when using Maximum-Likelihood estimation and assuming that the underlying data follow a Multivariate-Normal distribution, the value of the test statistic T given by

$$T = \chi_0 - \chi_1, \tag{5.2}$$

is distributed asymptotically to the χ^2 distribution with $df_0 - df_1$ degrees of freedom (Satorra and Bentler, 2001). As a result, one is able to test whether a more restrictive nested model has a statistically significant reduction in model fit as a result of the imposed restriction.

In Chapter 3, I found that many of the variables included in my analysis did not follow a normal distribution. Due to this, in our analysis I am using a robust form of Maximum-Likelihood estimation (see Chapter 4) and so the test statistic T cannot be used. Instead, a new test statistic must be used, TrD, which is given by

$$TrD = \frac{(\chi_0^2 c_0 - \chi_1^2 c_1)(df_0 - df_1)}{df_0 c_0 - df_1 c_1}, \tag{5.3}$$

where c_0 and c_1 are scaling corrections calculated by the robust Maximum-Likelihood estimation (Bentler and Satorra, 2010; Muthen and Muthen, 2022). As before TrD is asymptotically χ^2 distributed on $df_0 - df_1$ degrees of freedom.

5.2.2.2 Relative Fit Indexes

Another way of measuring model fit is to calculate a relative fit index, which compares the model with the worst possible fitting model. A frequently used relative fit index is the comparative fit index or CFI. The CFI is given by:

$$\text{CFI} = 1 - \frac{\max\left(\chi^2 - df\right), 0}{\max\left(\chi^2 - df\right), \left(\chi^2_{null} - df_{null}\right), 0'} \tag{5.4}$$

where χ^2 and df are the Chi-square value and degrees of freedom for the estimated model respectively (Little and Card, 2013). χ^2_{null} and df_{null} are then the Chi-square and degrees of freedom for null model. Since the CFI is a relative fit index, values closer to 1 indicate a better fit for the model. Historically a CFI value greater than 0.9 was necessary to consider a model to have good fit, however, some suggest a more stringent cut-off of 0.95 should be applied (Hu and Bentler, 1999).

5.2.2.3 Absolute Fit Indexes

Alongside relative fit indexes, one can also calculate absolute fit indexes, which compare the model with the best possible fitting model (one that has a theoretical χ^2 value of 0) (Little and Card, 2013). A commonly used absolute fit index is the root-mean-square error of approximation, or RMSEA. The RMSEA is given by:

$$\text{RMSEA} = \sqrt{\left(\frac{\chi^2 - df}{N - 1}\right) \Big/ \left(\frac{df}{g}\right)}, \tag{5.5}$$

where χ^2 is the estimated model's Chi-square value, df is the degrees of freedom of the model, N is the sample size, and g is the number of groups (Little and Card, 2013). The closer the RMSEA value is to zero the better the fit of the model. As with the CFI, judgements on how well a model fits are usually made based on cut-off values of the RMSEA value. There is no absolute consensus on this, however, a frequently followed guideline is that of (MacCallum, Browne and Sugawara, 1996) who suggest that RMSEA values below 0.01, 0.05 and 0.08 represent a model with excellent, good and mediocre fit respectively. An RMSEA above 0.08 is considered a poor fitting model.

Other fit indices can be calculated, however, since these were not used to assess the model fit in this chapter, they have been omitted from this section.

5.3 The Measurement Model

As was mentioned previously, when constructs are represented by latent variables, the structural equation model includes a measurement model that describes how those latent variables are represented by the observed variables (Little and Card, 2013). Covariances between the latent constructs are also represented, however, unlike the structural model no causal assumptions are imposed on these covariances.

The measurement model for a single latent construct X at time-point t is described in the equations in 5.6, where $x_{1,t}$, $x_{2,t}$ and $x_{3,t}$ are three measures of the same underlying construct all measured at time t. Each measure is regressed onto the latent construct with the regression co-efficients $\lambda_{1,t}$, $\lambda_{2,t}$ and $\lambda_{3,t}$ being called the factor loadings for $x_{1,t}$, $x_{2,t}$ and $x_{3,t}$ respectively. In this way, the values for $x_{1,t}$, $x_{2,t}$ and $x_{3,t}$ are separated out into the portion that is predicted by the latent construct X and error terms given by $\varepsilon_{x_{1,t}}$, $\varepsilon_{x_{2,t}}$ and $\varepsilon_{x_{3,t}}$ respectively (Little and Card, 2013).

$$\begin{aligned} x_{1,t} &= \lambda_{1,t}X + \varepsilon_{x_{1,t}} \\ x_{2,t} &= \lambda_{2,t}X + \varepsilon_{x_{2,t}} \\ x_{3,t} &= \lambda_{3,t}X + \varepsilon_{x_{3,t}}. \end{aligned} \tag{5.6}$$

Figure 5.2 shows a schematic of the measurement model for the same construct X measured at two time-points. In the diagram, ψ_{X_t} represents the variance of X at time point t and ψ_{X_1,X_2} represents the covariance between the constructs. When more time points and constructs are represented in the model, each construct is allowed to covary with every other construct, at every time point (Little and Card, 2013). For example, depression at time 1 co-varies with obesity at all time points, and vice versa.

Another thing to note is that the error terms for the same measures collected at different time-points are correlated. This is to represent the possibility that measurement errors on the same item are likely to come from the same source at different time-points, and so these error terms need to be correlated (Little and Card, 2013; Cole, Ciesla and Steiger, 2007). In a multiple group model, as is presented in this chapter, the measurement model is estimated separately in each group.

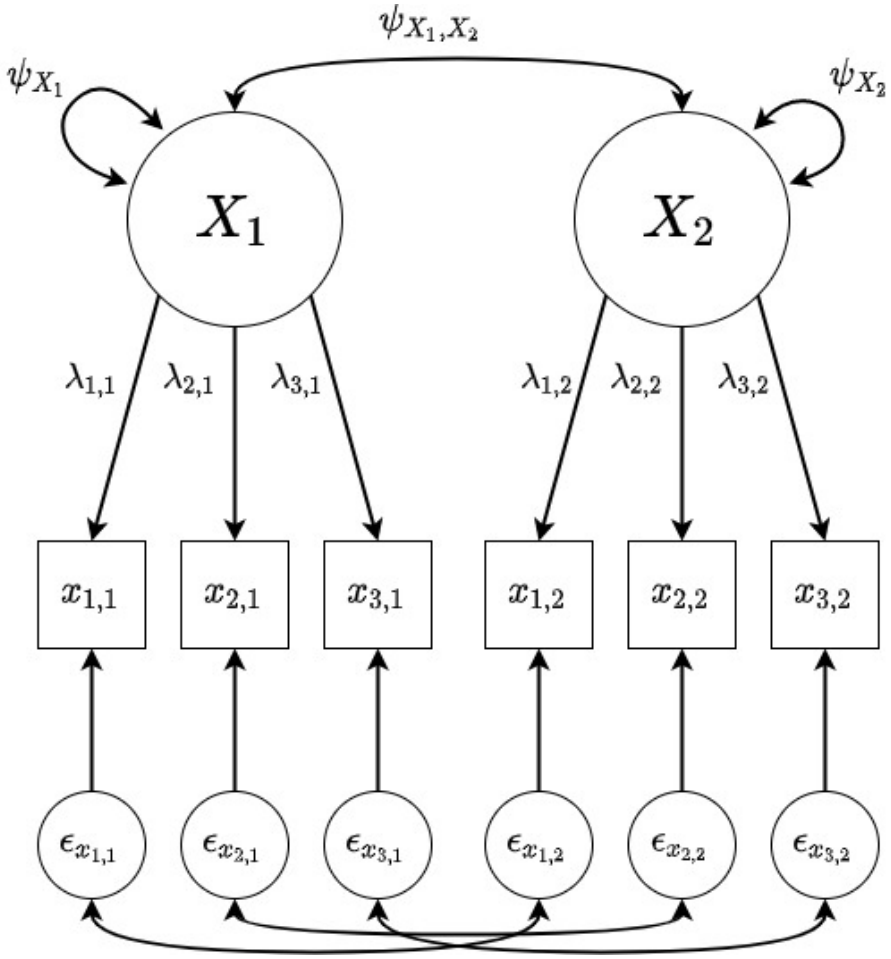

FIGURE 5.2: Schematic of measurement model for same construct measured at two time-points

5.3.1 Identification of the Measurement Model

In order for a Structural Equation Model to be estimable, it must be identified. Generally speaking, a model is identified if there is sufficient information to uniquely estimate all of the parameters in the model (Little and Card, 2013). In structural equation modelling, this condition is met if there is enough information in the variance-covariance matrix to estimate all of the free parameters in the model. Specifically, there needs to be more unique variances and covariances in the data's variance-covariance matrix than there are free parameters to be estimated in the model (Little and Card, 2013; Kenny, Kashy and Bolger, 1998). If both the structural model and measurement model are identified, the overall model is identified.

In the case of the measurement model, identification can be assessed by ensuring that the model meets the conditions which are described in more detail below. Firstly the latent variables in the model must have been scaled and

secondly there are sufficient indicators per latent variable. Other conditions on the correlations between indicators must also be met, however in practice, models that violate these conditions are rarely used and so I will not discuss these conditions beyond giving a brief description.

5.3.1.1 Setting the Scale of the Latent Variables

Since latent variables represent an unmeasured, perhaps unmeasurable concept, they have no inherent scale on which they are measured. As a result, we need to set a scale for the latent variable, so that variability in scores on the latent variable can be interpreted (Little and Card, 2013).

There are a two main ways of setting the scale, although an additional method is presented in Little, Slegers and Card (2006). The first method, known as the marker method, involves fixing the factor loading for one of the indicators on each latent variable in the model to 1. The other loadings are then freely estimated in the model. The choice of which indicator is used as the marker variable does not effect the overall fit of the model, however, it does effect how the scale of the latent variable is interpreted (Little and Card, 2013; Klopp and Klößner, 2021). Under the marker method, the latent variable inherits the metric of the marker variable, and the mean and reliable variance of the marker variable become the mean and variance of the latent variable. Given the fact an arbitrary choice of marker variable is able to alter the scale and interpretation of the latent construct, this method of scale setting is not recommended by some authors (Little and Card, 2013).

A second method of setting the scale, called the fixed factor method, involves setting the scale of the latent variable to 1, and allowing all the loadings of the indicators to be freely estimated. This method avoids the issue of the latent variable inheriting the scale of an arbitrary marker, and instead gives it a standardised metric. This means that covariances between the constructs are correlations (since both have a variance of 1) which can make interpretation of the model relationships easier (Little and Card, 2013).

The third additional method of scale setting introduced in Little, Slegers and Card (2006), is known as effects coding. In this method of scale setting, the factor loadings are constrained so that they average to 1. Under this method of scaling, the variance of the latent factor is the "average amount of reliable variance that each indicator contributes to the definition of the latent construct" (Little and Card, 2013; Klopp and Klößner, 2021).

In my analysis, I chose to use the fixed factor method of scale setting. This was chosen over the marker based method in order to reduce the arbitrarity of the metrics of the latent variables, and was chosen over the effects coding method for pragmatic purposes. The MPlus software has built in syntax for setting the scale of latent variables using the fixed factor method, whereas the effects coding method would have needed to be coded in manually. Not only would this require more time and effort for an arguably minimal gain, it would also have increased the likelihood of coding errors being introduced into the model.

5.3.1.2 Sufficient Indicators per Construct

The second main identification condition that the measurement model must satisfy is that there is enough information from the indicators to uniquely estimate all the parameters in each construct (Little and Card, 2013; Kenny, Kashy and Bolger, 1998). The information required is the unique variances and covariances from the indicators. For a construct with k indicators, there are k unique variances (one for each indicator), and $\binom{k}{2}$ unique covariances (one from each unique pair of indicators). Similarly, when a construct has k indicators, there are $2k + 1$ parameters to be estimated, comprising of k factor loadings, k residual variances, and 1 latent factor variance. Once the scale has been set by fixing the variance of the latent variable to 1 (or another scale setting method has been used), this constraint reduces the number of parameters that need to be estimated by 1, so that there are $2k$ parameters to estimate. Since we need more information than free parameters we require that

$$k + \binom{k}{2} \geq 2k \implies \binom{k}{2} \geq k, \tag{5.7}$$

where k is the number of indicators for a construct. It can be easily verified that this holds when $k \geq 3$. In fact, when there are three indicators there are exactly as many pieces of information as there are parameters to estimate, and the construct is said to be 'just identified', as there is no spare information from the indicators (Little and Card, 2013; Kenny, Kashy and Bolger, 1998). When there are more than three indicators the constructs are 'over identified', as there is now more information than is needed to estimate the parameters.

However, when two or fewer indicators are available for a latent construct, $\binom{k}{2} < k$ and so the construct is 'under-identified', as there is less information than there are parameters to be estimated (Little and Card, 2013). As a result, additional constraints must be placed on the loadings of the constructs

so that the construct is identified. If the construct has two indicators, there are three pieces of information (two variances and one covariance) and four parameters to be estimated. As a result a single constraint is needed to reduce the number of parameters by one. When adding this constraint was necessary in the model, the factor loadings were set to equality.

If only one indicator is available there is only one piece of information (the indicator's variance), and two parameters to estimate. Again, a constraint needs to be introduced which traditionally involves setting the indicator's residual variance to zero (Little and Card, 2013; Kenny, Kashy and Bolger, 1998). Another option is to make an assumption on how much of the indicators variance is 'reliable'. However, judging how much of the variance is reliable is somewhat arbitrary and so there is debate over whether this method is appropriate. In my model, when only one indicator was available for a construct, the indicator was included in the model as a manifest variable, as this is equivalent to having a latent variable with single indicator that has zero residual variance.

5.3.1.3 Other Identification Conditions

The first of the remaining conditions that needs to be satisfied for identification relates to the relationship between indicators on different constructs. In order for the measurement model to be identified, for every pair of constructs one of the following must hold (Kenny, Kashy and Bolger, 1998). Either:

1. there is, at least, one pair of indicators, one loading on one construct and one loading on the other, that does not have correlated measurement error between them or,

2. the correlation between the pair of constructs is specified to be zero (or some other *a priori* value).

This condition is necessary because if every pair of indicators has correlated measurement error, there will be insufficient information to estimate the correlation between constructs, as all the information from the covariances will be being used to estimate the size of the correlated errors.

The final condition needed for identification is that for every indicator, there is at least one other indicator with which it does not share correlated measurement error (Kenny, Kashy and Bolger, 1998).

Since indicators are usually assumed to have no correlated measurement error, these conditions almost always hold in practice. In the models presented in this chapter, the only indicators that have correlated measurement error are those that measure the same item but at different time-points e.g. SF-36 physical function score at time 1 has correlated measurement error with SF-36 physical function score at times 2 and 3 etc. As a result, both of the above identification conditions hold.

5.3.2 Estimation of the Measurement Model

Having outlined the general structure of the measurement model, here I will first describe the general testing procedure used to analyse the measurement model, before describing how this was implemented in the multiple stages of measurement model estimation. This testing procedure has been adapted form the procedures described in Little and Card (2013).

5.3.2.1 The Testing Procedure

Step 1: Estimating the Null Model

The first step in estimating the measurement model is to estimate an appropriate null model that provides a benchmark for how a reasonable 'worst fitting' model fits the data. As was seen in 5.2.2 the fit of the null model is important for calculating the fit measures of other models, namely the CFI.

Within this null model, the indicators in the model are assumed to have only a variance and all the observed covariances are assumed to be zero. In this way, it gives an indication of the amount of information contained in the covariances between indicators (Little and Card, 2013). When the sizes of the covariances between indicators are larger, restricting these covariances to zero in the model will result in a poorer model fit, and as such the null model's associated χ^2 value will be higher. Conversely, if the covariances of observed indicators are close to zero, the null model will fit less badly and have a lower χ^2 value.

In addition to the above constraints, in the case of a longitudinal multiple group panel model the variances and means of the indicators are assumed to be equal over time and across groups. For example, the variance of BMI would assumed to be the same at each time point and the same in men and women. These constraints represent the null expectation of there being no differences in the observed variables over time or across groups.

Step 2: Testing for Measurement Invariance

The second step in the procedure for analysing the measurement model is to test whether measurement invariance (also known as factorial invariance) holds. In a longitudinal Cross-Lagged Panel Model, we are essentially asking whether variability in a construct at one time point is associated with variability in a construct at a later time-point. When a construct is represented by a latent variable, we need to be sure that changes in the latent variable are due to true changes on the construct, and are not due to changes in the definition of the latent variable used to represent it. Testing whether latent variables have different definitions over time or across groups is called testing for measurement invariance.

There are three strengths of measurement invariance that are tested: configural invariance, weak invariance, and strong invariance. The test of configural invariance provides the weakest claim about measurement invariance, namely that the loadings all follow the same pattern over time and across groups. Weak invariance then posits that the factor loadings for the same measure are equal over time and across groups.

Suppose a construct $X_{t,g}$ with p indicators $x_{i,t,g}$ where $i = 1, \ldots, p$, and t and g denote the time-point and group. Suppose also that the factor loading of indicator $x_{i,t,g}$ is given by $\lambda_{i,t,g}$. Then weak factorial invariance is applied by setting $\lambda_{i,t,g} = \lambda_i$ separately for each value of i. As an example, to impose this constraint on the schematic from Figure 5.2 with two groups, we would set $\lambda_{i,1,1} = \lambda_{i,2,1} = \lambda_{i,1,2} = \lambda_{i,2,2}$ for $i = 1, 2, 3$, where the third index represents the group number.

Weak invariance is perhaps the most important type of invariance to assess in this particular project, as it allows us to be sure that the variance that the regression parameters are explaining comes from the same source in each time point and each group. If weak factorial invariance failed to hold, then the definitions of the constructs would be different over time and across groups, and therefore any relationships between them would no longer have the same meanings.

Strong invariance then claims that the intercepts for the same measure are equal over time and across groups. Using the same notation as above, the means of the indicators can be described by the following equations:

$$\bar{x}_{i,t,g} = \tau_{i,t,g} + \lambda_{i,t,g}\bar{X}_{t,g},$$

where $\tau_{i,t,g}$ is known as the intercept for indicator x_i at time t in group g. Strong invariance is then imposed by setting $\tau_{i,t,g} = \tau_i$ for each i separately, such that the means are now given by

$$\bar{x}_{i,t,g} = \tau_i + \lambda_i \bar{X}_{t,g},$$

since weak invariance is also assumed to hold whenever strong invariance is applied. This type of invariance is important when one wants to make comparisons between the means of the latent constructs over time or across groups. If strong factorial invariance does not hold, then differences in the means between the same latent construct at different time points, or in different groups, may not be down to changes in the underlying construct. This makes analysing those differences and deriving conclusions from these parameters tenuous.

To test whether measurement invariance held, first the configural invariance model was estimated, to provide a baseline against which the more restrictive measurement invariance claims could be tested. Configural invariance was deemed to hold if the CFI of the configural invariance model was greater than 0.95. If configural invariance was deemed to hold the model was then re-estimated with the weak invariance constraints imposed. The fit of this more restrictive model was then compared to the fit of the configural invariance model, with weak invariance being deemed to hold if the change in CFI between the models was less than 0.01 (Cheung and Rensvold, 2002). If weak invariance held then this process was repeated for the strong invariance model, with strong invariance being deemed to hold if the change in CFI between the weak invariance model and the strong invariance model was less than 0.01.

Step 3: Testing for Homogeneity of Variances and Co-variances

If strong measurement invariance was deemed to hold, the next step in the analysis procedure was to test whether the construct variances and covariances differed between the model's groups. These tests determine whether it is necessary to estimate the panel model separately in the groups or not. If the variances and covariances of the constructs are not substantially different in the groups then there is little need to estimate the model separately in each group and so the groups can be combined in the analysis.

The first test for the homogeneity of variances and covariances conducted was an omnibus test which examined the effect of constraining both the construct variances and covariances to be equal across the groups. A model with this constraint was compared to the strong invariance model using the robust Chi-square difference test from equation 5.3. A non-significant p-value from this test indicates that there has not been a significant decrease in model fit from the equality constraints, suggesting that homogeneity of variances and covariances holds and therefore, that the groups can be combined. Tests for significance were conducted at the 1% level. If this test indicated that there was a significant decrease in model fit, the covariances and variances were constrained to equality in separate models to test whether the variances alone, covariances alone, or both independently were responsible for the significant reduction in model fit.

5.3.2.2 Stages of the Measurement Model Estimation

The estimation of the measurement model was conducted in three main stages which are outlined below. In all three of the stages, the testing procedure outlined in section 5.3.2.1 was used to analyse the model.

Stage 1 - Checking for Model Estimation Issues and Construct Reliability

Before the above analysis of the measurement model could be completed, it is important to examine the measurement model for misspecifications that would result in estimation issues and convergence problems. To do this, I proceeded through the testing procedure outlined above until one or more estimation warnings or convergence failures occurred. Prior to giving any model parameter estimates the MPlus software displays messages to inform the user whether the model estimation has terminated normally, or whether any errors arose in the model estimation process. For example, a message may display to notify the user the model has failed to converge to a solution. Even when the model does converge to a solution, warning messages may be displayed that indicate an issue with some of the estimated parameters. For example, improper parameters such as negative variances and correlations of magnitude greater than one might have been estimated.

When such estimation issues arose, I investigated potential causes for the estimation issues, by both examining any parameter estimates that the model had produced, and referring to the SEM literature to understand why the

current specifications were causing these estimation issues. If this investigation suggested that the estimation issues were the result of the particular measures being used in the model, I respecified the measures being used and began the process again. Even when the MPlus software indicated that the model estimation had terminated normally, I scrutinised the parameter estimates to ensure that no improper estimates had been produced.

Once estimation issues were resolved, I examined the latent constructs' internal validity to ensure that the measures being used were adequately representing the hypothesised underlying latent factor. To assess the internal validity of the constructs, Average Variance Explained (AVE) and Composite Reliability (CR) were calculated for each of the constructs.

Suppose a latent construct X has p indicators, and that each indicator x_i has standardised loading given by λ_i for $i = 1, \ldots, p$. Standardised loadings are the factor loadings that are estimated when the variance of the latent construct is set to 1. Suppose also that the indicator x_i has an error term ϵ_{x_i}.

The R^2 value for an indicator shows the proportion of the indicator's variance that is reliably explained by the latent construct (Cohen, 1988). For a given indicator x_i, from 5.6, we have that

$$x_i = \lambda_i X + \epsilon_{x_i},$$

where the time subscript has been omitted here for convenience. The total variance for x_i is given by

$$\begin{aligned} Var(x_i) &= Var(\lambda_i X + \epsilon_{x_i}) \\ &= \lambda_i^2 Var(X) + Var(\epsilon_{x_i}) + Cov(\lambda_i X, \epsilon_{x_i}) \\ &= \lambda_i^2 + Var(\epsilon_{x_i}), \end{aligned}$$

since $Var(X) = 1$ and $Cov(\lambda_i X, \epsilon_{x_i}) = 0$. The R^2 value for indicator x_i, denoted R_i^2, is hence given by

$$R_i^2 = \frac{\lambda_i^2}{\lambda_i^2 + Var(\epsilon_{x_i})}.$$

The average variance explained is then calculated by taking the average of the R^2 values for each indicator on a construct (Fornell and Larcker, 1981).

As such,

$$\text{AVE} = \frac{\sum_{i=1}^{p} R_i^2}{p} = \frac{1}{p}\sum_{i=1}^{p} \frac{\lambda_i^2}{\lambda_i^2 + Var(\epsilon_{x_i})}.$$

Composite Reliability was calculated using the formula provided in (Netemeyer, 2003). Using the same notation as above the composite reliability is given below in equation 5.8:

$$\text{CR} = \frac{\left(\sum_{i=1}^{p} \lambda_i\right)^2}{\left(\sum_{i=1}^{p} \lambda_i\right)^2 + \left(\sum_{i=1}^{p} V(\epsilon_{x_i})\right)}. \tag{5.8}$$

A construct was considered to have good enough internal consistency if it had an AVE greater than 0.5, and a CR greater than 0.7 (Fornell and Larcker, 1981). Constructs failing these examinations were respecified using different measures from the Whitehall data.

Once new measures had been chosen, the process of checking for estimation errors and construct reliability was repeated until a set of reliable measures had been chosen that did not result in a model containing estimation issues.

Stage 2 - Testing for Moderation by Ethnicity

After the necessary revisions to the measurement model had been made to ensure that the constructs were suitably valid, and there were no issues due to convergence failure or improper solutions, the measurement model was analysed to examine any potential confounding by ethnicity. In the Whitehall II data, participants' ethnicity is recorded in two categories: "white" and "non-white" To test whether there was moderation by ethnicity a multi-group measurement model was estimated in which the model groups represented the two ethnic group categories.

The first step in this supplementary analysis was to test the multi-group measurement model for measurement invariance, to ensure that the latent constructs could be trusted to represent the same concept over time and across groups. These tests were then followed by the key tests for the homogeneity of the variances, covariances and means across ethnic groups. As was mentioned in section 5.3.2.1, these tests were carried out to examine whether it would be necessary to include ethnicity as an additional moderator variable by including it as a dimension in the grouping variable. In particular, if the test for homogeneity of covariances and variances suggested the model

fit was significantly better when allowing the variances and covariances to be different across the two ethnic groups, then ethnicity would need to be included in the grouping variable.

Stage 3: Analysing the Final Multi-group Measurement Model

Once the tests for moderation by ethnicity had been completed, the testing procedure was re-run for the final specification of the measurement model. Of particular interest in this stage is whether measurement invariance held, as this condition is fundamental to the reliability of the structural model's estimation. Details of the measures used in the final measurement model are outlined in the next section.

5.3.3 The Measurement Model - Results

Stage 1 - Re-specifications of the Measurement Model

Due to a combination of model estimation issues and lack of reliability for constructs, every construct except for physical function was respecified. Once all construct reliability and model estimation issues had been resolved, the final measurement model consisted of a mixture of latent variables and manifest variables.

Obesity was measured solely by BMI which was included in the model as a manifest variable, due to issues with model reliability when using both BMI and waist-hip ratio. In particular, using both measures as indicators of a latent construct led to an estimated model in which waist-hip ratio had a negative residual variance. This negative residual variance likely occurred from the confluence of two issues. Firstly, loadings for BMI and waist-hip ratio had to be fixed as equal in order to identify the construct, and secondly the two indicators were very highly correlated. Together, these issues led to a model being estimated in which the variance of the latent factor explained more than the total variance of the waist-hip ratio indicator, leaving it with a negative residual variance. This suggested the model was misspecified in this form and so needed re-specification.

Waist-hip-ratio had itself replaced waist-circumference due to similar estimation issues. As there were no additional measures in the Whitehall dataset that could be added to the construct in order to resolve the need for the identification constraint, I removed waist-hip ratio from the model and used BMI on its own as a measure of obesity.

I also re-specified the depression construct due to similar issues. In particular, GHQ-based depression score and SF-36 mental health score were replaced by four individual GHQ items due to a model being estimated in which GHQ-based depression score had a negative residual variance. The items I used in the final model are displayed in Table 5.1. I chose the four GHQ items based on the results of an Exploratory Factor Analysis (EFA) that I conducted on the GHQ-30 questionnaire data within Wave 5 of the Whitehall II dataset. Within the EFA, I examined factor solutions that extracted between three and seven factors, and chose items for a depression construct if they loaded consistently onto a factor that I felt could be easily interpreted as depression across the different factor solutions.

In all solutions, I used an oblique rotation to allow for extracted factors to be correlated with one another. I considered items to load onto a factor if the standardised loadings from the pattern matrix were above 0.4 (Stevens, 1992). Pattern loadings for each of the items in each solution can be found in Table 5.2. The four items I chose for the GHQ depression measure loaded exclusively onto the same factor across all of the factor solutions, except for the ghq-30 item which marginally missed the cut-off in the six factor solution. Also of note is that these four items were all originally included in the GHQ-based depression summary measure that I derived in section 3.3.2 of Chapter 3.

Item	Item Content
ghq-24	'Been thinking yourself as a worthless person'
ghq-25	'Felt that life is entirely hopeless'
ghq-29	'Felt that life isn't worth living'
ghq-30	'Found at times you couldn't do anything because your nerves were too bad'

TABLE 5.1: Items from GHQ included in our measure of depression

Due to poor construct validity when using multiple measures of physical activity, I replaced the three separate measures for physical activity with a measure of total derived met-hrs. Similarly, due to poor construct validity of the diet construct when indicated by fat, sugar and processed meat consumption frequency, new measures based on estimated daily calorie, fat and carbohydrate consumption were derived to be used as measures for the diet construct. Using more than one of these measures still proved problematic in

	Factor solution				
	3-factor	4-factor	5-factor	6-factor	7-factor
TLI	0.869	0.896	0.919	0.933	0.944
RMSEA	0.063	0.056	0.049	0.045	0.041
ghq-24 loading	0.66	0.66	0.63	0.58	0.50
ghq-25 loading	0.89	0.89	0.89	0.93	0.89
ghq-29 loading	0.83	0.83	0.85	0.83	0.83
ghq-30 loading	0.45	0.45	0.43	0.38	0.41

TABLE 5.2: Summary of fit statistics and pattern loadings for depression indicators with EFA solutions

the model estimation, and so in the final model only calorie intake (scaled to assist model convergence) was used.

The distributions of calorie consumption and physical activity can be seen in Figures 5.3 and 5.4 respectively. The distributions of calories appear to follow roughly a Normal distribution, albeit with a truncated lower tail due to the inability for people to consume very few, or negative calories. The positions of the peaks of the distributions perhaps suggest that the estimated calories are under-representative of calorie consumption on average, given that the peak is below the recommended daily intake for adult women of 2000 daily calories. This is not entirely surprising though, as the FFQ data from which these estimates are derived is not an exhaustive list of all the foods one could consume, and its retrospective nature makes it possible that participants might under-report the consumption of foods due to inaccuracies in their memory and social desirability biases (Hebert et al., 1995; Hebert et al., 1997). The distributions of physical activity level also, for the most part, follow a bell-shaped curve similar to that of a Normal distribution, however, in this case, the left tail of the distribution is non-existent, indicating some deviation from Normality. The positions of the peaks of the distribution suggest that the average level of physical activity is around 40 met-hrs a week, which is roughly equivalent to 10 or so hours of moderate walking a week.

In the final model I dropped the measure of social ladder position from the SEP construct, as including it resulted in a construct with unacceptable levels of internal validity. To avoid creating similar issues by adding another measure of SEP from the Whitehall II data set which had a different scale in place of social ladder position, I finalised the SEP construct to be indicated by frequency of money problems and frequency of problems paying bills. As a result, the the SEP construct used in the model might be best described as

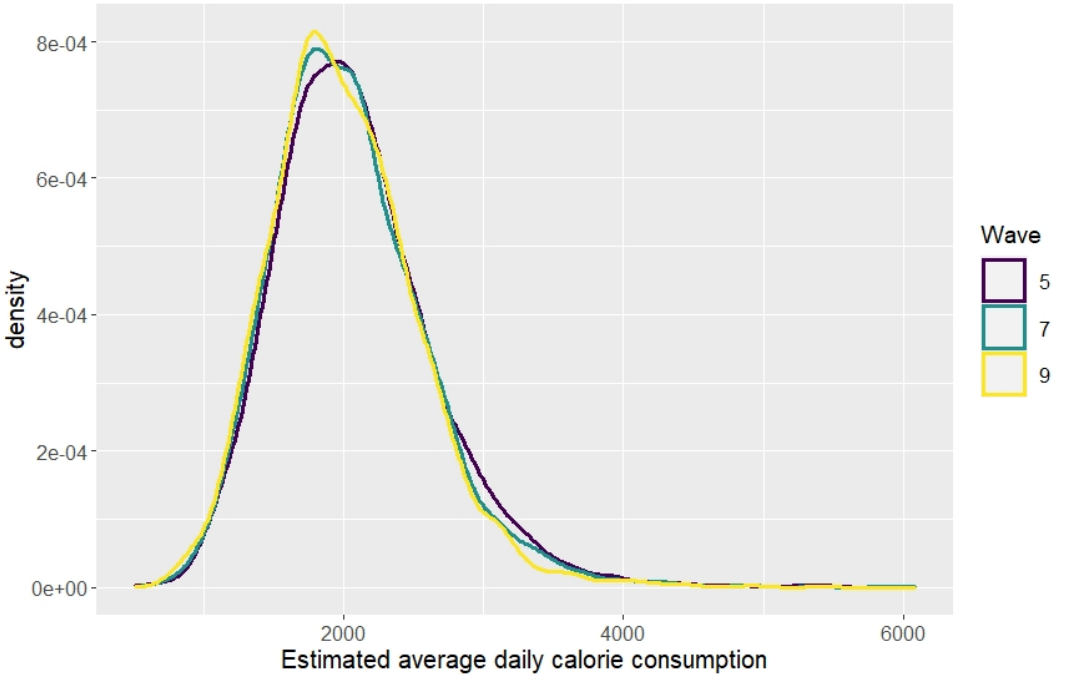

FIGURE 5.3: Distribution of estimated daily calorie consumption in waves five, 7 and nine of the Whitehall II data set

'extent of financial problems'.

Finally, in order to prevent convergence issues during the model estimation, I rescaled the measures of physical function, calories, and physical activity. The large variances these measures had on their original scale led to the software being unable to converge on a solution. This is a known phenomenon with the MPlus software which can be solved by rescaling the variables such that their variances are roughly between 0 and 10 (Muthen and Muthen, 2017). As a result, I divided the measures of physical function by 20, to put them on a scale of zero to five; the measures of calories by 100, such that it was a measure of how many hundreds of calories were consumed on average per day; and the measures of physical activity by 7, so that it was an estimate of daily met-hrs.

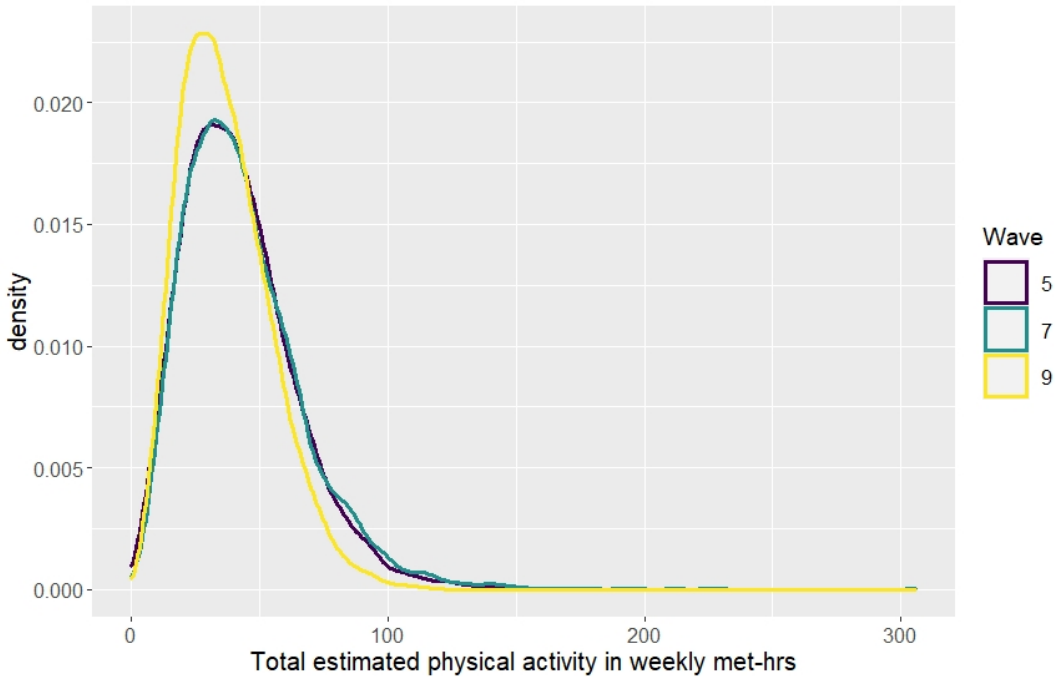

FIGURE 5.4: Distribution of estimated weekly physical activity in waves five, 7 and nine of the Whitehall II data set

Stage 2 - Testing for Moderation By Ethnicity

Table 5.3 gives the results for models analysing whether there was evidence of moderation by ethnicity. The configural, weak and strong invariance models, all passed their respective tests, allowing the analysis of the covariance, variance and means to continue. The omnibus test of differences between the variances and covariances suggested that there were significant differences between them across the two groups. When this test was repeated for the variances and covariances separately, both tests suggested that there were significant differences between the groups. Similarly, the test for homogeneity of means found evidence that the means were different between the two groups.

TABLE 5.3: Fit statistics for tests of moderation by ethnicity

Model Tested	χ^2	df	Scaling correction	TrD	Δdf	p	Cramer's V	RMSEA	RMSEA 90% C.I	CFI	ΔCFI	Pass
Null	109786.035	1526	1.3018					0.131	.130;.132			
Configural invariant	3218.972	924	1.1915					.025	.024;0.025	.978		Yes
Weak invariant	3350.205	949	1.2099					.025	.024;.026	.977	.001	Yes
Strong invariant	4238.905	979	1.2032					.028	.028;0.029	.968	.009	Yes
Homogeneity var-cov	4968.683	1187	1.2032	77.778	208	< 0.001	.021	.028	.027;.029	.963	.005	No
Homogeneity variances	4578.128	997	1.2186	232.788	18	< 0.001	.040	.029	.029;.030	.965	.003	No
Homogeneity covariances	4752.117	1169	1.1855	487.421	190	< 0.001	.018	.027	.026;.028	.965	.003	No
Homogeneity of Means	4585.096	997	1.2007	379.321	18	< 0.001	.051	.030	.029;.030	.965	.003	No

Despite these significant differences it is worth noting the very small effect sizes associated with the chi-square difference tests that were used to compare the model fit. Despite the small effect sizes, in taking a conservative approach to the analysis of this data, I decided to include ethnicity in the grouping variable alongside sex, so that any chance of bias was minimised.

Stage 3 - The Multigroup Measurement Model

Once I decided that ethnicity was to be included in the model as an additional grouping variable, the testing procedure was re-run on the final measurement model to assess its fit. Table 5.4 gives the fit statistics for those analyses.

Again the tests of measurement invariance passed for configural, weak and strong factorial invariance, suggesting that the meanings of the constructs were not different over time or across groups. As was the case in the test of moderation by ethnicity, the omnibus test examining the differences between the covariances and variances across the two groups found a significant reduction in model fit when compared to the strong invariance model suggesting that moderation effects by sex and ethnicity existed. The test of homogeneity of variances over time and between groups also found a significant reduction in model fit, suggesting that whilst the meaning of the constructs was the same over time and across groups, their underlying metrics were different. As a result of this test, phantom constructs were added into the model so that covariance parameters could be tested for equality across groups and over time. Finally, the test for the homogeneity of means across groups also found a significant reduction in the model fit.

TABLE 5.4: Fit statistics for the final measurement model

Model Tested	χ^2	df	Scaling correction	TrD	Δdf	p	Cramer's V	RMSEA	RMSEA 90% C.I	CFI	ΔCFI	Pass
Null	115392.226	2928	1.2837					.136	.136;.137			
Configural invariant	4051.703	1776	1.1524					.025	.024;.026	0.978		Yes
Weak invariant	4284.703	1831	1.1753					.025	.025;.026	0.976	.002	Yes
Strong invariant	5408.051	1897	1.1681					.030	.029;.031	.966	0.01	Yes
Homogeneity of means (omnibus)	8243.697	1963	1.1634	3183.447	66	< 0.001	.076	.039	.039;.040	.939	0.027	No
Homogeneity of Means (groups)	6525.806	1951	1.1639	1257.671	54	< 0.001	.053	0.034	.033;.035	.955	.011	No
Homogeneity var-cov	7996.702	2464	1.1837	2547.593	567	< 0.001	.023	.033	.032;.035	.946	.02	No
Homogeneity variances	6677.311	1951	1.1946	780.778	54	< 0.001	.042	.034	.033;.035	.954	.012	No

5.4 The Structural Model

After the measurement model had been estimated, and measurement invariance tests had been satisfied, the structural model was fit to the data. To examine the hypotheses in section 5.2, three different specifications of the structural model were analysed. In all models the relationships between the constructs were assumed to be linear. In model 1, the initial specification of the structural model, shown in Figure 5.1, it was assumed that all the effects could only occur over single time gaps; no paths existed directly between time 1 and time 3 in the model.

In model 2 direct auto-regressive paths between time 1 and time 3 were added to the model to take into account additional construct stability over time that might have been underestimated in model 1. In model 3, the assumption that all cross-lagged effects could only occur over a single time-step was relaxed, with paths being added between constructs in time 1 and constructs in time 3 if there was an indirect effect between the two constructs via a hypothesised mediator at time 2. For example, direct effects were added between BMI at time 1 and depression at time 3 since they are linked indirectly by the mediator physical function at time 2.

A digram showing the paths added to each model for the analysis of hypothesis 1 is displayed in Figure 5.5.

5.4.1 Identification of the Structural Model

Identification of the structural model can be determined by assessing whether the model meets two main rules (Kenny, Kashy and Bolger, 1998).These rules do not give an exhaustive set of conditions under which the structural model is identified, however, they do provide a guidelines that suggest which models may be identified.

Rule A: Minimum Condition of Identifiability

This identification rule compares the number of construct covariances with the number of paths and relationships that are estimated between them in the model. For a structural model with k constructs there are $\binom{k}{2} = k(k-1)/2$ pairs of constructs, with each pair associated with a unique covariance. If we set p as the total of:

1. The total number of paths from one construct to another,

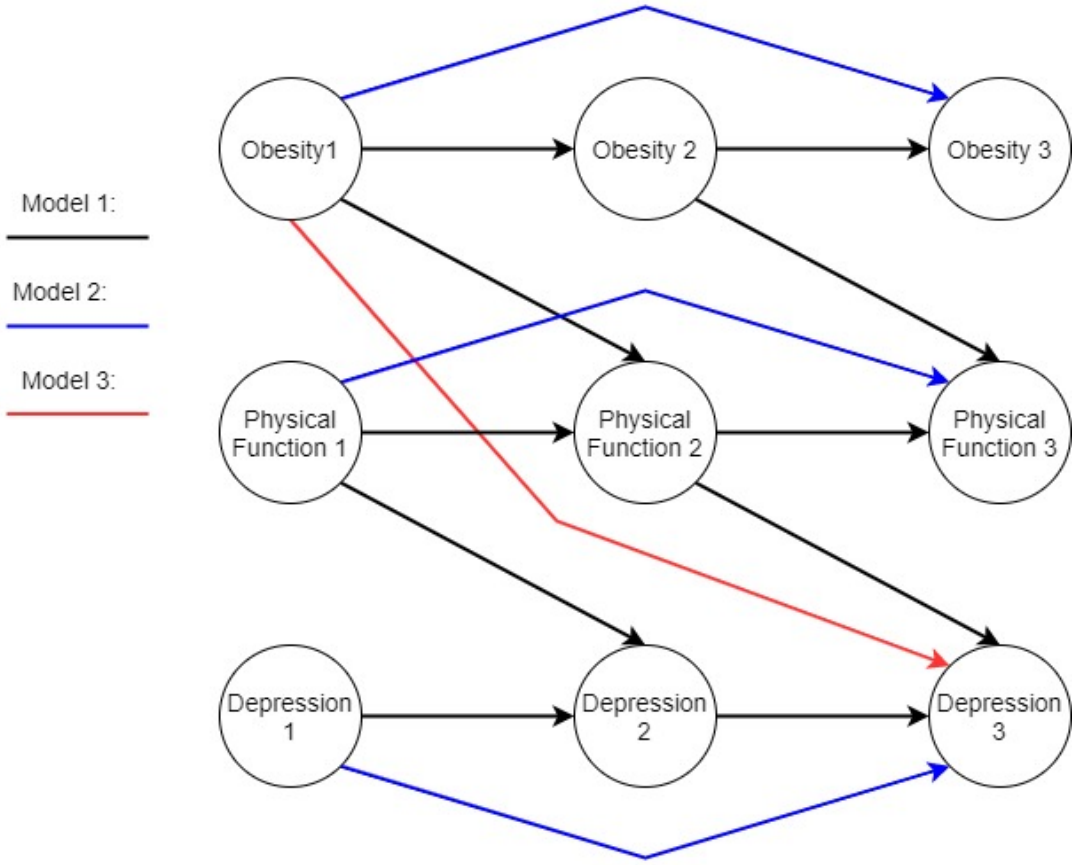

FIGURE 5.5: Schematic showing added in each specification of the Structural model

2. The number of correlations between exogenous constructs,
3. The number of correlations between the error variances, endogenous constructs and exogenous constructs,
4. the number of correlations between construct errors,

then the model may be identified if $p < k(k-1)/2$. Exogenous variables are those variables that have no paths leading to them in the structural model, whereas endogenous variables have paths leading them from other variables in the model.

Rule B: Apparent Necessary Condition of Identifiability

The second rule for identification of the structural model states that, if for any pair of constructs in the model X and Y, no more than one of the following is true:

1. X directly causes Y,

2. Y directly causes X,
3. X and Y have correlated errors, or if either X or Y is exogenous, it is correlated with the other's error,
4. X and Y are correlated exogenous variables,

then the model is identified. The second identification rule for structural models is not a proven rule for identification, however, there are no known exceptions. As such, all models that follow this rule appear to be identified.

Rule B can be relaxed under certain conditions if the model is using instrumental variables, however, since our model does not include instrumental variables, this condition is not relevant here.

5.4.2 Estimation of the Structural Model

5.4.2.1 The Structural Model Fit

To analyse the fit of the structural models, each model was estimated, and its associated fit statistics extracted. The Chi-square statistic was then compared with that of the strong invariance model using the Chi-square difference test for MLR method described earlier in equation 5.3.

5.4.2.2 Extracting Important Paths

Once each model had been tested against the strong invariance model for goodness of fit, 'important' parameters relevant to each of the four hypotheses were extracted from the model results. All the model parameters for each model were extracted into Rstudio using the MPlus Automation package (Hallquist and Wiley, 2018), where they were analysed for importance. Parameters were considered to be important if their associated p-value in the unstandardised model solution was less than 0.05.

5.4.3 The Structural Model - Results

5.4.3.1 Fit of the Structural Model

Model 1: The Initial Structural Model

The model fit statistics for the structural models examined are shown in Table 5.5. When compared to the strong invariance model the Chi-square difference test for model fit found a significant reduction in the model fit, suggesting that important paths were missing from model 1. This result was

significant despite the very small value of Cramer's V. The model also had a large drop in CFI when compared to the strong invariance model (0.012).

Model 2: Including Autoregressive Paths from Time 1 to Time 3

Adding in autoregressive paths between time 1 and time 3 resulted in a much improved model fit. The difference in CFI against the strong invariance model was 0.002, and the RMSEA estimate was better than that of the strong invariance model: .029 compared to .030 for the strong invariance model. However, given the large overlap in the RMSEA's confidence intervals there is little evidence to suggest this difference is meaningful. Once again, the Chi-square difference test suggested that there was a significant reduction in the model fit when compared to the strong invariant model.

Model 3: Including Partial Mediation Paths

Model 3 had a marginally better fit than model 2 as suggested by its slightly improved CFI, however the RMSEA and its confidence interval were same in the two models. The MLR chi-square difference test between the two models found a significant difference between the two models $(p = 0.0005)$. Again, when compared to the strong invariance model, the MLR chi-square difference test suggested there was a significant difference in the model's fit $(p < 0.001)$, in spite of the small value for Cramer's V.

TABLE 5.5: Fit statistics for fitted structural models

Model Tested	χ^2	df	Scaling correction	TrD	Δdf	p	Cramer's V	RMSEA	RMSEA 90% C.I	CFI	ΔCFI	Pass
Strong invariant	5408.051	1897	1.1681					.030	.029;.031	.966	0.01	Yes
Model 1 (initial)	6887.691	2145	1.1638	1502.111	248	< 0.001	.027	.033	.032;.034	0.954	0.012	No
Model 2 (+ AR2)	5788.349	2121	1.1587	361.243	224	< 0.001	.014	.029	.028;.030	.964	.002	No
Model 3 (+AR2 + partial mediation)	5619.621	2021	1.1660	207.548	124	< 0.001	.014	.029	.028;.030	.965	.001	No

5.4.4 Important Paths in the Structural Model

In general, the models found little evidence to support any of the hypotheses in any of the groups. Figures 5.6 to 5.9 show the paths that were identified as potentially important in each of the models within each group. In the diagrams, black lines indicate that the path was significant in all models where it was estimated and red lines indicate paths that were significant in some, but not all, models that they were estimated in.

Additionally, there was little evidence of a prospective relationship between obesity and depression in this cohort. There were no significant time 1 covariations between BMI and depression in any of the groups, and only in the white female group did any of the models find a significant direct path between BMI and depression over time. Even in this group, the over time association was only found between time 2 and time 3, and was not significant in every model. Evidence in support of an indirect association via the proposed mediators was also weak and in some cases non-existent. More details on the indirect associations are given in the following subsections.

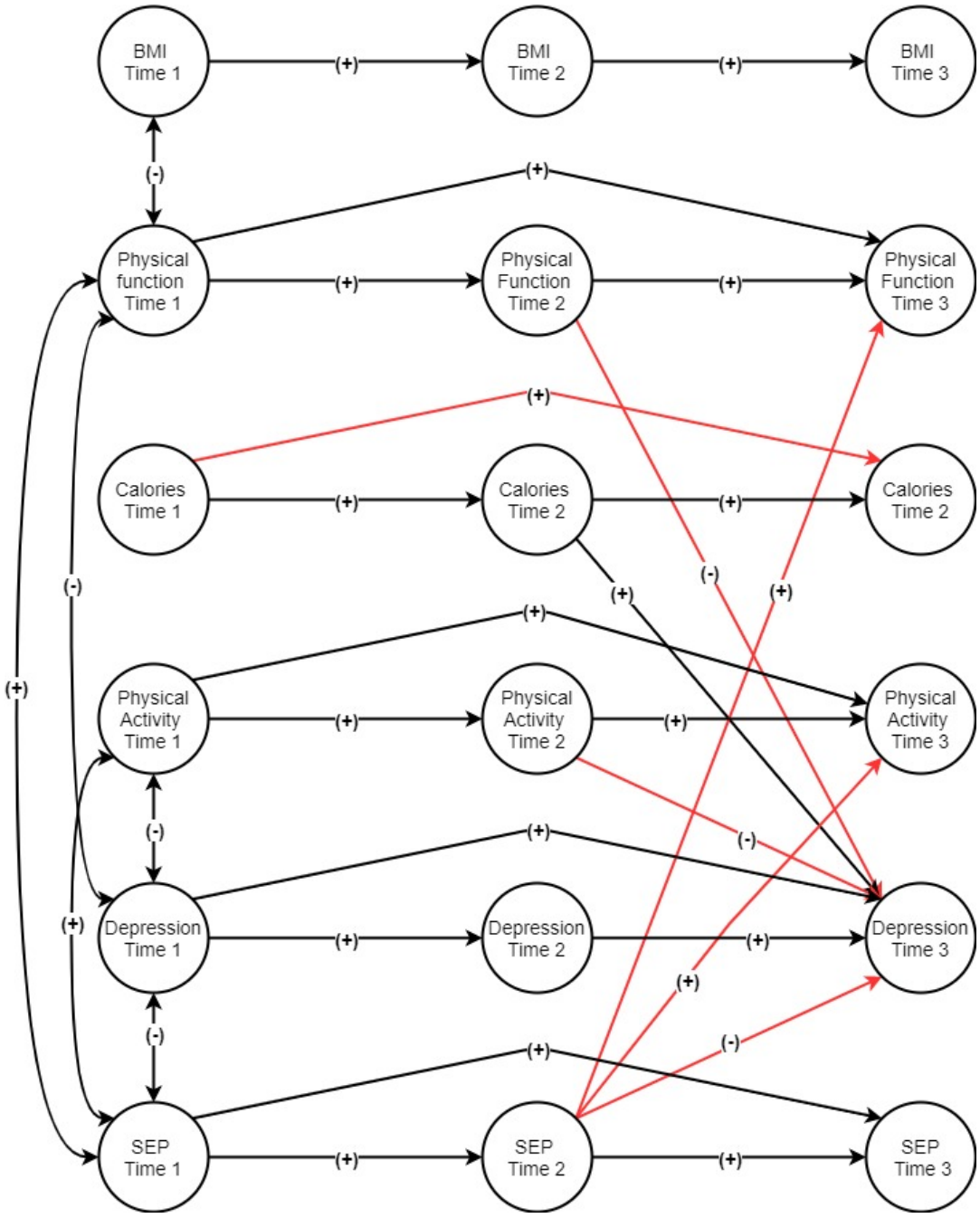

FIGURE 5.6: Path Diagram of significant paths and time 1 covariances in the non-white female group. Black lines indicate that the path was significant in all models where it was estimated and red lines indicate paths that were significant in some, but not all, models that they were estimated in. Double headed lines indicate significant covariances at time 1. Single headed arrows indicate significant longitudinal relationships.

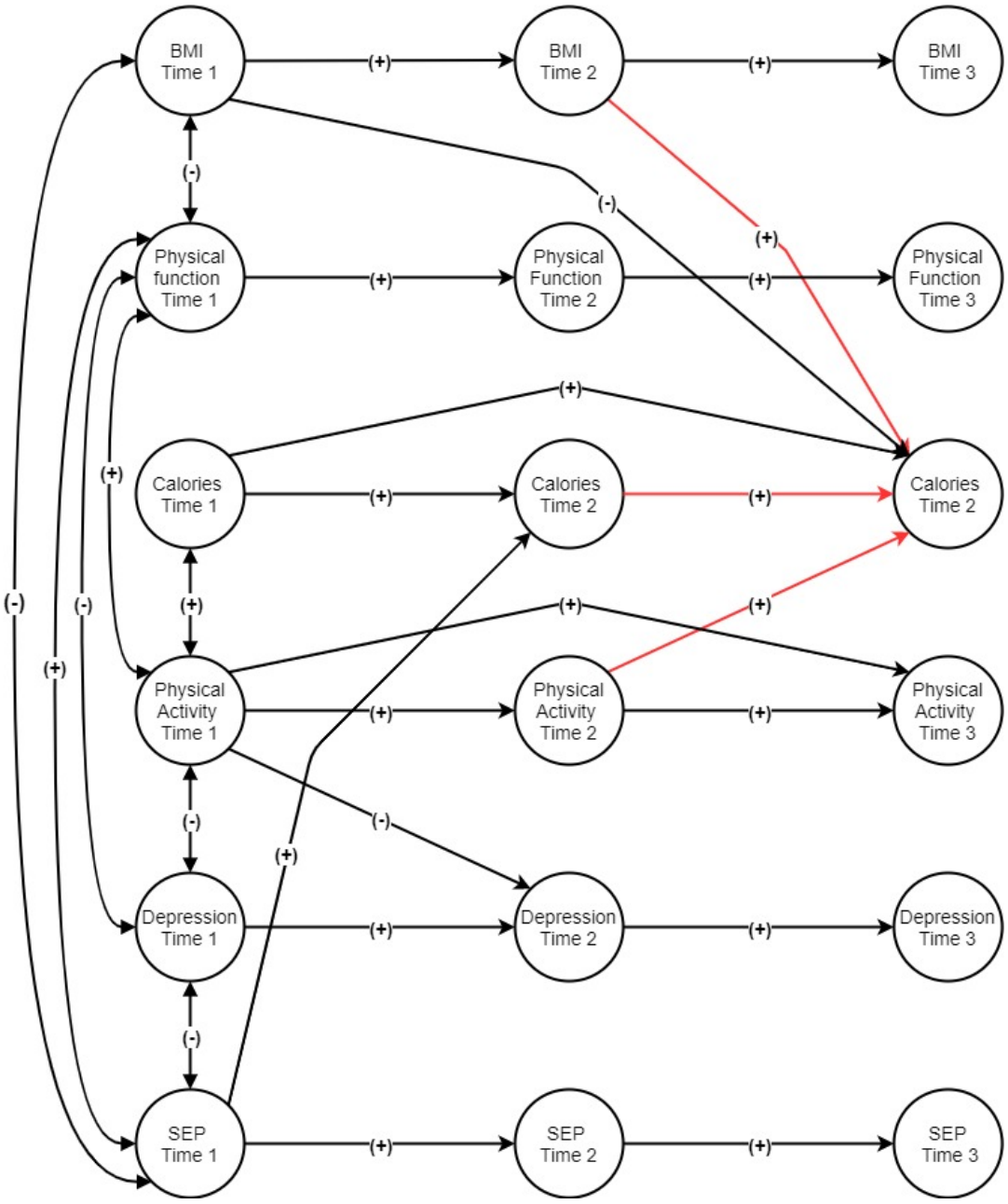

FIGURE 5.7: Path Diagram of significant paths and time 1 covariances in the non-white male group. Black lines indicate that the path was significant in all models where it was estimated and red lines indicate paths that were significant in some, but not all, models that they were estimated in. Double headed lines indicate significant covariances at time 1. Single headed arrows indicate significant longitudinal relationships.

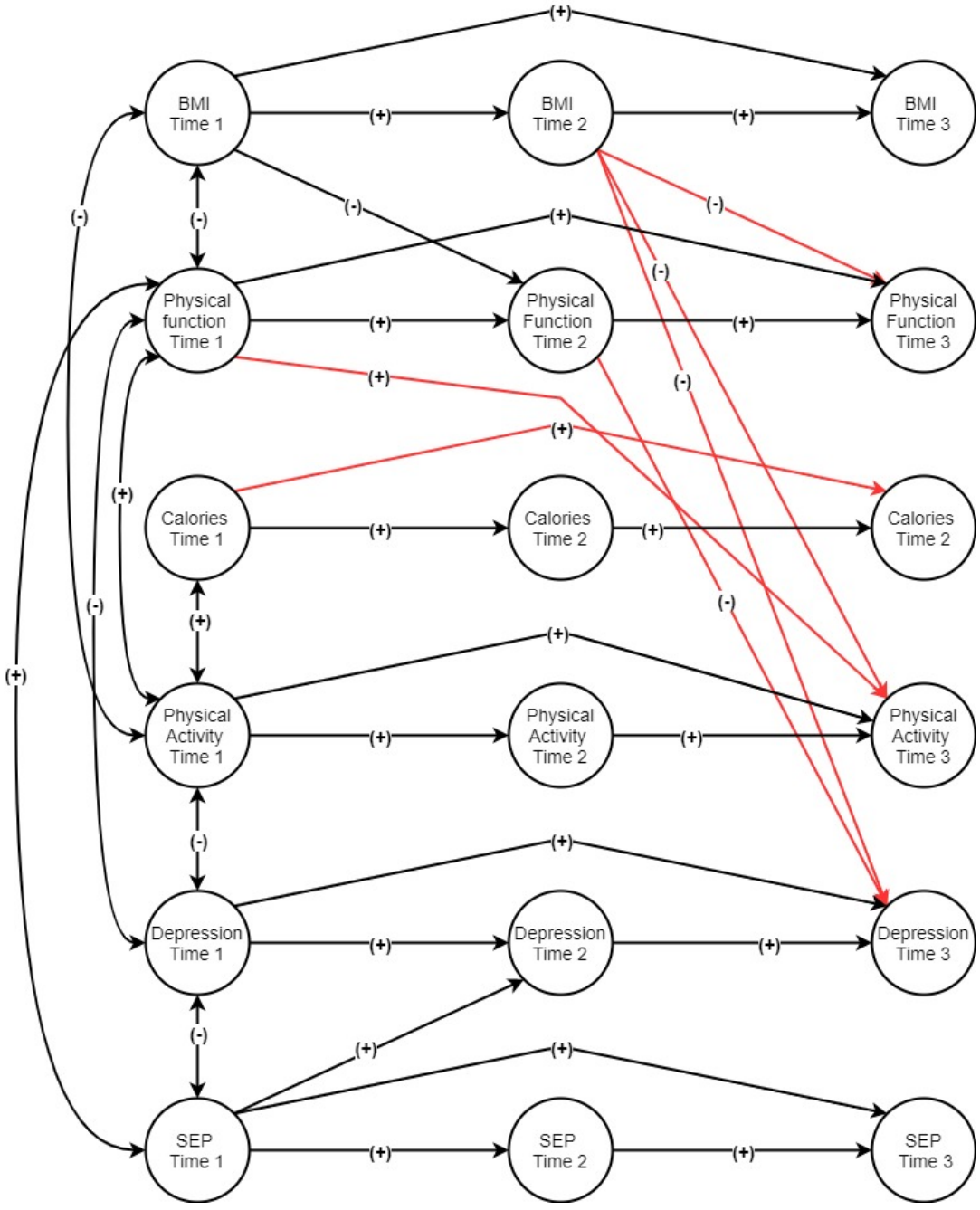

FIGURE 5.8: Path Diagram of significant paths and time 1 covariances in the white female group. Black lines indicate that the path was significant in all models where it was estimated and red lines indicate paths that were significant in some, but not all, models that they were estimated in. Double headed lines indicate significant covariances at time 1. Single headed arrows indicate significant longitudinal relationships.

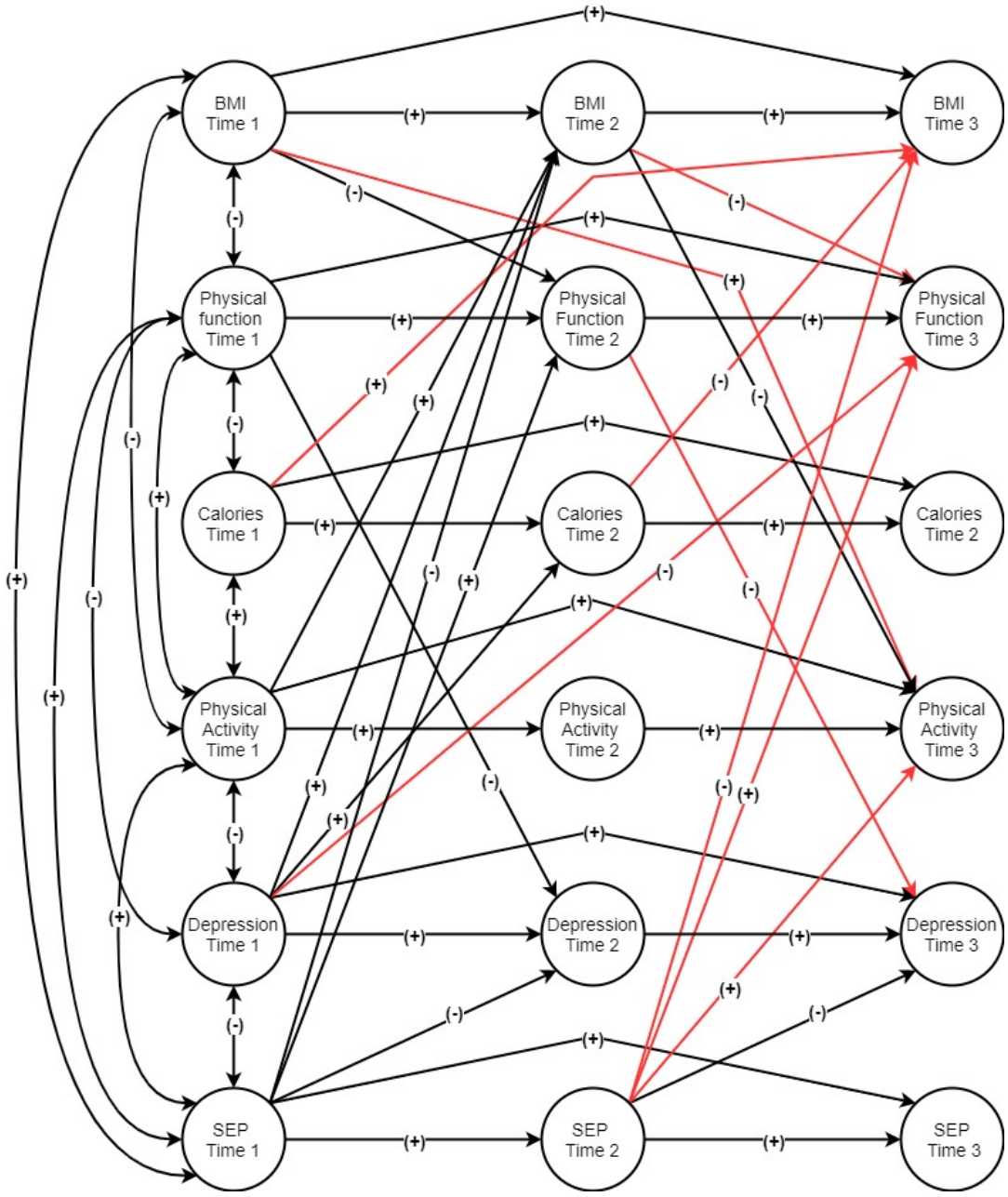

FIGURE 5.9: Path Diagram of significant paths and time 1 covariances in the white male group. Black lines indicate that the path was significant in all models where it was estimated and red lines indicate paths that were significant in some, but not all, models that they were estimated in. Double headed lines indicate significant covariances at time 1. Single headed arrows indicate significant longitudinal relationships.

Evidence for Hypothesis 1: That Physical Function Mediates the Association between Obesity and Future Depression

Overall, there was little consistent evidence to support an association between obesity and future depression via an effect on physical function. In the non-white female group, physical function at time 2 was predictive of depression at time 3 in model 1, but no other paths relating to this hypothesis were significant in any of models. In non-white males, there were no significant paths in any of the models that provided any support to this hypothesis. In white females, BMI was predictive of future physical function in the first time-step in model 3, and in both time-steps in models 1 and 2. In white males, there was stronger evidence in support of the hypothesis, with both model 1 and model 2 having significant paths from BMI to future physical function and from physical function to future depression in both time-steps. However, these relationships were only present in the first time-step of model 3, so the evidence is not unequivocal.

	Dependent Variable	Independent Variable	B	S.E	B/S.E	p
Non-White Females						
	DEP3	PF2	-0.174	0.071	-2.434	0.015
Non-White Males						
	-	-	-	-	-	-
White Females						
	DEP3	PF2	-0.079	0.031	-2.539	0.011
	PF3	BMI2	-0.013	0.004	-3.487	0.000
	PF2	BMI1	-0.016	0.004	-3.704	0.000
White Males						
	DEP3	PF2	-0.091	0.027	-3.398	0.001
	DEP2	PF1	-0.137	0.035	-3.961	0.000
	PF3	BMI2	-0.014	0.003	-5.156	0.000
	PF2	BMI1	-0.012	0.003	-4.262	0.000

TABLE 5.6: Parameter estimates for significant unstandardised parameters from Model 1 relevant to Hypothesis 1

	Dependent Variable	Independent Variable	B	S.E	B/S.E	p
Non-White Females						
	-	-	-	-	-	-
Non-White Males						
	-	-	-	-	-	-
White Females						
	PF3	BMI2	-0.013	0.004	-3.600	0.000
	PF2	BMI1	-0.016	0.004	-3.882	0.000
White Males						
	DEP3	PF2	-0.085	0.025	-3.380	0.001
	DEP2	PF1	-0.144	0.035	-4.114	0.000
	PF3	BMI2	-0.015	0.003	-5.398	0.000
	PF2	BMI1	-0.013	0.003	-4.414	0.000

TABLE 5.7: Parameter estimates for significant unstandardised parameters from Model 2 relevant to Hypothesis 1

	Dependent Variable	Independent Variable	B	S.E	B/S.E	p
Non-White Females						
	-	-	-	-	-	-
Non-White Males						
	-	-	-	-	-	-
White Females						
	PF2	BMI1	-0.017	0.004	-3.895	0.000
White Males						
	PF3	DEP1	-0.043	0.016	-2.744	0.006
	DEP2	PF1	-0.141	0.035	-4.068	0.000
	PF2	BMI1	-0.012	0.003	-4.336	0.000

TABLE 5.8: Parameter estimates for significant unstandardised parameters from Model 3 relevant to Hypothesis 1

Evidence for Hypothesis 2: That Dietary Energy Intake Mediates the Association between Obesity and Depression Bi-directionally

There was little to no evidence found in the three models that supported hypothesis 2. In non-white females, calories at time 2 had a positive association with depression at time 3 in model 1, but no other paths in the model supported the hypothesis. The same pattern was found in models 2 and 3.

In non-white males, there were no paths in support of the hypothesis in model 1 or 2, however, in model 3, BMI at both times 1 and 2 was associated with calorie intake at time three. However, the direction of this association was inconsistent: BMI at time 1 was negatively associated with calories at time 3, and BMI at time 2 was associated positively.

In white females, there were no significant paths in support of this hypothesis in any of the models. In white males, depression at time 1 was positively associated with calorie intake at time 2 in all three models, and in model 3, calorie intake at times 1 and 2 was associated with BMI at time 3. Again though, the direction of this association differed between the two time points: calorie intake at time 1 was associated positively with BMI at time 3, whilst calorie intake at time 2 was negatively associated with BMI at time 3.

	Dependent Variable	Independent Variable	B	S.E	B/S.E	p
Non-White Females						
	DEP3	CALS2	0.025	0.012	1.993	0.046
Non-White Males						
	-	-	-	-	-	-
White Females						
	-	-	-	-	-	-
White Males						
	CALS2	DEP1	0.307	0.139	2.205	0.027

TABLE 5.9: Parameter estimates for significant unstandardised parameters from Model 1 relevant to Hypothesis 2

	Dependent Variable	Independent Variable	B	S.E	B/S.E	p
Non-White Females						
	DEP3	CALS2	0.026	0.010	2.669	0.008
Non-White Males						
	-	-	-	-	-	-
Non-white male						
	-	-	-	-	-	-
White Females						
	-	-	-	-	-	-
White Males						
	CALS2	DEP1	0.300	0.138	2.175	0.030

TABLE 5.10: Parameter estimates for significant unstandardised parameters from Model 2 relevant to Hypothesis 2

	Dependent Variable	Independent Variable	B	S.E	B/S.E	p
Non-White Females						
	DEP3	CALS2	0.033	0.011	3.030	0.002
Non-White Males						
	CALS3	BMI2	0.701	0.337	2.082	0.037
	CALS3	BMI1	-0.791	0.353	-2.244	0.025
White Females						
	-	-	-	-	-	-
White Males						
	CALS2	DEP1	0.298	0.138	2.160	0.031
	BMI3	CALS2	-0.016	0.007	-2.470	0.013
	BMI3	CALS1	0.013	0.006	2.151	0.031

TABLE 5.11: Parameter estimates for significant unstandardised parameters from Model 3 relevant to Hypothesis 2

Evidence for Hypothesis 3: That Physical Activity Level Mediates the Association between Obesity and Depression Bi-directionally

Similar to hypothesis 1, there was little consistent evidence found in any of the models in support of hypothesis 3. In non-white females, physical activity was predictive of future depression in the second time-point but not the first in models 1 and 3 but not in model 2, providing inconsistent evidence of an effect. There were no other significant paths in the non-white female group in support this hypothesis.

In non-white males, evidence was similar. In all three models, physical activity at time 1 was predictive of depression at time 2 but the same association was not found from time 2 to time 3. Again, no other paths were significant. In white females, BMI at time 2 was predictive of physical activity at time 3 in models 1 and 2. Again, no other paths relevant to this hypothesis were significant meaning that overall there was little evidence to support it.

In white males, there was similarly sparse support for the hypothesis. In all three models, BMI at time 2 was negatively associated with physical activity at time 3, and in model 3, BMI at time 1 was positively associated with physical activity at time 3. Physical activity at time 1 was positively associated with BMI at time 2 in all three models, however this association was not consistent over time.

	Dependent Variable	Independent Variable	B	S.E	B/S.E	p
Non-White Females						
	DEP3	METS2	-0.045	0.023	-2.006	0.045
Non-white Males						
	DEP2	METS1	-0.039	0.016	-2.494	0.013
White Females						
	METS3	BMI2	-0.028	0.010	-2.812	0.005
White Males						
	METS3	BMI2	-0.023	0.009	-2.569	0.010
	BMI2	METS1	0.017	0.007	2.444	0.015

TABLE 5.12: Parameter estimates for unstandardised parameters from Model 1 relevant to Hypothesis 3

	Dependent Variable	Independent Variable	B	S.E	B/S.E	p
Non-White Females						
	-	-	-	-	-	-
Non-Whites Males						
	DEP2	METS1	-0.039	0.015	-2.503	0.012
White Female						
	METS3	BMI2	-0.024	0.009	-2.536	0.011
White Males						
	METS3	BMI2	-0.021	0.009	-2.417	0.016
	BMI2	METS1	0.017	0.007	2.389	0.017

TABLE 5.13: Parameter estimates for unstandardised parameters from Model 2 relevant to hypothesis 3

	Dependent Variable	Independent Variable	B	S.E	B/S.E	p
Non-White Females						
	DEP3	METS2	-0.070	0.030	-2.318	0.020
Non-White Males						
	DEP2	METS1	-0.041	0.015	-2.701	0.007
White-females						
	-	-	-	-	-	-
White Males						
	METS3	BMI2	-0.088	0.027	-3.224	0.001
	METS3	BMI1	0.077	0.030	2.578	0.010
	BMI2	METS1	0.017	0.007	2.462	0.014

TABLE 5.14: Parameter estimates for unstandardised parameters from Model 3 relevant to hypothesis 3

Evidence for Hypothesis 4: That Socioeconomic Position Affects the Relationship between Obesity and Depression via Multiple Effects on Obesity, Depression and their Mediating Variables

In non-white females, in models 1 and 2, SEP at time 2 was positively associated with physical activity at time 3, and negatively associated with depression at time 3. In model 1 SEP at time 2 was also negatively associated with physical function at time 3. In model 3 there were no significant paths between SEP and other constructs.

In non-white males, SEP at time 1 was positively associated with physical function at time 2 in all three models, but no other paths from SEP to the other constructs were significant. In white females, SEP at time 1 was negatively associated with depression at time 2 in all three models. No other paths

between SEP and the other constructs were significant.

In white males, some consistent associations were found. SEP at times 1 and 2 were negatively associated with depression at times 2 and 3 respectively in all three models. However, other paths involving SEP were either inconsistent over time or across models. In all three models, SEP at time 1 was negatively associated with BMI at time 2, and positively associated with physical function at time 2. In models 1 and 2, these same relationships were found between times 2 and 3, however in model 3 this was not the case. In models 1 and 2, SEP at time 2 was positively associated with physical activity at time 3, however this relationship was not found in model 3.

	Dependent Variable	Independent Variable	B	S.E	B/S.E	p
Non-White Females						
	DEP3	SEP2	-0.211	0.099	-2.134	0.033
	PF3	SEP2	0.198	0.083	2.371	0.018
	METS3	SEP2	0.405	0.152	2.667	0.008
Non-White Males						
	PF2	SEP1	0.164	0.053	3.087	0.002
White Females						
	DEP2	SEP1	-0.089	0.030	-2.963	0.003
White Males						
	DEP3	SEP2	-0.088	0.027	-3.254	0.001
	PF3	SEP2	0.072	0.023	3.207	0.001
	DEP2	SEP1	-0.050	0.023	-2.173	0.030
	PF2	SEP1	0.049	0.019	2.633	0.008
	METS3	SEP2	0.185	0.073	2.531	0.011
	BMI3	SEP2	-0.211	0.052	-4.039	0.000
	BMI2	SEP1	-0.131	0.046	-2.841	0.005

TABLE 5.15: Parameter estimates for unstandardised parameters from Model 1 relevant to Hypothesis 4

	Dependent Variable	Independent Variable	*B*	S.E	*B*/S.E	*p*
Non-White Females						
	DEP3	SEP2	-0.176	0.085	-2.065	0.039
	METS3	SEP2	0.367	0.140	2.624	0.009
Non-White Males						
	PF2	SEP1	0.167	0.053	3.128	0.002
White Females						
	DEP2	SEP1	-0.094	0.030	-3.095	0.002
White Males						
	DEP3	SEP2	-0.070	0.025	-2.824	0.005
	PF3	SEP2	0.065	0.021	3.141	0.002
	DEP2	SEP1	-0.050	0.023	-2.149	0.032
	PF2	SEP1	0.051	0.019	2.700	0.007
	METS3	SEP2	0.170	0.069	2.441	0.015
	BMI3	SEP2	-0.203	0.051	-3.984	0.000
	BMI2	SEP1	-0.134	0.046	-2.883	0.004

TABLE 5.16: Parameter estimates for unstandardised parameters from Model 2 relevant to Hypothesis 4

	Dependent Variable	Independent Variable	*B*	S.E	*B*/S.E	*p*
Non-White Females						
	-	-	-	-	-	-
Non-white Males						
	PF2	SEP1	0.159	0.054	2.963	0.003
Non-white Females						
	DEP2	SEP1	-0.093	0.030	-3.071	0.002
White Males						
	DEP3	SEP2	-0.082	0.039	-2.126	0.034
	DEP2	SEP1	-0.051	0.023	-2.186	0.029
	PF2	SEP1	0.049	0.019	2.600	0.009
	BMI2	SEP1	-0.132	0.046	-2.852	0.004

TABLE 5.17: Parameter estimates for unstandardised parameters from Model 2 relevant to Hypothesis 4

5.5 Discussion

Overall, the three SEMs found little evidence to support any of the hypotheses represented in the model. There was little to no evidence to support a direct relationship between obesity and depression in this sample, and weak evidence for indirect relationships via the proposed mediators. There were no significant co-variations between BMI and depression at time 1 in any of the groups, and there were no consistent cross-lagged effects between them in any of the groups, in any of the models.

Viewed in the context of the exploratory analysis in Chapter 3 however, the results are perhaps not surprising. In that analysis there was little evidence of a relationship between obesity and depression, both cross-sectionally and longitudinally, and the only mediator that was highlighted as having some evidence of an effect was physical function. The substantive conclusions from the models are hence broadly the same as those from Chapter 3, perhaps with the caveat that a moderating effect of physical function was only found in white males. Evidence supporting that effect in white males is still weak at best.

Overall, the findings from my analysis of the Whitehall II data conflict with the data presented in the review of reviews I conducted in Chapter 2. The data presented in those studies suggested that there was a relationship between obesity and depression, with reviews and meta-analyses of longitudinal data and cross-sectional data both finding evidence that supported an association between obesity and depression.

Two reviews pointed out that there was a risk of publication bias towards the reporting of positive results (Mannan et al., 2016a; Mannan et al., 2016b). It is possible therefore that the positive results found in reviews in Chapter 2 are skewed towards positive results, and that there is un-published evidence which also finds a lack of associations between obesity and depression. However, four other reviews presented in Chapter 2 found no significant evidence of publication bias, weakening the possibility that this is responsible for the contradictory conclusions (Luppino et al., 2010; Abou Abbas et al., 2015; Pereira-Miranda et al., 2017; Quek et al., 2017).

A perhaps better explanation, is that there is a previously unconsidered moderating effect of age. In a clustering analysis of individuals with obesity by Green et al. (2015), two clusters of elderly people were found, described as

the "affluent and healthy elderly", and the "physically sick but happy elderly". One cluster of middle-aged people was also found which was described as "the unhappy and anxious middle-aged". Since the Whitehall II data I have analysed contains adults from the age of 55 and older, it is possible that within the obese sub-population of the Whitehall sample, many of them may fit into the two categories of elderly people, which, based on their description are perhaps less likely to be both depressed and obese. This could be responsible for the weak associations found between body weight and depression in my analysis.

Evidence presented in Chapter 2 was broadly consistent in suggesting that obesity could be associated with depression via an effect on poor physical health and physical function. In this analysis, I focused on the effect of physical function, and only found weak evidence to support a mediating effect of physical function, and only in the white male group. Given that the white male group had the largest sample size (64.8% of the sample), this may indicate that an effect does exist between physical function and future depression, but that it is sufficiently small that it can only be detected in large sample sizes. However, it may also indicate that physical function concerns associated with obesity do not occur in older adults. Experiencing some sort of physical decline in later life is unavoidable and so in a cohort of older adults, having poorer physical function, even due to obesity, may not have such an impact on the participants mental health as it is a common experience in that demographic.

There was also little to no evidence to suggest that physical activity level is a mediator between obesity and depression in either direction. This finding is contradictory to the evidence presented in the reviews in Chapter 2 and the wider literature in general, which suggest that physical activity is associated with future depression, such that higher levels of exercise are associated with lower depression (Rebar et al., 2015; Teychenne, Ball and Salmon, 2008). This contrary result may be due to heterogeneity in the effect of physical activity on depression, such that the strength of the association varies across different groups within the population. Evidence from the literature suggests that the effect of physical activity on depression is lower in non-clinical populations than in clinical populations being treated for depression (Rebar et al., 2015; Cooney et al., 2014). We saw in the analysis in Chapter 3 that depression scores in the Whitehall II were heavily right skewed with the vast majority of participants having very low scores on the GHQ-30 depression items. If the

finding of lower associations found within non-clinical populations is influenced by there being lower levels of depressive symptoms present overall, then this might explain why no association was found in this study.

The lack of association may also be due to weaknesses in the measures used within the study. Whilst the measure of depression used has some overlap with the validated depression subscale from the GHQ-28, it is not a validated measure of depression itself. As a result, participants' levels of depression may not be adequately captured in the model, possibly leading to an attenuation of estimated associations. My measure of depression is missing three items from the GHQ-28 depression subscale. These items ask respondents the extent to which they have recently: 'thought of making away with themselves'; 'found themselves wishing they were dead and away from it all'; and 'found the idea of taking their own life kept coming into their mind' (Goldberg and Hillier, 1979). If physical activity has a stronger effect in reducing these symptoms than those included in this study, then this may be responsible for the lack of association between physical activity and depression in the model.

In terms of the reverse association, whereby depression was hypothesised to be associated with future obesity via an effect on physical activity, there were no studies presented in the reviews in Chapter 2 that examined depression as a predictor of future activity. Considering this, the results of my analysis might indicate that depression is not associated with future physical activity, and that any association between physical activity and depression that has been observed else-where in the literature is due to a reciprocal effect.

There was little evidence to suggest that calories consumed mediates either the relationship between obesity and future depression, or the relationship between depression and future obesity in this cohort. In particular, the models' results suggest that BMI is not associated with future calorie consumption, nor is calorie consumption associated with future BMI in this cohort. The relationship between BMI and future calorie consumption is rarely examined; research has instead focused on the impact of calorie consumption on future body weight. In the absence of other data, the findings from these models might suggest then that obesity is not associated with future calorie consumption. This lack of association may reflect a combination of two opposing effects. Whilst some individuals with obesity may be likely to eat an increased number calories in the future, higher BMI has also been shown to be associated with an increased likelihood of dieting (Jeffery, Adlis and

Forster, 1991). Individuals with higher BMIs who are engaging in calorie restriction may, in effect, balance out with other individuals with higher BMIs that are eating a high number of calories, thus leading to an overall null association.

A more surprising result here is that calories consumed was not associated with future BMI in the model. There are, however, caveats to consider before one would conclude that this is evidence that calories consumed is not associated with future levels of obesity. The first thing to consider is that whilst calorie consumption tells us how much energy an individual is consuming on average, it does not give us information on how much energy they have expended. Obesity is primarily a condition associated with energy imbalance, in which more energy is consumed than is expended. High energy consumption may therefore only be associated with future obesity in so far as it causes an individual to live with an energy imbalance. Calorie consumption co-varied positively with physical activity in all groups except non-white males so it is possible this positive covariation influenced the amount of energy imbalance in the sample whereby individuals who consumed more calories also performed more physical activity in compensation.

Another explanation for the lack of association between calorie intake and BMI is that limitations of the derived measures of calorie consumption may also have effected the strength of the detected relationships estimated by the model. As was discussed at the end of Chapter 3, diet measures derived from FFQ data are at risk of bias due to the over-reporting of foods considered 'good' and the under-reporting of foods considered 'bad' (Hebert et al., 1995). The presence of this social desirability bias might lead to an under-reporting of calories consumed if high calorie 'bad' foods were under-reported in the FFQ data. Additionally, this under-reporting of high-calorie foods, and the over-reporting of low calorie 'good' foods might also reduce the variability present in the diet measure, since reporting that is biased towards a 'healthy' diet might also bias participants derived calorie consumption closer to the calorie intake that is associated with a 'healthy' diet. This reduced variation in calories consumed would in-turn lead to attenuation of relationships between calories consumed and other variables in the model, as has been seen before with FFQ diet data (Kipnis et al., 2002).

The distribution of calories in Figure 5.3 suggests that very few participants are consuming very high numbers of calories. In fact, approximately half of the participants report consuming less than 2000 calories per day, which is

less than the recommended daily calorie intake for both men and women. Given around two thirds of the study participants are male, for the distribution of calorie intake to be accurate, a large number of the study participants would have to be consuming a calorie restricted diet. This is perhaps evidence of the under-reporting issues mentioned above, and as a result may have contributed to the lack of associations found between calories consumed and other variables in the model. At the very least, it suggests that some caution should be taken in making strong conclusions about the relationships between calorie consumption and obesity and depression using this model.

Within the lack of mediation by calorie consumption between BMI and depression, the models also found little evidence that calorie consumption and depression were associated with one another over time. Initially, this finding appears contrary to the evidence presented in Chapter 2 which suggested that poorer diet was associated with increased risk for depression. However, the reviews presented focused on the effect of eating behaviours such as binge eating (Preiss, Brennan and Clarke, 2013), disordered eating (Allison, 2003), and diet quality (Hoare et al., 2014). None of the studies investigated the relationship between calories consumed and depression and so it is possible that the differing relationship I have found is due to factors such as binge eating having a relationship with depression that is indepedent of calories consumed. For example, the binge eating may be related to depression through feelings of shame and being out of control over one's eating behaviour, which themselves may be un-related to the amount of calories one consumes in general. In summary, it may be the manner in which calories are consumed, rather than the exact quantities, that is related to depression.

It is also possible that the lack of associations I have found between diet and depression are evidence of this relationship being a complicated and inconsistent relationship. Reviews of the relationship between diet and depression have provided some suggestion that 'healthy' diets may be beneficial in protecting against depression, however, evidence of this effect is considered neither strong nor conclusive (Quirk et al., 2013; Rahe, Unrath and Berger, 2014; Lassale et al., 2019). The strongest evidence appears to be found for Mediterranean style diets providing protection against depression (Lassale et al., 2019), however, again evidence is not unequivocal (Quirk et al., 2013). There is also evidence suggesting that the consumption of particular food groups such as fruits, nuts and legumes may be specifically protective against

depression (Sanhueza, Ryan and Foxcroft, 2013).

'Healthy' diets such as the Mediterranean diet are not necessarily low calorie, (a Mediterranean diet can include high quantities of carbohydrates and fats through bread and oil) and so it is possible that participants may have been consuming a nutritious, balanced diet that confers protection against depression, whilst still consuming a higher amount of calories. In order to explore the true relationship between calories and depression, one would need to control for these protective dietary patterns. This was out of scope for this study, as doing so would begin to make the model too large and complicated. As a result, it is possible that any effect of calories on depression is confounded by the dietary patterns of the participants.

Socioeconomic position, based on the extent of financial problems, was also found to have a limited effect on this system, with the only consistent effect over time being an effect on depression. It co-varied significantly with depression in all groups and had a consistent longitudinal association in white males such that higher SEP was associated with lower depression. This may suggest that the only impact that SEP has on the obesity-depression relationship is via an impact on depression. This would be in line with literature presented in the review by Preiss, Brennan and Clarke (2013) that suggested financial problems were associated with higher prevalence of comorbid obesity and depression. Financial problems might lead to an increased rate of comorbidities between obesity and depression if it is associated with both depression and obesity concurrently and as such was a joint cause of the two conditions. Else, if depression leads to obesity via other mechanisms then the association between financial problems and depression may then in turn produce an association between financial problems and comorbidities of obesity and depression. However, neither of these relationships were supported by the model and so at this point there is not any firm evidence to support that hypothesis.

Once again, null associations may have been influenced by limitations in the measures of SEP. The vast majority of participants reported that they experienced little to no financial problems with more severe financial problems being less common. This low amount of variation in the extent of financial problems may have attenuated relationships between SEP and other constructs in the model, particularly if the underlying relationship is only small. However, no other SEP measures were deemed suitable for this analysis, as

they suffered with either very high missing data rates, or were not available at multiple waves in the study and so could not be used in the analysis without potentially introducing bias into the model parameters.

Although it was not a main focus of my analysis, through examining ethnicity as a confounder it was possible to examine the potential for ethnicity to have a moderating effect on this system. The model fit tests and pattern of estimated paths provided some evidence of a moderation effect by ethnicity, but overall the evidence was not sufficient to suggest that meaningful differences had been detected. In particular, the tests of homogeneity found significant differences between ethnicities, and in the estimated paths there were almost no paths that were significant concurrently in both the non-white and white groups: only the regression path from physical function at time 2 to depression at time 3 in model 1 was significant in both non-white females and white females.

The differences highlighted above might seem like corroborating evidence of ethnic differences, however, when viewed in the context of the overall substantive conclusions suggested by the model, the overall model sample size and the vastly different sample sizes in non-white and white groups it seems unlikely that these differences are sufficient evidence to conclude a true moderating effect of ethnicity has been found.

The main reason to doubt that there is a moderating effect due to ethnicity is that the substantive conclusions suggested by each model are largely the same in both ethnic groups, namely that there is little to no consistent evidence to support the hypotheses investigated by the model. If there was a true difference in the relationships between obesity and depression across the white and non-white groups, then one would expect that this would be detected in the model to an extent that was sufficient to lead to differing conclusions being drawn about the relationships that are present in each group. Since no differing conclusions have been drawn, it therefore does not make sense to suggest a moderating effect exists.

A second reason to doubt the existence of a moderating effect of ethnicity is that the detection of significant differences between models, and significant paths within models, may have been influenced by the sample size of the model. The overall model size is large, and the chi-square difference test used is sensitive to sample size, such that models with a larger sample size will return a significant result for a much smaller difference between the

models than do models with a smaller sample size. Hence the significant results in the tests conducted to detect confounding by ethnicity are perhaps a result of the large sample size, rather than the existence of true moderation by ethnicity.

Similarly, the differing significance of paths in the models between the white and non-white groups may also have been influenced by the differing sample sizes in the two groups. Paths in the model were considered 'important' if they had an associated p-value less than 0.05 in the model. This p-value will be influenced by the sample size such that p-values are more likely to be smaller when the sample size is large. The sample size in the non-white group ($N = 724$) was substantially smaller than was the case for the white groups ($N = 7534$). Therefore, paths are more likely to be considered important in the white groups than the non-white groups. This effect can perhaps be seen in Figures 5.6 to 5.9 whereby the model with the largest group (White males) has the most 'important' paths than the other groups, and particularly more so than the non-white groups.

The Chi-square difference tests conducted on all of the structural models suggested that each model had a significantly worse fit to the data than the strong invariance model, perhaps casting some doubt as to the reliability of their model parameters and the substantive conclusions that can be drawn. However, as has already been noted, this Chi-square difference test is sensitive to sample size and the decreases have small effect sizes in terms of Cramer's V. Furthermore, the model fit statistics suggest the models have a good or adequate fit when compared with general rules of thumb such as RMSEA < 0.05 and CFI > 0.95. Therefore, the model results are likely to be reliable.

5.6 Limitations

Despite conducting a comprehensive analysis of the Whitehall II data-set, and using state of the art missing data methods, there are limitations to this study which must be taken into account when viewing the results. I have already discussed limitations of the measures used in the study in the previous section, and so here I will focus on other potential issues.

Firstly, limitations in the data might affect the generalisability of the conclusions drawn from this study. The Whitehall II data set contains data from

a narrow range of the UK population, namely civil servants who were recruited from 20 London offices in the 1980s. As a result, this population is likely to have different demographics and socioeconomic characteristics to the U.K population as a whole. For example, in terms of socioeconomic position, one might expect a bigger range of socioeconomic circumstances to be found in the general population than was the case in this sample, and average socioeconomic circumstances may also be higher in the sample than the general population. These altered distributions on key concepts may affect the model estimates for associations, particularly if the relationships are not exactly linear across their entire domains.

However, it is important to note that longitudinal datasets that consistently collected data, using a consistent format, on all the necessary constructs that were needed for this model were lacking. In order for detailed quantitative research to investigate these potential mechanisms in a broader context, more considered data collection will be needed to be carried out to address the issues I faced in trying to find a suitable dataset that contained data capable of investigating my research questions.

The second main limitation comes as a result of the compromise between model depth, and model breadth. In this study, the model was designed to analyse multiple potential mediating mechanisms that have been suggested as important in the relationship between obesity and depression. In order to avoid creating an unwieldy model that was difficult to analyse and interpret, single constructs were used to represent each mechanism. However, in some of these cases, the true structure of the mechanism may be more complicated than could be represented by a single mediating construct. Future research may therefore benefit from analysing the mechanisms I investigated here in isolated models, with more detail included from theory detailing how that mechanism might operate.

5.7 Conclusions and Next Steps

In the review of the literature conducted in Chapter 2, I noted that there was a lack of evidence from longitudinal studies examining the relationship between obesity and depression, and where longitudinal evidence existed, it did not explore in detail the potential mechanisms by which obesity and depression might by related. In these last three chapters, I have addressed this gap by analysing longitudinal data from the Whitehall II study using

a Structural Equation Model to investigate whether there was evidence for some of the underlying mechanisms that were suggested to be important by previous research.

Overall, little evidence was found to support an association between obesity and depression, either directly, or via the hypothesised mechanisms of physical activity, diet and physical function. Additionally, socioeconomic position (as defined by the extent of financial problems) was not found to have a consistent effect on any of the other constructs examined.

However, it should be noted that limitations in the data may limit the generalisability of these conclusions and so the lack of support found for the mechanisms investigated here may not be the case in general. Despite this, the evidence here suggests that obesity and depression are not strongly related in this population and as such, any relationships found in previous studies indicates further that this relationship is heterogeneous. Additionally, if physical function, physical activity and diet do relate obesity to depression, then the results of this study suggest the effects may occur through aspects of these constructs that were not considered in this study, such as disordered eating, social impacts of exercise and development of diagnosed physical health conditions.

For the rest of this thesis, I will expand upon the analysis presented here, by investigating whether obesity stigma is a driving mechanism of the relationship between obesity and both depression and socioeconomic position.

Chapter 6

An Initial Agent-Based Model of Stigma in the Obesity-Depression-SEP System

6.1 Introduction

In Chapter 2, I found that one of the most consistently hypothesised mechanisms that may relate obesity and depression was that of obesity stigma. However, the suggestion that stigma may play a role in the development of depression in people with obesity was often made speculatively by authors of the reviews, with only scarce empirical evidence supporting the claims being presented. In the following two chapters, I will address this by developing an agent-based model to explore the role of stigma in generating both comorbidities in obesity and depression as well as the socioeconomic distribution of obesity in the population.

This chapter will be presented in five main sections. In Sections 6.2 and 6.3, I will outline the Agent-based Modelling (ABM) methodology and describe the steps that one takes in order to develop an ABM. This is followed in Sections 6.4 and 6.5 by the description and analysis of a simple agent-based model investigating whether obesity stigma is capable of generating observed relationships between body weight and both depression and socioeconomic position. The implications of the model results are then briefly discussed in Section 6.6.

In Chapter 7, I will then present an improved version of this model that aims to address some of the shortcomings of the model presented in this chapter such that more credible conclusions can be drawn from the model output.

6.2 Agent-based Modelling

An Agent-based Model (ABM) is a computational model in which a real world system is represented by a collection of autonomous 'agents' who may interact with each other and their environment based on a collection of specified behaviour rules (Epstein and Axtell, 1996; Epstein, 1999). This focus on describing the system at the micro-level is one of the most important and distinguishing features of Agent-based Modelling (Bonabeau, 2002). In contrast to other modelling methodologies like system dynamics, Agent-based Modelling focuses on describing a system using a micro-level representation, rather than explicitly representing higher macro-level processes and relationships. Once a micro-level specification of the system has been described one can then examine what macro-level phenomena can emerge from the behaviours and interactions of the micro-level specifications (Gilbert, 2008).

Agent-based Models are particularly well suited to investigating how macroscopic properties of systems are formed. When we are interested in explaining some macro-level property of a real world system, Agent-based Modelling approaches this by asking whether the macro-level phenomena can be generated by the behaviours and interactions of autonomous, heterogeneous agents at the micro-level (Salgado and Gilbert, 2013; Epstein and Axtell, 1996; Epstein, 1999). If a micro-level specification is able to generate the macro-level phenomena, then the micro-level description is considered a candidate explanation for the phenomena. This paradigm is summarised by Joshua Epstein in his extensive work on ABMs as: "if you didn't grow it, you didn't explain it" (Epstein, 2012). In other words, if the Agent-based description of the system cannot generate the phenomena of interest, then it cannot be an explanation for the formation of that phenomena. If it does generate the phenomena, it is a possible explanation (perhaps one of many). The quality of the explanation can then be examined using further supplementary research (Casini and Manzo, 2016).

ABMs provide a flexible way to investigate the properties of systems that are made up of multiple interacting entities. One way in which ABMs provide this flexibility is by allowing modellers to describe their model in varying amounts of detail. At the inception of a modelling project it may not be entirely clear what all the important features of agents and their interactions might be, so flexibility in how agents and their behaviour can be represented allows for changes to be more easily incorporated into the model as the

need arises (Bonabeau, 2002). Secondly, ABMs can be used to describe a wide range of entities and interactions: new agents and their rules of interaction can be straightforwardly added to an existing model without having to redesign the rest of the model from the ground up (Bonabeau, 2002). This means models investigating a similar system to an existing ABM may be able to save considerable time in the project by simply adapting or adding to the pre-existing model to answer the new research question.

Agent-based Modelling can explicitly incorporate feedback loops and adaptation, which may be difficult to include in other modelling paradigms (Gilbert, 2008; Bonabeau, 2002). As a result they are commonly used to model complex systems where these features are often present (Gilbert, 2008).

A typical Agent-based Model has three main components: the agents themselves, a collection of agent interactions and relationships, and an environment. These components are described in detail below.

6.2.1 Agents

Whilst there is no generally agreed upon definition for what an agent is, generally speaking agents are the pieces of the Agent-based Model used to represent the entities of interest that exist in the real world system (Gilbert, 2008; Epstein and Axtell, 1996). In our system the main agent of interest will be people, however in other applications agents may represent a wide range of entities such as businesses (Kant, Ballot and Goudet, 2020), cells (An et al., 2009), traffic and more (Chen, 2012).

Agents are often described as having four main characteristics (Gilbert, 2008):

- *Perception:* Agents are able to perceive their local environment, along with other agents that are considered nearby. This is important as it informs which agents are able to interact with others and what aspects of their environment they can interact with. Without perception agents would be entirely isolated from one another in the model.
- *Performance:* Agents have a set of actions they are able to perform, such as moving around the environment, or interacting with other agents and the environment. The ability to perform interactions with other agents in the simulation, and with the environment is a key feature of Agent-based Modelling.

- *Memory:* They have a memory, which records the agent's previous states and actions. This memory may be long or short, but at the very least agents need a way of incorporating their previous state into their next one.
- *Policy:* Agents have a set of rules or heuristics which describes how they decide which actions to take at any given point.

The above features can be implemeted in varying levels of detail, depending on the aim of the simulation model.

Within the description of the above four characteristics, agents are often represented with other important features. Firstly, agents are autonomous such that the actions they perform are not decided or dictated by another entity within the model (Macal and North, 2010; Bonabeau, 2002). Instead, agents make individual decisions on what actions to take at a given point in the simulation based on their policy. The policy by which agents make these decisions can be very simple, or they may involve more complicated decision rules reflecting influences from the environment, interactions with other agents, and their own internal psychology (Bonabeau, 2002). Agents may also be given the ability to adapt and modify their policy and performance in order to better achieve some overall goal (Epstein and Axtell, 1996; Macal and North, 2010).

Agents can also be highly heterogenous (Salgado and Gilbert, 2013). Each agent has a state which describes the agent in terms of a set of important characteristics (Bonabeau, 2002; Gilbert, 2008). These important characteristics are decided based on the model's objectives and are represented in the agents by a set of variables. Each agent in an ABM is a uniquely identifiable object within the model and so has their own set of values for each of these variables which describe the agent's state on each characteristic (Salgado and Gilbert, 2013; Macal and North, 2010; Gilbert, 2008).

As well as heterogeneity among the agents' states, agents can also be heterogeneous in terms of all four of the main characteristics listed above (Gilbert, 2008; Bonabeau, 2002). Agents may have heterogeneous perception such that some agents can perceive other agents who are far away, whilst others can only perceive those nearby (Epstein, 2002). Some agents may only be able to carry out a limited set of actions in comparison to others, and may also have a different memory length that can be used to record their previous states (Gilbert, 2008). Finally, agents may have different policy rules and goals, leading

them to make different behaviour choices in the model (Epstein and Axtell, 1996; Epstein, 2013).

6.2.2 Agent Interactions

Agent interactions are described in an ABM by two main model components. The first of these components is a description of which agents can interact with which others, and the second consists of the rules and dynamics of these interactions.

In order to describe which agents can interact with which others in the model, a model topology is described that states which agents in the model are connected to one another (Macal and North, 2010). There are multiple ways in which the topology can be described, and the representation is often tailored so that it recreates some key features of the interactions present in the real world system (Bonabeau, 2002).

For example, agents may be connected to other agents based on spatial relationships such as proximity to one another. In an ABM predator-prey model, predator agents might interact with prey agents by moving towards any prey within a certain distance of them in the model and killing those whom they are immediately adjacent to. Similarly, prey might move away from predators who are within a certain distance of them (Wilensky and Reisman, 2006)

The model topology can also include a social network component, for when spatial considerations are not of primary interest in the interaction (Bonabeau, 2002; Hamill and Gilbert, 2009; Hamill and Gilbert, 2010). Within a social network based topology, agents are represented by nodes in the network and links between the agents represent connections over which interactions are then able to occur. The network can be static so that agents always interact with the same set of neighbours over the course of a simulation, or dynamic to allow agents to interact with different subsets of the population as time progresses (Hamill and Gilbert, 2009; Hamill and Gilbert, 2010).

The mechanisms and dynamics of the interaction are primarily informed by the modelling objectives and the underlying theory surrounding the system of interest (Casini and Manzo, 2016). Interactions in the real world system are likely to be complicated and so the important features will likely have to be extracted so that a simplified representation can be modelled (Auchincloss and Garcia, 2015). These modelling choices are likely to influence and be

influenced by the topology of the model, as different topologies are better suited to representing different mechanisms of interaction.

Whatever the nature of the underlying system, interactions in an Agent-based Model are built around the capacity of agents to send and receive messages to and from other agents in the model and the environment (Gilbert, 2008). These messages may be simple and involve a simple transfer of data from one agent to another. For example, one agent might receive information about what the neighbouring agent's BMI value is. However, messages can also be more complex with agents sending and receiving information in a shared language (Gilbert, 2008). The interpretation of messages in this language does not necessarily have to be the same for both agents involved in the interaction. Information passed from agent to agent and between agents and the environment can then feed into the agents' individual behaviour decisions, creating feedback effects.

6.2.3 The Model Environment

The environment of an ABM describes the virtual world in which the agents are situated. This virtual world may represent a physical space, such as that in Esptein's artificial Anasazi model (Epstein, 2012), or it may represent a more abstract space. As with the agents themselves the specific detail of the environment will depend on the objectives of the model. In research settings where exploring effects of the environment are central to the research question, a detailed environment may need to be included in the ABM. However, in other applications, the spatial environment may not be of particular interest and so may only be minimally represented. Even when a spatial environment is represented, the level of detail can range from a largely abstract representation of physical space to a highly detailed recreation of a specific physical environment based on geographic data (Macal and North, 2010; Epstein, 2012).

From a practical standpoint, the model environment can also provide a useful medium through which agents can pass information to one another (Gilbert, 2008). It can make the monitoring of agents easier, and prevent messages being delivered out of turn between the agents. Specifically, agents all send messages to the environment, before these are then sent on to the recipients so that no agent receives a message before all other agents have sent theirs (Gilbert, 2008).

6.3 The Model Building Process

The process for building an ABM can be broadly broken down into three main stages: specification and formalisation; modelling, verification and experimentation; and calibration and validation (Salgado and Gilbert, 2013). I will briefly introduce each stage here, with further details being given below. In the specification and formalisation stage we are primarily concerned with the design of two models: the conceptual model and the computational model. The conceptual model for a simulation model can be defined as *"a non-software specific description of the computer simulation model (that will be, is, or has been developed), describing the objectives, inputs, outputs, content, assumptions, and simplifications of the model"* (Robinson et al., 2010). As such the, conceptual model describes generally how the real world system is to be represented in the form of a simulation model. The modeller then takes this conceptual model and translates it into the a specific computational design - in our case an Agent-based Model.

The modelling, verification and experimentation stage then broadly covers the activities of implementing the design of the computational model into a computer code, and performing some initial tests and experiments on the model to gain some first insights into its behaviour (Salgado and Gilbert, 2013). Calibration and validation then provides a more formal investigation of the model behaviour using data from the real world system.

Whilst these stages of analysis are done primarily in the order described, it is important to note that there is considerable iteration between the stages, as insights in later parts of the process can highlight changes that need to be made to other parts of the model's design (Robinson et al., 2010). For example, attempts to design the computational model may highlight gaps in the conceptual model, requiring the researcher to revisit the conceptual model before coming back again to the computational model. Similarly, model calibration may highlight issues with the design of the computational model that need resolving by further development. This again might require the conceptual model to be revisited and redeveloped.

Stage 1: Specification and Formalisation

As I mentioned above, the first stage of building an Agent-based Model involves designing two main models: the conceptual model and the computational model. Here I will describe in more detail what the designing of these

models aims to achieve.

Designing the Conceptual Model

The first stage of the developing an ABM involves creating a conceptual model that describes how the real world system is to be represented in such a way as it can be investigated with a simulation model. In understanding real world systems, researchers rely on implicit mental models that describe what features and mechanisms of the real world system they believe are important within the system (Auchincloss and Garcia, 2015; Epstein, 2008). However, due to their implicit nature, these mental models may include hidden assumptions and implications which cannot be examined or critiqued. The process of specifying a conceptual model thus serves to put the modelling on more solid ground by translating the mental models which describe the researcher's understanding of the system into an explicit description of how the real world system is to be represented in the simulation model (Epstein, 2008).

Of primary focus within this process is defining the model's research questions (Robinson et al., 2010). The model research questions have a large impact on the model scope and boundary, and thus provide an anchoring point on which decisions about what features are included within the conceptual model can be based. In line with the modelling rationale of ABMs, research questions are usually of the following structure: *"how do decentralised individual behaviours and interactions generate a certain macroscopic phenomenon?"* (Salgado and Gilbert, 2013; Epstein, 1999). Both the research questions and the conceptual model start broad, characterising the general problem the researcher wishes to investigate as well as some important characteristics from the system. From this broad conceptual model, gaps in the researcher's mental model and understanding can be explored and used to focus down the research questions, and the conceptual model itself, around a narrower part of the underlying system that can be explored more succinctly using an ABM (Auchincloss and Garcia, 2015).

This process of narrowing down the scope of the conceptual model also provides the opportunity to examine the evidence (or lack of evidence) on which the original mental models were based, and can also expose gaps in current literature surrounding the system of interest (Auchincloss and Garcia, 2015; Epstein, 2008).

Overall, the conceptual model should be as simple as possible whilst still representing the key features necessary for answering the research questions (Robinson, Sutin and Daly, 2017). The flexibility that Agent-based Modelling provides can often come with a temptation to represent every fine detail of the system and its components in the model. However, doing so would create a model that was difficult to investigate and understand (Casini and Manzo, 2016). As a result, detail should only be included if they are deemed sufficiently important to the model's research questions.

Various frameworks exist for how to describe a conceptual model that underpins a simulation project. In this thesis, I have chosen to use the framework provided by Robinson et al. (2010), which breaks down the description of the conceptual model into six sections: understanding the problem; determining the model objectives; identifying the model outputs; identifying the model inputs; determining the model content (scope and level of detail) and identification of any assumptions and simplifications.

The first stage of developing a conceptual model in the Robinson et al. (2010) framework involves developing an understanding of the problem situation that the simulations aim to address. This stage in the development often highlights gaps in what is currently known about the underlying system, and as such can provide a way to focus the research questions the model aims to address.

Having outlined the problem situation, the next stage of Robinson's framework for developing a conceptual model is to determine the modelling objectives. This stage sets out what one aims to achieve by creating and experimenting with the simulation model. This then sets out a reference point around which design choices in the model can be made (Robinson et al., 2010). In a research context this is where the specific research questions the model aims to address are given.

After the modelling objectives have been determined one can outline the modelling outputs and inputs. The modelling outputs are used to assess whether the modelling objectives have been achieved, which in this case is whether the model research questions have been answered (Robinson et al., 2010). In other words, this section describes what outputs produced by the model will be used to answer the research questions.

Determining the model inputs requires describing the experimental factors that are to be altered or tuned in order to achieve the model objectives. Broadly

speaking, these are variables and parameters in the model that are to be tweaked in order to alter the simulation output (Robinson et al., 2010). Another important input to consider is data from the real world system that will be used to inform how well the model fits the underlying reality it aims to represent. Whilst this input data falls outside of Robinson's definition of an input, as it is not a variable factor in the model, data is a major source of information the model draws upon to investigate the model research questions (Casini and Manzo, 2016; Thiele, Kurth and Grimm, 2014; Fagiolo, Moneta and Windrum, 2007).

In the fifth stage of the framework one then focuses on determining the model content. This requires two main steps: determining the model scope; and determining the level of detail. The process of determining the model scope involves setting out what parts of the system of interest are to be represented in the model, and which are omitted. The process of determining the level of detail then sets out how detailed the representation of in-scope components will be (Robinson et al., 2010).

Whenever a real-world system is represented in a model, parts of reality are often simplified in order to prevent creating an intractable model (Fagiolo, Moneta and Windrum, 2007). Similarly, assumptions are made to cover gaps in understanding of the real world system. In the final section of the conceptual model design, these assumptions and simplifications are outlined explicitly so that their credibility can be assessed (Robinson et al., 2010).

Designing the Computational Model

Once a complete conceptual model has been developed it is translated into a computational model that describes how the conceptual model will be represented quantitatively. Within the design of the computational model, important features of the model highlighted by the conceptual model need to be represented in a manner that is amenable to being programmed on a computer, whilst still remaining true to any assumptions and theories that are being represented in the model (Salgado and Gilbert, 2013; Gilbert, 2008).

It is at this point that the modeller decides how to represent the agents' characteristics, their interactions and the environment of the model in a concrete mathematical and logical description (Salgado and Gilbert, 2013; Epstein and Axtell, 1996). For example, agents in the model are assigned state variables to represent their key characteristics, and agent decisions are represented as

equations that describe the way an agent calculates how it will behave and interact with other agents and its environment.

Defining the components of the computational model can also highlight gaps in the design of the conceptual model such as assumptions that have been made but not stated and model content that was not previously considered in scope (Robinson et al., 2010; Epstein, 2008). As a result there is often much iteration between the designing of the two models within this stage of development for the ABM.

As was the case for the development of the conceptual mode, there are frameworks for the design of the computational component of an Agent-based Model, the most notable being the Overview, Design concepts and Details (ODD) protocol (Grimm et al., 2006). The ODD protocol aims to provide a consistent structure for reporting the design of ABMs such that models can be more easily critiqued and recreated by a reader. When describing the computational model of my ABM in section 6.4.2, I will use the ODD protocol and so here I will briefly introduce the sections to provide some context for that description. Full details of the protocol can be found in Grimm et al. (2006) and Grimm et al. (2020).

The ODD protocol covers seven elements under the three main sections. The overview section of the protocol aims to give a summary of the overall purpose and structure of the model and consists of three subsections: purpose; state variables and scales; and process overview and scheduling. The purpose subsection describes the overall aims and objectives of the model such as why an ABM is necessary for the project and what precisely the model is going to be used for (Grimm et al., 2006). This is important as it highlights why certain aspects of reality are included and others are not. The state variables and scales subsection then gives a description of the entities that are represented in the model. In particular, what are the agents included in the model and how are their characteristics represented by variables in the model. Similarly, how are entities such as the environment and any social networks described in the model. Finally, the process overview and scheduling subsection aims to give a general conceptual description of the the main processes that are included in the model, as well as their effects. The order of processes is also explained as well as how time is represented in the model. This is important as the order of processes within the model can impact on the model dynamics (Grimm et al., 2006).

The design concepts section of the ODD stands alone as its own element and describes the key model features within the design. (Grimm et al., 2006) outlines a checklist of features around which the model design can be described, in order to provide a consistent framework to describe the model features. The checklist includes the following items: emergence, adaptation, fitness, prediction, sensing, interaction, stochasticity, collectives, and observation; though not all these items need be described if they are not represented in the model.

The details section of the ODD protocol gives a more thorough technical description of the model. This section is broken down into three subsections: initialisation; input; and sub-models. The initialisation subsection in the ODD explains how the initial conditions of the model are set prior to each simulation. The input subsection deals with environmental conditions and dynamics that are imposed on the model, rather than being derived internally within the model as result of agent behaviours and interactions. The Submodels subsection then describes in more detail the processes outlined in the process overview and scheduling section. In particular, the logical rules and mathematical equations of the model are described and explained so that the underlying structure of the model can be analysed and underlying assumptions implied by the model structure can be identified. The description of the model rules should be complete enough that it is possible to be reproduced.

Choosing the Modelling Software

Between Stages 1 and 2 of the model building process the modeller must decide which software will be used to program the model (Salgado and Gilbert, 2013). Bespoke software can be created in an object-oriented language such as Java or C++, however, doing so adds a large amount of work to the modelling process. Instead of building software from scratch, one can instead make use of existing packages and libraries that have been designed for Agent-based Modelling (Gilbert, 2008).

Within the existing Agent-based Modelling software options, a popular choice with Agent-based Modellers is the NetLogo software program (Wilenski, 1999). NetLogo is a free to use open source program that was designed based on the simple programming language Logo. As such it has a simple syntax that makes it intuitive to use and easy to learn (Thiele, Kurth and Grimm,

2014). It also has a built in integrated development environment for programming and simulating Agent-based Models and a graphical user interface (GUI) for visualising the outcomes. Additionally, it has a large user base that can be used for support and troubleshooting.

NetLogo includes three built in standard agent types: turtles, patches and links. In general, turtles are used to represent the agents in the model, patches are used to describe the model's physical environment and links are used to describe the topology of interactions between Agents. Alongside these standard agents, NetLogo also includes pre-defined commands that can be used to design their behaviour rules. These building blocks, along with the software features listed above mean that building Agent-based Models in NetLogo is more straightforward than is the case for other software options such as Repast (Gilbert, 2008).

As a result of this, the models presented in this chapter and in Chapter 7 were both programmed in NetLogo (ver 6.20). Further details on the NetLogo programming environemnt can be found in its documentation (Wilenski, 2021).

Stage 2: Modelling, Verification and Experimentation

Once the computational model has been described, the model can be formalised into computer code, programmed and verified (Salgado and Gilbert, 2013). It's important to note at this point that whilst this is listed as a different stage, there is again considerable overlap between this stage and the design of the computational model, as the capabilities and limitations of the modelling software might need to be considered in the design of the mathematical relationships in the model.

Verification of the model's code is then necessary to ensure that the model design has been properly implemented, such that there are no coding errors and the mechanisms inscribed are working in line with their design (Salgado and Gilbert, 2013). Failure to properly verify the model may lead to erroneous conclusions about the system being drawn as the model may be failing to properly represent the conceptual model on which it was built.

The process of verifying an agent-based model can be complicated, as the exact state of a model at any one time point is usually not predictable, making it difficult to form an expected state against which the model results can be compared. However, some formal methods for verification do exist (Kefalas et al., 2003; Niazi, Hussain and Kolberg, 2009; Ammar and Abdallah, 2011)

and tests can still be designed and run within the model code to ensure as much as possible that the code is doing what was intended.

Alongside tests, experimentation can also form a vital part of the verification process. Running a broad range of parameter inputs aimed to produce extreme circumstances that might be likely to throw up errors during the model's execution can sometimes highlight hard to spot errors and weaknesses in the code. Exploring the parameter space is also useful to examine whether the model is able to generate the macroscopic patterns of interest in the model (Salgado and Gilbert, 2013). When a model can generate a pattern of interest, it is said to have reached 'generative sufficiency' (León-Medina, 2017). If the model is unable to generate the phenomenon of interest then the current specification is either not an adequate representation of the underlying system, or the hypothesised mechanisms are not a candidate explanation for the phenomena (Epstein and Axtell, 1996).

Stage 3: Calibration and Validation

The final stage of analysing an ABM is to calibrate and validate the computational model using data. There are two ways which we can view the process of calibration. The first is that by calibrating the model we are tuning the model parameters to find the values that make the model more closely match patterns that are observed in the real world system. The second way to view the process of calibration is that we are using data from the real world system to perform Bayesian inference in order to generate posterior distributions for the models' input parameters (Kennedy and O'Hagan, 2001). In other words, given some data from the real world system, the calibration allows us to describe the probability that each input parameter takes a given value in its input range.

Calibration can be done using both qualitative and quantitative data (Auchincloss and Garcia, 2015). Calibrating to qualitative data usually involves assessing the accuracy of a model based on its ability to reproduce a set of 'stylised facts' that are chosen to represent phenomena that are present in the real world (Gilbert, 2008). These stylised facts are derived from literature on the social theory that is relevant to the model's content. For example, in Epstein's model of civil violence, the model was considered a promising fit to the real world based on its ability to reproduce 'noteworthy phenomena' (Epstein, 2002). In particular, deceptive behaviours of agents emerged within the model, in which agents hide rebellious sentiments when police

agents are nearby, but then engage in rebellious activity once the police move away. Similarly, the completeness of Epstein's Artificial Anasazi model has sometimes been questioned due to its inability to reproduce the large scale population drop that occurred in the real world history (Epstein, 2012).

Quantitative calibration involves assessing how closely data derived from the simulation matches equivalent data from the real system (Auchincloss and Garcia, 2015; Thiele, Kurth and Grimm, 2014). For example, the distribution of a characteristic such as BMI within the simulated population could be compared with the distribution of BMI in observed data. This matching involves two main steps. Firstly, a set of targets from the real world data need to be chosen that will be used as comparators for the simulation output. Secondly, a method for calculating the distance between this target and the simulated data is decided in order to give a metric on which 'closeness' can be measured (Salecker et al., 2019).

Once parameters that provide a good fit to available data have been identified through calibration, the model then may go on to be externally validated (Casini and Manzo, 2016; Fagiolo, Moneta and Windrum, 2007). This process involves comparing the output of the calibrated ABM to additional data that was not used in the model building or calibration process. For example, the model design may be validated by examining the validity of the social theory that under-pins it, as well as by scrutinising how faithfully this theory has been reproduced (Auchincloss and Garcia, 2015). Additionally, the computational model could be validated by comparing the output of the calibrated model to additional quantitative data - if the calibrated model matches closely to new data, this provides more evidence of the model's usefulness and veracity.

6.4 A Simple ABM of Obesity Stigma

Having outlined the process for developing an Agent-based Model in the previous section, for the remainder of this chapter I will present a simple, initial ABM that aims to explore how stigma of obesity might generate associations between body weight and both depression and socioeconomic position. The structure of this section will roughly follow the stages of development outlined in the previous section.

Firstly, I will outline the conceptual model that I developed, which is described using the framework for conceptual modeling outlined in Chapter

4 of Robinson et al. (2010). The description of the conceptual model is then followed by the description of the computational model, which is outlined using the Objective, Design concepts and Details (ODD) protocol in (Grimm et al., 2006), which aims to provide a consistent structure for reporting the design of ABMs such that models can be more easily critiqued and recreated by a reader.

After the design of the model has been outlined, the model experimentation and calibration procedures are described, followed by the results. The implications of the model results are then discussed in terms their ability to shed light on the underlying system.

6.4.1 The Conceptual Model

This section describes the development of a conceptual model of the role of stigma in generating comorbidities in obesity and depression, as well as socioeconomic inequality in obesity. As mentioned in section 6.3 the description follows the framework for developing a conceptual model outlined in Chapter 4 of Robinson et al. (2010) and so is split into the following sections: understanding the problem; determining the model objectives; identifying the model outputs; identifying the model inputs; determining the model content (scope and level of detail) and identification of any assumptions and simplifications.

6.4.1.1 Understanding the Problem

Since this simulation model is being conducted in a research context, this problem situation can be broadly placed within the aim of improving the understanding of the current research on what impact stigma has in the obesity-depression-SEP system. To that end, in this section, I will briefly outline the relevant literature on which the simulation model aims to build.

The Stigma of Obesity

Stigmatising attitudes and behaviours towards people with obesity are widespread, and pervade all manner of situations. Reviews of research into the prevalence and consequences of stigma have shown that people with obesity can expect negative treatment in personal relationships (Chen and Brown, 2005), employment settings (Fikkan and Rothblum, 2012; Godfree, 2020),

education settings (Puhl and King, 2013) as well as negative portrayal in the media (Ata and Thompson, 2010).

In healthcare settings, people with obesity suffer from provider-based stigma from medical professionals who describe individuals with obesity as lazy, lacking will power and non-compliant with lifestyle changes (Puhl and King, 2013; Fruh et al., 2016). In schools, young people with obesity are subjected to provider based stigma from teachers who consider them less able in a variety of skills (Puhl and King, 2013) and in employment settings, having obesity has been associated with denial of promotions and mistreatment from co-workers (Fikkan and Rothblum, 2012; Godfree, 2020). In terms of public stigma, in a systematic review conducted by Sikorski et al. (2011), it was found that, in the only sample analysing public stigma prevalence, around a quarter of individuals expressed definite stigmatising attitudes, and only a fifth of individuals expressed no stigmatising attitudes. In the studies that examined attitudes towards different causes of obesity, participants in multiple studies were more supportive of explanations based on individual behaviour such as overeating and sedentary activity than was the case for environmental or social causes, and one study in particular found that will power was considered a cause over and above the food environment (Taylor, Funk and Craighill, 2006).

Portrayals of individuals with obesity in media could also be seen to reflect levels of public stigma. In a systematic review of weight bias within media consumed by children, adolescents and adults, Ata and Thompson (2010) found that people with obesity were presented unfavourably in a variety of media including books, film, television and news. In books, films and television, characters with obesity were more likely to be associated with negative behaviours and characteristics and receive more negative treatment from others. In news media, the issue of obesity was often framed in a light that emphasised the personal responsibility that people had for their condition.

Consequences of Obesity Stigma

From the above it is clear that having obesity increases one's chances of being subject to stigma. However, at this point, it is not clear what the precise effects of being in this stigmatised group are, and how stigma effects might lead onto differential health outcomes between stigmatised and unstigmatised groups. In this section, I will discuss the potential consequences of stigma, focusing specifically on the consequences of obesity stigma.

Impacts on Depression

The reviews presented in Chapter 2 showed that one of the best supported mechanisms linking obesity to the onset of depression was that of weight based stigma and body-image dissatisfaction. All six of the included conceptual reviews presented evidence that weight-based stigma may be a mechanism by which obesity confers greater risk for depression, and all but two of the non-conceptual reviews also discussed stigma as a likely mediator in the obesity-depression relationship.

More recent studies have also found evidence to support the notion that a consequence of weight based stigma is an increased likelihood of developing depression. In a study of 300 female undergraduate students, Stevens et al. (2017) found that weight stigma mediated the relationship between both childhood and current BMI and depressive symptoms, as well as the relationship between BMI and body image satisfaction. Further, in three nationally representative samples Robinson, Sutin and Daly (2017) found that Class II obesity (BMI between 35 and 39.9) and Class III obesity (BMI of 40 or more) were consistently associated with depression prospectively, and that this association was mediated by perceived weight-discrimination.

Worryingly, the association between obesity stigma and depression may not be combatable using individual attempts to cope or resist. In a study of German adults with obesity, perceived weight discrimination was still associated with increased depression in those who attempted to cope with using problem-solving strategies, and the association was exacerbated in those who resorted to avoidant strategies such as self-blame, denial and venting (Spahlholz et al., 2016).

Impacts on Obesity

The onset of depression is not the only adverse health outcome that being exposed to stigma might lead to; evidence also suggests that the experience of being stigmatised can lead to subsequent weight gain. Sutin and Terracciano (2013) showed prospectively that individuals who are stigmatized are more likely to become obese and stay obese than those who were not stigmatised. This effect was found uniquely for weight based discrimination, such that other forms of discrimination did not experience greater weight gain. Tomiyama (2014) and Brewis (2014) both presented models that point to stigma

as a cause and perpetuator of obesity. Brewis (2014) suggested that stigma contributed to weight gain via 4 main mechanisms:

1. Direct effects on diet, exercise and help-seeking behaviour;
2. Indirect effects via psychosocial stress;
3. Indirect effects from changes to social relationships; and
4. Indirect structural effects of discrimination such as layering of stigma.

Tomiyama (2014) made the cyclic nature of the relationship more explicit in their Cyclic Obesity Weight-Based Stigma (COBWEBS) model. In this model, greater body weight lead to greater levels of experienced stigma, which was presented as a chronic stressor to those who were stigmatised. The stress of stigma was then proposed to link stigma to increased body weight via effects on eating behaviour, physiological changes and emotional responses to shame.

Other Potential Impacts

Outside of health consequences, there is also evidence of socioeconomic consequences for those who are subjected to obesity related stigma. Studies of youth populations have found that those who are subjected to weight based stigma form weaker social ties, experience greater social isolation and can even suffer a damage to their academic achievement (Puhl and King, 2013). In adults, the experiences of weight-based discrimination in employment settings may lead to missed promotions and mistreatment from co-workers (Godfree, 2020). These consequences may then lead onto lower levels of income and achievement for people with obesity.

In two meta-analyses looking at prospective associations between body weight and socioeconomic position, consistent evidence was found for an effect of body weight on future income and education level, such that higher BMI was associated with lower future income and education level (Kim, Roesler and von dem Knesebeck, 2017; Kim and Von Dem Knesebeck, 2018). This effect was most pronounced in women and the authors suggested that the overall negative associations they found could be down to the consequences of stigma for higher weight individuals. Given the above, it is possible that obesity stigma might play a generative role in the formation of the inverse relationship found between obesity and socioeconomic position in high income countries such as the U.K (Cohen et al., 2013).

Sex Differences in the Effect of stigma

There is also evidence of sex differences in the associations between obesity and both SEP and depression. In terms of the relationship between obesity and depression, my review in Chapter 2 found that, in general, obesity is more strongly associated with depression in females than in males. Evidence from cross-sectional data was consistently supportive of an increased risk for developing comorbid obesity and depression in females compared to males, however, longitudinal data was less supportive. That being said, some longitudinal studies did find age sex interactions, and no studies found an increased risk for males over females.

Similarly, studies of the relationship between obesity and socioeconomic position have also found more consistent inverse associations in women than in men despite inverse associations affecting both sexes (Sobal and Stunkard, 1989; McLaren, 2007; Newton, Braithwaite and Akinyemiju, 2017).

One possible explanation posited for the differing strengths of these relationships between men and women is the differing extent to which men and women are subject to obesity stigma. Whilst both sexes are vulnerable to the effects of obesity stigma, there is extensive evidence that women face a greater burden of obesity stigma than men (Fikkan and Rothblum, 2012). Within this, women face higher rates of obesity stigma than men in general and women with obesity suffer more severe penalties than men with obesity in employment settings, education settings, romantic relationships (Fikkan and Rothblum, 2012).

It has been suggested that the different experiences of obesity in men and women may be driven by differing perceptions of the ideal body size in the two sexes. The prevailing image of the 'ideal' female body is one of thinness (MacNeill and Best, 2015; Cohn and Adler, 1992), whereas in males there is a belief that larger, more physically powerful body sizes are more desirable (Cohn and Adler, 1992). As a result, men with obese bodies are less deviant from the ideal body than females with obesity, perhaps resulting in increased vulnerability to stigmatisation for women.

Summary

The research on the association between obesity and depression presented in Chapter 2, in addition to that presented above, suggests that stigma may

be a potential mechanism through which obesity is related to both depression and socioeconomic position. However, to date there are few studies that have explored quantitatively how stigma might generate these relationships. Instead, the majority of the research looking into obesity stigma has focused on finding associations between obesity stigma and various outcomes in isolated populations, with these associations being used to make causal hypotheses about the underlying relationships. Whilst studies of association might provide some insight into causal mechanisms, the causal relationships are included in these studies more as untestable assumptions, or causal conclusions are usually given as tentative explanations of the data rather than being explicitly modelled and tested.

The hypotheses suggested by this literature are displayed diagrammatically in Figure 6.1. Ovals represent key concepts highlighted in the literature and the arrows between them are interpreted broadly as 'has an effect on'. For example, one's sex has an effect on one's ideal body size which in turn effects the amount of obesity stigma one is subjected to. Within the diagram, dashed arrows represent hypothesised relationships, and solid lines represent relationships with more concrete evidence supporting them.

From the above, it is clear that there is a need to find a way to more explicitly test and examine the causal implications of obesity stigma outlined in this section. Namely, are they sufficient for generating the relationships between obesity and both depression and socioeconomic position seen in the wider population. Exploring this general question forms the foundation on which the development of the ABM presented in this chapter. The remaining sections of this conceptual model will reduce this broad question down into a more specific formulation that can be investigated using an agent-based model, before section 6.4.2 then outlines the computational model itself.

6.4.1.2 Determining the Modelling Objectives

Overall, this simulation model will address some of the gaps in knowledge highlighted in the previous section. Namely, it will explore whether obesity stigma is capable of generating the relationships between BMI and both depression and socioeconomic position. On top of this it will also seek to explore whether deviation from a socially imposed ideal body size is responsible for driving stigma of obesity and its consequences. These modelling objectives can be summarised by saying that we are interested in answering the following research questions:

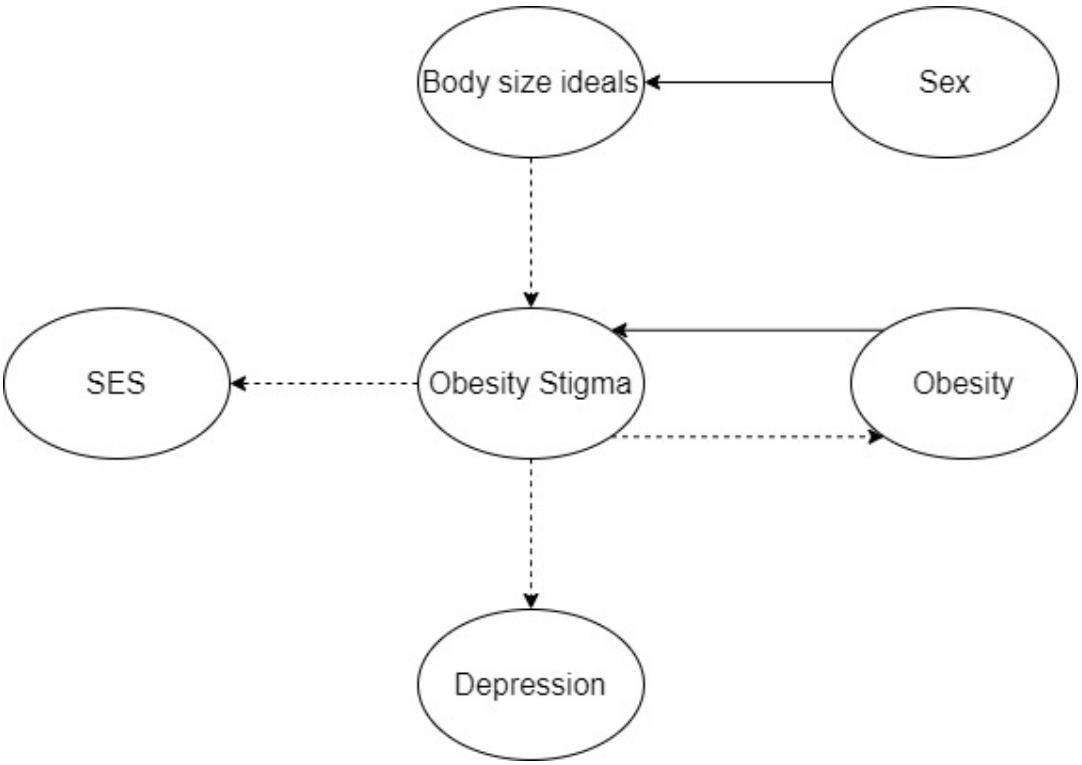

FIGURE 6.1: Diagrammatic representation of hypothesised relationships from the literature. Dashed arrows represent hypothesised relationships, and solid lines represent relationships with more concrete evidence supporting them.

The Model Research Questions

1. Is workplace obesity stigma in the form of exclusion from opportunities to progress capable of generating the macroscopic relationship between BMI and socioeconomic position that is observed in the real world?

2. Similarly, is obesity stigma capable of generating the association between BMI and depression observed in the population?

3. To what extent are different body size ideals responsible for the sex differences in the relationships between BMI and both socioeconomic position and depression.

6.4.1.3 Determining the Model Inputs

As described in section 6.3, the main modelling inputs include two things: the model parameters to be varied across the simulations, and the real world data used to compare the model outputs to the real world system.

In terms of model parameters, in order to investigate the first of our research questions the model will need to include parameters that define the scale of

the obesity penalty to future improvement to socioeconomic position, so that the model is able to investigate how the generated relationship between BMI and SEP differs under different sizes of obesity penalty.

Secondly, in order to examine the second research question, the model will need to include parameters that describe the impact of being stigmatised on future BMI and levels of depression. Lastly, it will also need to include parameters of the ideal body size for each sex as well as what is considered deviant from such an ideal body. These parameters will allow the model to explore the impact of differing ideal body sizes between the sexes on the model's outputs.

Data used in the model is combined from a mixture of sources, including the Whitehall II data set, the Health Survey for England and the ONS Effects of Taxes and Benefits Survey. The detail of how these data sources were used in the model will be described in more detail in sections 6.4.2 and 6.4.3.

6.4.1.4 Determining the Modelling Outputs

In order to investigate the model research questions, the model will need to produce three main types of output. Firstly, the model will need to generate outputs that determine whether the model has satisfied the condition of generative sufficiency. In particular, is the model able to reproduce dynamics that are realistic enough for us to trust it as a representation of the real world system. Since we are interested in the relationships between BMI and both depression and SEP the model will need to produce population summaries of these characteristics so that they can be compared with equivalent summaries in the real world system. In particular, the model will need to produce time series of the average BMI, level of depression and SEP.

Secondly, the model will need to produce output for investigating substantively the research questions. In order to investigate the first research question the model will need to produce a measure of association between BMI and the levels of SEP in the population. This implies that it will also need to generate both a distribution of BMI and a distribution of SEP so that the relationships between these two characteristics can be investigated. Similarly, to investigate the relationship between BMI and depression, the model will need to generate a distribution of depression so that the relationship between BMI and depression can be analysed. Lastly, so that the sex differences in the

third research question can be considered, the model will also need to generate these distributions separately for each sex, so that intersex comparisons can be made.

The last type of output that is important to the answering of the research questions is the posterior distributions of the input parameters associated with the experimental factors listed above. This output is more closely associated with the model calibration than the model itself, and so more detail on the generation of these posteriors will be given in section 6.4.3. However, at this point it's important to state that these posteriors give estimates of credible ranges within which the input parameters sit, and as such give us important information about which values of the experimental factors are most likely to produce output that closely matches reality.

6.4.1.5 Determining the Model Content

Within the relationships between obesity, depression and socioeconomic position there is almost endless detail that could be included within a model. As a result, in this section I will not systematically go through every aspect of the system and justify its inclusion or exclusion. Instead, here I will outline the key features of the model, and the level of detail that they are described in, guided by an overall aim to select components and a level of detail that is able to answer the model research questions whilst keeping the model as simple as possible.

Representing Stigma in the Model

One of the most important features of the model that needs to be designed within the model is the representation of stigma. Stigma is a complex and multi-faceted phenomenon, and as a result, much work had been conducted that attempts to create a firm definition of it. Most of the early credit for conceptualising stigma is attributed to Goffman, who described stigma as a process that evolves via social interaction when those considered normal and abnormal meet, in which the abnormal is one in possession of a "deeply discrediting attribute" and is considered "less than human" (Goffman, 1963). The one who is in possession of the discredited attribute, as a result, is denied the same basic privileges, dignity and respect that 'normals' are afforded, which has various consequences for that individual including psychological struggles, difficulty with social integration and more.

A plethora of work on stigma followed the work of Goffman, and it was during this period of growth in stigma research that the criticism of the vague and varying definitions used for stigma was raised. To address this criticism, Link and Phelan (2001) attempted to conceptualise stigma from a sociological perspective. They noted that the precise definition used in practice will likely differ due to the wide ranging contexts that stigma arises in, that span many disciplines of research. Despite this, they argued that when researching stigma one should be clear about what aspects of stigma one is researching, and described an overarching definition of stigma based on the co-occurrence of interrelated components they deemed relevant. In particular, they described stigma as the "co-occurrence of labelling, stereotyping, separation, status loss and discrimination all in a context in which power can be exercised".

Having a clear and well-defined definition of stigma is especially important when attempting to define it within a quantitative simulations model. A vague stigma concept cannot be translated into model code, as the process of coding will automatically restrict and refine the definition of stigma in ways that may not match up with the modeller's original concept. Hence, here I will outline which aspects of obesity stigma this model will aim to represent in order to make the translation into model code as seamless as possible.

1. Obesity Stigma in the Workplace:

Since research question 1 is focused on the effect of obesity stigma on socioeconomic opportunities, the model will need to represent clearly what these socioeconomic opportunities are, as well as what kind of stigma impact we are interested in. People with obesity have been suggested to receive stigma in both employment and education settings (Fikkan and Rothblum, 2012), however, since my research has focused on adult populations, this model will be limited to investigating employment situations. On the whole, adults will have completed the majority of their formal education and so stigma in these settings is less likely to be as impactful as stigma in employment settings for these populations. Additionally, some of the data that will be used to calibrate the model comes from the Whitehall II study (see Chapter 3), which is a population of working age individuals, who at the time of recruitment were all working for the civil service. This again makes it more likely that the SEP related stigma impacts will more impactful for this population in workplace settings than in education.

Within employment settings, evidence suggests that individuals with obesity suffer stigma within multiple situations. For example, in hiring situations they are less likely to be considered a suitable candidate for a job, and once in a job they are often rated more negatively in performance evaluations, disciplined more harshly and less likely to be recommended for supervisory responsibilities (Fikkan and Rothblum, 2012; Godfree, 2020). All of the above may contribute overall to the reduced likelihood of being hired and reduced wages that people with obesity receive (Fikkan and Rothblum, 2012).

The majority of literature investigating obesity stigma in the workplace has focused on recruitment practices, with less research done on other areas such as training and day to day treatment (Godfree, 2020). From this, there is more data available that supports the idea that differential treatment in hiring practices for people with and without obesity may generate socioeconomic inequality between these groups of people. As a result, this model will focus on obesity stigma in hiring situations. Specifically, the model will be interested in hiring situations where individuals are applying for a promotion, such that they will implicitly take on more responsibility for increased remuneration. The model is not examining job changes where individuals move 'horizontally' to a different job with the same pay.

Once again, the representation of the promotion application process will be kept as simple as possible to avoid creating an overly complicated model. The main key components that will need to be considered is the scale of the 'penalty' to one's likelihood of being promoted for being obese and a representation of one's likelihood of being promoted in general. No distinction is made between applying for promotions within one's current organisation and promotions sought in an external organisation. In this way, the process of being promoted is boiled down to a single interaction between an applicant and a hirer. Starting from this base, additional components may be added if this basic level of detail does not achieve 'generative sufficiency'.

Obesity Stigma in Social Situations:

In order to investigate the effects of obesity stigma on depression and BMI it will also be important to represent more common forms of everyday stigma that might contribute to the consequences of obesity stigma outlined in section 6.4.1.1.

The literature outlined in Section 6.4.1.1 suggested that people with obesity were subject to stigma from a number of sources in multiple settings including personal relationships, workplaces, healthcare settings and the media. Representing all of these contexts individually in a sufficient level of detail to give them credibility is likely to create an intractable model, and so as a result a more general representation of this stigma will be necessary in the model.

One way of doing this is to look at common features of situations in which individuals with obesity perceive they have been stigmatised. One study of the phenomenology of weight stigma found that, despite stigmatising experiences occurring in a variety of settings, and coming from a variety of sources, the vast majority of stigmatising experiences involved the recipient being given a stigmatising 'message' via verbal comments, body language, written communication or a mixture of modalities (Vartanian, Pinkus and Smyth, 2014). Hence, in order to keep this model simple I will include a representation of the delivery of this 'message' from one person to another.

Previous Agent-based Models looking at the relationship between obesity and depression have focused on rejection and social exclusion as the modality of stigma (Mooney and El-Sayed, 2016). This model will differ in this respect as social exclusion will not be represented in the model. Instead the focus is on the delivery of stigmatising 'messages' from one individual to another.

Since the focus of this study is around the generation of macroscopic patterns of BMI and depression in the population, the hypothesised casual mechanisms presented in the models of Brewis (2014) and Tomiyama (2014) will not be explicitly represented, as this would add further complexity to the model without necessarily adding to the answering of our specific research questions. Instead, being stigmatised in the model will have a direct impact upon the individuals' BMIs and levels of depression.

Effect of Ideal Body Size:

The representation of the ideal body size is also an important feature of the model. The construction of an societal 'idealised' body is a complicated social process and so again will need to be simplified for use in this model, since these processes are not central to the answering my research questions. In light of this, the ideal body in this model will be represented exogenously,

with the mechanisms by which this ideal was created being considered out of scope.

Representing Body Size, Depression and SES

Up until now in this conceptual model description, the concepts of body size, depression and SES have been discussed generally, however, in order to describe these characteristics in the computational model, precise descriptions of how body size, depression and SES are to be represented are necessary. Given these are key characteristics for individuals in the model, data on these characteristics need to be available throughout the simulation's time-horizon in order to aid the model's calibration. As such, the representations were chosen for the model such that they provided a description of these concepts that was consistent with available data.

In the case of body size, an individual's body size is represented using BMI. As such, in this model an individual with a higher BMI is considered to have a larger body size. Depression is represented in the model by the extent of depressive symptoms. As such, higher depression for an individual in the model indicates that they are suffering a higher severity of depressive symptoms. Hence, diagnosis of depression, or depression caseness in the model are not represented explictly in the model.

Lastly, socioeconomic position is represented in the model by individual's income. Since the model is focused on stigma in employment settings (specifically, the process of applying for a promotion) the representation of SEP needs to be appropriate to this context. For example, occupational status could be used to represent an individual's level of seniority in the model, or an individual's level of income could be used to represent the remuneration one receives from wages. Of these representations, income was chosen to represent SEP in the model due to the greater availability of data on income that could be used to calibrate the model.

Non-stigma Effects

Stigma is unlikely to be the only contributor to individuals' weight gain and depression characteristics. However, including a wide range of additional behaviours and mechanisms into the model in great detail will again risk the model becoming intractable, particularly given that both obesity and depression have complex aetiologies (Vandenbroeck, Goossens and Clemens, 2007;

Goldberg, 2006; Bargiotas, 2017). In light of this, non-stigma based mechanisms will not be explictly represented in the model. Instead, random variation will be included into the model which aims to recreate the additional variability in characteristics lost during the exclusion of these mechanisms.

6.4.1.6 Assumptions and Simplifications:

In determining the model scope and detail, a number of assumptions and simplifications have been made. Below are the model assumptions and simplifications, presented with justification (where possible) from relevant literature.

Model Assumptions:

General Assumptions

1. Individuals can accurately assess the BMI of others so that they are able to see how deviant others are from the ideal body size. This may not be the case in reality, as studies have found that people are often unable to recognise their own obesity (Truesdale and Stevens, 2009). However, this study was focused on the assessment of one's own body weight, which may be subject to different biases than the assessment of another's body weight. In addition, the study was conducted in a small unrepresentative sample and so its conclusions may not be suitable for extrapolation to wider populations. Hence, in the absence of additional evidence assuming individuals can accurately assess others BMI seems a reasonable assumption to make.

2. An idealised body is able to be captured by a single, ideal value for one's body mass index. Whilst individuals may not have an ideal BMI in mind when they are thinking about their ideal body, their visual descriptions will have physical characteristics that would allow the rough calculation of an associated BMI. Hence it stands to reason that the body ideals in the model will have an associated value of BMI which one can consider, in some sense, ideal.

Assumptions for Hiring Situations 'Submodel'

1. As stated previously, the hiring mechanism aims to be as simple as possible, excluding detail that is not relevant to the model research questions. Within this simple representation it is assumed that individuals are qualified

to do the job they are applying for so that differential likelihood of being hired can only occur due to obesity stigma.

2. It is also assumed that, unlike stigma received in social situations, workplace stigma does not have an impact on individuals' BMIs or levels of depression. Whilst this may not be strictly the case, the study of the phenomenology of weight stigma conducted by Vartanian, Pinkus and Smyth (2014) suggested that stigma in employment situations made up only a small proportion of stigmatising experiences. As a result, it seems reasonable to assume that the effect of these experienced is small in scale when compared to that of social situations, and as a result is worth omitting for the sake of model parsimony.

3. Since the model is not representing individuals' occupations in a detailed way it is implicitly assumed that people with obesity are stigmatised the same amount irrespective of where they are in the workplace i.e. the stigma one receives is independent of the job being applied for. This may not be the case as one study has suggested that workplace stigma is more present as individuals attempt to move into jobs of higher prestige and compensation (Fikkan and Rothblum, 2012). However, since the evidence on these potential differences in limited, this is a reasonable assumption to make at this time.

Assumptions for Social Stigma 'submodel'

1. Since only a general stigmatising interaction is being considered in the model, we are assuming that the effect of social stigma is the same independent of the type of social connection. This means that being stigmatised by a stranger has the same impact as being stigmatised by a close family member or friend. Whilst in reality, this may not be the case, there is little evidence available to support what differential effects would exist and as a result, multiple assumptions would need to be made in order to represent such situations in the model, leading to a more complicated model overall. As a result, making this assumption seems reasonable.

2. Whilst current literature suggests that body size ideals differ between the sexes, there is less clear cut evidence on whether there is a reduced tolerance for deviations from the ideal body size. As a result, in this initial model, I will assume that the differential levels of stigma received between men and women are due only to differences in the ideal body size; the tolerance for deviations from this ideal body size is the same for both sexes.

Model Simplifications

1. All individuals subscribe to the same exogenously defined ideal body size; individuals are not heterogeneous in their belief about what constitutes an ideal body. Whilst there is likely some variation in ideal body sizes, this simplification is made to reduce complexity in the model so that every individual has an ideal body size that is externally defined and represents the idealised body of the agents' culture. The reasonableness of this simplification is hard to examine, as there are multiple theories as to how body size ideals develop (Heinberg, 2001). However, this assumption is in line with one of the dominant theoretical frameworks through which the formation of body image is examined: the sociocultural perspective (Tiggemann, 2012). This perspective sees culturally held ideals of body image as central to individuals' formation of their own ideas of what an idealised body is. As a result assuming that all individuals in the model will subscribe to the imposed ideal BMI is analogous to the ideas put forward by the theory.

2. Body size ideals are static in the model. Whilst there has likely been some variation in the sociocultural ideal body size over the last few decades, overall ideals have been largely static. Thinness ideals and their effects on womens' bodies in particular have been investigated since the 1970s showing the pervasiveness of these ideals in society. Hence making the simplification of considering these ideals as static is a reasonable starting point in the model.

3. Successful promotions increase an individual's income by a fixed percentage. In reality, the size of a pay rise from a new job will depend on many factors, however, since these factors are not being investigated in the model, using a fixed percentage is a reasonable simplification of the amount individuals might receive from a pay rise. This fixed percentage could be interpreted as the average pay rise that a worker receives in a promotion.

6.4.1.7 Summary

In the preceding sections I have outlined the design of a conceptual model that will form the basis of an Agent-based simulation model. Figure 6.2 provides a diagrammatic representation of the conceptual model, which has expanded upon the representation of the literature provided by Figure 6.1. Once again, ovals represent key concepts included in the model, and arrows

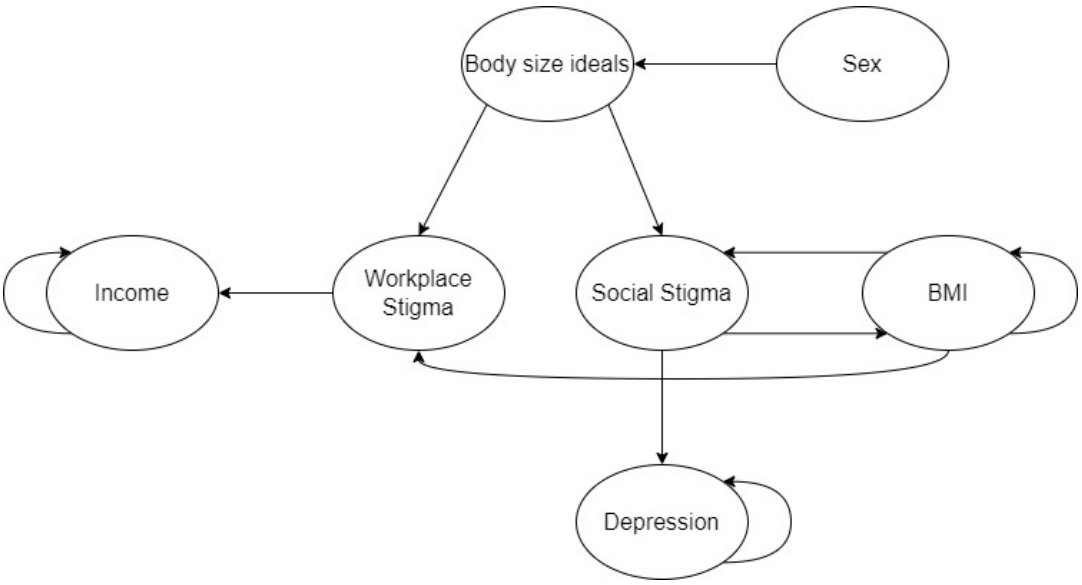

FIGURE 6.2: Schematic of the conceptual model for the Agent-based simulation

represent direct effects that will be included between these concepts. Overall, this model has been designed to remain as simple as possible whilst including key relationships that are important to the analysis of our research questions. In the next section, this conceptual model will be developed into a computational model.

6.4.2 Detailed Model Design

In this section I will outline the detailed design of the computational model. As mentioned earlier the description of the computation model will follow the structure of the sections outlined in the ODD protocol (Grimm et al., 2006). Since the ODD protocol aims to describe an ABM in its entirety, there may be some overlap between the sections here and those in the conceptual model description. To avoid repetition, when this is the case I will simply refer back to the relevant section of the conceptual model description, rather than restating the same information.

6.4.2.1 Purpose

The purpose of this model has been outlined in detail in the conceptual modelling section, so here I will briefly summarise what the purpose of the model is. Broadly, the model aims to explore the role of stigma in generating the relationship between BMI and depression, as well as whether stigma is capable

of generating observed patterns of BMI with respect to socioeconomic position.

6.4.2.2 State Variables and Scales

Agents

There is one main type of agent in this model, 'people', whose characteristics are described in this section.

'People' Variables

People in the model possess the following characteristics: BMI, depression level, biological sex, body-deviance, stigmatised?, stigmatising?, stigma-count; income and time-since-promoted. Table 6.1 contains the descriptions of these variables and what they represent. In this version of the model, state variables were chosen in order to represent the key features of the conceptual model using as simple a representation as possible.

These characteristics can be split into three main groups which when taken together describe an agent's state at a given point in the simulation. The first group includes the agent's sex, BMI, depression and income and describes the agent's main attributes of interest as well as their demographic information. In line with the representation in the conceptual model, an individual's depression level represents the extent of depressive symptoms they are experiencing. Since the model will eventually be calibrated to GHQ data (see section 6.4.3.3) an individual's level of depression can in essence be thought of the level of depressive symptoms that the individual would report via a GHQ-30 questionnaire. Similarly, in line with the representation of income in the conceptual model, an agent's income here describes their current weekly wage income.

The second grouping then describes the agent's involvement with obesity stigma within the current time point including their vulnerability to stigma, their stigma behaviours and how much stigma they have received. The final group then keeps track of information relevant to the agent's ability to get promoted, which in this model is just a count of the number of time-steps since the agent was last promoted.

Age and ethnicity characteristics were not added to the model as they were not central to the answering of the model's research questions. Representing age in particular would also have added in additional complexity in how

Variable	Description
Core Variables	
Sex	The agent's biological sex
BMI	The agent's body mass index
Depression	Agents's depressed affect, this is measured as a continuous variable between 0 and 1, with 1 representing maximum depressed affect, and 0 representing minimum depressed affect.
Income	The agent's weekly wage income
Stigma Variables	
body-deviance	The difference between the agent's BMI and the ideal BMI for their sex
stigmatised? (Yes/no)	A boolean variable that indicates whether an agent has been stigmatised or not in the current time-point
stigmatising? (Yes/no)	a boolean variable that indicates whether an agent is stigmatising or not
stigma-count	An integer that describes the number of times that an agent has been stigmatised in the current time-point
Promotion related variables	
time-since-promoted	An integer representing the number of time-steps since the agent was last promoted

TABLE 6.1: Variables representing the characteristics of agents in the model

to represent the ageing process in the model, as well as whether births and deaths should be represented. Including such a process would come at a cost of adding multiple extra parameters into the model, without helping the model better fulfil its purpose. Similarly, as the model is not intended to explore ethnic differences in the effects of obesity stigma, ethnicity was also not included.

Environmental/Global Conditions

Variable	Description
Ideal body size Variables:	
male-ideal-bmi	The BMI associated with the 'ideal' male body size
female-ideal-bmi	The BMI associated with the 'ideal' female body size
body-tolerance	Threshold of deviance from ideal BMI before an individual is stigmatised for their weight
Stigma effect Variables:	
stigma-to-depression	Effect of being stigmatised on an individual's depression
stigma-to-obesity	Effect of being stigmatised on an individual's BMI
meeting-benefit	Effect of having a meeting without stigma on an individual's depression
Employment Conditions:	
promotion-prob	probability of achieving a promotion when applying
obesity-promotion-penalty	obesity promotion penalty scale constant

TABLE 6.2: Variables representing the input parameters to be calibrated in the model

Table 6.2 lists the global variables in the model. These are the input variables that represent the experimental factors outlined in the conceptual model and as such will be calibrated based on data from the real-world system. Again these fit into three groups. The first group of variables describes the ideal body size within the artificial society, along with what is considered deviant from that body size. The second describes the effects individuals receive when they interact with other agents in the model, and the third describes inputs related to individuals' ability to achieve a promotion.

The Social Network

The people in the model were embedded in a social network in which each agent has at least two social ties with other agents in the model. This social network is static and does not update. Since there is not enough available evidence to say whether different types of relationship are more or less important in terms of their effects on stigma, links represent a 'general' social tie which could be anything from a close family member to a stranger.

The social network is described using a set of links between the agents. Links have one main feature: a 'meeting' variable which describes whether the two people joined by the link are currently meeting up or not.

6.4.2.3 Process Overview and Scheduling

In this model, each time step in the model represents one month in real time. In each time step, each agent first updates their `body-deviance` to incorporate changes to their BMI that occurred in the previous time step. Once this has been updated, the rest of the model is run using three sub-models, which are described in more detail within the 'sub-models' section below.

The first of these sub-models evaluates the stigma effects for the model. Within this sub-model each individual decides which of their social connections they will be meeting in the current time point. Individuals meet with two other individuals in the model in a given time-point.

Once the meetings for the current time-step have been allocated, the stigmatise procedure is run to see what the outcome of those meeting is with respect to whether individuals are stigmatised or not. Each time an individual is stigmatised by another person they are meeting they receive a small increase to their BMI and depression level. However, when a person receives no stigma in a given time-point, they experience a small decrease in their depression level. All the effects of stigma are considered for every person in the model prior to the incorporation of beneficial meetings' effects.

The second sub-model then simulates which people in the model are promoted in the current time-step. The final sub-model then incorporates random variation into people's BMI and depression level to include effects that are not related to stigma.

6.4.2.4 Design Concepts

The following items from Grimm's checklist apply to this model:

Emergence: Emergence refers to the system level phenomena that emerge from individual traits, rather than being imposed. In this model the population levels of obesity and depression emerge from the effects of interactions between the agents in the model. Similarly, the prevalence of comorbidities in obesity and depression also emerge from these interactions.

Sensing: Sensing refers to which internal and environmental state variables are individuals assumed to 'sense' or 'know' so that they can use them within their decision making. Within the model people are assumed to know the values of all global variables so that they may use them to update internal characteristics and evaluate others' BMIs. Individuals are also assumed to know all of their internal state variables so that they can apply the effects of interactions with other agents to update their internal state. For example, people in the model are assumed to know their own sex, as well as the ideal body sizes associated with their sex so that they can evaluate how 'far' they are from the ideal body size associated with their sex. People can also sense the `body-deviance` of other people that they meet in order to decide whether they stigmatise them or not.

Individuals also know the value of a promotion and how long ago they were last promoted so that they can apply for a promotion and incorporate any successful promotions into their income.

Interaction: This item explains what kinds of interactions are assumed in the model. The main type of interaction that is explicitly modelled here is social interaction between people. Interactions between people and an implicit 'job market' are also made when individuals apply for promotions. When this implicit interaction results in a successful promotion, people are unable to interact with the job market for the next six months.

Stochasticity: This part of the design checklist describes the reasons for any stochasticity included in the model. Stochasticity is present in the model from two main sources. The first source of stochasticity is in the initial condition for each individual's `time-since-promoted`, which is given by an integer drawn uniformly from the range zero to five. This means each individual has been promoted sometime in the last 6 months and was done so that individuals in the model would not all be applying for promotions concurrently in the early stages of the simulation. The second source of stochasticity is

the random variation incorporated in each time point into peoples' BMI and depression characteristics. This is included to ensure that individuals' characteristics can vary in the model without input from stigma.

Observation: The observation item explains how data are collected from the simulation in order to analyse it. In this model the output of the simulations were extracted into the R software package using the NLRX package. Summary statistics of the population level BMI, depression and income were extracted and analysed. Full details of this process can be found in the model calibration section.

6.4.2.5 Initialisation

The model is initialised with a total of 1000 agents. People's initial BMI, depression levels and sex were sampled jointly using a random sample taken from wave 1 of the Whitehall II data set. Initial incomes were then sampled from the 1987 effects of taxes and benefits survey. Since the effects of taxes and benefits survey had no data on respondents sex, BMI or depression these sampled incomes were assigned to agents independently of these other characteristics. The sample was generated in R prior to simulation in Netlogo and so the same sample was used in each separate simulation.

Agents were assumed to neither be stigmatising, nor stigmatised initially, so that these characteristics could be determined by the model's procedures. As mentioned above people were also assigned a random `time-since-promoted` between zero and five to represent having received a promotion some time in the last 6 months. When simulation experiments use the same random seed these sampled characteristics will remain the same, however, for experiments with a varying random seed, these characteristics will vary across the simulations.

6.4.2.6 Input

Since the environment was only minimally represented in this model, there were very few external conditions and dynamics imposed on the model outside of the model's global variables. Within the model, the population size was set to 1000 for each simulation, and the values of global variables were sampled using a latin hypercube sample (McKay, Beckman and Conover, 1979) as described in more detail in section 6.4.3. There were no other inputs into the model.

6.4.2.7 Submodels

As was mentioned in the process overview and scheduling within each time-step the model runs through four main sections. In the first section all the people in the model compare their BMI with the ideal BMI associated with their sex in order to calculate their new value of `body-deviance`.

If we denote by $y_{i,t}$ the `body-deviance` of person i at time step t, $x_{i,t}$ the BMI of person i at time t and k the ideal BMI associated with the biological sex of person i, then $y_{i,t}$ is calculated using the following equation:

$$y_{i,t} = \max\{0, x_{i,t-1} - k\}.$$

In this way, `body-deviance` gives an indication of how much bigger than the ideal body size the agent is. Note that the BMI value is taken from the previous time-point rather than the current time point. This is because this value of BMI has incorporated all of the previous time-step's stigma effects, but no effects from the current time-point.

The Social Stigma Submodel:

Once individuals have updated their `body-deviance` the first of the three main sub-models is run: the social stigma submodel. This sub-model consists of three processes, the first of which decides which agents are to meet in a given time-point. Within this procedure each agent chooses two of their neighbours to meet up with in each time-point. To do this, people randomly select two of the links that they have with other people in the model and set the meeting variable of those links to 'TRUE' signifying that a meeting will take place. In this way, people meet up with at least two other people in each time-step, as people can both select others to meet and be selected themselves.

The second procedure then describes which agents are stigmatised within the model. Within this procedure people 'meet up' with the other people they are connected to by the 'meeting' links defined in the meet-up procedure. Within meetings, there is a possibility of stigmatising comments being made.

Suppose that person i in time step t has `body-deviance` $x_{i,t}$, and that the `body-tolerance` in the society is given by k. If we denote by $p_{i,t}$ the probability that person i is stigmatised in a single meeting during time-step t,

then $p_{i,t}$ is given by the following equation:

$$p_{i,t} = \begin{cases} \max\{\frac{x_{i,t}}{25}, 1\}, & \text{if } x_{i,t} \geq k \\ 0, & \text{otherwise.} \end{cases} \tag{6.1}$$

The above formulation means that people are only a possible victim of stigmatising comments in these interactions if their BMI deviates from the 'ideal' BMI of their sex by more than the value of the global variable `body-tolerance`. Once a person's `body-deviance` reaches the `body-tolerance` threshold the probability they will be stigmatised jumps up to $k/25$. This probability of being stigmatised then increases linearly with `body-deviance`, and hence to a maximum of 1 when a person has a BMI that is 25 units above the sum of the ideal BMI associated with their sex and the societal body tolerance.

Each time a person is stigmatised they receive a small increase to their level of depression equal to the `stigma-to-depression` global variable and small increase to their BMI equal to the `stigma-to-obesity` global variable.

The third process then applies the benefit of positive social interaction to those individuals who have not been stigmatised in the current time-point. If a person is not stigmatised they receive a small decrease to their depression given by the value of `meeting-to-depression`. This represents the assumption that social contact is in general beneficial to individuals' mental health, unless that contact is poor quality as in the context of stigma.

The Promotion Submodel:

Following the social-stigma submodel is the promotion submodel. This submodel consists of one main procedure in which people apply for a promotion if they have not been promoted within the last 6 time-steps. Once individuals are eligible to apply for a promotion, they apply in each time-step until they are successful. Once a person is successfully promoted, their income is multiplied by the value of the global variable `promotion-value`.

The probability that an individual gets promoted is influenced by three main factors. First, it depends on the value of the `promotion-prob` global variable, which one can think of as the unbiased chance of the individual getting promoted if their weight status was ignored. Second, it depends on the scale of bias against people with obesity in hiring situations, defined by the global variable `obesity-promotion-penalty`. Third the individual's `body-deviance`

value impacts their probability of promotion such that the individuals with a higher `body-deviance` have a reduced chance of promotion that is proportional to the size of the`obesity-promotion-penalty`.

Denote $q_{i,t}$ be the probability that person i with a `time-since-promoted` value of n and `body-deviance` $x_{i,t}$ is successfully promoted in time-step t. Let π denote the value of the global variable `promotion-prob` and let α denote the value of `obesity-promotion-penalty`. If k again denotes the value of `body-tolerance` then $q_{i,t}$ is given by

$$q_{i,t} = \begin{cases} \pi(1 - r\alpha), & \text{if } n \geq 6 \\ 0, & \text{otherwise,} \end{cases} \tag{6.2}$$

where

$$r = \max\{0, x_{i,t} - k\}. \tag{6.3}$$

As an example, suppose that $\pi = 0.2$, $\alpha = 0.05$ and $k = 5$ and that person i has a `body-deviance` in time point t of $x_{i,t} = 10$. Then person i would receive a 50% drop in their probability of getting promoted such that their promotion probability $q_{i,t} = 0.1$ rather than 0.2. A case when this might arise would be for an individual with BMI 40 when the ideal BMI of their sex was equal to 25.

The Update Submodel:

The final submodel doesn't represent anything substantive from the underlying system, but instead is simply used to update peoples' characteristics in the model ready for the next time-step. Two main processes are executed, the first of which updates individuals' BMI and the second updates their depression level.

Individuals' new values of BMI are given by the sum of three components: their previous BMI, the impacts of stigma they received in the current time point and random variation. Denote by $z_{i,t}$ the BMI of person i at time t who was stigmatised $n_{i,t}$ times and suppose that the effect of a single stigmatising encounter on an individual's BMI is denoted by β. Then $z_{i,t}$ is given by

$$z_{i,t} = z_{i,t-1} + s_{i,t} + \epsilon_z, \tag{6.4}$$

where ϵ_z is drawn from a $N(0, 0.1)$ random variable, and

$$s_{i,t} = n_{i,t}\beta. \tag{6.5}$$

The above setup implies that the effects of stigma combine additively within a single time-point such that individuals who receive more stigma receive more deleterious effects.

Depression is updated analogously to BMI, and is the sum of a persons previous depression level, the effects of any stigma received, and random variation. The equations used to update depression are hence identical to those listed above for the update to BMI, so I wont repeat them here. However, the parameters for the ϵ_z term are different. This time random variation is drawn from a $N(0, 0.05)$ distribution. This is due to the scale of the depression level being much smaller than the the scale for BMI and so to avoid random noise overly influencing the dynamics a smaller value was chosen so that the random variates were smaller in magnitude.

6.4.3 Calibrating the Model

Once the detailed design of the model had been implemented in NetLogo, the process of calibrating the model to observed data could begin. The model analysis and calibration was conducted using the NLRX package in R (Salecker et al., 2019), which allows Netlogo models to run via a java virtual machine from within R.

6.4.3.1 Model Parameters to Calibrate

As was highlighted in the previous section, all of the global variables included in Table 6.2 were calibrated to data. Table 6.3 displays the prior ranges for each of these parameters, where a range of [x,y] indicates that the parameter can take any value from the Real numbers between x and y inclusive. Within these ranges a uniform prior distribution was used. The range for `female-ideal-bmi` was chosen to cover a lower range than for `male-ideal-bmi` to reflect the fact that female ideal bodies have almost invariably been depicted as thin, whereas in some settings it has been desirable for men to have a larger BMI (Fikkan and Rothblum, 2012). The upper limit of 0.05 for the stigma effect variables was chosen based on initial experimentation of the model which suggested that unrealistic dynamics arose when these parameters were set at higher values.

Variable	Parameter range
Ideal body size Variables:	
male-ideal-bmi	[20,30]
female-ideal-bmi	[15,25]
body-tolerance	[0,10]
Stigma effect Variables:	
stigma-to-depression	[0, 0.05]
stigma-to-obesity	[0, 0.05]
meeting-benefit	[0, 0.05]
Employment Conditions:	
promotion-prob	[0, 0.5]
promotion-value	[1,1.25]
obesity-promotion-penalty	[0, 1]

TABLE 6.3: Model parameters to be calibrated with prior ranges

6.4.3.2 Calibration Targets

The model was calibrated using mean BMI, mean GHQ-based depression, and mean income. For each simulation the mean BMI, mean GHQ-based depression score and mean income were collected at every twelfth time-point, representing 12 month intervals. The target values for these statistics were then derived as follows. Mean income between 1987 and 2009 was derived from estimates of gross income from the ONS Effects of Taxes and Benefits Survey. The data were available every year in the simulation's time horizon giving a total of 23 values of mean income.

Mean BMI was calibrated to mean BMI values from the Health Survey for England between 1992 and 2009, and mean depression was calibrated to the mean GHQ-based depression score from waves 1, 2, 3, 5, 6, 7, 8 and 9 from the Whitehall II data set. The GHQ-based depression score used was the same as that derived in 3.3, however, in order to match it to the depression variable used in the model, the targets were rescaled in order to put the score in the range [0,1]. Specifically, each target was divided by a scale factor of 15 since this is the maximum score possible on the measure. This gave a total of 18 BMI targets and eight depression targets. Whilst data from a single wave of the Whitehall II data set were collected across multiple years, each waves' data were assigned to a single year for the purpose of making the calibration procedure simpler. Table 6.4 shows how the years, model time-steps and waves of the Whitehall II data were matched up in the model calibration.

Year	Time-step	Whitehall Wave
1987	0	1
1990	36	2
1993	72	3
1998	132	5
2001	168	6
2003	192	7
2006	228	8
2008	252	9
2013	312	11

TABLE 6.4: Years and Model time-steps that align with Whitehall II data waves

6.4.3.3 Calibrating the Model Parameters

To calibrate the model parameters using the target data, a Latin hypercube sample (LHS) with 80000 combinations of the input parameters was first generated. Latin hypercube samples aim to generate a sample of the entire parameter space in an efficient way, i.e without requiring very large numbers of samples (as would be needed in a full factorial design).

The LHS is generated as follows (from McKay, Beckman and Conover (1979)). Suppose we want a sample of size N from $\mathbf{X} = (X_1, ..., X_k)$. Then the parameter range for each of the $X_i, i = 1, \ldots, k$ is split into N strata with marginal probability $1/N$ and one sample is drawn from each of the strata. Hence we have a sample $X_{i,j}$, $j = 1, \ldots, N$ which describes the sampled components for X_i. These components are then matched randomly with components from the other input variables X_j, $j \neq i$ to form N complete samples from $\mathbf{X}$. In this model we have $N = 80,000$ and $k = 9$.

The model was then run for each of these 80000 parameter specifications and the results collected into the R software for analysis. Approximate posterior distributions were calculated using rejection sampling methods provided by the 'abc' package in R (Csilléry, François and Blum, 2012).

The rejection method works by accepting or rejecting a simulation input based on how far its associated model output is from a selection of target summary statistics. Summary statistics are calculated based on both the model output and the real world data. The difference between the observed summary statistics and the simulated summary statistics is then calculated in some metric (in this model the euclidean distance was used), and simulations

are accepted or rejected based on the size of this distance. In some settings this distance is set specifically, however, in the 'abc' package, a tolerance that describes the proportion of points that are to be accepted is defined instead. In my analysis, I used a tolerance proportion of 0.005 so that the best fitting 0.5% of points were used to estimate the posterior distributions. This is equivalent to using the best fitting 400 samples from the 80,000 simulations. The posterior distributions were then estimated from the univariate densities of the input parameters associated with the accepted simulations.

Alongside the rejection sampling, simulations were also compared to the targets manually, by examining whether the simulation output was within a specified distance of the targets. Specifically, simulations were considered a 'close' fit to the data if:

- the mean simulated BMI was within 1 BMI point of the observed BMI at all calibrated time-points;
- the simulated mean depression score was within 0.1 of the observed depression score at all calibrated time-points, and
- the simulated mean income was within £100 of the observed mean gross income at all calibrated time-points.

6.4.3.4 Sensitivity Analysis

Alongside the model calibration, a global sensitivity analysis was conducted to examine which model parameters were most influential on the model output. The relationship between each of the inputs and the following outputs were examined: The mean and standard deviation of the mean BMI, GHQ and income across the simulation. The strength of these relationships was assessed using partial correlation coefficients which estimate the linear relationship between each input parameter and the model output, whilst discounting the effect of the other inputs (Marino et al., 2008).

The partial correlation coefficient ranges from -1 to 1 and is interpreted similarly to a correlation coefficient, with the additional note that confounding introduced from correlation between the model inputs has been controlled for. As such positive values of the partial correlation coefficient indicate that an input parameter has a positive linear relationship with an output, whereas a negative value indicates that there is a negative linear relationship. In the context of these simulations, a positive value of the partial correlation coefficient hence means that, in general, simulations with a higher value of that

input parameter produced simulations with a higher value of the output, with larger coefficients indicating a more consistent linear association.

6.5 Model Results

6.5.1 Sensitivity Analysis

The sensitivity analysis showed that the separate model outputs were not affected by every model input parameter. For example, the average mean BMI in the model, and variability of the mean BMI was not affected by the values of the employment related variables, nor the variables describing the impact on depression from stigma and meetings where no stigma occurred. Similarly, the depression output was not affected by the employment conditions input parameters, and the income output was not affected by the depression effects parameters.

As a whole, this gives confidence that the conceptual model outlined in section 6.4.1 has been accurately reproduced in the detailed description of the model. Effects (and omitted effects) proposed in the conceptual model match up consistently with the effects highlighted in Figure 6.3. Specifically, the summary diagram of the model displayed in 6.2 suggested that income would not have a causal influence on BMI and depression in the model, which these sensitivity results are consistent with. Additionally, within the conceptual model income is not influenced by individuals' levels of depression and all three characteristics are influenced by social stigma and body size ideals.

In terms of the impacts on the model outputs, the partial correlation coefficients suggest that the average BMI within the simulations, and the variation of mean BMI in the simulations are more sensitive to changes in the effect of stigma on body weight than they are to changes in the ideal body size variables. The average depression and variation in average depression in the simulations is then most sensitive to changes in the strength of non-stigmatising meetings on depression, followed by changes to the ideal body size variables. The strength of stigma effects has the least impact on depression output within the variables that have influence. For income outputs, the model is highly sensitive to changes in the value and probability of promotion in the model, with changes to stigma effects and ideal body size variables having only a minimal impact on the income output.

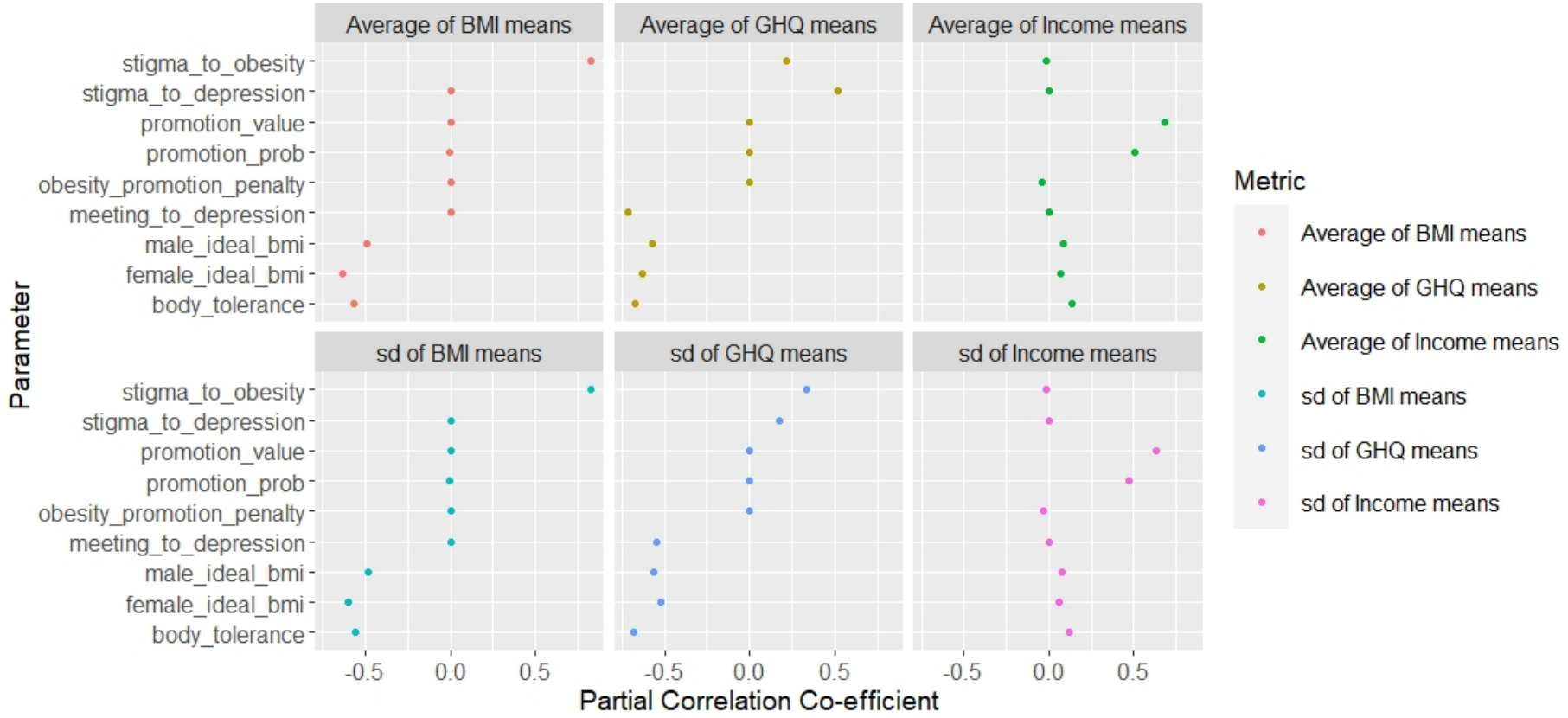

FIGURE 6.3: Partial Correlation Coefficients for effect of model inputs on model targets

6.5.2 Rejection Sampling

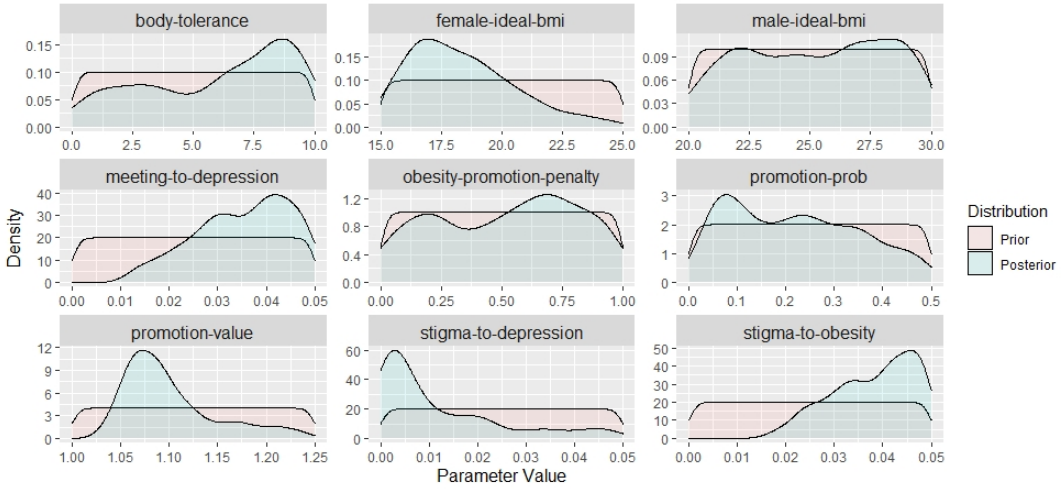

FIGURE 6.4: Posterior distributions of parameters in the model based on rejection sampling

Figure 6.4 shows the estimated posterior distributions from the rejection sampling procedure. The posterior densities are estimated by plotting how frequently each value of that input parameter features in the set of simulations that were kept after the rejection procedure had finished. In this way, areas of higher density indicate that these parameter values more frequently resulted in the model producing output that was close to the model targets.

Within these posteriors, `male-ideal-bmi` and `obesity-promotion-penalty` had only small differences from the prior distributions, such that slightly more models fit with higher values of these variables than was the case for lower values in the parameter range. However, in `male-ideal-bmi` this pattern of deviation form the prior was particularly minute.

All the other posterior distributions showed clearer patterns. The posterior for `promotion-value` suggest that models more frequently fit the targets when promotions that individuals achieve in the model are of low value. The majority of the distributions mass falls between 1.025 and 1.15. The `promotion-prob` posterior has a slight trend towards lower values in the parameter range suggesting the model might also fit better when there is a low chance of being promoted, before obesity penalties are considered. This is in line with the results from the sensitivity analysis that suggested the `promotion-value` had more influence on mean income than did the `promotion-prob`.

In terms of stigma effects the posteriors suggest that the model more frequently fits closely with the targets when there is a moderate to low prevalence of stigma. The posterior for `body-tolerance` peaks between 7 and 10 although some models still fit at lower values. The posteriors for `male-ideal-bmi` and `female-ideal-bmi` had differing patterns. The peak value of `female-ideal-bmi` sits at the lower end of the parameter range, whereas for `male-ideal-bmi` there was little to no difference between the prior and posterior. This means that in the best fitting models women were more frequently vulnerable to stigma, whereas it made little difference how vulnerable men were. Also of note is that the posterior distribution for female ideal BMI peaked below 18.5, meaning that in many of the best fitting models, female ideal BMI was within the underweight category.

The posteriors for the two main stigma effects give contrasting pictures as to their optimal size within the model. Namely, the model calibration suggests that the impact of obesity stigma on depression is small (and close to zero), whilst the the impact of obesity stigma on BMI is larger. The peak of the posterior distribution for `stigma-to-depresion` is between 0 and 0.005, with the bulk of the density's mass being below 0.01. The posterior distribution for the effect of stigma on BMI, however, peaks around 0.045 with nearly all of the distribution's mass falling above 0.02. Taken together with the above implications for stigma prevalence, this means that the model more frequently fits better when there is a low, sometimes moderate, prevalence of stigma that only impacts substantively on future body weight, and has a minimal impact on depression.

Finally, outside of stigma based effects, the posterior for `meeting-to-depression` suggested that in the best fitting models, non-stigmatising interactions have a substantial benefit on an individual's level of depression. The posterior peaks for higher values in the parameter range, above 0.3, with values in the lower end of the parameter range being less probable.

6.5.3 Close Fitting Models

Figures 6.5, 6.6 and 6.7 compare the output of the close fitting simulations with the target BMI, GHQ and income data respectively. The blue region depicts the area of tolerance within which simulations were accepted. 74 simulations out of 80,000 met the criteria to be considered a close fit, and even these simulations did not have dynamics that matched well to the target data qualitatively.

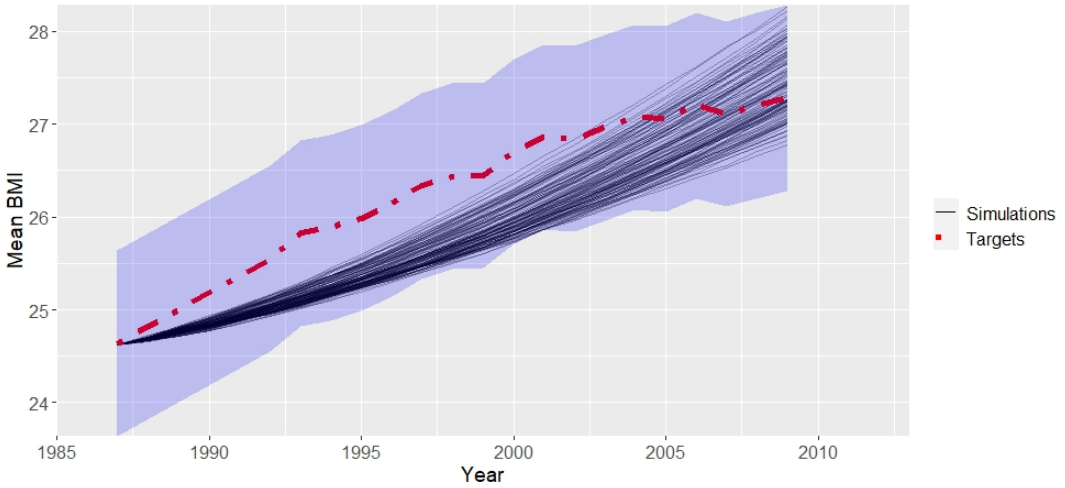

FIGURE 6.5: Trajectories of Mean BMI for 'close' fitting simulations (solid lines) compared to the observed data (dashed red line). The blue ribbon represents the region within which simulations are considered a close fit.

Both the mean BMI and the mean income in the simulations increased at an increasing rate over the duration of the simulation, however, neither target changed in this manner. Observed mean BMI increased at a decreasing rate over the time-horizon and between some time points dropped. This drop in mean BMI in particular was not seen in any of the simulations, suggesting the model is incapable of recreating potentially important features that can occur in the real system.

The mean BMI dynamics seen in the model are likely a result of individuals becoming increasingly vulnerable to stigma as their BMI increases from previous stigmatising encounters. As individuals in the model are stigmatised, their BMIs, and hence `body-deviances` increase. This increases the probability that they will be stigmatised again in the remainder of the simulation, thus exposing them to further increases in BMI. The model appears to have insufficient detail in non-stigma based mechanisms to prevent this positive feedback loop from dominating the overall trajectories.

Observed mean income also increased at an approximately constant rate, although there were some fluctuations around this roughly linear increase. The deviations from the target data also appear less severe than those found in the BMI trajectories. However, the trajectory of the target data still suggests that there may be mechanisms missing from the model's generating process. Specifically, mechanisms that influence income in such a way as to slow down the growth of mean income are potentially missing from the model.

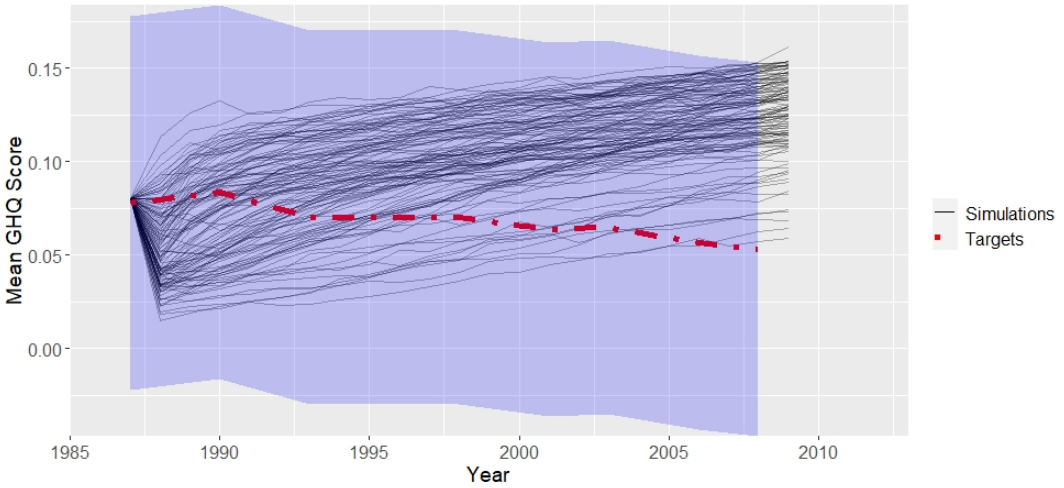

FIGURE 6.6: Trajectories of Mean depression score for 'close' fitting simulations (solid lines) compared to the observed data (dashed red line). The blue ribbon represents the region within which simulations are considered a close fit.

Depression dynamics within the close models appeared to have a particularly poor fit to the data. After a sharp deviation from the model initial conditions, all of the close fitting models feature a steady increase in mean GHQ levels for the remainder of the simulation; a feature that is not seen in the target data. Figure 6.6 shows that this overall behaviour is common to all the simulations that fit closely to the GHQ target, suggesting that it is something in the description of the model's mechanisms that is causing this qualitative behaviour.

The initial deviations in mean GHQ are likely caused by individuals who are initially vulnerable to stigma experiencing very different outcomes to those who are not vulnerable to stigma. Those who are not vulnerable but have a higher initial depression score will experience the benefit of meeting with other people in the model and hence will receive consistent drops to their depression level at early time-points in the model. Conversely, those vulnerable to stigma at the beginning of the simulation will receive a mixture of increases and decreases to their stigma. These effects combine to create early initial variation.

Once individuals who are not vulnerable to stigma have reached a level of zero depression in the model, the mean GHQ then increases due to the increasing effect of stigma on depression in the model. Individuals who are vulnerable to stigma will continue to be stigmatised, thus increasing their level of depression, and since mean BMI increases in the model over time

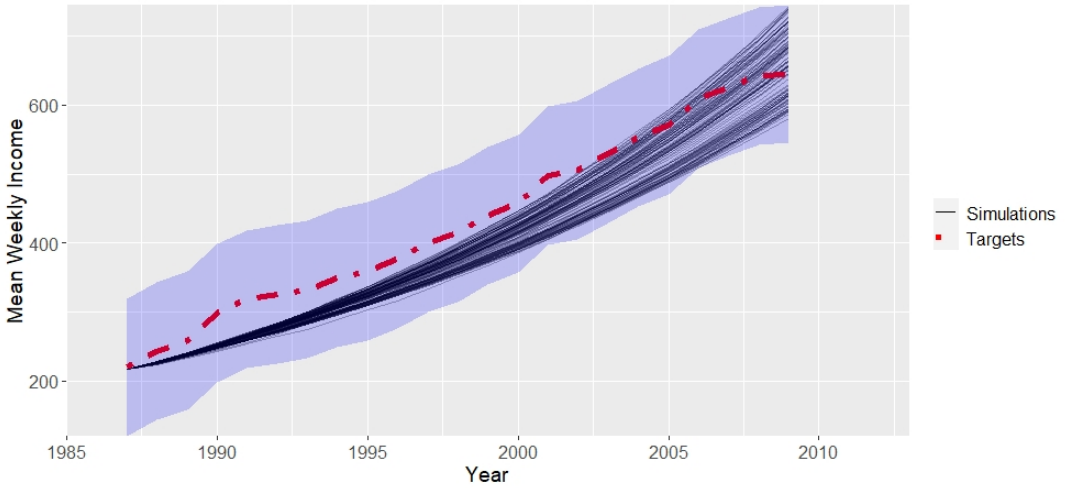

FIGURE 6.7: Trajectories of Mean income for 'close' fitting simulations (solid lines) compared to the observed data (dashed red line). The blue ribbon represents the region within which simulations are considered a close fit.

more people will be vulnerable to stigma and hence will experience increases to their depression from their social interactions. This again suggests that the model is perhaps too restrictive and is not adequately representing the data generating procedure of the real system as it suggests that a dichotomy is being created in the model, in which individuals who are not subject to obesity stigma have little to no depression, whereas individuals who are subject to stigma have higher levels of depression. In essence, the absence of detailed non-stigma based mechanisms for depression has resulted in an overestimation of the importance of obesity stigma in causing depression in the population.

6.6 Discussion

In this chapter I have designed and analysed a simple Agent-based Model of obesity stigma and its effects on individuals' body weights, depression and socioeconomic position. Despite initial rejection sampling based calibrations finding some clear peaks in the posterior distributions of parameters, further investigation of the model output suggests that it is not possible to yet make substantive conclusions about what the model says about its associated research questions.

In particular, investigations of 'close' fitting simulations showed that the model dynamics for BMI, depression and income differed qualitatively from

the target data, suggesting that the model has not achieved generative sufficiency. This lack of generative sufficiency means that the model can't be considered a sufficiently accurate representation of the real-world system, and as such making any conclusions about what its output can tell about the real-world system would be inappropriate (León-Medina, 2017; Epstein, 1999).

Despite not yet being able to answer the model research questions, the lack of generative sufficiency has important implications for Agent-based Modelling within the context of obesity. Firstly, it shows how considerations of generative sufficiency can be highly influenced by the choice of calibration target and features that the model aims to reproduce. For example, A previous Agent-based Model of obesity used similarly simple update rules for BMI, in which BMI was calculated as a sum of previous obesity, a small constant and random variation (Mooney and El-Sayed, 2016). However, rather than calibrating to individual BMI and depression trends the authors calibrated to the percentage of obese individuals who were depressed in two different U.S. states and found that the model was sufficient to produce the observations found in the real data (Mooney and El-Sayed, 2016). The authors concluded that stigma was sufficient to generate the observed patterns in the data (Mooney and El-Sayed, 2016).

Since the structure of our BMI update rule was somewhat similar to that presented in Mooney and El-Sayed (2016) it is possible that using different calibration targets in my analysis may have provided a model that appeared to meet generative sufficiency. Similarly, the analysis in Mooney and El-Sayed (2016) may have demonstrated worse generative sufficiency had the model been intending to explain a wider range of phenomena than a simple prevalence. Calibrations to BMI targets specifically may have highlighted unrealistic patterns of obesity in the model dynamics, casting doubt on the generative sufficiency.

The above comparison also highlights the increased challenge that is faced when attempting to validate more complex Agent-based Models. In comparison to the model developed by (Mooney and El-Sayed, 2016), my model is attempting to explain a larger number of phenomena. In addition to explaining the relationship between obesity and depression, we have included potential effects of stigma on future obesity, as well as effects of obesity stigma on income. This additional complexity hence requires additional calibration

targets to be met before we can be satisfied generative sufficiency has been met (Epstein, 1999).

Considering all the above, the model will need to be improved upon, as in its current state it cannot be seen as a candidate model that explains the real world system. In the current specification of the model, in order to keep the model as simple as possible, only stigma based mechanisms were represented in detail, effects of non-stigma based mechanisms were covered by random variation. Since this representation of the real-world mechanisms has not been able to produce realistic dynamics, additional detail will need to be added to this portion of the model.

Hence, in the next chapter, I will present an improved version of the model that adds additional detail to the non-stigma based mechanisms in an attempt to create a model that can more credibly investigate its associated research questions.

Chapter 7

A Revised Agent-Based Model of Stigma in the Obesity-Depression-SEP system

7.1 Introduction

In Chapter 6, I presented an initial Agent-based Model that aimed to explore the role of obesity stigma in the obesity-depression-SEP system. This initial model was shown to be unable to replicate important qualitative features of target data generated in the real world system, and as a result was not considered suitable for drawing conclusions on my research questions.

In particular, the model was not able to produce trajectories for mean BMI and mean income that were able to drop or level off from one time-point to the next. Instead, mean BMI and mean income increased at an increasing rate over the course of the simulations. Additionally, the 'close' simulations all contained mean depression trajectories that, after initial sharp deviations, increased gradually throughout the rest of the simulation. This behaviour was contrary to that of the depression targets within which mean depression slightly decreased over the model time-horizon.

To address the limitations of that model, in this chapter I will present an updated version of the Agent-based Model, that aims to better represent the underlying system by describing a more complex range of features that are potentially important for the system's dynamics.

The revised model will be presented in four main sections. In the first section (7.2) I present an updated version of the conceptual model. Since large amounts of the conceptual model have remained unchanged, this section will

only outline changes and additions to the conceptual model to avoid repeating large amounts of material.

The second section (7.3) will then give the full detailed design of the Agent-based model which, as in Chapter 6, is described using the ODD protocol outlined in Grimm et al. (2006). The results of this model are then outlined and discussed in sections 7.5 and 7.6 respectively.

7.2 Updating the Conceptual Model

The vast majority of the conceptual model outlined in Chapter 6 has remained unchanged for this iteration of the model. Namely, the understanding of the problem, the modelling objectives, the modelling outputs and the majority of the model detail is identical to that described previously. However, changes have been made to the non-stigma based effects in the model, with additional detail being added to these parts of the model. As a result, in this section I will describe the additional detail added to the non-stigma based mechanisms in the model, along with the additional model inputs and assumptions that these mechanisms will require.

7.2.1 Additional Model Detail

7.2.1.1 The Effect of Lifestyle on BMI

In the previous iteration of the model, changes to individuals' BMIs came as a result of two things: stigma and random variation. The random variation in the model aimed to provide a simple representation of non-stigma based effects on BMI and could result in both increases and decreases to an agent's BMI. However, this random variation was mean-centred at zero, and so on average would not generate changes to the population BMI. As a result, the mean BMI in the population was more heavily driven by the amount of stigma in the model, which was only able to result in increases to BMI, thus creating a too narrow range of BMI dynamics in the model.

To generate a population level BMI that can both increase and decrease it will therefore be necessary to include additional mechanisms that allow agents to actively resist weight gain and lose weight. For example, individuals may resist weight gain by attempting to change their diet and levels of physical activity.

Weight change is a complex process that involves the combination of a wide range of factors (Vandenbroeck, Goossens and Clemens, 2007). It would not be possible to represent all these factors that might influence weight change in the model and so a simple representation of non-stigma based effects is required; one which is capable of generating more complex behaviour than simple random variation.

To do this, a simple representation of how obesogenic the individuals' lifestyles are will be included in the model. Amongst the many factors that have been implicated in the development of excess weight, individual lifestyle behaviours such as how much physical activity people do, and how many calories people consume have been shown to play a central role. In general, if an individual is able to maintain a lifestyle in which they live in a calorie deficit, they will lose weight, and conversely, if an individual has a lifestyle in which they consume more calories than they expend, they will gain weight. Again for the sake of parsimony, a detailed representation of energy consumption and expenditure in an individuals lifestyle will not be included in the model. Instead individuals' lifestyles will be described in a way that summarises the impact of said lifestyle on their future weight.

7.2.1.2 The Effect of Inflation and Retirement on Income

In order to better align the model output with the dynamics of the mean income target, additional mechanisms that influence individuals' incomes in the model will be needed. The mean income generated in the previous model in Chapter 6 increased at an increasing rate throughout the simulation, whereas target data suggests that the mean income increases at a variable rate that more closely approximates a linear increase. Furthermore, the target mean income dropped towards the end of the simulation period, and this drop was not observed in any of the simulated mean incomes.

As with the BMI dynamics this suggests that additional mechanisms need to be included in the model so that it can be considered an adequate representation of the underlying system. The first of these mechanisms to be included is a retirement mechanism. As indicated in Table 7.5 wages are highest in the 40 to 49 age group in men and in the 30 to 39 age group in women. One potential contributor to this drop is retirement. Individuals with higher wages are more likely to have saved up sufficient resources to retire at a younger age than those in lower paid work. As a result, individuals with

higher wages may drop out of the workforce younger, meaning that the average wage found in the data of older age groups is more heavily influenced by workers with lower paid jobs in that age group. Individuals in their 40s are perhaps unlikely to retire even if they have high wages, as the cost of doing so would likely be very high. Hence, in this age group high earners will still be fully represented and have more influence on the mean income.

The second additional dynamic that needs to be considered is that of inflation. In the previous model, the only way for individuals to increase their income was through a promotion, however, wages may also increase via inflation. Failing to account for inflation might therefore overestimate the impact of a promotion on individual income which could in turn bias the estimated inequality observed between people with obesity and those without. In order to keep the representation of inflation in the model as simple as possible, the wage inflation rate will be set exogenously, with the mechanisms that influence inflation such as price rises being considered out of scope.

7.2.1.3 Non Stigma-based Depression Risk and Recovery

As with BMI, in the previous model, the depression dynamics were heavily driven by the effects of stigma. Whilst meetings without stigma were able to compensate somewhat for the effects of stigma the simulated mean GHQ was still qualitatively different to the target data. This was likely because in each of the simulations, either the effect of stigmatising actions would dominate the trajectory of the mean depression, or the beneficial effect of meetings would dominate.

Again this suggests that additional detail needs to be added for describing ways in which individuals are able to develop and recover from depression in the model. To this end, a general representation of non-stigma based depressive symptom onset and depression recovery will be added to the model. A general representation of these mechanisms is chosen due to the inherent complexity of these processes. As with the other mechanisms in the model, including large amounts of detail on how individuals develop and recover from depression would detract from the ability of the model to give clear answers to its research questions. Hence a simplified, general representation of these processes is more suitable.

With the inclusion of a mechanism through which individuals can recover from depression, the beneficial effect of social interaction without stigmatisation is now moved out of scope. I have done this for two reasons. Firstly,

this effect is encompassed in the new depression recovery mechanism, in that one can see the benefit of positive social contact as one of the hidden contributory mechanisms behind the general depression recovery mechanism. Secondly, understanding the effect of positive social contact on the relationship between obesity and depression is not central to the model research questions and so including it in this version now adds unnecessary detail to the model.

7.2.2 Additional Model Inputs

The addition of the above mechanisms requires the inclusion of additional input parameters in the model that can be varied across simulations. Without these, assessing the relative importance of stigma in comparison to these new mechanisms would not be possible. However, whilst the following inputs are important to the model, they are perhaps only of secondary importance in comparison to the inputs surrounding the stigma mechanisms outlined in Chapter 6, since those inputs are the ones that are most central to the model research questions.

In order to examine the relative impacts of lifestyle and stigma on BMI, two main factors will need to be parametrised within the model: the effect of lifestyle on BMI when individuals are actively attempting to lose weight and the impact of lifestyle when not doing so. Similarly, in order to understand the scale of stigma's impact on depression versus other causes, parameters describing the effect of non-stigma related causes on depression and the rate at which individuals can recover from depressive symptoms will need to be included. Lastly, in order to understand fully the relationship between income and BMI in the model, and to what extent stigma influences this relationship, parameters that describe the effect of income on one's likelihood to retire and the effect of retirement on income will also need to be included in the model.

I have also made changes to the data inputs of the model. Specifically, data used in this iteration of the model was taken exclusively from the Whitehall II dataset, the New Earnings Survey and the Annual Survey of Household Earnings. This was done to align the data generating mechanisms of the model more closely to the data generating mechanisms of the target data used to compare the model with the real world.

In the previous iteration of the model, real world data were taken from a

mixture of sources, some of which took a repeated cross-section of the population (Effects of Taxes and Benefits Survey and Health Survey for England) and others that examined a cohort (Whitehall II). Specifically, mean BMI targets for the model were derived from the Health Survey for England and mean income targets were derived from the Effects of Taxes and Benefits Survey. Since these datasets examine repeated cross-sections, the population on which data are collected will change over time. For example, the population examined by the Effects of Taxes and Benefits Survey will change as young adults enter the workforce whilst others retire and eventually pass away. However, the model population more closely represented a cohort: since births and deaths were not included in the model, the population of agents more closely represented a single closed group of individuals moving forward in time.

The result of this was that the data generating mechanisms for BMI and income in the target data differed from those in the model. Since the cross-sectional data at each year should capture a representative sample of individuals from across the entire spectrum of work, increases within the income targets over time will not be primarily due to promotions, but instead will be a result of other factors such as workforce wide labour changes and inflation. However, in the model, mean income increases as a result of people receiving promotions as they progress through their careers, and additionally from the impact of inflation. Additionally, as the cohort of agents progress forward in the simulation, they will achieve a greater number of promotions and so do not provide a representation of the entire working population, since less qualified younger workers are no longer included in the model at later time-points.

Hence, upon noticing these discrepancies, the input data sources were changed such that both the data targets and model were aligned with the mechanisms present in a closed cohort. Specifically, all the data targets were estimated using a sample of the Whitehall II population. Depression and BMI targets were derived directly from observations in this Whitehall II sample. Income was derived by combining employment data in the Whitehall sample with wage income data from the New Earnings Survey and Annual Survey of Household Earnings, alongside pension income data from the Effects of Taxes and Benefits Survey. More detail can be found on this in section 7.4.2.

7.2.3 Additional Model Assumptions

As with the inputs, the inclusion of additional mechanisms in the model means that additional assumptions and simplifications are also made in order to keep the representations of these mechanisms as simple as possible.

7.2.3.1 Assumptions and Simplifications for the Effect of Lifestyle on BMI

Within the representation of weight loss in the model, it will be assumed that all individuals are equally able to engage in a lifestyle that promotes weight loss. In other words, the ability of an agent to either begin or continue a lifestyle that promotes weight loss is not influenced by the success of previous attempts, the individual's BMI, or any of their other characteristics. Whilst this may seem like a strict simplification to make, adding in more detail to describe the potential predictors of successful weight control would add complexity to the model that is not directly relevant to the model research questions.

In addition to the above simplification, the decision to engage in weight loss will be assumed to depend only on whether an individual is above the ideal body size associated with their sex or not. As with the ability to engage in weight loss, it will not depend on the outcome of previous weight loss attempts. This assumption implies that individuals only seek to lose weight in an attempt to fit their body in line with the ideal body size in society.

7.2.3.2 Assumptions and Simplifications for the Inflation and Retirement Mechanisms

The representation of inflation in the model is simplified so that the rate of inflation is fixed for the entire duration of the model simulation. For the representation of retirement, it is assumed that the decision to retire depends only on one's age and current income. Since characteristics like health and private wealth are not explicitly modelled here, they are implictly assumed to have no impact on the individual's decision to retire. Once an individual retires, it is also assumed that they receive a retirement income that is a fixed proportion of their final income. This proportion will be the same for all individuals regardless of their age and final income. Whilst not all workers receive a retirement income based on their final salary, the cohort on which this model is based comprises of UK civil servants. The UK Civil Service

provides a pension to its workers based on the salary they earn whilst working. Hence, whilst the model representation suggested is somewhat crude, it still provides a reasonable representation of the real-world cohort's retirement incomes.

7.2.3.3 Assumptions and Simplifications for Depression Risk and recovery

In order to keep the representation of depression risk and recovery as simple as possible, each individual's non-stigma based depression risk and recovery chance will be assumed to be the same. People's characteristics such as their sex, BMI and income and current level of depression will not influence their risk of developing additional depressive symptoms, nor their chances of improving their current state.

7.2.4 Summary

In this section I have presented an update to the conceptual model which has largely aimed to include additional detail to the non-stigma based dynamics that are to be represented in the model. A schematic of this new conceptual model can be seen in Figure 7.1. As with the schematic presented in the previous chapter, ovals represent key micro-level components that are included in the model, and arrows between these ovals are loosely interpreted as 'has an effect on'. For example, within the model, an individual's sex has an effect on their ideal body size, which in turn has an effect on how much social stigma they receive.

As with the previous schematic the aim was to keep the model as simple as possible, however, it is clear to see that despite this there is still complexity within the model structure, such as feedback between obesity, social stigma and lifestyle. This level of complexity is there despite many potential relationships being omitted, such as direct relationships between depression and income, and between depression and lifestyle to name just a few. In the following section I describe an updated computational model based upon this schematic.

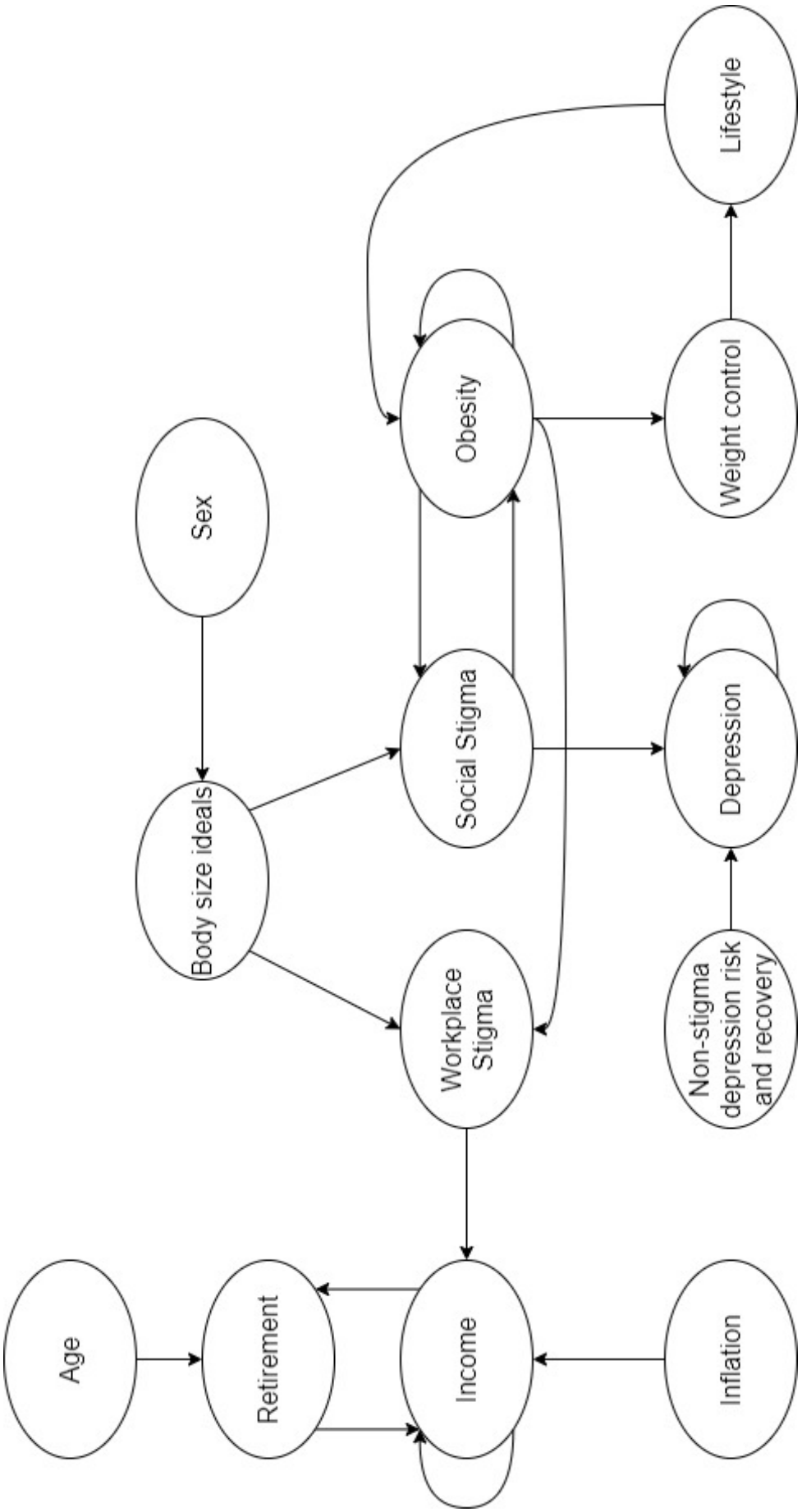

FIGURE 7.1: Updated Schematic of the conceptual model for the Agent-based simulation

7.3 Updating the Computational Model

Having outlined the additional features that are to be implemented in this version of the model, here I will outline the final model description, again using the ODD protocol as a guide. Since the purpose of the model is unchanged, this section has been omitted.

7.3.1 State Variables and Scales

As with the previous model, the only entities included in the model are people and links describing the potential meetings between agents in the model. In order to accommodate the new mechanisms that I have added to the model, additional global and person level variables were added to the model. The full list of variables used in the model are shown in Tables 7.1, 7.2 and 7.3. New variables that were added into the model are shown in italics.

7.3.1.1 People Variables

The main change to the people variables is the addition of a new group of 'health risk' variables: `weight-risk-stigma`, `weight-risk-lifestyle` and `depression-risk-stigma`. The `weight-risk-stigma` and `weight-risk-lifestyle` variables together represent the influence that stigma and lifestyle have respectively on the person's BMI, and the `depression-risk-stigma` similarly describes the increased risk of developing a new depressive symptom in the current time-point due to stigma. The only other change is the inclusion of the `retired?` variable to the employment related variables group, that keeps track of whether the person has dropped out of the workforce or not.

7.3.1.2 Global Variables

Similarly to the changes in people variables, the main changes to the global variables surround the inclusion of 'health risk' variables and employment related variables. In terms of health risk variables, parameters describing both weight-risk and depression-risk have been included, that represent the general non-stigma based mechanisms highlighted in the conceptual model.

Additionally, three parameters have been added to the model in the 'employment conditions' group. The first two of these (`pension-weight` and `retirement-threshold`) influence the conditions under which agents in the model retire, and what size pension they receive when they do retire. The

Variable	Description
Core Variables	
Sex	The agent's biological sex
Age	The Agent's age
BMI	Individual's body mass index
Depression	Individual's depressed affect, this is measured as an integer between 0 and 15 with 0 representing no depressive symptoms and 15 representing multiple severe symptoms
Income	The agent's yearly income
Health Risk Variables	
weight-risk-stigma	The additional BMI an agent will gain in a given time-point as a result of being stigmatised
weight-risk-lifestyle	The change in BMI that an agent will have in a given time-point from non-stigma based factors
Depression-risk-stigma	The risk developing developing additional depressive symptoms due to the effects of stigma in a given time point
Stigma Variables	
body-deviance	The difference between the individual's BMI and the ideal BMI for their sex
stigmatised?	A boolean variable that indicates whether an agent has been stigmatised or not in the current time-point
stigmatising?	a boolean variable that indicates whether an agent is stigmatising or not
stigma-count	An integer that describes the number of times that an individual has been stigmatised in the current time-point
Employment related variables	
time-since-promoted	The number of time-steps since the person was last promoted
retired?	Whether the agent is retired or not

TABLE 7.1: Variables agents in the model are endowed with

third variable `promotion-wait` was added to the model in order to relax the constraint that individuals wait six months after a successful promotion before applying for a another. Initial experimentation with the updated model suggested using a six month wait-time was too restrictive. Specifically, the

model struggled to fit all the calibration targets simultaneously and so this constraint was relaxed to give the model more flexibility to meet the targets.

The `body-tolerance-work` variable was added for similar reasons. Initial experimentation suggested that different values of the stigma parameters were required to produce simulations that fit the income targets well than those that were required to fit the BMI targets well. Hence, this variable was added to allow the model to implement different stigma conditions in social and employment settings, to see whether this would allow the targets to be fit simultaneously, thus aiding model fit.

Outside of the additional model input parameters to be calibrated, two inflation parameters were also added to set the rate of inflation for both employment and pension incomes in the model.

Variable	Description
Ideal body size Variables:	
male-ideal-bmi	The BMI associated with the 'ideal' male body size
female-ideal-bmi	The BMI associated with the 'ideal' female body size
body-tolerance	Threshold of deviance from ideal BMI before an individual is stigmatised for their weight in social situations
body-tolerance-work	Threshold of deviance from ideal BMI before an individual is stigmatised for their weight in employment situations
Weight risk variables	
Obesogenic-environment	Parameter used to describe influence of the obesogenic environment on changes to an individual's BMI when not engaging in weight control
Weight-control-impact	The effect of weight control on individual BMI each time point when an agent is engaging in active weight control
Diet-success-prob	Gives the probability of an agents' weight-control attempt being maintained to the next time-point

Depression risk variables	
depression-risk-base	Gives the probability that an individual develops a new depressive symptom in the current time-step
depression-decay-prob	Gives the probability that a depressive symptom is relieved in a give time-step
Stigma effect Variables:	
stigma-to-depression	Effect of being stigmatised on an individual's risk of developing additional depressive symptoms
stigma-to-obesity	Effect of being stigmatised on an individual's BMI in a given time-point
Economic Conditions:	
promotion-prob	probability of achieving a promotion when applying
promotion-value	scale-factor that income increases by when individuals are successfully promoted
obesity-promotion-penalty	obesity promotion penalty scale constant
pension-weight	The proportion of an agent's final wage income that they receive when retiring
retirement-threshold	Factor which scales the extent to which an agent with above average income is likely to retire
promotion-wait	The length of time and agent must wait after being successfully promoted before they can apply for a new promotion

TABLE 7.2: Global variables in the models that serve as model inputs to be calibrated

Variable	Description
Inflation parameters:	
wage-inflation	The rate of monthly inflation of wages in model
pension-inflation	The rate of monthly inflation of pension income in the model

TABLE 7.3: Global variables in the models that serve as model constants and are set prior to the simulation

7.3.2 Process Overview and Scheduling

As with the previous model, each time step represents one month in real time. Before individuals update their body deviance, global variables describing the correlation between BMI and income, and BMI and depression are updated. These calculations are included so that the model outputs can be more easily collected for analysis after the simulations have finished running.

Once these global variables have been updated, the rest of the model is effectively run using four main submodels. The outline of these submodels are given here though the full details are presented in the 'submodels' section. The first submodel decides the stigma effects, the second decides the impacts of promotions and retirements, the third decides effects of weight-control attempts, and the final section combines the effects of these three sections into the agents characteristics for the next time point.

At the beginning of the 'stigma-effects' section the people in the model update their body-deviance to examine how far they are above the ideal BMI for their sex. Meetings are then allocated to decide which pairs of agents will be meeting up in the current time-step. As with the previous model, the stigmatise procedure is then run to see whether agents who are meeting up with others in the simulation engage in stigmatising behaviours during the meetings. The agents then update their values of `weight-risk-stigma` and `depression-risk-stigma`, based on the outcomes of these meetings.

In the second submodel, three procedures are run. The first queries agents who are eligible to potentially retire to see whether they will retire in the current time-step. Once any retirements have been allocated, those who are still employed apply for a promotion. This is followed by the application of income inflation.

In the 'weight control' submodel, agents who are not currently controlling their weight are queried to see whether they will begin a weight control attempt. After any new weight-control attempts have been started, those who are currently controlling their weight, including those who have just started a weight control attempt are queried to see whether their diet will end in the current time point.

In the final 'effects update' submodel, the effects of the previous submodels are combined to update the agent's BMI and depression characteristics.

7.3.3 Design Concepts

The design concepts for this iteration of the model are predominantly the same as those outlined in Chapter 6. For example, their were no changes to the *emergence, sensing, interactions and observation* design concepts. However, some small changes have been made which are outlined below.

Stochasticity: In this version of the model, stochasticity from random variation incorporated into the agents depression characteristics has been removed from the model. In the previous model this random variation was used to represent non-stigma based effects on depression, which has now been represented more explictly. Therefore the random variation has been replaced by the non-stigma based depression mechanisms described in the updated conceptual model. Random variation incorporated into individuals' BMIs has remained, and additionally, random variation is also incorporated into individuals `weight-risk-lifestyle` to represent small fluctuations around individuals' habits.

Fitness: Whilst fitness is still not explicitly represented in the model, in this model `body-deviance` can perhaps be seen to represent an implicit fitness, since individuals use this value to decide whether they engage in weight control or not. When body deviance is greater than zero, agents are inclined to try and reduce their BMI by engaging in weight-control so long as they haven't recently finished a diet. In this way, `body-deviance` can perhaps be seen to represent a measure of 'unfitness' that individuals are trying to minimise in the model.

7.3.4 Initialisation

As with the previous model, each simulation was populated with 1000 agents, whose characteristics were generated from a mixture of observed and simulated data. Agents' age, sex, BMI and depression score, are sampled directly from a subsample of the Whitehall II dataset. In order for the input data to closely represent a closed cohort, individuals in the Whitehall II data set were only included in the subsample if they participated in every wave of data collection. This was done so that the data targets associated with the observed cohort were not influenced by effect of drop-out, as this was not represented in the model. The final subset of individuals contained 4655 of the original participants.

Peoples' initial incomes were sampled from an exponential distribution with a mean parameter derived from the mean income of full-time workers from the 1987 New Earnings Survey, split by age and sex. This survey provides individuals' income from wages and so as such has not been adjusted for inflation. For example, an agent with male sex aged 44 had an income given by a random draw from an exponential distribution with rate parameter $\lambda = 1/256.5$, such that the mean is 256.5. The rate parameters for the exponential distributions by age and sex are given in Table 7.4.

Sex	Age	Mean income, $(1/\lambda)$	Rate parameter, λ
Males	(30-39)	241.9	1/241.9
	(40-49)	256.5	1/256.5
	(50-59)	233.9	1/233.9
Female	(30-39)	167.0	1/167.0
	(40-49)	156.9	1/156.9
	(50-59)	152.5	1/152.5

TABLE 7.4: Initial mean incomes and rate parameters for income sample distribution

The exponential distribution was chosen since its qualitative properties match more closely to features of real-world income distributions than do others such as the Normal distribution. For example, its values are always non-negative and higher incomes are increasingly less likely to be sampled corresponding to the fact that the large numbers of people in the working populations have incomes below the mean, and that very high incomes are rare. This can be seen more clearly in Figure 7.2, which plots the incomes of participants from the ONS Effects of Taxes and Benefits Survey in 1987 against

an exponential distribution fit to the same data. One thing to note is that this choice was based on data from the whole population, rather than an analysis of each age-sex group separately. Such data was not available and so I have implicitly assumed that a similar pattern of incomes exists with each age-sex group as is observed in the population as a whole.

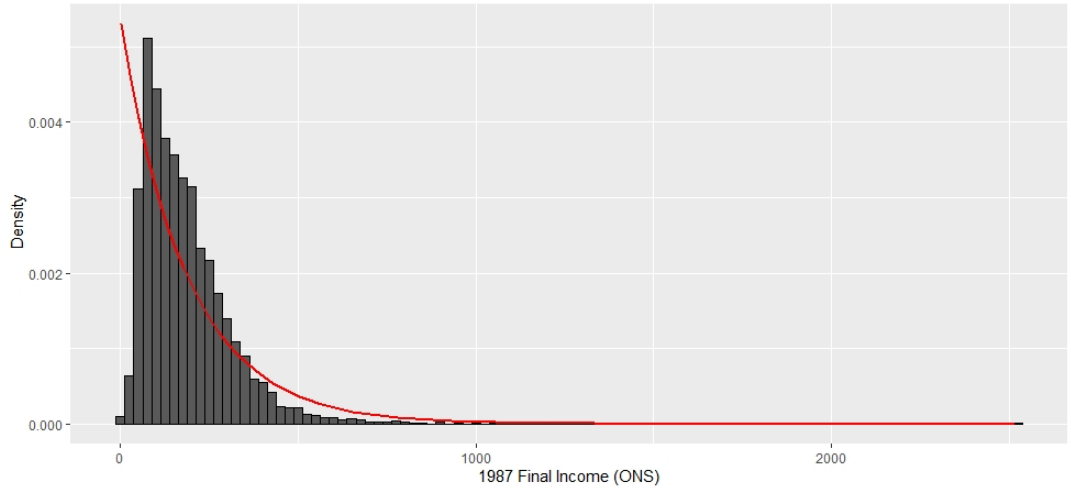

FIGURE 7.2: Histogram of income in 1987 compared with exponential distribution (red curve)

7.3.5 Inputs

7.3.5.1 Inflation

The two inflation parameters `wage-inflation` and `pension-inflation` were derived from estimates of the average increase in wages and pensions respectively between 1987 and 2013. The `wage-inflation` was calculated using data from the 1987 New Earnings Survey and the 2013 Annual Survey of Hours and Earnings (ASHE). Due to identifiability of the subjects in the dataset, I was unable to obtain the raw data for use, however, summaries of the mean income for full-time employees in different age groups by sex have been made available. These are displayed below in Table 7.5. Wages for full-time employees were used for the sake of consistency, as data for part-time employees from the New Earnings Survey was not available.

The value of the `wage-inflation` parameter was calculated by using the average rise in mean wages across age and sex groups. Specifically, within each age-sex group the mean wage in 2013 was divided by the mean wage in 1987 to find the scale factor by which wages had increased over the time period. These scale factors were then averaged to get an average increase

	Age-group			
	30-39	40-49	50-59	60+
1993				
Males	241.9	256.5	233.9	200.3
Females	167	156.9	152.5	143.5
2013				
Males	678.9	772.6	755.3	659.8
Females	587.1	576.8	551.6	496.0

TABLE 7.5: Average weekly wage for full-time UK employees by age-group and sex

in wages. This average increase was then taken to the power of 1/312 (26 years of twelve months) to find the average monthly inflation over the 20 year period. This resulted in a wage inflation each month of 0.39% which equates to a wage inflation of 4.7% per year.

To make this concrete, suppose $x_{s,g,t}$ denotes the average weekly wage in people of sex $s : s \in \{male, female\}$, in age group $g : g \in \{30-39, 40-49, 50-59, 60+\}$ in year $t : t \in \{1987, 2013\}$. Then the scale factor of wage increase in sex s and group g is given by

$$r_{s,g} = \frac{x_{s,g,2013}}{x_{s,g,1987}},$$

and monthly wage inflation γ_w (averaged over the eight age-sex groups) is given by

$$\gamma_w = \left(\frac{\sum_{s,g} r_{s,g}}{8} \right)^{1/312},$$

where the power 1/312 is due to there being 312 months over the 26 year time-period.

Pension inflation was derived similarly, but rises were based on the increases in average weekly pension received. Data on the mean value of a weekly pension was derived from the Effects of Taxes and Benefits Survey taken in 1987 and 2013. From the survey, two variables were used to assess the mean value of a pension: the weekly value of state pensions received and the weekly value of occupational pensions received. From these, the increase in the mean pension received was used to derive the pension inflation. As with wage inflation, the mean value for a pension in 2013 was divided by the value of a pension in 1993 to get a total pension inflation for the 20 year

period. This value was then taken to the power of (1/312) to get the monthly value for `pension-inflation`. This led to a monthly pension inflation in the model of 0.48% which equates to 6.0% per year.

Again, to make this concrete, suppose that $\mu_{j,t}$ represents the total value of the pensions participant j from the Effects of Taxes and Benefits Survey received across both types of pension in year $t : t \in \{1987, 2013\}$, excluding participants who received no pension from the sample. Then the mean value for a pension in year t, μ_t, excluding those who did not receive a pension, is given by

$$\mu_t = \frac{\sum_j \mu_{j,t}}{n_t},$$

where n_t is the number of people who received a pension in the data set during year t. The monthly pension inflation rate γ_p is then given by

$$\gamma_p = \left(\frac{\mu_{2013}}{\mu_{1987}}\right)^{1/312}.$$

Once again the power 1/312 is used to calculate the monthly inflation over the 26 year period.

7.3.6 Submodels

As with the previous chapter, in this section, I will describe in more detail the processes included within the submodels introduced in the process overview and scheduling section.

The first step that the model performs in each time step is to update the model's global variables. This is done for two main purposes. Firstly, some of the procedures described below rely on up-to-date global information in order for them to be implemented accurately and secondly, for the purpose of gathering summary information about the behaviour of the model. For example, at this point the mean BMI of agents is calculated so that comparisons can be made with target data.

Once the global variables have been updated each person in the model updates their body deviance using the method described in Chapter 6, after which, the model is run in four main submodels.

7.3.6.1 The Social Stigma Submodel

The stigma sub-model is mostly unchanged from that which was outlined in Chapter 6. However, the three procedures outlined previously have been modified slightly to account for the new representations of the model's mechanisms.

As in Chapter 6, in the first part of the social stigma submodel, people choose two of their neighbours at random to meet up with during the current time point. This is performed identically to the previous iteration of the model and so I will not repeat the description here. Similarly, once all the meetings in the model have been decided, each agent decides whether they will stigmatise the other agents they are meeting with. This is again performed identically to the procedure described in Chapter 6 whereby an agent is stigmatised by someone they meet with probability $p_{i,t}$ from equation 6.1.

Finally, once all the decisions on whether to stigmatise or not have been taken, each person counts the number of times they have been stigmatised and uses this to incorporate the effects of stigma into their state. If person i is stigmatised k times during time-point t, then their values of `weight-risk-stigma` and `depression-risk-stigma`, denoted $x_{i,t}$ and $y_{i,t}$ respectively are given by the following equations

$$x_{i,t} = kx \tag{7.1}$$

$$y_{i,t} = ky, \tag{7.2}$$

where x and y represent the values of the global variables `stigma-to-obesity` and `stigma-to-depression` respectively. This means that individuals who are stigmatised more in the model receive more deleterious effects than those who are only stigmatised a few times in a given time-point.

7.3.6.2 Employment Submodel

Once the social stigma submodel has been run, the next submodel examines the employment mechanisms in the model. Within the employment submodel, there are three main procedures: 'apply-retirements', 'apply-promotions' and 'apply-inflation'.

'apply-retirements'

In the first procedure agents who are potentially eligible for retirement have the chance to retire in the given time-point. Agents become eligible for retirement at age 55, and all agents retire with certainty when they reach the age of 65 in the simulation. Between the ages of 55 and 65, agents are more likely to retire the higher their income is.

Specifically, when deciding whether to retire early or not, people compare their income with an estimate of the mean income of the working population, and are more likely to retire the further above this estimate their income is. The value of this estimate m is based on the initial income of the cohort, combined with inflation, so that the effect of retirements in the model does not unduly influence the estimate of the mean income in workers.

Suppose the initial income of the cohort in the model is given by μ_0. Denote also by β the value of `wage-inflation` in the model. Then at time step t the estimated mean wage is given by

$$m = \beta^t \mu_0. \tag{7.3}$$

Then if τ denotes the value of the global variable `retirement-threshold`, at time-step t an individual with income $z_{i,t}$ retires with probability $p_{ret,t}$ given by

$$p_{ret,t} = \begin{cases} 0, & \text{if } z_{i,t} \leq m \\ \frac{z_{i,t}-m}{\tau m}, & \text{if } m < z_{i,t} < m(\tau+1) \\ 1, & \text{if } z_{i,t} \geq m(\tau+1) \end{cases}. \tag{7.4}$$

From the above, agents become more likely to retire early the further above m their income is. For example, if an agent has an income that is twice the estimated mean wage, the probability of retiring in that time-point will be $1/\tau$. In general, for an employed agent whose income is n-times the mean wage, the probability of retiring will be $(n-1)/\tau$.

When an agent retires, their income is multiplied by the value of `pension-value`. In this way, all people in the model receive a pension that is proportional to their final salary. Once an agent has retired they are no longer able to apply for promotions, and their income can only increase via pension inflation.

'apply-promotions'

The second main procedure of the employment submodel is the 'apply-promotions' procedure. This procedure is unchanged in its structure from the previous model and so its full description can be found in Chapter 6. However, in this iteration of the model only people who are not yet retired can apply for a promotion. Furthermore, since the wait time between promotions is now given by the value of the global variable `promotion-wait`, equations 6.2 become

$$q_{i,t} = \begin{cases} \pi(1 - r\alpha), & \text{if } n \geq w \\ 0, & \text{otherwise,} \end{cases} \tag{7.5}$$

where w is the value of `promotion-wait` and all other definitions are the same as in section 6.4.2.7.

In summary, agents who have not been promoted in the last w months and are not yet retired apply for a promotion. The application is successful with a given probability $q_{i,t}$ that is determined by the global variables `promotion-prob` and `obesity-promotion-penalty`. If a promotion is successful the agent's income is multiplied by the value of `promotion-value`.

'apply-inflation'

Once promotions and retirements have been applied in the model, the final procedure in the employment submodel applies the effect of inflation to the agent's incomes. Specifically, individuals who are employed have their income multiplied by the value of `wage-inflation`, and individuals who are retired have their incomes multiplied by the value of `pension-inflation`.

7.3.6.3 Weight Control Submodel

The weight control submodel examines the effect of any weight control attempts that the people in the model are partaking in and comprises of two main procedures: 'check-diet-status' and 'apply-new-diets'.

'check-diet-status'

This procedure checks whether agents who are actively engaging in weight control in the model successfully maintain their diet in the current timepoint. If currently on a diet, an agent successfully maintains their diet with probability set by `diet-success-prob`. This setup means that the length of an

agent's diet is distributed as a geometric random variable, and the probability that an individual's diet lasts k time-steps is given by $p^{k-1}(1-p)$ where p is given by the value of `diet-success-prob`. The mean length of a diet is then given by $1/p$.

If an agent successfully maintains their diet in the current time-point, their `weight-risk-lifestyle` variable is set to the value of `weight-control-impact`. If the diet is unsuccessful, then a new value for the agent's `weight-risk-lifestyle` is set as the maximum of either zero or a number drawn from a Normal distribution with mean given by the value of `obesogenic-environment` and standard deviation given by the value of `obesogenic-environment /2`.

These values were chosen to reflect the assumption that the kind of lifestyle an individual will revert to after a weight control attempt will be influenced by the extent of the obesogenic environment, such that a more obesogenic environment will result in more obesogenic lifestyles on average, outside of active weight control. However, the fact that the standard deviation also scales with the size of the `obesogenic-environment` variable means that there will be more variation in the the lifestyles people have when the environment is more obesogenic. This was included so that in all scenarios it was possible for individuals to have a lifestyle that had little to no impact on their weight. The values of `weight-risk-lifestyle` were bounded below at zero so that individuals could not systematically lose weight outside of engaging in active weight control.

This can be summarised mathematically in the following. Supposing that person i is engaging in active weight control at the beginning of time-point t, and denote their weight control status at the end of time point t by $d_{i,t}$ where $d_{i,t} = 1$ if they are dieting, and $d_{i,t} = 0$ if they are not. Then

$$P(d_{i,t} = 1) = p, \tag{7.6}$$

$$P(d_{i,t} = 0) = 1 - p. \tag{7.7}$$

Then, if $r_{i,t}$ denotes `weight-risk-lifestyle` of person i at time t, we have

$$r_{i,t} = \begin{cases} \omega, & \text{if } d_{i,t} = 1 \\ \max\{0, \epsilon\} & \text{if } d_{i,t} = 0 \end{cases} \tag{7.8}$$

where ϵ is drawn from an $N(v, v/2)$ distribution, ω is the value of the global

variable `weight-control-impact`, and v is the value of the `obesogenic-environment` global variable.

'attempt-new-diet'

After the agents currently engaging in weight control have checked whether they can successfully maintain their diet, some agents who are not currently on a diet can attempt a new diet. All agents who have not finished a previous diet attempt in the last 6 time-steps and have a `body-deviance` greater than zero attempt a new diet. Agents who attempt a new diet have their `weight-risk-lifestyle` set to the value of `weight-control-impact`. Agents who do not attempt a new diet have a random perturbation added to their `weight-risk-lifestyle` drawn from a $N(0, 0.01)$ distribution, which represents small fluctuations in individuals' lifestyles.

7.3.6.4 Effects Update Submodel

In the final submodel, the effects of the stigma submodel and the weight control submodel are incorporated into the agents' BMI and depression levels. This is done by two main procedures: update-weight and update-mood.

'update-weight'

In the update weight procedure, each agent's new BMI is calculated as the sum of four components: the agent's previous BMI, the agent's `weight-risk-stigma` value, the agent's `weight-risk-lifestyle` value and small random perturbation drawn from a $N(0, 0.1)$ distribution. Once the BMI has been updated, the `weight-risk-stigma` values are set back to zero before the beginning of the next time-point.

'update-mood'

In the update-mood procedure, individuals first have a chance of developing an additional depressive symptom, in which the agent's depression score increases by 1. The probability with which agents develop an additional symptom is given by the sum of two components: the agent's value of `depression-risk-stigma` and the value of the global variable `depression-risk-base`. Once any increase in an agent's depressive symptoms have been applied the agent also has a chance for a depressive symptom to resolve, resulting in their depression score decreasing by 1. The probability with which an agent's depression score decreases by 1 is given by the value of `depression-decay-prob`.

Again this can be expressed mathematically in the following. Denote by $g_{i,t}$ the number of depressive symptoms person i has at the end of time point t, and denote by λ and γ the values of the global variables `depression-risk-base` and `depression-decay-prob` respectively. Then,

$$g_{i,t} = g_{i,t-1} + u - t, \tag{7.9}$$

where

$$u = \begin{cases} 1, & \text{with probability } \lambda, \\ 0, & \text{with probability } (1-\lambda), \end{cases} \tag{7.10}$$

and

$$t = \begin{cases} 1, & \text{with probability } \gamma, \\ 0, & \text{with probability } (1-\gamma), \end{cases} \tag{7.11}$$

Finally, after the impacts on BMI and depression have been incorporated into the agents' characteristics, the effects of ageing were applied. Within this, each agent's age was increased by (1/12) to account for the fact they had aged by 1 month during the time-step.

7.4 Re-calibrating the Model

7.4.1 Model Parameters to Calibrate

All of the global variables listed in Table 7.2 were calibrated to data. In general, prior parameter ranges were chosen to be as uninformed as possible by any prior information and so a uniform prior distribution was assumed for all parameters. The prior ranges for the parameters are given in Table 7.6. The parameter range for the stigma effect variables `stigma-to-obesity` and `stigma-to-depression` have been expanded from the ranges used for in the previous model from Chapter 6. This is so that there was no *a priori* assumption made about the relative impact of lifestyle versus stigma on BMI and depression.

Variable	Parameter range
Ideal body size Variables:	
male-ideal-bmi	[20,30]
female-ideal-bmi	[15,25]
body-tolerance	[0,10]
body-tolerance-work	[0,10]
Weight risk variables:	
obesogenic-environment	[0,1]
weight-control-impact	[-2,0]
diet-success-prob	[0,1]
Depression risk variables:	
depression-risk-base	[0,1]
depression-decay-prob	[0,1]
Stigma effect Variables:	
stigma-to-depression	[0, 1]
stigma-to obesity	[0, 1]
Economic Conditions:	
promotion-prob	[0,0.9]
promotion-value	[1,1.1]
obesity-promotion-penalty	[0,1]
pension-weight	[0.1,0.9]
income-retirement-threshold	[1,150]
promotion-wait	[6,36]

TABLE 7.6: Model parameters to be calibrated with prior ranges

7.4.2 Revised Model Targets

As was mentioned previously, the data targets used to calibrate the model were modified so that they more closely matched the model output which was based on a closed cohort of agents. As before, the model was calibrated using the mean BMI, depression score, and income of the population. Table

7.7 lists the targets used in full, along with their corresponding year and time-step from the model. Overall, there were 24 targets: 6 BMI means, 9 GHQ means and 8 income means.

Year	Time-step	BMI mean	Depression mean	Income mean
1987	0	24.31	1.09	-
1990	36	-	1.23	-
1993	72	25.05	0.98	333.35
1995	96	-	-	320.51
1998	132	25.99	0.97	339.12
2001	168	-	0.88	356.39
2003	194	26.57	0.90	355.55
2006	218	-	0.77	360.74
2008	242	26.66	0.73	358.27
2013	312	26.62	0.76	386.41

TABLE 7.7: Data targets used for the model calibration

Mean BMI was calculated using the same subsample of the Whitehall II dataset used to populate the agents' initial characteristics. In the Whitehall II data, BMI was available in waves 1, 3, 5, 7, 9 and 11, corresponding to the years 1987, 1993, 1998, 2003, 2008 and 2013. Mean depression was calculated using the GHQ-based depression score presented in Chapter 3. This score was available at waves 1, 2, 3, 5, 6, 7, 8, 9 and 11 corresponding to the years 1987, 1990, 1993, 1998, 2001, 2003, 2006, 2008.

The new income targets were calculated in each year as a weighted average of the estimated average salary of those still working in the Whitehall II cohort, and the estimated average retirement income of those who were retired. Suppose that $p_{emp,t}$ denotes the proportion of the Whitehall II sample who are working at time point t and that $p_{ret,t}$ denotes the proportion who are retired in time-point t. Then the mean income for the sample at time t, $\bar{I}_t$ is given by:

$$\bar{I}_t = p_{emp,t} \times \bar{I}_{emp,t} + p_{ret,t} \times \bar{I}_{ret,t}, \tag{7.12}$$

where $\bar{I}_{emp,t}$ and $\bar{I}_{ret,t}$ represent the average working income and average retirement income at time t respectively. In order to calculate $\bar{I}_{emp,t}$ and $\bar{I}_{ret,t}$, the Whitehall sample was separated out in each wave into those who had retired and those who hadn't. The subsample of those who hadn't retired were then split into categories by age and sex and the mean income of this group

$\bar{I}_{emp,t}$ was calculated by multiplying the average wage within each age-sex group in the observed data by the proportion of the working Whitehall II sample in this age-sex group. The mean wage in years 1993 and 1995 was taken from summaries for full time workers in the New Earnings Survey, and mean wages in 1998, 2001, 2003, 2006, 2008 and 2013 were taken from the summaries of the Annual Survey of Household Earnings. The products of these mean wages and proportions were then summed to get $\bar{I}_{emp,t}$.

To make this more explicit, if we denote again by $x_{s,g,t}$ the average weekly wage in people of sex $s : s \in \{male, female\}$, in age group $g : g \in \{30-39, 40-49, 50-59, 60+\}$ in year $t : t \in \{1993, 2013\}$, and denote by $n_{emp,s,g,t}$ the number of people in group g of sex s who are still employed in the Whitehall II sample at time t, then

$$\bar{I}_{emp,t} = \left(\sum_{s,g} x_{s,g,t} \times n_{emp,s,g,t} \right) \Big/ \sum_{s,g} n_{emp,s,g,t} \tag{7.13}$$

$\bar{I}_{ret,t}$ was calculated using data from the Effects of Taxes and Benefits survey. Specifically, a variable representing total pension benefits was derived by summing together the values of state pension earnings and occupational pension earnings. $\bar{I}_{ret,t}$ was then calculated as the mean of this total pension benefits variable, where the mean was taken over individuals in the survey who received some form of pension benefits. In other words, those who did not receive a pension were not included in the average. Table 7.8 gives a summary of newly derived income targets by year, along with the mean employment and mean retirement incomes used to construct them.

Year	N (working)	N (retired)	$\bar{I}_{emp,t}$	$\bar{I}_{ret,t}$	$\bar{I}_t$
1993	4282	373	351.87	120.73	333.35
1995	3627	1028	373.30	134.27	320.51
1998	3072	1583	432.49	157.92	339.12
2001	2737	1918	475.71	186.10	356.39
2003	2357	2298	506.85	200.37	355.55
2006	1738	2917	566.41	238.20	360.74
2008	1397	3258	588.96	259.35	358.26
2013	758	3897	624.79	340.04	386.41

TABLE 7.8: Table giving revised targets for each year of the Whitehall II data collection. *Note: values may differ slightly due to rounding*

7.4.3 The Calibration Procedure

The same procedure used to calibrate the model presented in Chapter 6 was re-used for the calibration of this model i.e. the model was calibrated using both rejection sampling, and a direct assessment of 'closeness' to the targets. In this model a larger number of samples was used in order to explore the parameter space in more depth. As such, 200,000 samples were generated from the latin hypercube design.

As in Chapter 6, the tolerance proportion was set at 0.005 such that the posterior distributions for the input parameters were based on the closest 0.5% of simulation runs to the targets. This meant that the posteriors are based on the 'best' 1000 simulations.

Manual comparison with the targets was performed similarly to in Chapter 6. However, due to the change in the model targets, the ranges within which models were considered a 'close' fit were amended. Namely, simulations were considered a close fit if to the targets if:

- the mean simulated BMI was within 1 BMI point of the observed BMI at all calibrated time-points;
- the simulated mean depression score was within 0.5 of the observed mean GHQ-based depression score at all calibrated time-points, and
- the simulated mean income was within £25 of the estimated mean income $\bar{I}_t$ at all calibrated time-points.

7.4.3.1 Sensitivity Analysis

As with the previous model, partial correlation coefficients were used to conduct a global sensitivity analysis. The partial correlation coefficients were calculated based on the mean and standard deviation of the mean BMI, GHQ and income produced by the model.

7.5 Model Results

7.5.1 Sensitivity Analysis

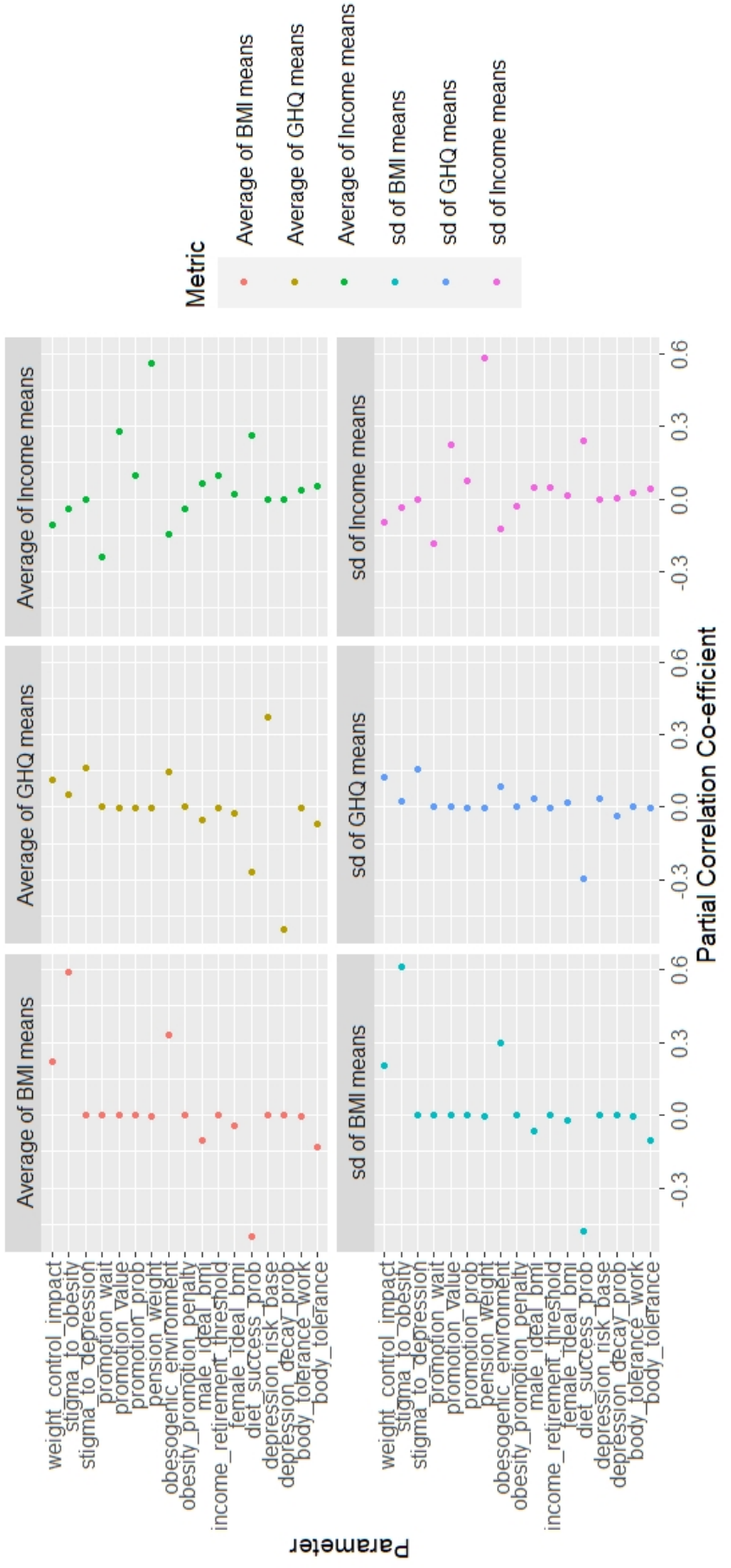

FIGURE 7.3: Partial Correlation coefficients for input parameters with model output

Figure 7.3 displays the partial correlation coefficients that summarise the relationship between the model input parameters and the simulation outputs. Within this, the results suggest that mean BMI and the variation in mean BMI within the simulations is most sensitive to changes in the impact of obesity stigma on future body weight. Weight control attempts and non-stigma based lifestyle risks are then the next most influential, followed by variables that describe how vulnerable to stigma individuals are in the model.

Depression is the only output for which there is a different pattern of influence on its mean value than is the case for its variation. Specifically, the probability of having a successful weight control attempt is the most influential input for variation in mean depression, whereas the `depression-risk-base` and `depression-decay-prob` were the most influential for the average of mean GHQ. This suggests that these depression risk variables, and the model mechanisms that use them are not a large source of stochasticity in the model.

Income output is then most sensitive to changes in the `pension-weight`, `promotion-value` and `diet-success-prob` variables. This sensitivity to the `diet-success-prob` suggests that all three model outputs are sensitive to this input, but in opposing directions. A higher value of `diet-success-prob` is associated with a higher mean income in the model, but a lower value of mean BMI and mean GHQ.

7.5.2 Generative Sufficiency

Figures 7.4, 7.5 and 7.6 display the trajectories for BMI, GHQ and income in simulations that ran close to the separate model targets, but were not necessarily close to all targets simultaneously. For example the simulations displayed in Figure 7.4 were close to the BMI targets, but might not have produced output that was close to the GHQ and income targets.

Overall, these plots suggest that there have been some improvements to the generative sufficiency of the model. In the case of BMI, trajectories are able to fluctuate both up and down over time, whereas in the previous model in chapter 6, mean BMI consistently increased throughout all simulations. Similarly, the mean GHQ trajectories produced by this model much more closely match the qualitative behaviour of the target data than was the case for the previous model. Both the early initial rise in mean GHQ and the

subsequent decay over the remainder of the simulation are captured by the model dynamics.

The plot of income trajectories displayed in Figure 7.4 shows that the model is able to much more closely match to the income targets than was the case in the previous model. Early initial rises in mean income between the initial conditions and the first model target can be replicated, as well as a diminished increase in the middle stages of the simulation time horizon. The model is also able to replicate the up-tick in the target mean income seen at the end of the time-horizon.

However, there are still some concerns about the generative sufficiency raised by these plots. In the mean BMI trajectories, but also somewhat in the GHQ trajectories, extreme initial changes can be seen before a period of more steady change emerges for the remainder of the simulation. Within the close BMI runs, this is typically present in the form of steep initial rises followed by a drop or levelling out of mean BMI which then slowly begins to increase again later on in the simulation. For GHQ there is similarly a sharp increase in initial mean depression score, although it is usually less pronounced and does not last as long as is the case for BMI.

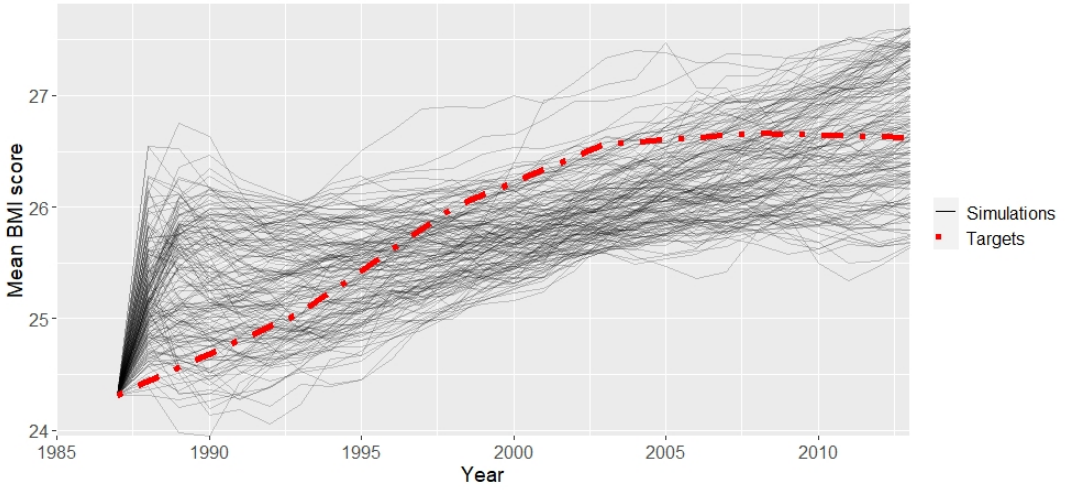

FIGURE 7.4: Demonstration of generative sufficiency for BMI trajectories

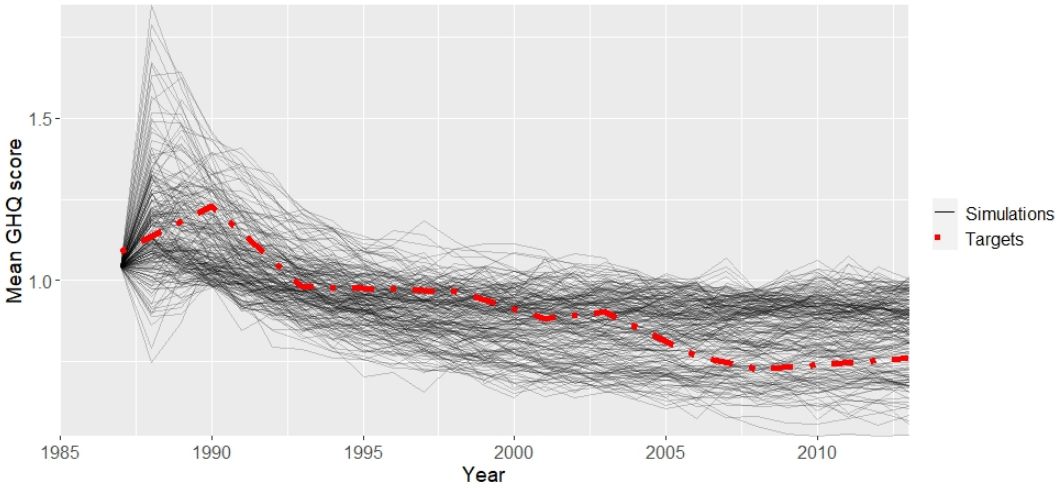

FIGURE 7.5: Demonstration of Generative sufficiency for GHQ trajectories

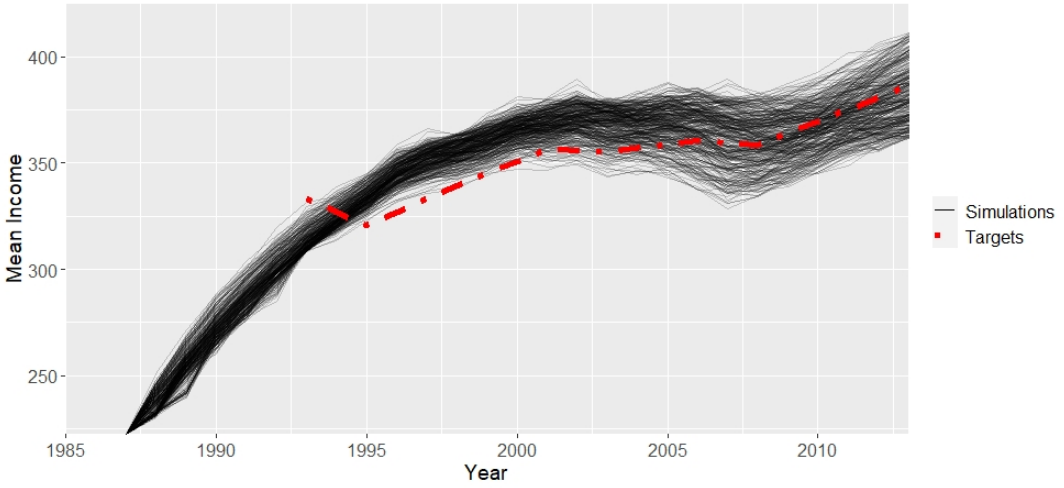

FIGURE 7.6: Demonstration of Generative sufficiency for Income trajectories

7.5.3 Rejection Sampling

7.5.3.1 Trends in the Posteriors

Figures 7.7 to 7.10 display the posterior distributions of the input parameters derived from rejection sampling. Examining first the posteriors based on calibration to all the targets similarly in Figure 7.7, three main patterns surrounding the model parameters emerge.

The first of these patterns is that the closest fit models are in general associated with scenarios where there is a lower chance of being stigmatised, and only small effects from being stigmatised when it does occur. The low chance of being stigmatised is evident from the skew in the posterior distribution of

body-tolerance towards higher values of the parameter range. This means that models more often fit better when people in the model are allowed to be further from the ideal BMI before they are vulnerable to being stigmatised, hence reducing the amount of stigma present in the model. The posterior distributions for stigma-to-obesity and stigma-to-depression then suggest that the effects of stigma are low when it does occur. The peaks of these distributions both fall at the bottom end of their parameter ranges indicating the model is more likely to fit well to the target data if the effects if stigma in the model are small.

The second main pattern that can be seen from Figure 7.7 is that the posteriors suggest the model fits best to the targets in scenarios that minimise weight gain and where weight loss is effective. The effect of lifestyle on BMI is primarily set in the model by the variable obesogenic-environment and so as the posterior for this parameter peaks in the lower part of the parameter space, this suggests that the model fits better when individuals do not have strongly weight-gain-promoting lifestyles. Similarly, the posteriors for diet-success-prob and weight-control-impact suggest that the model fits best when weight control attempts have a greater impact and are highly likely to succeed. In combination, these parameter settings will combine to create scenarios in which there are not high levels of obesity, which would further reduce the amount of stigma present in the model.

The third pattern in the model posteriors is that models with parameter settings that have a high likelihood of individuals recovering from depression, and a low likelihood of individuals developing depression more frequently fit closely to the targets. This can be seen from the fact that the depression-decay-prob posterior peaks at higher values of the parameter space, whereas, both depression-risk-base and stigma-to-depression have peaks at the lower end of their parameter ranges.

It also appears that the model targets have not been able to give information on all of the parameters in the model. For example, body-tolerance-work, female-ideal-bmi,male-ideal-bmi, income-retirement-threshold, promotion-prob and obesity-promotion-penalty all have posterior distributions that are only marginally different to their priors.

On the surface, this would suggest that these parameters are not influential for how well the model fits to the targets, however, when looking at the calibrations based on a single group of targets, a more nuanced picture emerges. In the calibration based on the BMI targets, the posteriors for

both `male-ideal-bmi` and `female-ideal-bmi` have a negative skew, such that higher values in the parameter range more frequently fit the target data well than do lower values. Both distributions have peaks at the upper end of the parameter range: between 23 and 25 for females and between 27.5 and 30 in males. This again supports the idea that the model more frequently produces output that fits the targets well when there is a lower prevalence of obesity stigma.

Similarly, in the calibration based on income, there is a skew towards higher values of both `male-ideal-bmi` and `promotion-prob`, with lower values in the parameter ranges being less likely to fit well. There is also a skew towards lower values of `income-retirement-threshold` in the income based posteriors, suggesting that the model more frequently fits better when there is lower levels of pressure on the wealthy members of the cohort to retire early. The other three parameters, however, still only show small changes from their prior distributions when calibrated to each group of targets separately, perhaps suggesting that they have minimal impact on the outcome of the simulations, at least in comparison to the other inputs.

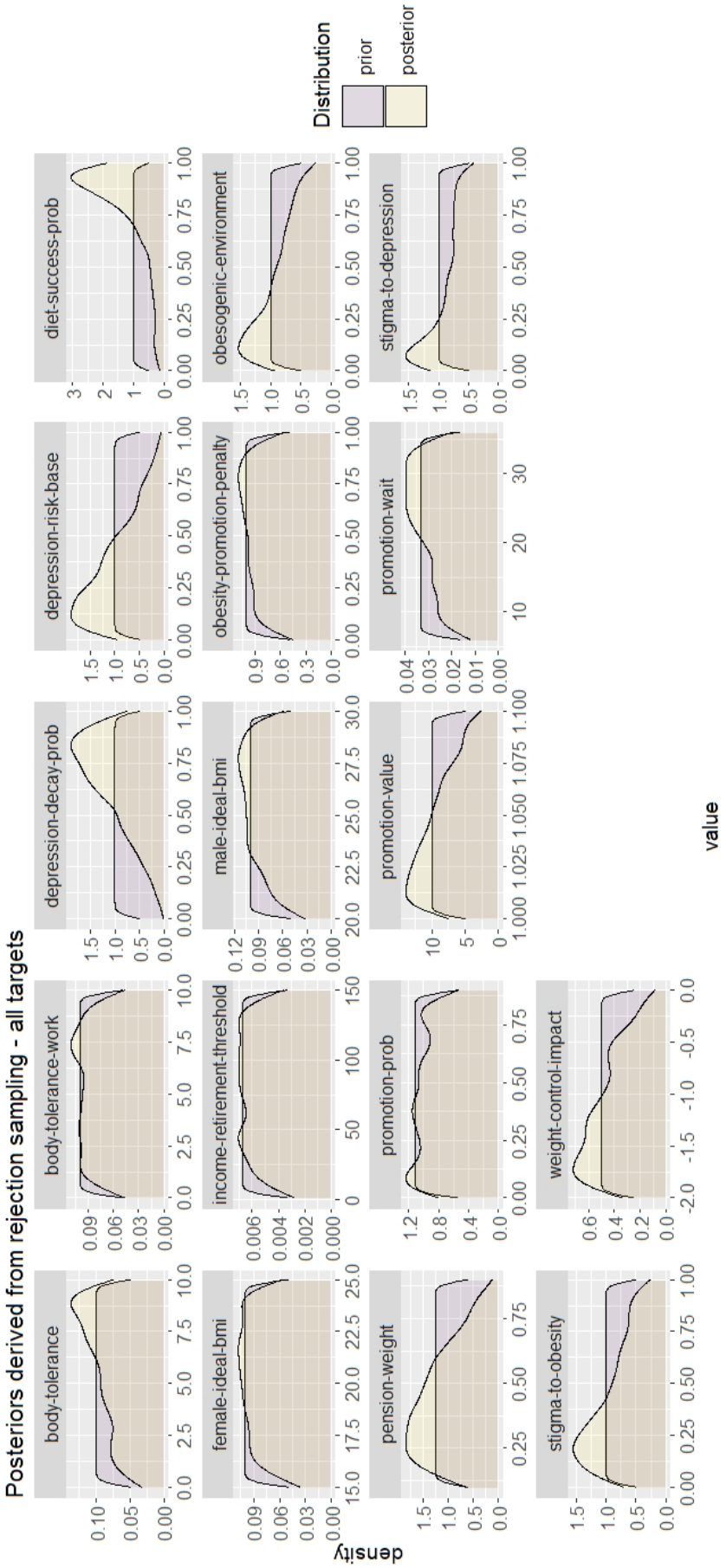

FIGURE 7.7: Prior and posterior distributions generated via calibration to all model targets

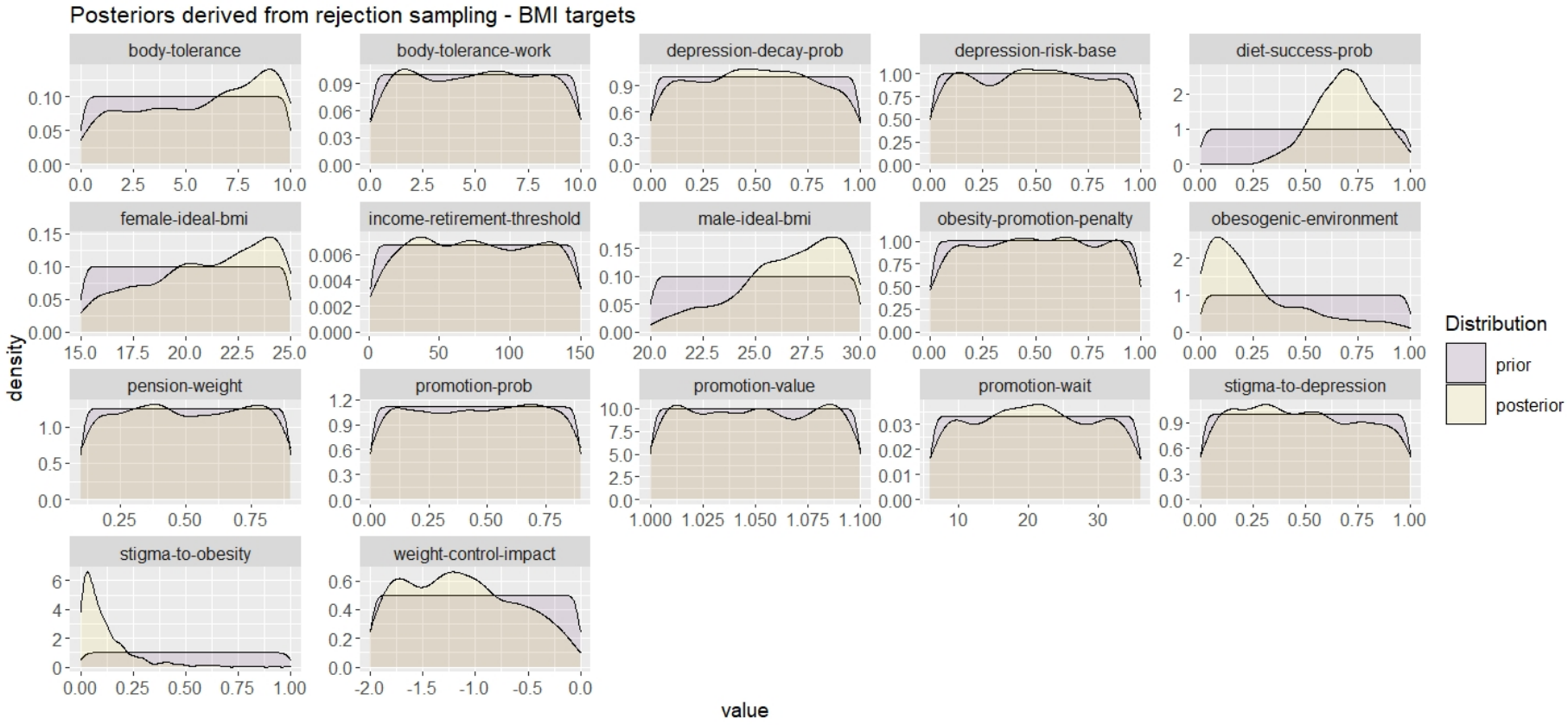

FIGURE 7.8: Prior and posterior distributions generated via calibration to BMI targets

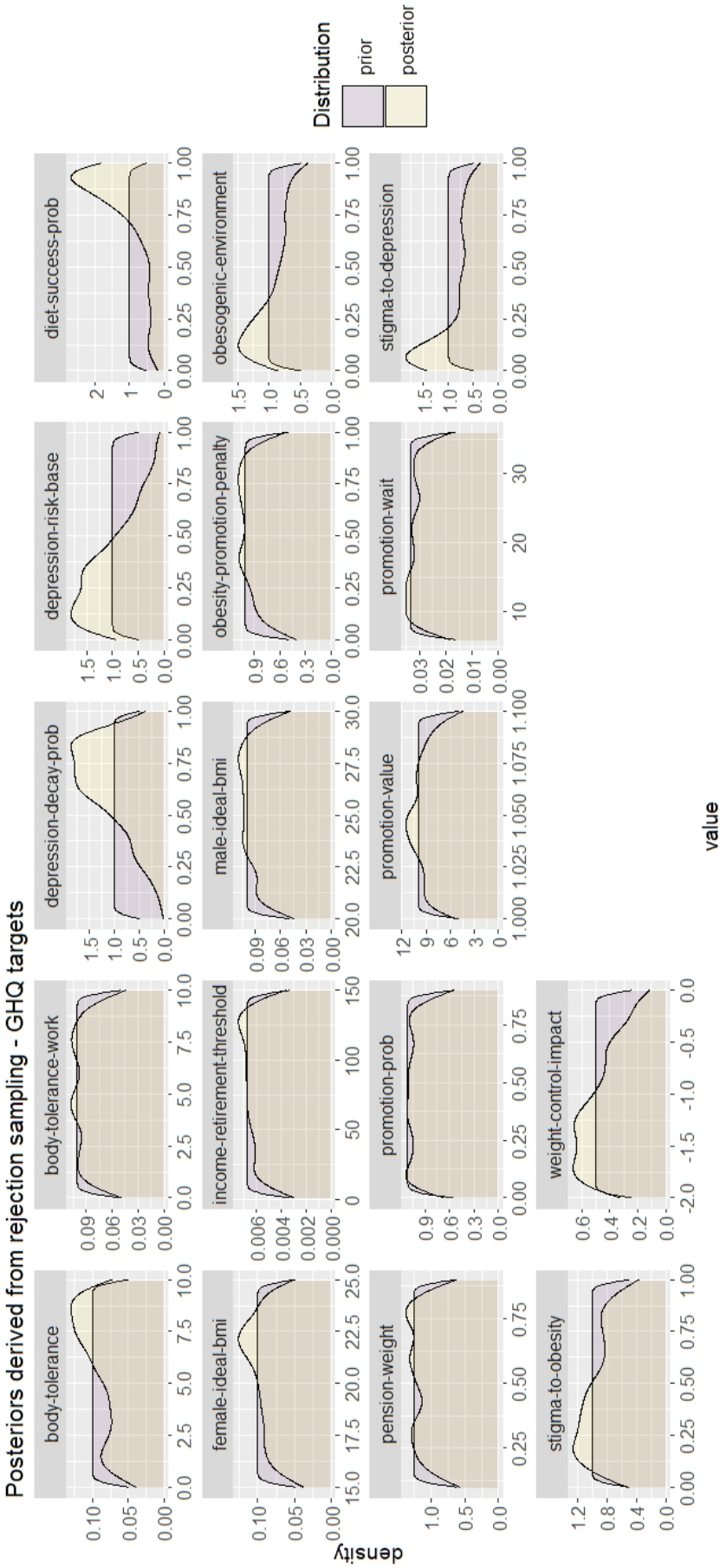

FIGURE 7.9: Prior and posterior distributions generated via calibration to GHQ targets

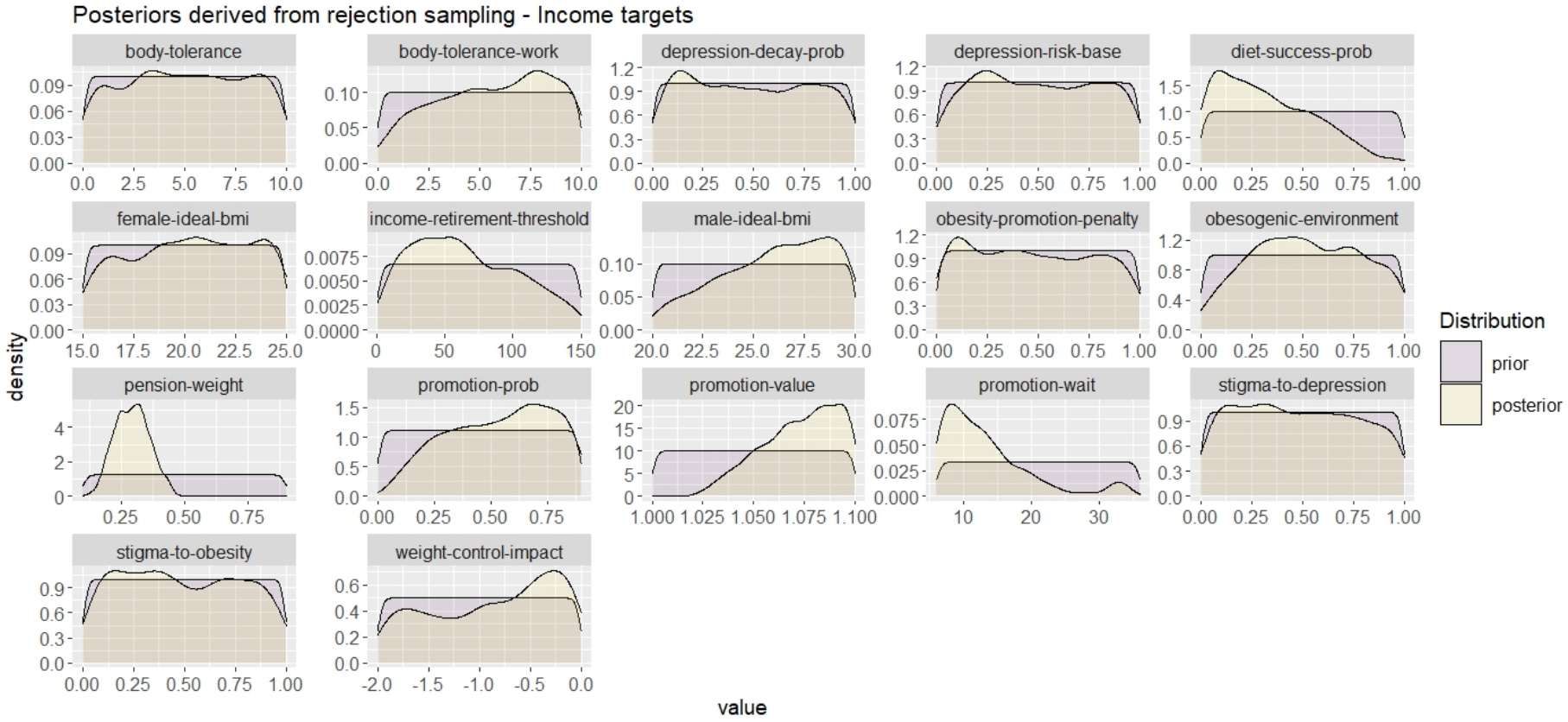

FIGURE 7.10: Prior and posterior distributions generated via calibration to income targets

Discrepancies between the Posteriors based on Different Targets

Across the posteriors generated by calibrations to the different model targets two significant discrepancies can be seen. The first is the discrepancy between the posteriors generated based on the income targets and those generated using the GHQ and BMI targets, particularly surrounding the variables that influence non-stigma related weight change: `diet-success-prob`, `weight-control-impact` and `obesogenic-environment`. Overall, the posteriors based on income targets suggest that scenarios in which there is ineffective weight loss and a potentially moderate effect of the obesogenic environment fit best to the data, whereas the model is more likely to fit the BMI and GHQ targets better when weight loss is effective, and lifestyles are not very obesogenic on average.

High obesity scenarios perhaps fit better to the income targets due to the impact these scenarios have on promotions in the model. Within the model, there are two ways in which downward pressure on incomes can be applied: through retirements and through restriction of promotions. Retirements can create downward pressure on income by making wealthy individuals more likely to retire early, although this pressure only applies to at most half of the cohort who are still working. Restriction of promotions, however, affects the entire portion of the cohort who are working and so is perhaps a more effective means of limiting incomes in the model than are retirements.

The impact of promotions can be limited in two main ways. The first is the direct effect of the employment condition variables `promotion-value`, `promotion-prob` and `promotion-wait`. However, the posteriors for these variables suggest that the model fits best in scenarios with easy-to-obtain frequent high value promotions. This is likely due to the rapid rise in mean income between the initial conditions and the first income target not being fully explained by the effect of inflation in the model. The second way that the model can restrict promotions is to make the penalty for having high body weight in the model severe in such a way as to make large numbers of the model cohort unlikely to be promoted. It is this that likely explains why the posteriors based on the income calibration suggest the model fits better in scenarios that promote weight gain and hinder weight loss. In this scenario peoples' BMI values are likely to rise over the course of the simulation, making them more likely to be subject to obesity stigma in the workplace. This has the impact of making promotions harder and harder to obtain as the model progresses, allowing the early rises in mean income to be curtailed

later on in the simulation.

The second discrepancy is between the posteriors of the promotion variables based on calibration to all targets and those based on calibration to just income. Specifically, in the calibration based on all targets, the posteriors for `promotion-value` and `promotion-wait` are skewed to the left and right respectively, such that the best fitting models more frequently occur in scenarios where promotions are of low value and there is a long wait between them. However, in the income based calibration, these posteriors are skewed in the opposite directions, such that the best fitting models to the income targets more frequently occur in scenarios where promotions are of high value and there is only a short wait between them. There is little to no change in between the priors and posteriors for these variables in the GHQ and BMI based calibration.

This discrepancy is likely a result of the procedure through which the rejection algorithm decides which models have output that is closest to the target data. Comparing the posteriors based on all targets in Figure 7.7 to the posteriors based on a single group of targets in Figures 7.8, 7.8, 7.9 and 7.10, the posteriors based on all targets most closely resemble the posteriors based on the GHQ targets, and are less similar to those based on the BMI and income targets. This suggests that poor fit to the GHQ targets has a greater impact on the model fit assessment than do BMI or income.

This can be seen clearly in Figures 7.11, 7.12, 7.13, which display the model trajectories for the 1000 simulations deemed closest to the targets overall according to the rejection sampling. Besides some deviations at the beginning of the model, these model runs track closely to the GHQ targets for the duration of the simulation. However, in the case of BMI and income, the model runs exhibit signs of poor fit, particularly at later years in the simulation. Mean BMI is able to diverge to impossible values and mean income diverges substantially from the targets in many of the 'best' fitting simulations. Across the included simulations the variability in BMI is so severe that it makes the trajectories of the model and the targets appear linear, as the scale of variability is substantially greater than the changes seen within each simulation and within the BMI targets. As a result, the posteriors based on all targets do not fully reflect the parameter sets that fit best to the BMI and income targets, hence allowing the aforementioned discrepancies to occur.

The fact that poor fitting models with respect to mean BMI and income can be included in the 'best' fitting models overall suggests that differences in these

outputs are not having a substantial effect on the distance score used to decide which models fit best to the target data. As mentioned in section 6.4.3.3, the distance between the model output and model targets is calculated using the Euclidean distance. Within this procedure, both the model output and the data targets are standardised using the median absolute deviation of the output across the simulations (Bonabeau, 2002). For example, the mean BMI in 1993 is divided by the median of all simulated mean BMIs at time 1993. This process aims to put all the model targets and output in a standardised metric to avoid one type of target biasing the overall fit. However, since BMI and income can diverge during the simulation to a much greater extent than can GHQ, this may have had the effect of reducing their impact on the overall model fit assessment.

In light of the above, caution must be taken when examining the posteriors based on all targets, though the overall patterns analysed previously still hold when examining the posteriors based on each target group separately. In the case of the discrepancy between the income based posteriors and overall posteriors analysed here, it suggests that the income based posteriors likely paint the more trustworthy picture of the model behaviour, and that the model will fit better in general when there is high value promotions with short wait times overall.

As well as casting doubt over the model posteriors based on all targets, the severe deviations found in the BMI trajectories highlight scenarios in the model that are unable to fit to the real world. Scenarios in which there is large amounts of weight gain in the model occur when the impact of obesity stigma is high, individuals have highly obesogenic lifestyles, and weight control attempts have little impact and are unlikely to succeed. Additionally, scenarios in which there is extreme weight loss in the model occur when the impact of stigma is low, individuals have lifestyles that do not promote weight gain, weight loss is highlight impactful, and is very likely to succeed.

Within the model, there are no exogeneous limits placed on what values of BMI an individual can attain, since it was hoped that the mechanisms described would generate reasonable values without the need for such limits. However, without these limits, models with parameter specifications that fit within the above two scenarios are able to produce runaway BMI trajectories leading to features such as negative BMI values (which is physically impossible) and BMI values that are far higher than a human could possibly attain. For example, for a person 1.8m in height to have a BMI of 500 they

would need to weigh 1620Kg. As a result, if the scenarios that generate these trajectories exist in the real world system, this would imply that the model's mechanisms need modifying so that impossible trajectories are not produced.

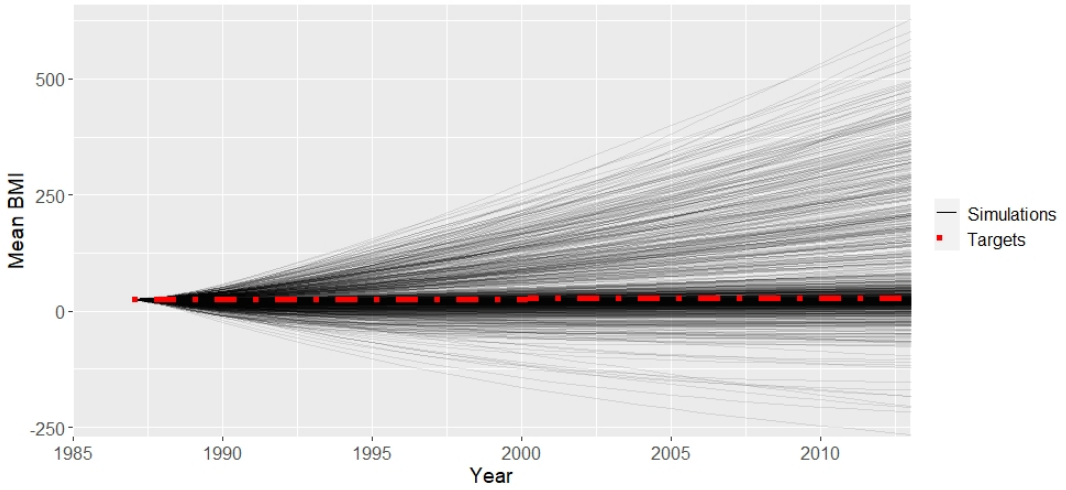

FIGURE 7.11: Mean BMI trajectories for 1000 'best' model runs (black) compared to the target mean BMI (red)

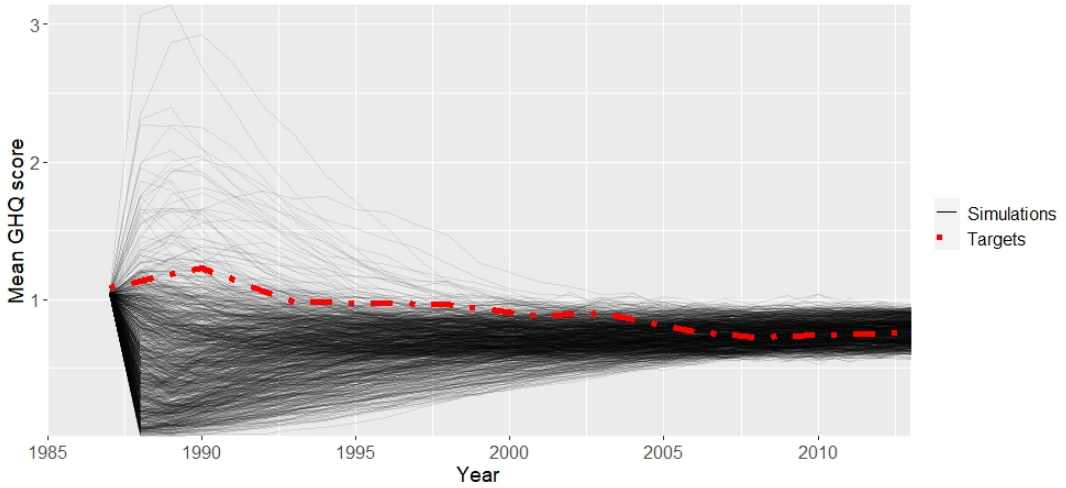

FIGURE 7.12: Mean GHQ trajectories for 1000 'best' model runs (black) compared to the target mean GHQ (red)

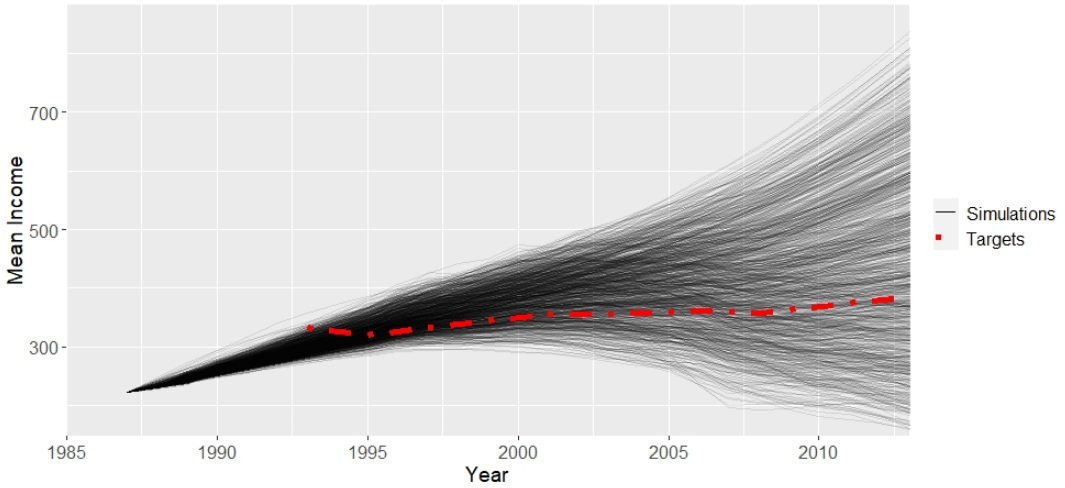

FIGURE 7.13: Mean income trajectories for 1000 'best' model runs (black) compared to the target mean income (red)

7.5.4 Close Fitting Models

The effect of the discrepancies highlighted in the previous subsection can be seen clearly in the analysis of close simulations. Of the 200,000 parameter combinations analysed, there were no parameter settings that produced simulations capable of matching 'closely' to all three targets simultaneously.

This is likely because of the tension created by trying to fit to the income targets and the BMI and GHQ targets simultaneously. Since the model fits these targets best on different sections of the parameter space, finding parameter sets that fit all three targets simultaneously is likely to be difficult. In particular, the discrepancy between the model posteriors for weight-change-related input parameters highlighted in the previous section suggested that the model fit to the income targets in scenarios that were directly in contrast to those that fit best to the BMI and GHQ targets.

Attempting to widen the tolerance around which runs were considered a close fit still resulted in very few model 'close' model runs being found. Specifically, changing the closeness thresholds to include simulations that were: within 1.5 BMI points of each BMI target; within 0.5 of each GHQ target and within £100 of each income target still resulted in only one simulation being considered a close simulation to all three targets. In addition, expanding the tolerance in this way meant that more simulations with undesirable qualitative patterns began to be included in the simulations that were close to a single group of targets, suggesting that meeting all three targets comes at the cost of good fit to the each target individually.

7.6 Discussion

In this chapter I have presented a revised agent based model that aimed to explore how obesity stigma might be a generative mechanism of the relationship between obesity and both depression and socioeconomic position.

In comparison to the model presented in Chapter 6, the model displayed an improved level of generative sufficiency in some aspects. In particular, it was capable of separately producing mean BMI, depression, and income trajectories that more closely matched the qualitative features of the trajectories seen in the observed target data produced by the real-world system (see Figures 7.4, 7.5 and 7.6). This provides further evidence that the representation of non-stigma-based mechanisms was previously too simple.

Despite the improved generative sufficiency in the revised ABM, issues and questions remain over its accuracy that prevent me from investigating the substantive research questions in detail. Firstly, of the 200,000 parameter combinations simulated, no parameter sets produced mean levels of body weight, depression, or income in the population that simultaneously matched closely to those found in the real world.

This absence of close fitting model runs was likely caused by tensions within the model mechanisms creating a situation in which different targets required contrasting parameter specifications for the simulation to match them closely. Specifically, it was likely a result of the tension between the fit to the income targets on one hand, and the BMI and GHQ targets on the other. This tension arises from the fact that the model fits the BMI and GHQ targets best when there is high impact of weight-control and a high chance of successfully maintaining a weight control, whereas the income targets are matched to best in the opposite scenarios. Naturally these two scenarios can't both exist simultaneously and so some compromise between the scenarios must be made in order to try and fit all the model targets. However, this may have the effect of reducing the overall credibility of the model, as any such compromise will reduce how well the model fits to each target, meaning the model perhaps does not fit the data well overall. At the very least, this tension has likely contributed to the fact that no parameter sets were considered a close fit to the targets, as finding parameters which satisfy this compromise without also providing poor fit themselves is difficult.

Additionally, even within the model dynamics that matched to the individual targets, there were still some qualitative features that cast doubts over the accuracy of the described mechanisms. For example, in the BMI and GHQ trajectories there are sharp deviations away from the model's initial conditions, followed by a period of more consistent behaviour. This suggests that at first the models mechanisms are having a strong effect on the population, before some sort of new balance is achieved, from which only smaller changes can occur.

For example, for sharp initial increases in the model outputs' mean depression are likely the result of agents who are vulnerable to stigma, but were not previously scoring high on the depressions scale rapidly increasing there level of depression as a result of stigma. Similarly, when the impact of stigma on depression is low, the sharp initial drop may be caused by people who currently score highly on the depression scale becoming less depressed through the depression recovery mechanism. Similarly, in the BMI trajectories, there was frequently a sharp initial rise in mean BMI, followed by a period of more steady increases. This is likely due to agents who do not have a BMI that is greater than the socially held ideal BMI increasing their BMIs as a result of the obesogenic environment until they reach a point at which they are above the ideal BMI and begin deciding to try and control their weight. This would suggest that there is not sufficient detail in the weight control mechanisms in the model to accurately represent the real world.

Overall, the limitations of the model highlighted by the model analysis suggest that the model still has a questionable level of generative sufficiency, and as such cannot reliably be used to answer its intended research questions (León-Medina, 2017; Epstein, 1999). Despite this, there are still some useful insights that can be extracted from the process of designing and developing the model. In exploring the potential uses for building a quantitative model, Epstein (2008) highlights 16 reasons other than prediction why one might build a model:

1. To explain
2. To guide data collection
3. To illuminate core dynamics
4. To suggest dynamical analogies
5. To discover new questions

6. To promote a scientific habit of mind
7. To bound outcomes to plausible ranges
8. To illuminate core uncertainties
9. To offer crisis options in near-real time
10. To demonstrate trade-offs or suggest efficiencies
11. To challenge the robustness of prevailing theory through perturbations
12. To expose prevailing wisdom as incompatible with available data
13. To train practitioners
14. To discipline the policy dialogue
15. To educate the general public
16. To reveal the apparently simple (complex) to be complex (simple).

Originally, this model was built with the first aim in mind: to explain the relationships between obesity and both depression and socioeconomic position. Whilst we have not been able to explore this in depth due to the lack of generative sufficiency achieved, this lack of generative sufficiency does itself tell us something about the underlying system. Namely, that the mechanisms described in the model are not sufficient to generate the realistic trajectories of Body weight, depression and income, and hence aren't able to explain the relationships between these concepts (Epstein, 1999).

At first glance, this lack of generative sufficiency would seem to conflict with previous literature that suggested obesity stigma was a mechanism that might relate obesity to both depression and socio-economic position. Studies presented in both Chapter 2 and 6 suggested that depression was a likely consequence of obesity stigma (Puhl, Moss-Racusin and Schwartz, 2007; Faith, Matz and Jorge, 2002; Markowitz, Friedman and Arent, 2008). In particular, Faith, Matz and Jorge (2002) presented evidence suggesting that obesity stigma might lead to depression via an effect on body image. Similarly, stigma has been shown to impact multiple socio-economically important factors such as eduction (Puhl and King, 2013) and employment (Fikkan and Rothblum, 2012). As the models constructed based upon this literature in the past two chapters have not generated realistic data patterns, this could be interpreted as evidence that obesity stigma does not explain the relationships between obesity and both depression and socioeconomic position.

However, given the weight of evidence suggesting obesity stigma is related to psychological outcomes including depression (Puhl and King, 2013; Papadopoulos and Brennan, 2015; Pearl and Puhl, 2018) what is more likely is that key parts of the mechanisms' descriptions are missing and that more detail would be required to reproduce the patterns of BMI, dperession and income seen in the real world data. Despite the current knowledge surrounding obesity stigma, the model was built using multiple assumptions and simplifications to deal with uncertainties from the literature, and the flexibility of Agent-based modelling provided many different ways of addressing these assumptions (Bonabeau, 2002). As such, the lack of generative sufficiency may in fact be reflective of the way gaps in the literature have been represented, and the specific modelling choices made, rather than a reflection of stigma's true impact on the obesity-depression-SEP system.

Further to the above, the model building process has also fulfilled three other main uses from the list in Epstein (2008). Specifically, the model can be used to guide data collection, discover new questions and illuminate core uncertainties.

Of these, perhaps the main additional insights that have been gained from the model development process is the discovery of new questions highlighted by the need to make assumptions whilst building the conceptual model. One group of questions that have been highlighted surrounds the phenomenology of weight stigma, for which there exist only a handful of studies (Vartanian, Pinkus and Smyth, 2014; Vartanian, Pinkus and Smyth, 2018). Whilst these studies have provided some useful insight into how stigma is actually experienced within peoples' lives, there still remains unanswered questions that would aid our understanding of the consequences of obesity stigma. For example, despite stigma being commonly perpetrated in multiple different relationships, as of yet it is unclear whether the type of relationship the recipient of stigma has with their stigmatiser influences the effects they are likely to receive (Vartanian, Pinkus and Smyth, 2014; Vartanian, Pinkus and Smyth, 2018).

Likewise it isn't currently clear whether receiving stigma in different contexts results in only context specific consequences, or whether there are generalised impacts to being stigmatised regardless of the context. Within the literature there is consistent evidence that obesity stigma in social situations can affect body weight, weight related behaviours and mental health (Puhl, Moss-Racusin and Schwartz, 2007; Puhl and Brownell, 2003; Vartanian, Pinkus and

Smyth, 2018) and obesity stigma socioeconomic contexts has been shown to impact on socioeconomic outcomes (Godfree, 2020; Fikkan and Rothblum, 2012). As such, in this model we assumed that the impacts of stigma were context specific, such that stigma in social situations only impacted depression and BMI, and stigma in hiring situations only impacted income. However, it is possible that stigma in social situations has socioeconomic consequences for the stigmatised individual and similarly, stigma in socioeconomic contexts may impact individuals mental health. Hence, future research could seek to examine this in more detail to create a more thorough description of the consequences of obesity stigma.

A second group of new questions that has been highlighted surrounds the formation of body image ideals and how these body image ideals influence, and are perhaps influenced by obesity stigma. Within the model, it was assumed that deviations from the ideal body size were equally likely to be stigmatised in males and females. This was assumed due to a lack of evidence available on whether different sexes experience different levels of tolerance for deviations from socially held images of ideal bodies. Whilst there is some evidence that women experience greater levels of obesity stigma than men (Sattler et al., 2018; Fikkan and Rothblum, 2012), men are still vulnerable to obesity stigma (Himmelstein, Puhl and Quinn, 2018) and the precise reasons for any differences in stigma are not fully explained.

In this model differences in stigma experiences between males and females were controlled entirely by the ideal body size associated with each sex. This was due to evidence of difference between the idealised bodies in men and women (Cohn and Adler, 1992; MacNeill and Best, 2015; Stanford and Mccabe, 2002). However, given the evidence on the large amount of scrutiny females receive about their bodies (MacNeill and Best, 2015; Fikkan and Rothblum, 2012), it is possible that women are more likely to be stigmatised for deviations from this idealised body than are men. As such, further research on whether such differences exist, and how these impact experiences of obesity stigma would therefore improve the understanding of the underlying system, allowing more credible models to be made using fewer assumptions.

Alongside the question of differing levels of body tolerance between males and females, the development of the model has also highlighted the potential for examining the theory on how idealised body images develop both within society, and within individuals. In this model, the description of ideal body

sizes was based on the sociocultural theory of body size ideals (Tiggemann, 2012). However, other theoretical perspectives exist (Frederick and Reynolds, 2022; Krayer, Ingledew and Iphofen, 2007; Grogan, 2006), and given modelling can be used to 'challenge the robustness of prevailing theory' future simulation models could attempt to examine and test these theories. Insights gained from the testing of these model theories might then improve understanding of the relationship between body image ideals and stigma.

In terms of providing guidance on data collection, the building of the model has highlighted the need to collect rich, detailed data on obesity stigma and its consequences in order to properly investigate its effects. Agent based models in particular can have potentially large data requirements due to the number of parameters that can be included in the model (Bonabeau, 2002; Casini and Manzo, 2016; Hazelbag et al., 2020). The models presented in these chapters are no different: multiple features were included in the model including stigma, promotions, retirements, depression development, obesity development and weight control. Parametrising these features, and then calibrating these parameters required a combination of multiple data sources as well as assumptions to fill in gaps in the available data.

In light of this, richer data from the obesity-depression-socioeconomic position system would allow a more comprehensive investigation of this relationship to take place. For example, in the model, due to a lack of available data it was assumed that individuals waited six months between diet attempts. Similarly, in order to calibrate the model, multiple data sources had to be combined in order to generate targets against which the model could be compared and income targets were generated using summary statistics of mean wages rather than being based on individual incomes. Data for the other targets were also drawn from an unrepresentative sample of the UK population meaning any model conclusions would have perhaps not been generalisable.

Resolving these data scarcities would put the model development on firmer ground. Basing parameters on data rather than assumptions gives the model additional credibility and reduces the need to verify the impact of model assumptions on the model results (Casini and Manzo, 2016; Fagiolo, Moneta and Windrum, 2007; Hazelbag et al., 2020). Similarly, being able to derive calibration targets from a rich dataset of representative individuals rather than piecing them together from multiple sources provides flexibility in the kinds of model targets that can be calibrated to (Liu et al., 2017b), and one

can also have greater confidence that the data generating mechanisms in the model match those that generated the targets (Moss, 2008).

7.7 Conclusions

In the last two chapters I have presented the iterative development of an Agent-based Model that aimed to explore the role of stigma in generating the relationship between BMI and both depression and socioeconomic position. Despite combining data from both literature and real world data into the development and analysis of the model, the final model was unable to answer its intended research questions due to concerns about how well the model was representing the underlying real world system.

Despite this, the modelling process has still provided useful insights that can be used to aid future research. Practically speaking, it has highlighted one of the key challenges of applying Agent-based Modelling in research, specifically, the challenge of dealing with the inherent flexibility that the method provides. The flexibility of Agent-based Modelling is often cited as an advantage of the method, as one can use it to investigate a wide variety of phenomena (Bonabeau, 2002). However, with that flexibility comes the challenge of finding adequate data and evidence to inform the model description (Casini and Manzo, 2016; Moss, 2008; Liu et al., 2017b). Such data is not always available, and assumptions and simplifications become necessary to fill the gaps.

Making such assumptions and simplifications makes the already difficult task of satisfying generative sufficiency even more so. Myriad different choices could be made in the face of such uncertainty, each of which will have different implications for the dynamics of the model. As such, finding a model specification that fits the real world data in the face of such uncertainties presents a significant challenge to investigators using Agent-based Modelling in their research (Casini and Manzo, 2016; Fagiolo, Moneta and Windrum, 2007). This is perhaps particularly so in a model such as the one presented here that aims to look at multiple outcomes, as each output adds more complexity into the challenge of attaining generative sufficiency.

With this in mind, the model building process also highlighted some particular areas of research that could be examined to reduce the need to rely on assumptions and simplifications such as increased research on the phenomonology of obesity stigma, the formation and maintenance of socially

held body-size ideals and habits around frequency of dieting. Conducting such research would allow future Agent-based Models of obesity stigma to begin on a stronger platform, with less uncertainty surrounding how to appropriately describe the mechanisms present in the underlying system.

Chapter 8

Discussion

In this thesis, I have aimed to quantitatively explore the relationship between obesity and both depression and socioeconomic position, focusing on investigating mechanisms that had been hypothesised as important but had not yet been explored in detail. My review of the literature in Chapter 2 highlighted a range of mechanisms that might be important in the relationship between obesity and depression, including obesity stigma, poor physical function, diet and physical activity. It was also noted that despite these hypotheses there was very little causal evidence supporting the existence of these mechanisms; most of the studies presented relied on cross-sectional data and those that did examine longitudinal data stopped short of examining any causal implications.

As a result of this, in the remainder of the thesis, I have built upon this literature by exploring the effects of these mechanisms and how they might be related to socioeconomic position using two different modelling techniques: Structural Equation Modelling and Agent-based Modelling. The specific details of these models and their results have already been discussed in the previous chapters, hence, here the discussion will take a different focus. Firstly, I will breifly summarise what these studies showed and what they add to the literature more broadly. The remainder of this chapter I will then discuss what causal information these studies can tell us about the mechanisms in the obesity-depression-socioeconomic position system, and how additional modelling work using these methods could supplement what I have produced in this thesis.

In Chapters 3, 4 and 5 I analysed a Structural Equation Model that explored whether the relationship between obesity and depression was mediated by physical function, diet and physical activity. Additionally I explored whether socio-economic position influenced the obesity-depression relationship via

effects on these mechanisms. Overall, the study found little evidence to support the existence of a relationship between obesity and depression, either directly or through the hypothesised mechanisms.

This study has contributed to the literature surrounding the obesity-depression relationship in three main ways. Firstly, the findings have contributed to the literature by expanding on knowledge of what mechanisms might obesity and depression. Specifically, the findings suggest that should obesity be related to depression via diet, physical activity and physical function, that these mechanisms may operate through different parts of these constructs than those that I have examined in this study. For example, in the case of diet, evidence presented by Markowitz, Friedman and Arent (2008) and Preiss, Brennan and Clarke (2013) in Chapter 2 suggested eating behaviours such as repeated dieting and binge eating may be important in the relationship between obesity and depression. Taken together with the lack of association found in my study, this may imply that any diet mediated relationship between obesity and depression operates via these constructs rather than consumption itself. Alternatively, the relationships may follow a different structure to the linear patterns I have examined. These implications are discussed in more detail in section 8.2.

Secondly, it has expanded on the evidence analysed in Chapter 2 by using longitudinal data to analyse mechanisms that had previously been investigated in primarily cross-sectional studies. Studies presented in Hoare et al. (2014), Preiss, Brennan and Clarke (2013) and Markowitz, Friedman and Arent (2008) all provided evidence that poor diet and eating behaviours were associated with depression, but the evidence was not sufficient to conclude on the direction of association. Similarly Markowitz, Friedman and Arent (2008) provided evidence that physical function issues in the obese might lead to depression, however, this assertion was not based on longitudinal data. Where longitudinal evidence was presented, as in the case of physical function, it was often limited in scope such that only one part of a mechanism had been investigated (Hoare et al., 2014). As a result, this thesis has begun to fill the gap in longitudinal evidence so that a better evidence base will be available for making conclusions on the direction of association within the relationship between obesity and depression.

Thirdly, this study has also expanded on the literature presented in Chapter

2 by examining mechanisms in their entirety, rather than relying on an amalgamation of different studies to hypothesise existence of a mechanism. Authors of the reviews presented in Chapter 2 often hypothesised the existence of mechanisms between obesity depression by amalgamating results from studies investigating separate parts of a mechanism. For example, Markowitz, Friedman and Arent (2008) based their assertions that physical function might relate obesity to depression using separate observations about the relationship between physical function and both obesity and depression. Similarly, longitudinal data presented by Hoare et al. (2014) to suggest that physical activity was associated with reduced risk of future depression did not also include an examination of how obesity or body weight affect physical activity. As a result, this thesis has expanded on the current literature by developing the evidence base in which mechanisms have been explored in full.

In Chapters 6 and 7, I then presented the iterative development of an Agent-based Model that explored the role of obesity stigma in generating observed relationships between BMI and both depression and income.

This model has contributed to the literature in two main ways. Similarly to the SEM study, this study has built on previous literature by investigating previously hypothesised mechanisms relating obesity, depression and socioeconomic position that as of yet had not been investigated in detail. Much of the evidence surrounding the impact of obesity stigma reviewed in Chapter 2 and Chapter 6 was based on examinations of individual obesity stigma effects (Markowitz, Friedman and Arent, 2008; Puhl and King, 2013); few studies to date have attempted to empirically examine multiple potential outcomes of obesity stigma simultaneously. Instead separate studies have been drawn together to conceptualise models of obesity stigma that as of yet have not been fully empirically tested (Brewis, 2014; Tomiyama, 2014). Overall, the development of my Agent-based Model has highlighted the challenges associated with empirically testing these proposed models, such as creating a reliable model in spite of literature gaps and finding adequate data for empirical calibration and validation.

The development of my Agent-based Model has also contributed to the expanding use of Agent-based simulation within obesity research. Previous Agent-based models within obesity research have mostly focused on analysis of food and activity environments, social network influences of obesity

and physiology (Morshed et al., 2019). Those that have explored the relationship between obesity and depression have focused singly on depression outcomes, without examining potential feedback loops to obesity, or additional effects on socioeconomic position (Mooney and El-Sayed, 2016). Hence, to my knowldege the models presented in my thesis are the first models to examine the potential for feedback effects to exist between BMI and obesity stigma as well as impacts on depression and socio-economic position.

Having described the general contributions of this thesis to the literature, the remainder of this chapter will focus on discussion what information my SEM and ABM studies can give about causality in relationships between obesity depression and socioeconomic position.

8.1 Causality in SEM and ABM

Developing causal evidence is often seen as the principal aim in science (León-Medina, 2017). However, despite the interest in developing causal explanation, there are differing accounts as to what constitutes a causal explanation, and the notion of causality is sometimes criticised for being vague and unspecific (Cartwright, 2004). Causal relationships can be conceived of in two different ways which are important for how Structural Equation Models and Agent-based Models approach causality.

The first way that causality can be conceived is through 'dependence' accounts of causality (Hall, 2004). The intuition behind this account of causality is that a cause is such that a difference in the cause results in a difference in the effect (Casini and Manzo, 2016; Hall, 2004). This account often aligns closely with the counterfactual framework of causality, and experiments that attempt to manipulate a suspected cause to observe whether this is associated with a different outcome (Morgan and Winship, 2014). In the counterfactual framework of causality, "an event A may be considered a cause of event Y if, contrary to fact, had A not occurred then Y would not have occurred" (Arnold et al., 2019). As such, within this account of causality, a caused outcome can often be seen as 'depending' on the occurrence of the causing event.

Another example of the dependence account of causality in research is the view of causation as robust dependence (Goldthorpe, 2001). Within this, it is recognised that whilst association is not sufficient to imply there is a causal effect, the presence of a causal effect will itself result in an association. As

such, one argues that X is a genuine cause of Y if the association is robust such that the introduction of other variables into the analysis does not eliminate the association between X and Y (Goldthorpe, 2001).

The second way in which causality can be conceived is through 'production' accounts (Hall, 2004). The intuition behind this account of causality is that a cause generates its effects (Casini and Manzo, 2016). Counterfactual dependence may still apply in this account of causality, however, it is not necessary for an outcome to be dependent on the event that caused it. For example, Paul and Hall (2013) present a hypothetical situation in which two children, Suzy and Billy, both throw rocks at a bottle and have perfectly accurate throws. Suzy's rock hits first as she throws her rocks faster, breaking the bottle, and so 'produces' the caused effect. However, had she not thrown, Billy's throw would still have broken the bottle and so the bottle being broken does not depend on the cause that 'produced' it.

Another example that highlights the distinction between production and dependence accounts of causality is in considering rainy weather as a cause of forest fire (Hall, 2004). Clearly, under the production account, we would say that rain is not a cause of forest fires as it does not help produce a fire, however, under dependence accounts we might argue that they do, due to the fact that forest fires robustly depend on rainfall.

Causality is often examined by searching for mechanistic explanations of observed phenomena (Hedström and Ylikoski, 2010). Up until now, in this thesis the concept of a mechanism has been talked about loosely, however, in order to aid the causal interpretation and discussion of my models' results, it is necessary to more precisely define what is meant by a mechanism within the context of my research methods. In light of this, before discussing the causal implications of the studies presented in this thesis, first I will briefly outline how Structural Equation Models and Agent-based Models represent and examine mechanisms and causality.

Similarly to causality, there are multiple approaches to describing what constitutes a mechanism. Two approaches that will be of particular use to this discussion are the 'horizontal' and 'vertical' views of mechanisms (Casini and Manzo, 2016). In the horizontal view, mechanisms are interpreted as sequences of variables that have robust relationships between them (Casini and Manzo, 2016; Woodward, 2002). This view of mechanisms often coincides with the dependence account of causality in that, the presence of robust relationships between variables such that differences in one variable

leads to differences in another gives supporting evidence for the presence of a causal mechanism (Woodward, 2002; Morgan and Winship, 2014; Casini and Manzo, 2016). Returning to the forest fire example from earlier, differences in rainfall could be viewed as a horizontal causal mechanism of forest fires in that the amount of rainfall a forest receives will be robustly related to the occurrence of forest fires (Hall, 2004).

In the vertical view of mechanisms, a mechanism is envisaged as a complex system that consists of entities and interactions that over time combine to generate some behaviour of the system (Glennan, 2002). This view of mechanisms coincides clearly with the production account of causality such that when a certain collection of entities and interactions that represents a mechanism generates a behaviour in the system, this mechanism can be considered a cause of the system behaviour (Casini and Manzo, 2016; Hedström and Ylikoski, 2010). Again within the forest fires example, a vertical view of the causal mechanisms may look at human behaviours (such as having camp fires and barbecues), and their interaction with periods of dry weather and characteristics of the forest ecosystem as describing a causal mechanism of forest fires (Hall, 2004).

Each of these accounts of causal mechanisms have importance in this thesis. Structural Equation Modelling studies generally examine mechanisms in the horizontal view and within a dependence account of causality (Goldthorpe, 2001; Casini and Manzo, 2016). For example, in line with the intuition of the dependence account of causality, the SEM presented in Chapter 5 examines whether inter-individual differences in one construct cause inter-individual differences in another construct.

When specifying a Structural Equation Model, the structural portion of the model encodes a combination of strong and weak causal assumptions (Bollen and Pearl, 2013). Strong causal assumptions are represented by the omission of a path in the model, implying that there is no effect of one construct on the other. Weak assumptions are then represented by the included paths in the model, which imply that there is some effect of one construct on another. If the specified model fits well to the data, this gives credibility to the causal assumptions used to construct it (Bollen and Pearl, 2013), however, if the model fits poorly, this suggests that there are weaknesses in the causal assumptions encoded in the model.

The counterfactual framework for causality also features heavily within the causal analysis of Structural Equation Models. Many Structural Equation

Models, including the Cross-Lagged Panel Model presented in Chapter 5, can be represented using a Directed Acyclic Graph (DAG) (Arnold et al., 2019; Tu, 2012). A DAG is a graphical causal model containing nodes and edges in which nodes represent variables and edges represent direct causal relationships between them (Shrier and Platt, 2008). These edges are uni-directional, and no path exists in the graph from a node back to itself.

DAGs can be used to estimate counterfactual quantities from observed data by highlighting which variables are sufficient to remove biases in causal effects due to confounding (Arnold et al., 2019). Assuming that all common causes $\mathbf{Z}$ of a factor X and an outcome Y are represented in a DAG, then counterfactual quantities can be estimated directly by regressing Y on both X and $\mathbf{Z}$, since $\mathbf{Z}$ are sufficient for removing bias from confounding (Arnold et al., 2019). A representation of this using a linear regression is displayed in Equation 8.1 below:

$$Y = \beta_0 + \beta_X X + \beta_{\mathbf{Z}} \mathbf{Z} + \epsilon. \tag{8.1}$$

In this regression model, the beta co-efficient for X, β_X is an estimate of the total causal effect of X on Y such that, for individuals with the same values of $\mathbf{Z}$ a unit difference in X is expected to cause a difference in Y of size β_X, on average (Arnold et al., 2019).

Agent-based models then examine mechanisms in the vertical view, in line with a production account of causality (Casini and Manzo, 2016). When macro-level patterns of interest are 'produced' by the behaviours and interactions of micro-level entities in the model, this is taken as evidence that the micro-specification is a candidate cause of the macro-level phenomena (Epstein, 1999). Dependence type causal relations may also feature in an Agent-based Model within the micro-level specification, however these are not usually of primary interest in the model (Casini and Manzo, 2016).

Analysis of counterfactuals can also be conducted within an agent based model. Specifically, the modeller can alter features of the model such as parameter input values or agent interactions between simulation runs to examine how the model output differs under these alternative scenarios. Provided the initial population remains unchanged in each simulation, this provides the exchangeable units required for a counterfactual analysis (Arnold et al., 2019).

Within the production account of causality, Casini and Manzo (2016) outlines the conditions under which Agent-based Models can give causally relevant

information. Primarily, this is based on three main conditions. Firstly, the micro-specification must be built on solid theory. In other words, the conceptualisation of the agents in the model, their behaviours and interactions should be backed up by evidence, rather than simply being based on the modeller's intuition alone. Secondly, the macro-patterns produced by the model should be compared systematically with quantitative data from the real-world system and thirdly, empirical information should be used to calibrate the model's low level specification. For example, characteristics of the agents should, where possible, be derived from empirical data rather than sampled at random from an assumed distribution.

Given the flexibility of Agent-based Models, fully satisfying these conditions is a challenge as finding available data for the many model designs that Agent-based Models are capable of investigating is likely not possible (Casini and Manzo, 2016). This limits our ability to always fully calibrate every aspect of a model. However, theoretical explorations of the model dynamics can still somewhat make up for this inability to base every aspect of the model on empirical evidence (Casini and Manzo, 2016). For example, sensitivity analysis can estimate how changes to the parameter space influence the output of the model (Thiele, Kurth and Grimm, 2014), and dispersion analysis can investigate how much stochasticity in the model influences the output when the same parameters sets are run multiple times (Manzo, 2013). The above can be used to assess the robustness of the model output and hence the causal information it can provide.

Having outlined the way in which Structural Equation Models and Agent-based Models approach causal mechanisms, we can now examine what causal information can be extracted from the studies presented in this thesis.

8.2 Causal Implications of my SEM Study

The hypotheses outlined at the beginning of Chapter 5, represented diagrammatically in Figure 5.1, can be translated into a collection of equivalent causal assumptions which are articulated below. Note that in each of the assumptions listed below differences are between individuals, rather than within individuals, in line with the description of the Cross-Lagged Panel Modelling method from Chapter 5. As such, when we state that differences in one construct are a cause of differences in another, we are in fact stating that differences between individuals' levels on the causing construct lead to differences

between individuals on the caused construct (Selig and Little, 2012).

Another way of interpreting these assumptions is to say that if there were no differences between individuals on the causing construct, then (assuming all other causes have been accounted for), there would be no differences between individuals on the caused construct (Pearl, 2001). This interpretation more clearly lines the assumptions up with the counterfactual descriptions of causality.

1. **Weak Assumptions:**
    - Differing obesity is a cause of differing future depression, physical function, calories consumed and physical activity performed.
    - Differing depression is a cause of differing future obesity, calories consumed and activity performed.
    - Differing physical function is a cause of differing future depression
    - Differing calories consumed is a cause of differing future depression and obesity
    - Differing physical activity is a cause of differing future depression and obesity
    - Differing socioeconomic position is a cause of differing obesity, depression, physical function, physical activity and calories consumed
    - Differing levels of each construct is a cause of differing future levels of the same construct
2. **Strong Assumptions:**
    - Differing depression does not cause differing physical function
    - Differing physical function does not cause differing future obesity
3. **Other assumptions:** (see Bollen and Pearl (2013))
    - All causal effects are linear
    - Effects are exclusive. This means that when holding a 'causing' variable constant, changes in other variables do not impact the value of the 'caused' variable (unless it is itself another cause of that variable)

- All effects are homogeneous i.e all individuals in the population are subject to the same causal effect

In addition to the above assumptions, these causal effects were not assumed to be constant over time. Within the model this assumption was implemented by allowing the parameters that described relationships between constructs across different time points to take different values during the model estimation. In other words, the path coefficients for arrows in Figure 5.1 were not constrained to be equal over time during the model estimation. For example, the effect of BMI at time 1 on depression at time 2 was not constrained *a priori* to be equal to the effect of BMI at time 2 on depression at time 3.

Overall the above assumptions mean that for each individual, having a different value to others on the causing construct leads to them having a different value to others on the caused construct (all else being equal), and that the impact of this difference is the same for all individuals in the population (due to the assumption that the causal effects are homogeneous) (Arnold et al., 2019; Morgan and Winship, 2014; Bollen and Pearl, 2013). This expected effect of one construct on another is analogous to the counterfactual causal effects introduced in Equation 8.1.

These assumptions are subtly different to a description of causal effects based on intra-individual differences, which state that a change within an individual on a construct leads to a change within the same individual on another. Other SEM methods, such as Random-Intercept Cross-Lagged Panel Models are able to incorporate this kind of causal assumption into the model, however, the standard Cross-Lagged Panel Model is not (Mulder and Hamaker, 2021).

Each of the model specifications that I analysed in Chapter 5 showed good fit to the data overall, giving credibility to the causal assumptions used in the models' formulations. However, the size of effects in the models suggested there was little evidence to support any of the hypotheses the model represented. This was evidenced by the presence of very few 'important' paths being highlighted between different constructs in the model, and those that were highlighted rarely had a consistent effect over time. As a result, the good model fit may simply be due to the inclusion of the weak assumption that causal effects can be different over time. Strengthening this assumption to demand that the effects be the same over time may result in a poor model fit, which would then be more consistent with the results implied by the path coefficients.

The inconsistency in the path coefficients over time in the models is perhaps evidence that the hypothesised mechanisms are not causal mechanisms that relate obesity and depression, at least in the population I have examined. In this SEM we are searching for evidence using the dependence account of causality, which requires the identifying of robust relationships between variables (Woodward, 2002; Morgan and Winship, 2014). The relationships found in this study were not robust over time and as such are perhaps too spurious to be considered evidence of a causal mechanism.

This does however, present the opportunity for some causal assumptions to be strengthened in future examinations of the relationship between obesity and depression, in that *a priori* one could now remove these paths from the structural model to make the stronger assumption that differences in the included constructs do not cause differences in the other constructs.

Additionally, we may wish to challenge the additional causal assumptions implied by the model structure. In particular, we may wish to challenge the assumptions of linearity and homogeneity. With regards to linearity, if any true causal effect was severely non-linear, a linear representation may be unable to properly detect such an association. For example, supposing that there is a value of BMI, τ, below which inter-individual differences in obesity do not cause inter-individual differences in depression, but that having a BMI above τ does cause a difference in depression of size β. Then, depending on the size of β and the position of τ on the underlying BMI distribution, assuming a linear causal effect might have the impact of underestimating the effect β, since its effect will be averaged out over the whole BMI range which includes sections where there is no causal effect. In light of this, future studies may wish to examine the possibility for the hypothesised mechanisms to contain non-linear effects.

The second additional causal assumption that could be challenged is that of homogeneity of causal effects. In my Structural Equation Model, the causal effects are assumed to be homogeneous across the units of the study. In my review of the literature in Chapter 2, it was suggested that there may be significant heterogeneity within the relationship between obesity and depression. In particular, 12 reviews found evidence of heterogeneity in either methodology or effects (Atlantis and Baker, 2008; de Wit et al., 2010; Faith et al., 2011; Jung et al., 2017; Korczak et al., 2013; Luppino et al., 2010; Mannan et al., 2016b; Mannan et al., 2016a; Mühlig et al., 2016; Pereira-Miranda et al., 2017;

Quek et al., 2017; Rooke and Thorsteinsson, 2008). This heterogeneity is potentially influenced by heterogeneity within the obese population (Green et al., 2015), differences between those with depression (Fried and Nesse, 2015), and by the large number of potential mechanisms and risk factors that relate the two conditions.

I attempted to account for this by running the model in multiple groups that separated out participants by sex and ethnicity. Despite this, additional heterogeneity may exist within the included participants that results in heterogeneous causal effects being present within the population. If in some groups there are no causal effects, whilst in others there is a causal effect, this could have the impact of attenuating the average causal effect estimated in the groups combined. For example, it is possible that subgroups in the population of individuals with obesity may have different exposures to the mechanisms that might lead to depression (Green et al., 2015).

My review of the literature, however, did not find strong evidence for groupings beyond those I have studied that could be responsible for such heterogeneity of effects. Similarly, conducting a clustering analysis such as that conducted by (Green et al., 2015) to search for additional subgroups was beyond the scope of this thesis. As a result, studies examining in detail under what conditions, and in which people obesity and depression are related would help to improve our understanding of the relationship between the two conditions and provide firmer ground upon which causal mechanisms could be investigated.

The lack of support for a causal mechanism provided by my SEM study may also suggest that any mediating effect of diet, physical activity and physical function between obesity and depression comes via different components of these constructs than I have investigated in this thesis. For example, in the case of diet, disordered eating and repeated dieting were both hypothesised to be potential mechanisms that might relate obesity and depression (Markowitz, Friedman and Arent, 2008; Preiss, Brennan and Clarke, 2013). As such, future SEM studies could examine to what extent obesity and depression are robustly related to disordered eating over time, and as such gain evidence as to whether obesity is a cause of disordered eating which in turn causes depression or vice versa (or both).

Similarly, of physical acitivity presented in 2 were often conducted in treatment seeking populations, or have often focused on aspects of physical activity such as sports club participation (Hoare et al., 2014). Once again, this

may suggest that any relationship between obesity and depression via physical activity may operate through the social aspects of exercise, rather than simply being a function of how much activity individuals partake in.

However, in order to address such questions, additional data on these concepts, collected regularly on a consistent cohort of individuals is needed. Within this thesis, concepts such as disordered eating and repeated dieting were not examined due to lack of data availability and so filling this data gap is a pre-requisite to exploring this potential causal mechanism using SEM.

8.3 Causal Implications of my ABM Study

As with the Structural Equation Modelling study, within the design and analysis of the Agent-based Models presented in Chapters 6 and 7, multiple causal assumptions were encoded. Overall, since this modelling method subscribes to a production account of causality, the causal assumptions of the model are expressed in terms of how features and outputs of the model are generated by entities and interactions (Casini and Manzo, 2016; Hall, 2004).

By calibrating to the mean BMI, GHQ-based depression and income, the models have effectively examined whether the described mechanisms included in the models are a cause of average levels of obesity, depression and income in the population. Examining whether the model could generate realistic patterns of these variables was a necessary prerequisite to exploring the generated relationships between them (Epstein, 1999). The relationship between two variables will be contingent on the patterns of the individual variables themselves generated by the model. Hence, if these generated patterns are unrealistic this will in turn cast doubt on the relationship between the variables that has been produced by the model.

Within the description of the model, two broad categories of mechanisms were described. The first category described obesity stigma and its various impacts, and the second category described ways in which non-stigma dynamics contribute to the outcomes. In describing how these mechanisms might be causes of population level characteristics, a number of causal claims were made. The first, and perhaps most important, is that vulnerability to obesity is entirely caused by deviations from a socially defined ideal body size. The omission of other factors implicitly states that they have no causal effect. Similarly, it was also claimed that the sole cause of engaging in weight

control behaviour is deviation from the ideal body size, and that the obesogenicity of the environment is the sole cause of how weight promoting individuals lifestyles are.

In addition to the above assumptions about causal pathways, we have also made claims about the homogeneity of causal effects in the population. In general, effects within the model were homogeneous across individuals unless specifically stated otherwise. For example, it was claimed that the impact of obesity stigma was homogeneous across the population, such that two individuals who receive the same amount of obesity stigma would suffer the same impact.

Having outlined the causal claims and assumptions that were embedded into the model design, we can now discuss what the model can tell us about these causal claims. In both of the models presented, there were significant doubts over their generative sufficiency, suggesting that the mechanisms as described in the models are not candidate causes of macroscopic patterns of BMI, depression and income in the real world. The model presented in Chapter 6, in which only obesity stigma was represented in explicit detail struggled to generate realistic macroscopic patterns in any simulation, suggesting that on its own, obesity stigma is not a cause of these macroscopic patterns. Including additional non-stigma based causes into the model improved the generative sufficiency, such that the model was now able to generate more realistic macroscopic patterns for each target separately. However, the model still struggled to generate all three macroscopic patterns simultaneously, and there were also qualitative features in the individual trajectories that cast doubt on the accuracy of the model as a representation of the real world system.

Taken literally, this could suggest that the model specification describes a candidate causal explanation for some of the targets, but not all. For example, the specification may accurately capture the causal process through which obesity and depression patterns are generated, but not the pattern of income, or vice versa. However, what is more likely, is that the model has highlighted the need for a more nuanced representation of all the causal processes included in the model.

As mentioned above, in describing the model we made a collection of strong causal claims. The inability for the model to generate multiple macroscopic patterns in the same simulation may indicate that these causal claims are too strong, and need to be weakened or adapted in order to be able to explain

more system features simultaneously. For example, the model supposed that the extent of obesity stigma individuals are subjected to, and hence the effects they receive from it, are caused solely by how much the individual differs from the ideal body size. This ideal body size was also static throughout the simulation. Hence, the model results potentially imply that, if stigma is a contributory cause of macroscopic patterns of obesity, depression and income in the real world, that the extent of stigma received and the effects it causes depends on more than just deviation from an ideal body size; other factors also impact upon this.

Further to this point, Additional causal effects may also need to be included in the model. Authors researching obesity stigma have examined links between individuals perceptions of obesity stigma (Lewis et al., 2011; Jackson and Steptoe, 2017) as well as the different methods individuals use to cope with stigma when it is experienced (Spahlholz et al., 2016; Puhl and Brownell, 2003). Overall, individuals may perceive and cope with stigma in a variety of ways, which may have implications for the impact stigma has on their mental health (Puhl and Brownell, 2003). Different coping strategies may have different implications for whether individuals internalise stigma or not, which has been shown to be particularly harmful to mental health (Pearl and Puhl, 2016). Whilst examining these factors was beyond the current scope of my model, examination of these effects in future models may add important nuance to the causal descriptions, allowing the models to more adequately reproduce observed data.

Within my Agent-based Modelling study, I tried to satisfy the conditions under which an ABM can produce reliable causal information as much as possible (Casini and Manzo, 2016). Firstly, the model structure was designed with reference to the underlying literature on obesity stigma, its effects and how body image ideals may influence the likelihood of receiving stigma. Secondly, the macroscopic BMI, depression and income patterns generated in the model were compared systematically with targets based on data from the real-world system. Thirdly, where possible, the microlevel specification of the model was calibrated to real world data. For example, the rate of inflation was derived from data on the average weekly wage and pensions received in the population, and agents initial values for BMI, depression, income, age and sex were all derived from data.

Additionally, the model output was explored via sensitivity analysis, and the models' output trajectories and posterior distributions were scrutinised

thoroughly in order to understand what features of the model description were responsible for its behaviour. For example, in the simulations displayed in Figure 7.4, the model internals were examined to see if the initial steep rises in mean BMI could be explained.

The above process gives credibility to the model's ability to provide causal information (Casini and Manzo, 2016). Specifically, it gives confidence to the conclusions one can draw about the causal claims that are made by the model's design. As such we can be confident that the causal claims encoded in the model currently are insufficient to explain population trends in average obesity, depression and income.

8.4 Combining Research from ABMs and SEM to Investigate Causality

Whilst both Structural Equation Models and Agent-based Modelling can shed light on causal relationships between obesity, depression and socioeconomic position, each has its limitation on what causal information it can provide. Within the dependence account of causality used in Structural Equation Modelling, claims can be made about to what extent differences in a cause lead to differences in an effect (Woodward, 2002; Goldthorpe, 2001). However, some critics of this account of causality argue that this does not give an explicit account of how these differences are generated (Hedström and Ylikoski, 2010). Whilst mediation studies can perhaps add additional detail to the collection of variables that describes the causal mechanism, the question of how one variable causes differences in the next variable in the chain still holds.

Similarly, whilst Agent-based Models more explicitly represent how a phenomena is caused by low level entities and interactions, it is often more difficult to fully calibrate and validate the model using empirical evidence (Casini and Manzo, 2016; Fagiolo, Moneta and Windrum, 2007). The flexibility of the method allows complex models to be created, which can result in assumptions and relationships being included in the model for which data is not necessarily available. Such challenges can undermine the ability of an Agent-based Model to provide causal evidence (Casini and Manzo, 2016).

It has been argued that Agent-based modelling and data-driven methods such as SEM can be used conjointly to improve the quality of causal evidence available (Casini and Manzo, 2016). For example, claims implied by

Agent-based modelling studies could be examined by data driven studies, and conversely, Agent-based Models could be used to examine in more detail how causal relationships highlighted in Structural Equation Models might be generated.

In this thesis, separate mechanisms have been investigated using separate modelling techniques and as such further research may benefit from examining how the alternative modelling technique could improve our understanding of the hypothesised causal mechanisms examined in this thesis. In particular, the mechanisms examined in my Structural Equation Modelling study could be examined within an ABM such that intra-individual differences rather than just inter-individual differences could be examined. In the case of physical function, one would be able to explore whether increases in an individual's level obesity leads to increases in their depression via a change to their physical function, and in turn whether such intra-individual differences are able to generate population level associations between the diseases.

ABM studies such as those proposed above could also be supported by additional SEM work that is capable of examining intra-individual differences such as Random-Intercept Cross-lagged Panel Modelling (Mulder and Hamaker, 2021). The result of this is that we would able to say more than just statements about whether differences in obesity lead to differences in depression in the population. We would also be able to explain more about the process within individuals that leads to this difference.

As has been noted before, the flexibility of Agent-based Modelling allows one to incorporate heterogeneity and non-linearity into the model in a way that is more straightforward than is the case in SEM (Bonabeau, 2002; Gilbert, 2008). Hence, supplementing my SEM work with Agent-based Modelling would also present the opportunity relax the causal assumptions of linearity and homogeneity that are imposed within the SEM model.

Similarly, Structural Equation Modelling or a similar data driven method that relies on a dependence account of causality could supplement the analysis of obesity stigma I conducted in Chapters 6 and 7. First and foremost, analogously to the study presented in chapter 5, SEM models could be used to investigate whether differences in obesity are a cause of differing levels of received obesity stigma, and whether this in turn is a cause of differing levels of depression. Some studies have attempted to do this in the case of internalised obesity stigma (Decker, Thurston and Kamody, 2018; Spahlholz et al., 2016),

however these studies are frequently limited to analyses of cross-sectional data, restricting the causal interpretation of results.

Second, individual causal pathways included in the conceptual model could be examined in order to improve the evidence base informing these pathways. For example, Structural Equation Modelling could be used to examine the components of body image that are causally relevant for obesity stigma. Within this, SEM could be used to address some of the gaps in the literature highlighted in Chapters 6 and 7. For example, one could investigate to what extent differing deviations from an ideal body size in males and females has on the amount of stigma individuals receive, and whether this impacts on individuals' mental health.

Additionally, SEM studies could be used to examine whether stigma effects depend on the stigma context or not. Stigma is a complex phenomenon with multiple variants and contexts within which it can operate (Pescosolido and Martin, 2015; Vartanian, Pinkus and Smyth, 2014; Puhl and King, 2013). However, as was mentioned in Chapter 7, the current literature did not provide conclusive evidence of general stigma effects. Instead, research has largely explored context specific effects (Puhl and King, 2013; Godfree, 2020; Puhl, Moss-Racusin and Schwartz, 2007; Fikkan and Rothblum, 2012). As a result structural equation modelling could be used to explore whether differences in obesity stigma lead to differences in mental health outcomes in contexts such as the workplace and education.

Overall, Agent-based Models and Structural Equation models provide different but complementary causal evidence (Casini and Manzo, 2016; Bollen and Pearl, 2013). Therefore, despite the lack of support for causal relationships found in my own studies both of these methods still have the potential to improve understanding on the mechanisms that relate obesity and depression.

8.5 Concluding Remarks

This thesis has built upon existing knowledge of the relationship between obesity and depression by examining causal evidence for proposed mechanisms that might relate the two conditions. Specifically, I have examined the role of physical function, physical activity, diet and social stigma as mechanisms that relate obesity with depression. Additionally, we have examined how socioeconomic status interacts with these proposed mechanisms.

Overall, the models presented have not provided evidence that can support the existence of these causal mechanisms, nor have they presented evidence of interactions with socioeconomic position. However, in the case of stigma, the model used did not reach a stage where it was sufficiently reliable to conclusively rule out stigma as a generative cause of relationships within the obesity-depression-socioeconomic system. As a result, this mechanism still has the potential to be of importance in the relationship between obesity and both depression and socioeconomic position. Furthermore, many additional proposed mechanisms between obesity and depression remained unexplored. Given the lack of evidence found in my studies, additional research exploring such mechanisms will likely be important in explaining the association between obesity and depression.

Bibliography

Abdullah, A. et al. (2010). 'The magnitude of association between overweight and obesity and the risk of diabetes: A meta-analysis of prospective cohort studies'. In: *Diabetes Research and Clinical Practice* 89.3, pp. 309–319. ISSN: 01688227. DOI: 10.1016/j.diabres.2010.04.012.

Abou Abbas, L. et al. (2015). 'Obesity and symptoms of depression among adults in selected countries of the Middle East: a systematic review and meta-analysis'. In: *Clinical Obesity* 5.1, pp. 2–11. ISSN: 17588103. DOI: 10.1111/cob.12082. URL: http://doi.wiley.com/10.1111/cob.12082.

Ainsworth, B. E. et al. (2011). '2011 compendium of physical activities: A second update of codes and MET values'. In: *Medicine and Science in Sports and Exercise* 43.8, pp. 1575–1581. ISSN: 01959131. DOI: 10.1249/MSS.0b013e31821ece12.

Ali, A. M. et al. (2011). 'Comparison of methods for handling missing data on immunohistochemical markers in survival analysis of breast cancer'. In: *British Journal of Cancer* 104.4, pp. 693–699. ISSN: 00070920. DOI: 10.1038/sj.bjc.6606078. URL: http://dx.doi.org/10.1038/sj.bjc.6606078.

Allison, P. D. (2003). 'Missing Data Techniques for Structural Equation Modeling'. In: *Journal of Abnormal Psychology* 112.4, pp. 545–557. ISSN: 0021843X. DOI: 10.1037/0021-843X.112.4.545.

American Psychological Association (2013). *Diagnostic and statistical manual of mental disorders*. Fifth edit. Arlington, VA: Arlington, VA : American Psychiatric Association, 2013. ISBN: 9780890425558.

Ammar, B. and Abdallah, K. (2011). 'Towards the Formal Specification and Verification of Multi-agent Based Systems'. In: *International Journal of Computer Science Issues* 8.4, pp. 200–210. ISSN: 1694-0784.

An, G. et al. (2009). 'Agent-based models in translational systems biology'. In: *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 1.2, pp. 159–171. ISSN: 1939-5094. DOI: 10.1002/wsbm.45. URL: http://link.springer.

com/10.1007/978-1-4614-8008-2{_}3https://onlinelibrary.wiley.com/doi/10.1002/wsbm.45.

Anderson, J. C. and Gerbing, D. W. (1988). 'Structural equation modeling in practice: A review and recommended two-step approach.' In: *Psychological Bulletin* 103.3, pp. 411–423. ISSN: 1939-1455. DOI: 10.1037/0033-2909.103.3.411. arXiv: /dx.doi.org/10.1111/j.1748-8583.2005.tb00154[http:]. URL: http://www.cob.unt.edu/slides/Paswan/BUSI6280/Anderson{_}Gerbing{_}1988.pdf{\%}5Cnhttp://www.fdewb.unimaas.nl/meteor/EDEN/MikeBrady/AndersonandGerbing1988.pdf{\%}5Cnhttp://doi.apa.org/getdoi.cfm?doi=10.1037/0033-2909.103.3.411http://doi.apa.org/getdoi.cfm?doi=.

Andridge, R. R. and Little, R. J. A. (2010). 'A Review of Hot Deck Imputation for Survey Non-response'. In: *International Statistical Review* 78.1, pp. 40–64. ISSN: 03067734. DOI: 10.1111/j.1751-5823.2010.00103.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1751-5823.2010.00103.x.

Angermeyer, M. C. et al. (2002). 'Depression and quality of life: results of a follow-up study.' In: *The International journal of social psychiatry* 48.3, pp. 189–199. ISSN: 0020-7640. DOI: https://dx.doi.org/10.1177/002076402128783235.

Arnold, K. F. et al. (2019). 'DAG-informed regression modelling, agent-based modelling and microsimulation modelling: a critical comparison of methods for causal inference'. In: *International Journal of Epidemiology* 48.1, pp. 243–253. ISSN: 0300-5771. DOI: 10.1093/ije/dyy260. URL: https://academic.oup.com/ije/article/48/1/243/5231935.

Asparouhov, T. and Muth, B. (2008). 'Auxiliary Variables Predicting Missing Data'. In: *Mplus Technical Appendices*, pp. 1–4. URL: https://www.statmodel.com/download/AuxM2.pdf.

Ata, R. N. and Thompson, J. K. (2010). 'Weight Bias in the Media: A Review of Recent Research'. In: *Obesity Facts* 3.1, pp. 41–46. ISSN: 1662-4033. DOI: 10.1159/000276547. URL: https://www.karger.com/Article/FullText/276547.

Atlantis, E. and Baker, M. (2008). 'Obesity effects on depression: Systematic review of epidemiological studies'. In: *International Journal of Obesity* 32.6, pp. 881–891. ISSN: 03070565. DOI: 10.1038/ijo.2008.54.

Auchincloss, A. H. and Garcia, L. M. T. (2015). 'Brief introductory guide to agent-based modeling and an illustration from urban health research'. In: *Cadernos de Saúde Pública* 31.suppl 1, pp. 65–78. ISSN: 0102-311X. DOI: 10.1590/0102-311X00051615. URL: http://www.scielo.br/scielo.php?script=sci{_}arttext{\&}pid=S0102-311X2015001300065{\&}lng=en{\&}tlng=en.

Bandura, A (1977). 'Self-efficacy: Toward a unifying theory of behavioral change.' In: *Psychological Review* 84.2, pp. 191–215. ISSN: 0033-295X. DOI: 10.1037/0033-295X.84.2.191.

Bargiotas, T. (2017). *The Aetiology of Depression*. Vol. 1. Oxford University Press. DOI: 10.1093/oso/9780198801900.003.0005. URL: https://oxford.universitypressscholarship.com/view/10.1093/oso/9780198801900.001.0001/oso-9780198801900-chapter-5.

Bartlett, J. W. et al. (2015). 'Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model'. In: *Statistical Methods in Medical Research* 24.4, pp. 462–487. ISSN: 14770334. DOI: 10.1177/0962280214521348. arXiv: 1210.6799.

Bentler, P. M. and Satorra, A. (2010). 'Testing model nesting and equivalence.' In: *Psychological Methods* 15.2, pp. 111–123. ISSN: 1939-1463. DOI: 10.1037/a0019625. arXiv: NIHMS150003. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdfhttp://doi.apa.org/getdoi.cfm?doi=10.1037/a0019625.

Beretta, L. and Santaniello, A. (2016). 'Nearest neighbor imputation algorithms: a critical evaluation'. In: *BMC Medical Informatics and Decision Making* 16.S3, p. 74. ISSN: 1472-6947. DOI: 10.1186/s12911-016-0318-z. URL: http://dx.doi.org/10.1186/s12911-016-0318-zhttp://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-016-0318-z.

Bianchini, F., Kaaks, R. and Vainio, H. (2002). 'Overweight, obesity, and cancer risk.' In: *The Lancet. Oncology* 3.9, pp. 565–74. ISSN: 1470-2045. DOI: 10.1016/S1470-2045(02)00849-5. URL: http://www.ncbi.nlm.nih.gov/pubmed/12217794.

Blaine, B. (2008). 'Does Depression Cause Obesity?' In: *Journal of Health Psychology* 13.8, pp. 1190–1197. ISSN: 1359-1053. DOI: 10.1177/1359105308095977. URL: http://journals.sagepub.com/doi/10.1177/1359105308095977.

Bollen, K. A. (2002). 'Latent Variables in Psychology and the Social Sciences'. In: *Annual Review of Psychology* 53.1, pp. 605–634. ISSN: 0066-4308. DOI: 10.1146/annurev.psych.53.100901.135239. URL: https://www.annualreviews.org/doi/10.1146/annurev.psych.53.100901.135239.

Bollen, K. A. and Pearl, J. (2013). 'Eight Myths About Causality and Structural Equation Models'. In: *Handbook of Causal Analysis for Social Research*. Ed. by S. Morgan. Handbooks of Sociology and Social Research. Dordrecht: Springer. Chap. 15, pp. 301–328. ISBN: 978-94-007-6093-6. DOI: 10.1007/978-94-007-6094-3. arXiv: arXiv:1011.1669v3. URL: http://link.springer.com/10.1007/978-94-007-6094-3.

Bonabeau, E. (2002). 'Agent-based modeling: Methods and techniques for simulating human systems'. In: *Proceedings of the National Academy of Sciences* 99.Supplement 3, pp. 7280–7287. ISSN: 0027-8424. DOI: 10.1073/pnas.082080899. arXiv: 1709.03423. URL: http://www.ncbi.nlm.nih.gov/pubmed/12011407http://www.pnas.org/cgi/doi/10.1073/pnas.082080899.

Bray, G. A. and Popkin, B. M. (1998). 'Dietary fat intake does affect obesity!' In: *The American Journal of Clinical Nutrition* 68.6, pp. 1157–1173. ISSN: 0002-9165. DOI: 10.1093/ajcn/68.6.1157. URL: https://academic.oup.com/ajcn/article/68/6/1157/4666148.

Brewis, A. A. (2014). 'Stigma and the perpetuation of obesity'. In: *Social Science & Medicine* 118.C, pp. 152–158. ISSN: 02779536. DOI: 10.1016/j.socscimed.2014.08.003. URL: http://dx.doi.org/10.1016/j.socscimed.2014.08.003https://linkinghub.elsevier.com/retrieve/pii/S0277953614005206.

Broom, I. (2006). 'Review: Thinking around abdominal obesity and cardiovascular risk'. In: *The British Journal of Diabetes & Vascular Disease* 6.2, pp. 58–61. ISSN: 1474-6514. DOI: 10.1177/14746514060060020301. URL: https://doi.org/10.1177/14746514060060020301.

Brunner, E. et al. (2001). 'Dietary assessment in Whitehall II: comparison of 7 d diet diary and food-frequency questionnaire and validity against biomarkers'. In: *British Journal of Nutrition* 86.3, pp. 405–414. ISSN: 0007-1145. DOI: 10.1079/BJN2001414. URL: https://www.cambridge.org/core/product/identifier/S0007114501002069/type/journal{_}article.

Carpenter, K. M. et al. (2000). 'Relationships between obesity and DSM-IV major depressive disorder, suicide ideation, and suicide attempts: Results from a general population study'. In: *American Journal of Public Health* 90.2, pp. 251–257. ISSN: 00900036. DOI: 10.2105/AJPH.90.2.251.

Cartwright, N. (2004). 'Causation: One Word, Many Things'. In: *Philosophy of Science* 71.5, pp. 805–820. ISSN: 0031-8248. DOI: 10.1086/426771. URL: https://www.journals.uchicago.edu/doi/10.1086/426771.

Casini, L. and Manzo, G. (2016). 'Agent-based models and causality: a methodological appraisal'. In: *The IAS Working Paper Series*, pp. 1–80. URL: http://liu.diva-portal.org/smash/get/diva2:1058813/FULLTEXT01.pdf{\%}0Ahttp://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-133332.

Chen, E. Y. and Brown, M. (2005). 'Obesity Stigma in Sexual Relationships'. In: *Obesity Research* 13.8, pp. 1393–1397. ISSN: 10717323. DOI: 10.1038/oby.2005.168. URL: http://doi.wiley.com/10.1038/oby.2005.168.

Chen, L. S. et al. (2000). 'Understanding the heterogeneity of depression through the triad of symptoms, course and risk factors: A longitudinal, population-based study'. In: *Journal of Affective Disorders* 59.1, pp. 1–11. ISSN: 01650327. DOI: 10.1016/S0165-0327(99)00132-9.

Chen, L. (2012). 'Agent-based modeling in urban and architectural research: A brief literature review'. In: *Frontiers of Architectural Research* 1.2, pp. 166–177. ISSN: 20952635. DOI: 10.1016/j.foar.2012.03.003. URL: http://dx.doi.org/10.1016/j.foar.2012.03.003https://linkinghub.elsevier.com/retrieve/pii/S2095263512000167.

Cheung, G. W. and Rensvold, R. B. (2002). 'Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance'. In: *Structural Equation Modeling: A Multidisciplinary Journal* 9.2, pp. 233–255. ISSN: 1070-5511. DOI: 10.1207/S15328007SEM0902_5. URL: http://www.tandfonline.com/doi/abs/10.1207/S15328007SEM0902{_}5.

Cohen, A. K. et al. (2013). 'Educational attainment and obesity: a systematic review'. In: *Obesity Reviews* 14.12, pp. 989–1005. ISSN: 14677881. DOI: 10.1111/obr.12062. URL: https://onlinelibrary.wiley.com/doi/10.1111/obr.12062.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd. Routledge, p. 567. DOI: 10.4324/9780203771587. URL: https://www.taylorfrancis.com/books/9780203771587.

Cohn, L. D. and Adler, N. E. (1992). 'Female and Male Perceptions of Ideal Body Shapes: Distorted Views Among Caucasian College Students'. In: *Psychology of Women Quarterly* 16.1, pp. 69–79. ISSN: 0361-6843. DOI: 10.1111/j.1471-6402.1992.tb00240.x. URL: https://doi.org/10.1111/j.1471-6402.1992.tb00240.x.

Cole, D. A., Ciesla, J. A. and Steiger, J. H. (2007). 'The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis.' In: *Psychological Methods* 12.4, pp. 381–398. ISSN: 1939-1463. DOI: 10.1037/1082-989X.12.4.381. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/1082-989X.12.4.381.

Cole, D. A. and Maxwell, S. E. (2003). 'Testing Mediational Models With Longitudinal Data: Questions and Tips in the Use of Structural Equation Modeling.' In: *Journal of Abnormal Psychology* 112.4, pp. 558–577. ISSN: 1939-1846. DOI: 10.1037/0021-843X.112.4.558. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0021-843X.112.4.558.

Collins, L. M., Schafer, J. L. and Kam, C.-M. (2001). 'A comparison of inclusive and restrictive strategies in modern missing data procedures.' In: *Psychological Methods* 6.4, pp. 330–351. ISSN: 1082-989X. DOI: 10.1037/1082-989x.6.4.330.

Cooney, G. et al. (2014). 'Exercise for depression'. In: *Advances in Psychiatric Treatment* 20.1, pp. 2–2. ISSN: 1355-5146. DOI: 10.1192/apt.20.1.2. URL: https://www.cambridge.org/core/product/identifier/S1355514600011305/type/journal{_}article.

Csilléry, K., François, O. and Blum, M. G. B. (2012). 'abc: an R package for approximate Bayesian computation (ABC)'. In: *Methods in Ecology and Evolution* 3.3, pp. 475–479. ISSN: 2041210X. DOI: 10.1111/j.2041-210X.2011.00179.x. arXiv: 1106.2793. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.2041-210X.2011.00179.x.

De Wit, L. et al. (2010). 'Depression and obesity: A meta-analysis of community-based studies'. In: *Psychiatry Research* 178.2, pp. 230–235. ISSN: 01651781. DOI: 10.1016/j.psychres.2009.04.015. URL: http://dx.doi.org/10.1016/j.psychres.2009.04.015.

Decker, K. M., Thurston, I. B. and Kamody, R. C. (2018). 'The mediating role of internalized weight stigma on weight perception and depression among emerging adults: Exploring moderation by weight and race'. In: *Body Image* 27, pp. 202–210. ISSN: 17401445. DOI: 10.1016/j.bodyim.2018.10.004. URL: https://doi.org/10.1016/j.bodyim.2018.10.004https://linkinghub.elsevier.com/retrieve/pii/S1740144517305454.

Department of Health (2011). 'No health without mental health: Delivering better mental health outcomes for people of all ages. Supporting document - The economic case for improving efficiency and quality in mental health'. In: February, p. 101. ISSN: 1474-547X. DOI: 10.1016/S0140-6736(07)61238-0. arXiv: NIHMS150003.

Dong, Y. and Peng, C.-Y. J. (2013). 'Principled missing data methods for researchers'. In: *SpringerPlus* 2.1, p. 222. ISSN: 2193-1801. DOI: 10.1186/2193-1801-2-222. URL: https://springerplus.springeropen.com/articles/10.1186/2193-1801-2-222.

Doshi, T. et al. (2016). 'Socioeconomic Status and Cardiovascular Risk Control in Adults With Diabetes.' In: *The American journal of the medical sciences* 352.1, pp. 36–44. ISSN: 1538-2990. DOI: 10.1016/j.amjms.2016.03.020. URL: https://linkinghub.elsevier.com/retrieve/pii/S0002962916302221http://www.ncbi.nlm.nih.gov/pubmed/27432033http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4955859.

Emery, E. M. et al. (1993). 'A Review of the Association between Abdominal Fat Distribution, Health Outcome Measures, and Modifiable Risk Factors'. In: *American Journal of Health Promotion* 7.5, pp. 342–353. ISSN: 0890-1171. DOI: 10.4278/0890-1171-7.5.342. URL: https://doi.org/10.4278/0890-1171-7.5.342.

Enders, C. K. (2001). 'A Primer on Maximum Likelihood Algorithms Available for Use With Missing Data'. In: *Structural Equation Modeling: A Multidisciplinary Journal* 8.1, pp. 128–141. ISSN: 1070-5511. DOI: 10.1207/S15328007SEM0801_7. URL: https://doi.org/10.1207/S15328007SEM0801{_}7.

Enders, C. K. (2008). 'A note on the use of missing auxiliary variables in full information maximum likelihood-based structural equation models'. In: *Structural Equation Modeling* 15.3, pp. 434–448. ISSN: 10705511. DOI: 10.1080/10705510802154307.

Enders, C. K. and Bandalos, D. L. (2001). 'The relative performance of full information maximum likelihood estimation for missing data in structural equation models'. In: *Structural Equation Modeling* 8.3, pp. 430–457. ISSN: 10705511. DOI: 10.1207/S15328007SEM0803_5.

Epstein, J. M. and Axtell, R (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. A Bradford book. MIT Press. ISBN: 9780262050531. URL: https://books.google.co.uk/books?id=xXvelSs2caQC.

Epstein, J. M. (1999). 'Agent-based computational models and generative social science'. In: *Complexity* 4.5, pp. 41–60. ISSN: 10990526. DOI: 10.1002/(SICI)1099-0526(199905/06)4:5<41::AID-CPLX9>3.0.CO;2-F.

Epstein, J. M. (2002). 'Modeling civil violence: An agent-based computational approach'. In: *Proceedings of the National Academy of Sciences* 99.Supplement 3, pp. 7243–7250. ISSN: 0027-8424. DOI: 10.1073/pnas.092080199. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.092080199.

Epstein, J. M. (2008). 'Why Model?' In: *Journal of Artificial Societies and Social Simulation* 11.4, p. 12. ISSN: 1460-7425. URL: https://www.jasss.org/11/4/12.html.

Epstein, J. M. (2012). *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton University Press. ISBN: 9781400842872. DOI: 10.1515/9781400842872. URL: https://doi.org/10.1515/9781400842872https://www.degruyter.com/document/doi/10.1515/9781400842872/html.

Epstein, J. M. (2013). *Agent zero : toward neurocognitive foundations for generative social science*. Ed. by ProQuest. Princeton, NJ: Princeton University Press. ISBN: 0-691-15888-6.

Everson, S. A. et al. (2002). 'Epidemiologic evidence for the relation between socioeconomic status and depression, obesity, and diabetes'. In: *Journal of Psychosomatic Research* 53.4, pp. 891–895. ISSN: 00223999. DOI: 10.1016/S0022-3999(02)00303-3.

Fabricatore, A. and Wadden, T. A. (2004). 'Psychological Aspects of Obesity'. In: *Clinics in Dermatology* 22, pp. 332–337. DOI: 10.1016/j.clindermatol.2004.01.006.

Fagiolo, G., Moneta, A. and Windrum, P. (2007). 'A Critical Guide to Empirical Validation of Agent-Based Models in Economics: Methodologies, Procedures, and Open Problems'. In: *Computational Economics* 30.3, pp. 195–

226. ISSN: 0927-7099. DOI: 10.1007/s10614-007-9104-4. URL: http://link.springer.com/10.1007/s10614-007-9104-4.

Faith, M. S. et al. (2011). 'Evidence for prospective associations among depression and obesity in population-based studies'. In: *Obesity Reviews* 12.501, pp. 438–453. ISSN: 14677881. DOI: 10.1111/j.1467-789X.2010.00843.x.

Faith, M. S., Matz, P. E. and Jorge, M. a. (2002). 'Obesity-depression associations in the population.' In: *Journal of psychosomatic research* 53.4, pp. 935–942. ISSN: 00223999. DOI: 10.1016/S0022-3999(02)00308-2.

Ferketich, A. K. et al. (2000). 'Depression as an Antecedent to Heart Disease Among Women and Men in the NHANES I Study'. In: *Archives of Internal Medicine* 160.9, p. 1261. ISSN: 0003-9926. DOI: 10.1001/archinte.160.9.1261. URL: http://dx.doi.org/10.1001/archinte.160.9.1261http://archinte.jamanetwork.com/article.aspx?doi=10.1001/archinte.160.9.1261.

Fikkan, J. L. and Rothblum, E. D. (2012). 'Is Fat a Feminist Issue? Exploring the Gendered Nature of Weight Bias'. In: *Sex Roles* 66.9-10, pp. 575–592. ISSN: 0360-0025. DOI: 10.1007/s11199-011-0022-5. URL: http://link.springer.com/10.1007/s11199-011-0022-5.

Fitzgibbon, M. L., Stolley, D. S. and Kirschenbaum, M. R. (1993). 'Obese People Who Seek Treatment Have Different Characteristics Than Those Who Do Not Seek Treatment'. In: *Health Psychology* 12.5, pp. 342–345.

Ford, E. S. et al. (1991). 'Physical activity behaviors in lower and higher socioeconomic status populations'. In: *American Journal of Epidemiology* 133.12, pp. 1246–1256. URL: http://aje.oxfordjournals.org/content/133/12/1246.full.pdf.

Fornell, C. and Larcker, D. F. (1981). 'Evaluating Structural Equation Models with Unobservable Variables and Measurement Error'. In: *Journal of Marketing Research* 18.1, p. 39. ISSN: 00222437. DOI: 10.2307/3151312.

Fox, K. R. and Hillsdon, M. (2007). 'Physical activity and obesity'. In: *Obesity Reviews* 8.s1, pp. 115–121. ISSN: 1467-7881. DOI: 10.1111/j.1467-789X.2007.00329.x. URL: http://doi.wiley.com/10.1111/j.1467-789X.2007.00329.x.

Frederick, D. A. and Reynolds, T. A. (2022). 'The Value of Integrating Evolutionary and Sociocultural Perspectives on Body Image'. In: *Archives of*

Sexual Behavior 51.1, pp. 57–66. ISSN: 0004-0002. DOI: 10.1007/s10508-021-01947-4. URL: https://doi.org/10.1007/s10508-021-01947-4https://link.springer.com/10.1007/s10508-021-01947-4.

Fried, E. I. (2017). 'The 52 symptoms of major depression: Lack of content overlap among seven common depression scales'. In: *Journal of Affective Disorders* 208.October 2016, pp. 191–197. ISSN: 15732517. DOI: 10.1016/j.jad.2016.10.019. URL: http://dx.doi.org/10.1016/j.jad.2016.10.019.

Fried, E. I. and Nesse, R. M. (2015). 'Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study'. In: *Journal of Affective Disorders* 172, pp. 96–102. ISSN: 15732517. DOI: 10.1016/j.jad.2014.10.010. arXiv: 15334406. URL: http://dx.doi.org/10.1016/j.jad.2014.10.010.

Friedman, M. A. and Brownell, K. D. (1995). 'Psychological correlates of obesity: moving to the next generation'. In: *Psychol Bull* 117.1, pp. 3–20.

Fruh, S. M. et al. (2016). 'Obesity Stigma and Bias'. In: *The Journal for Nurse Practitioners* 12.7, pp. 425–432. ISSN: 15554155. DOI: 10.1016/j.nurpra.2016.05.013. URL: http://dx.doi.org/10.1016/j.nurpra.2016.05.013https://linkinghub.elsevier.com/retrieve/pii/S155541551630188X.

Galimard, J. E. et al. (2018). 'Heckman imputation models for binary or continuous MNAR outcomes and MAR predictors'. In: *BMC Medical Research Methodology* 18.1, pp. 1–13. ISSN: 14712288. DOI: 10.1186/s12874-018-0547-1.

Gilbert, N. (2008). *Agent-Based Models*. SAGE Publications, Inc. ISBN: 9781412949644. DOI: 10.4135/9781412983259. URL: https://methods.sagepub.com/book/agent-based-modelshttp://methods.sagepub.com/book/agent-based-models.

Giles-Corti, B. and Donovan, R. J. (2002). 'Socioeconomic status differences in recreational physical activity levels and real and perceived access to a supportive physical environment'. In: *Preventive Medicine* 35.6, pp. 601–611. ISSN: 00917435. DOI: 10.1006/pmed.2002.1115.

Glennan, S. (2002). 'Rethinking Mechanistic Explanation'. In: *Philosophy of Science* 69.S3, S342–S353. ISSN: 0031-8248. DOI: 10.1086/341857. URL: https://www.journals.uchicago.edu/doi/10.1086/341857.

Global Burden of Disease Collaborative Network. (2019). 'Global Burden of Disease Study 2019 (GBD 2019) GBD results tool. Institute for Health Metrics and Evaluation (IMHE)'. In:

Godfree, K. (2020). '" I always feel that I'm underestimated": obesity stigma in the workplace'. PhD thesis.

Goffman, E. (1963). *Stigma; notes on the management of spoiled identity*. Englewood Cliffs, N.J. : Prentice-Hall. URL: https://search.library.wisc.edu/catalog/999472649302121.

Goldberg, D. P. and Hillier, V. F. (1979). 'A scaled version of the General Health Questionnaire'. In: *Psychological Medicine* 9.1, pp. 139–145. ISSN: 14698978. DOI: 10.1017/S0033291700021644.

Goldberg, D. (1972). *The detection of psychiatric illness by questionnaire: a technique for the identification and assessment of non-psychotic psychiatric illness*. London: London : Oxford University Press, 1972. ISBN: 0197121438.

Goldberg, D. (2006). 'The aetiology of depression'. In: *Psychological Medicine* 36.10, pp. 1341–1347. ISSN: 00332917. DOI: 10.1017/S0033291706007665.

Goldthorpe, J. H. (2001). 'Causation, Statistics, and Sociology'. In: *European Sociological Review* 17.1, pp. 1–20. ISSN: 02667215. DOI: 10.1093/esr/17.1.1. URL: https://academic.oup.com/esr/article-lookup/doi/10.1093/esr/17.1.1.

Gorwood, P. et al. (2008). 'Toxic effects of depression on brain function: Impairment of delayed recall and the cumulative length of depressive disorder in a large sample of depressed outpatients'. In: *American Journal of Psychiatry* 165.6, pp. 731–739. ISSN: 0002953X. DOI: 10.1176/appi.ajp.2008.07040574.

Graham, J. W. (2003). 'Adding Missing-Data-Relevant Variables to FIML-Based Structural Equation Models'. In: *Structural Equation Modeling: A Multidisciplinary Journal* 10.1, pp. 80–100. ISSN: 1070-5511. DOI: 10.1207/S15328007SEM1001_4. URL: https://doi.org/10.1207/S15328007SEM1001{_}4.

Green, M. et al. (2015). 'Who are the obese? A cluster analysis exploring subgroups of the obese'. In: *Journl of Public Health*.

Grimm, V. et al. (2006). 'A standard protocol for describing individual-based and agent-based models'. In: *Ecological Modelling* 198.1-2, pp. 115–126. ISSN: 03043800. DOI: 10.1016/j.ecolmodel.2006.04.023. arXiv: NIHMS150003.

Grimm, V. et al. (2020). 'The ODD protocol for describing agent-based and other simulation models: A second update to improve clarity, replication, and structural realism'. In: *Jasss* 23.2. ISSN: 14607425. DOI: 10.18564/jasss.4259.

Grogan, S. (2006). 'Body Image and Health'. In: *Journal of Health Psychology* 11.4, pp. 523–530. ISSN: 1359-1053. DOI: 10.1177/1359105306065013. URL: http://journals.sagepub.com/doi/10.1177/1359105306065013.

Hall, K. D. and Kahan, S. (2018). 'Maintenance of Lost Weight and Long-Term Management of Obesity'. In: *Medical Clinics of North America* 102.1, pp. 183–197. ISSN: 00257125. DOI: 10.1016/j.mcna.2017.08.012. URL: https://linkinghub.elsevier.com/retrieve/pii/S0025712517301360.

Hall, N. (2004). 'Two concepts of causation'. In: *Causation and Counterfactuals*. Ed. by J. Collins, N. Hall and L. Paul. MIT Press, pp. 225–276.

Hallquist, M. N. and Wiley, J. F. (2018). 'MplusAutomation : An R Package for Facilitating Large-Scale Latent Variable Analyses in M plus'. In: *Structural Equation Modeling: A Multidisciplinary Journal* 25.4, pp. 621–638. ISSN: 1070-5511. DOI: 10.1080/10705511.2017.1402334. URL: https://doi.org/10.1080/10705511.2017.1402334https://www.tandfonline.com/doi/full/10.1080/10705511.2017.1402334.

Hamill, L. and Gilbert, N. (2009). 'Social circles: A simple structure for agent-based social network models'. In: *Jasss* 12.2, p. 3. ISSN: 14607425.

Hamill, L. and Gilbert, N. (2010). 'Simulating large social networks in agent-based models: A social circle model'. In: *Emergence: Complexity and Organization* 12.4, pp. 78–94. ISSN: 15213250.

Harel, O. et al. (2007). 'Population inference with mortality and attrition in longitudinal studies on aging: A two-stage multiple imputation method'. In: *Experimental Aging Research* 33.2, pp. 187–203. ISSN: 0361073X. DOI: 10.1080/03610730701239004.

Hazelbag, C. M. et al. (2020). 'Calibration of individual-based models to epidemiological data: A systematic review'. In: *PLOS Computational Biology* 16.5. Ed. by R. D. Kouyos, e1007893. ISSN: 1553-7358. DOI: 10.1371/journal.

pcbi.1007893. URL: http://dx.doi.org/10.1371/journal.pcbi.1007893https://dx.plos.org/10.1371/journal.pcbi.1007893.

Hebert, J. R. et al. (1995). 'Social Desirability Bias in Dietary Self-Report May Compromise the Validity of Dietary Intake Measures'. In: *International Journal of Epidemiology* 24.2, pp. 389–398. ISSN: 0300-5771. DOI: 10.1093/ije/24.2.389. URL: https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/24.2.389.

Hebert, J. R. et al. (1997). 'Gender differences in social desirability and social approval bias in dietary self-report'. In: *American Journal of Epidemiology* 146.12, pp. 1046–1055. ISSN: 00029262. DOI: 10.1093/oxfordjournals.aje.a009233.

Hedström, P. and Ylikoski, P. (2010). 'Causal Mechanisms in the Social Sciences'. In: *Annual Review of Sociology* 36.1, pp. 49–67. ISSN: 0360-0572. DOI: 10.1146/annurev.soc.012809.102632. URL: https://www.annualreviews.org/doi/10.1146/annurev.soc.012809.102632.

Heinberg, L. J. (2001). 'Theories of body image disturbance: Perceptual, developmental, and sociocultural factors.' In: *Body image, eating disorders, and obesity: An integrative guide for assessment and treatment.* Washington: American Psychological Association, pp. 27–47. ISBN: 1-55798-324-0 (Paperback). DOI: 10.1037/10502-002. URL: http://content.apa.org/books/10502-002.

Himmelstein, M. S., Puhl, R. M. and Quinn, D. M. (2018). 'Weight Stigma in Men: What, When, and by Whom?' In: *Obesity* 26.6, pp. 968–976. ISSN: 19307381. DOI: 10.1002/oby.22162. URL: https://onlinelibrary.wiley.com/doi/10.1002/oby.22162.

Hoare, E. et al. (2014). 'Associations between obesogenic risk factors and depression among adolescents: A systematic review'. In: *Obesity Reviews* 15.1, pp. 40–51. ISSN: 14677881. DOI: 10.1111/obr.12069.

Hoebel, J. et al. (2017). 'Social inequalities and depressive symptoms in adults: The role of objective and subjective socioeconomic status'. In: *PLoS ONE* 12.1, pp. 1–18. ISSN: 19326203. DOI: 10.1371/journal.pone.0169764.

Holmes, J. (2021). 'Tackling obesity The role of the NHS in a whole-system approach'. In:

Hooper, L. et al. (2012). 'Effect of reducing total fat intake on body weight: systematic review and meta-analysis of randomised controlled trials and cohort studies'. In: *BMJ* 345.dec06 1, e7666–e7666. ISSN: 1756-1833. DOI: 10.1136/bmj.e7666. URL: http://www.bmj.com/cgi/doi/10.1136/bmj.e7666.

Hox, J. and Bechger, T (1998). 'Introduction to Structural Equation Modeling'. In: *Family Science Review* 11, pp. 354–373.

Hu, L. and Bentler, P. M. (1999). 'Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives'. In: *Structural Equation Modeling: A Multidisciplinary Journal* 6.1, pp. 1–55. ISSN: 1070-5511. DOI: 10.1080/10705519909540118. URL: http://www.tandfonline.com/doi/abs/10.1080/10705519909540118.

Huppert, F. A. et al. (1989). 'The factor structure of the General Health Questionnaire (GHQ-30). A reliability study on 6317 community residents'. In: *British Journal of Psychiatry* 155.AUG. Pp. 178–185. ISSN: 00071250.

Hwang, Y. C. et al. (2015). 'Visceral abdominal fat accumulation predicts the conversion of metabolically healthy obese subjects to an unhealthy phenotype'. In: *International Journal of Obesity* 39.9, pp. 1365–1370. ISSN: 14765497. DOI: 10.1038/ijo.2015.75.

Istvan, J, Zavela, K and Weidner, G (1992). 'Body weight and psychological distress in NHANES I'. In: *International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity* 16.12, pp. 999–1003. URL: http://europepmc.org/abstract/MED/1335980.

Jackson, S. E. and Steptoe, A. (2017). 'Association between perceived weight discrimination and physical activity: a population-based study among English middle-aged and older adults'. In: *BMJ Open* 7.3, e014592. ISSN: 2044-6055. DOI: 10.1136/bmjopen-2016-014592. URL: https://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2016-014592.

Jeffery, R. W., Adlis, S. A. and Forster, J. L. (1991). 'Prevalence of dieting among working men and women: The healthy worker project.' In: *Health Psychology* 10.4, pp. 274–281. ISSN: 1930-7810. DOI: 10.1037/0278-6133.10.4.274. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-6133.10.4.274.

Jia, F. (2016). 'Methods for Handling Missing Non-Normal Data in Structural Equation Modeling'. In: *ProQuest Dissertations and Theses*, p. 111. URL: http://ezphost.dur.ac.uk/login?url=https://search.proquest.com/docview/1830471445?accountid=14533{\%}0Ahttp://openurl.ac.uk/ukfed:dur.ac.uk?genre=dissertations+{\%}26+theses{\&}issn={\&}title=Methods+for+Handling+Missing+Non-Normal+Data+in+Structural+Equation+Modelin.

Jonsson, P. and Wohlin, C. (2004). 'An evaluation of k-nearest neighbour imputation using likert data'. In: *10th International Symposium on Software Metrics, 2004. Proceedings.* IEEE, pp. 108–118. ISBN: 0-7695-2129-0. DOI: 10.1109/METRIC.2004.1357895. URL: http://ieeexplore.ieee.org/document/1357895/.

Jung, S. J. et al. (2017). 'Association between body size, weight change and depression: systematic review and meta-analysis'. In: *British Journal of Psychiatry* 211.1, pp. 14–21. ISSN: 0007-1250. DOI: 10.1192/bjp.bp.116.186726. URL: http://bjp.rcpsych.org/lookup/doi/10.1192/bjp.bp.116.186726https://www.cambridge.org/core/product/identifier/S0007125000279920/type/journal{_}article.

Kaiser, K. A. et al. (2014). 'Increased fruit and vegetable intake has no discernible effect on weight loss: a systematic review and meta-analysis'. In: *The American Journal of Clinical Nutrition* 100.2, pp. 567–576. ISSN: 0002-9165. DOI: 10.3945/ajcn.114.090548. URL: https://academic.oup.com/ajcn/article/100/2/567/4576504.

Kanavos, P. and Aardweg, S. V. D. (2012). 'Diabetes expenditure , burden of disease and management in 5 EU countries'. In:

Kant, J.-D., Ballot, G. and Goudet, O. (2020). 'WorkSim: An Agent-Based Model of Labor Markets'. In: *Journal of Artificial Societies and Social Simulation* 23.4, pp. 1–39. ISSN: 1460-7425. DOI: 10.18564/jasss.4396. URL: http://jasss.soc.surrey.ac.uk/23/4/4.html.

Katon, W. J. (2003). 'Clinical and health services relationships between major depression, depressive symptoms, and general medical illness'. In: *Biological Psychiatry* 54.3, pp. 216–226. ISSN: 00063223. DOI: 10.1016/S0006-3223(03)00273-7. URL: https://linkinghub.elsevier.com/retrieve/pii/S0006322303002737.

Kefalas, P et al. (2003). 'A Formal Method for the Development of Agent-Based Systems'. In: *Intelligent Agent Software Engineering*. IGI Global, pp. 68–98. DOI: 10.4018/978-1-59140-046-2.ch004. URL: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-59140-046-2.ch004.

Kennedy, M. C. and O'Hagan, A. (2001). 'Bayesian calibration of computer models'. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.3, pp. 425–464. ISSN: 13697412. DOI: 10.1111/1467-9868.00294. URL: http://doi.wiley.com/10.1111/1467-9868.00294https://onlinelibrary.wiley.com/doi/10.1111/1467-9868.00294.

Kenny, D. A., Kashy, D. A. and Bolger, N. (1998). 'Data analysis in social psychology.' In: *The handbook of social psychology, Vols. 1-2, 4th ed.* New York, NY, US: McGraw-Hill, pp. 233–265. ISBN: 0-19-521376-9 (Hardcover).

Kim, T. J., Roesler, N. M. and von dem Knesebeck, O. (2017). 'Causation or selection - examining the relation between education and overweight/obesity in prospective observational studies: a meta-analysis'. In: *Obesity Reviews* 18.6, pp. 660–672. ISSN: 14677881. DOI: 10.1111/obr.12537. URL: https://onlinelibrary.wiley.com/doi/10.1111/obr.12537.

Kim, T. J. and Von Dem Knesebeck, O. (2018). 'Income and obesity: What is the direction of the relationship? A systematic review and meta-analysis'. In: *BMJ Open* 8.1. ISSN: 20446055. DOI: 10.1136/bmjopen-2017-019862.

Kipnis, V. et al. (2001). 'Empirical Evidence of Correlated Biases in Dietary Assessment Instruments and Its Implications'. In: *American Journal of Epidemiology* 153.4, pp. 394–403. ISSN: 1476-6256. DOI: 10.1093/aje/153.4.394. URL: https://academic.oup.com/aje/article-lookup/doi/10.1093/aje/153.4.394.

Kipnis, V. et al. (2002). 'Bias in dietary-report instruments and its implications for nutritional epidemiology'. In: *Public Health Nutrition* 5.6a, pp. 915–923. ISSN: 1368-9800. DOI: 10.1079/PHN2002383. URL: https://www.cambridge.org/core/product/identifier/S1368980002001234/type/journal{_}article.

Klopp, E. and Klößner, S. (2021). 'The Impact of Scaling Methods on the Properties and Interpretation of Parameter Estimates in Structural Equation Models with Latent Variables'. In: *Structural Equation Modeling: A Multidisciplinary Journal* 28.2, pp. 182–206. ISSN: 1070-5511. DOI: 10.1080/10705511.

2020.1796673. URL: https://doi.org/10.1080/10705511.2020.1796673https://www.tandfonline.com/doi/full/10.1080/10705511.2020.1796673.

Korczak, D. J. et al. (2013). 'Are Children And Adolescents With Psychiatric Illness At Risk For Increased Future Body Weight? A Systematic Review'. In: *Developmental Medicine and Child Neurology* 55.11, pp. 980–987. ISSN: 00121622. DOI: 10.1111/dmcn.12168.

Kowarik, A. and Templ, M. (2016). 'Imputation with the R Package VIM'. In: *Journal of Statistical Software* 74.7. ISSN: 1548-7660. DOI: 10.18637/jss.v074.i07. URL: http://www.jstatsoft.org/v74/i07/.

Kraemer, H. C. et al. (2001). 'How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors'. In: *American Journal of Psychiatry* 158.6, pp. 848–856. ISSN: 0002953X. DOI: 10.1176/appi.ajp.158.6.848.

Krayer, A., Ingledew, D. K. and Iphofen, R. (2007). 'Social comparison and body image in adolescence: a grounded theory approach'. In: *Health Education Research* 23.5, pp. 892–903. ISSN: 0268-1153. DOI: 10.1093/her/cym076. URL: https://academic.oup.com/her/article-lookup/doi/10.1093/her/cym076.

Kristal, A. R. et al. (2000). 'Precision and bias of food frequency-based measures of fruit and vegetable intakes'. In: *Cancer Epidemiology Biomarkers and Prevention* 9.9, pp. 939–944. ISSN: 10559965.

Lassale, C. et al. (2019). 'Healthy dietary indices and risk of depressive outcomes: a systematic review and meta-analysis of observational studies'. In: *Molecular Psychiatry* 24.7, pp. 965–986. ISSN: 14765578. DOI: 10.1038/s41380-018-0237-8. URL: http://dx.doi.org/10.1038/s41380-018-0237-8.

Lavie, C. J., Milani, R. V. and Ventura, H. O. (2009). 'Obesity and Cardiovascular Disease. Risk Factor, Paradox, and Impact of Weight Loss'. In: *Journal of the American College of Cardiology* 53.21, pp. 1925–1932. ISSN: 07351097. DOI: 10.1016/j.jacc.2008.12.068. URL: http://dx.doi.org/10.1016/j.jacc.2008.12.068.

Lee, K. J. and Carlin, J. B. (2012). 'Recovery of information from multiple imputation: A simulation study'. In: *Emerging Themes in Epidemiology* 9, pp. 1–10. ISSN: 17427622. DOI: 10.1186/1742-7622-9-3.

Leite, W. and Beretvas, S. N. (2010). 'The performance of multiple imputation for Likert-type items with missing data'. In: *Journal of Modern Applied Statistical Methods* 9.1, pp. 64–74. ISSN: 15389472. DOI: 10.22237/jmasm/1272686820.

León-Medina, F. J. (2017). 'Analytical Sociology and Agent-Based Modeling: Is Generative Sufficiency Sufficient?' In: *Sociological Theory* 35.3, pp. 157–178. ISSN: 0735-2751. DOI: 10.1177/0735275117725766. URL: http://journals.sagepub.com/doi/10.1177/0735275117725766.

Lewis, S. et al. (2011). 'How do obese individuals perceive and respond to the different types of obesity stigma that they encounter in their daily lives? A qualitative study'. In: *Social Science & Medicine* 73.9, pp. 1349–1356. ISSN: 02779536. DOI: 10.1016/j.socscimed.2011.08.021. URL: http://dx.doi.org/10.1016/j.socscimed.2011.08.021https://linkinghub.elsevier.com/retrieve/pii/S0277953611005284.

Liem, E. et al. (2008). 'Association between depressive symptoms in childhood and adolescence and overweight in later life: review of the recent literature.' In: *Archives of Pediatrics & Adolescent Medicine* 162.10, pp. 981–988.

Lincoln, J. E. (1972). 'Calorie intake, obesity, and physical activity'. In: *The American Journal of Clinical Nutrition* 25.4, pp. 390–394. ISSN: 0002-9165. DOI: 10.1093/ajcn/25.4.390. URL: https://academic.oup.com/ajcn/article/25/4/390/4818963.

Link, B. G. and Phelan, J. C. (2001). 'Conceptualizing Stigma'. In: *Annual Review of Sociology* 27.1, pp. 363–385. ISSN: 0360-0572. DOI: 10.1146/annurev.soc.27.1.363. URL: http://www.annualreviews.org/doi/10.1146/annurev.soc.27.1.363.

Little, T. D. and Card, N. A. (2013). *Longitudinal Structural Equation Modeling*. Guildford Publications, p. 410. ISBN: 9781462510160.

Little, T. D., Slegers, D. W. and Card, N. A. (2006). 'A Non-arbitrary Method of Identifying and Scaling Latent Variables in SEM and MACS Models'. In: *Structural Equation Modeling: A Multidisciplinary Journal* 13.1, pp. 59–72. ISSN: 1070-5511. DOI: 10.1207/s15328007sem1301_3. URL: http://www.tandfonline.com/doi/abs/10.1207/s15328007sem1301{_}3.

Liu, A. G. et al. (2017a). 'A healthy approach to dietary fats: understanding the science and taking action to reduce consumer confusion'. In: *Nutrition Journal* 16.1, p. 53. ISSN: 1475-2891. DOI: 10.1186/s12937-017-0271-4. URL: http://nutritionj.biomedcentral.com/articles/10.1186/s12937-017-0271-4.

Liu, Z. et al. (2017b). 'A simulation and optimization based method for calibrating agent-based emergency department models under data scarcity'. In: *Computers & Industrial Engineering* 103, pp. 300–309. ISSN: 03608352. DOI: 10.1016/j.cie.2016.11.036. URL: http://dx.doi.org/10.1016/j.cie.2016.11.036https://linkinghub.elsevier.com/retrieve/pii/S0360835216304685.

Logue, J. et al. (2011). 'Obesity is associated with fatal coronary heart disease independently of traditional risk factors and deprivation'. In: *Heart* 97.7, pp. 564–568. ISSN: 1355-6037. DOI: 10.1136/hrt.2010.211201. URL: https://heart.bmj.com/lookup/doi/10.1136/hrt.2010.211201.

Łojko, D. and Rybakowski, J. K. (2017). 'Atypical depression: Current perspectives'. In: *Neuropsychiatric Disease and Treatment* 13, pp. 2447–2456. ISSN: 11782021. DOI: 10.2147/NDT.S147317.

Lomax, R. G. (1986). 'The Effect of Measurement Error in Structural Equation Modeling'. In: *The Journal of Experimental Education* 54.3, pp. 157–162. ISSN: 0022-0973. DOI: 10.1080/00220973.1986.10806415. URL: http://www.tandfonline.com/doi/abs/10.1080/00220973.1986.10806415.

Lorant, V. et al. (2003). 'Socioeconomic inequalities in depression: A meta-analysis'. In: *American Journal of Epidemiology* 157.2, pp. 98–112. ISSN: 00029262. DOI: 10.1093/aje/kwf182.

Luppino, F. S. et al. (2010). 'Overweight, Obesity, and Depression: a systematic review and meta-analysis of longitudinal studies'. In: *Archives of General Psychiatry* 67.3, p. 220. ISSN: 0003-990X. DOI: 10.1001/archgenpsychiatry.2010.2. URL: http://archpsyc.jamanetwork.com/article.aspx?doi=10.1001/archgenpsychiatry.2010.2.

Macal, C. M. and North, M. J. (2010). 'Tutorial on agent-based modelling and simulation'. In: *Journal of Simulation* 4.3, pp. 151–162. ISSN: 1747-7778. DOI: 10.1057/jos.2010.3. URL: https://www.tandfonline.com/doi/full/10.1057/jos.2010.3.

MacCallum, R. C., Browne, M. W. and Sugawara, H. M. (1996). 'Power analysis and determination of sample size for covariance structure modeling.' In: *Psychological Methods* 1.2, pp. 130–149. ISSN: 1939-1463. DOI: 10.1037/1082-989X.1.2.130. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/1082-989X.1.2.130.

MacNeill, L. P. and Best, L. A. (2015). 'Perceived current and ideal body size in female undergraduates'. In: *Eating Behaviors* 18, pp. 71–75. ISSN: 14710153. DOI: 10.1016/j.eatbeh.2015.03.004. URL: http://dx.doi.org/10.1016/j.eatbeh.2015.03.004https://linkinghub.elsevier.com/retrieve/pii/S1471015315000367.

Maguire, E. R. and Monsivais, P. (2015). 'Socio-economic dietary inequalities in UK adults: an updated picture of key food groups and nutrients from national surveillance data.' In: *The British journal of nutrition* 113.1, pp. 181–9. ISSN: 1475-2662. DOI: 10.1017/S0007114514002621. URL: http://www.ncbi.nlm.nih.gov/pubmed/25399952http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4351901.

Malik, V. S., Schulze, M. B. and Hu, F. B. (2006). 'Intake of sugar-sweetened beverages and weight gain: A systematic review'. In: *American Journal of Clinical Nutrition* 84.2, pp. 274–288. ISSN: 00029165. DOI: 10.1093/ajcn/84.2.274.

Malik, V. S. et al. (2013). 'Sugar-sweetened beverages and weight gain in children and adults: a systematic review and meta-analysis'. In: *The American Journal of Clinical Nutrition* 98.4, pp. 1084–1102. ISSN: 0002-9165. DOI: 10.3945/ajcn.113.058362. URL: https://academic.oup.com/ajcn/article/98/4/1084/4577098.

Mannan, M. et al. (2016a). 'Is there a bi-directional relationship between depression and obesity among adult men and women? Systematic review and bias-adjusted meta analysis'. In: *Asian Journal of Psychiatry* 21.2016, pp. 51–66. ISSN: 18762026. DOI: 10.1016/j.ajp.2015.12.008. URL: http://dx.doi.org/10.1016/j.ajp.2015.12.008.

Mannan, M. et al. (2016b). 'Prospective Associations between Depression and Obesity for Adolescent Males and Females- A Systematic Review and Meta-Analysis of Longitudinal Studies'. In: *PLOS ONE* 11.6. Ed. by J. Homberg, e0157240. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0157240. URL: http://journals.plos.org/plosone/article/file?id=10.1371/

journal.pone.0157240{\&}type=printablehttps://dx.plos.org/10.1371/journal.pone.0157240.

Manzo, G. (2013). 'Educational Choices and Social Interactions: A Formal Model and a Computational Test'. In: *Comparative Social Research*. Vol. 30. 2013. Emerald Group Publishing Limited, pp. 47–100. ISBN: 9781781905371. DOI: 10.1108/S0195-6310(2013)0000030007. URL: http://dx.doi.org/10.1108/S0195-6310(2013)0000030007https://www.emerald.com/insight/content/doi/10.1108/S0195-6310(2013)0000030007/full/html.

Marino, S. et al. (2008). 'A methodology for performing global uncertainty and sensitivity analysis in systems biology'. In: *Journal of Theoretical Biology* 254.1, pp. 178–196. ISSN: 00225193. DOI: 10.1016/j.jtbi.2008.04.011. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022519308001896.

Markowitz, S., Friedman, M. A. and Arent, S. M. (2008). 'Understanding the Relation Between Obesity and Depression: Causal Mechanisms and Implications for Treatment'. In: *Clinical Psychology: Science and Practice* 15, pp. 1–20. ISSN: 09695893. DOI: 10.1111/j.1468-2850.2008.00106.x.

Marmot, M. (2010). 'The Marmot Review - Fair Society, Healthy Lives: A Strategic Review of Health Inequalities in England post-2010'. In:

Marmot, M. and Brunner, E. (2005). 'Cohort Profile: The Whitehall II study'. In: *International Journal of Epidemiology* 34.2, pp. 251–256. ISSN: 1464-3685. DOI: 10.1093/ije/dyh372. URL: http://academic.oup.com/ije/article/34/2/251/746997/Cohort-Profile-The-Whitehall-II-study.

Marshall, A. et al. (2009). 'Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines'. In: *BMC Medical Research Methodology* 9.1, p. 57. ISSN: 1471-2288. DOI: 10.1186/1471-2288-9-57. URL: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-9-57.

McCrone, P et al. (2008). *Paying the price: the cost of mental health care in England to 2026*. London: The King's Fund. ISBN: 9781857175714 BT - Paying the price: the cost of mental health care in England to 2026.

McKay, M. D., Beckman, R. J. and Conover, W. J. (1979). 'Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code'. In: *Technometrics* 21.2, pp. 239–245. ISSN:

0040-1706. DOI: 10.1080/00401706.1979.10489755. URL: http://www.tandfonline.com/doi/abs/10.1080/00401706.1979.10489755.

McLaren, L. (2007). 'Socioeconomic Status and Obesity'. In: *Epidemiologic Reviews* 29.1, pp. 29–48. ISSN: 0193-936X. DOI: 10.1093/epirev/mxm001. URL: https://academic.oup.com/epirev/article-lookup/doi/10.1093/epirev/mxm001.

Mcmanus, S. et al. (2016). 'Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014'. In: *Leeds: NHS Digital*, pp. 1–405. URL: http://content.digital.nhs.uk/catalogue/PUB21748/apms-2014-full-rpt.pdf.

McPhee, J. S. et al. (2016). 'Physical activity in older age: perspectives for healthy ageing and frailty'. In: *Biogerontology* 17.3, pp. 567–580. ISSN: 15736768. DOI: 10.1007/s10522-016-9641-0.

Milanović, Z. et al. (2013). 'Age-related decrease in physical activity and functional fitness among elderly men and women'. In: *Clinical Interventions in Aging* 8, pp. 549–556. ISSN: 11769092. DOI: 10.2147/CIA.S44112.

Mooney, S. J. and El-Sayed, A. M. (2016). 'Stigma and the etiology of depression among the obese: An agent-based exploration'. In: *Social Science & Medicine* 148, pp. 1–7. ISSN: 02779536. DOI: 10.1016/j.socscimed.2015.11.020. URL: http://dx.doi.org/10.1016/j.socscimed.2015.11.020https://linkinghub.elsevier.com/retrieve/pii/S0277953615302288.

Moore, M. E., Stunkard, A. J. and Srole, L. (1997). 'Obesity, social class, and mental illness'. In: *Obesity Research* 5.5, pp. 503–508. ISSN: 10717323. DOI: 10.1002/j.1550-8528.1997.tb00678.x.

Morenga, L. T., Mallard, S. and Mann, J. (2013). 'Dietary sugars and body weight: Systematic review and meta-analyses of randomised controlled trials and cohort studies'. In: *BMJ (Online)* 345.7891, pp. 1–25. ISSN: 17561833. DOI: 10.1136/bmj.e7492.

Morgan, S. L. and Winship, C. (2014). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. 2nd ed. Cambridge: Cambridge University Press. ISBN: 9781107065079. DOI: DOI:10.1017/CBO9781107587991. URL: https://www.cambridge.org/core/books/counterfactuals-and-causal-inference/5CC81E6DF63C5E5A8B88F79D45E1D1B7.

Morshed, A. B. et al. (2019). 'A systematic review of system dynamics and agent-based obesity models: Evaluating obesity as part of the global syndemic'. In: *Obesity Reviews* 20.S2, pp. 161–178. ISSN: 1467-7881. DOI: 10.1111/obr.12877. URL: https://onlinelibrary.wiley.com/doi/10.1111/obr.12877.

Moss, S. (2008). 'Alternative approaches to the empirical validation of agent-based models'. In: *Jasss* 11.1. ISSN: 14607425.

Moussavi, S. et al. (2007). 'Depression, chronic diseases, and decrements in health: results from the World Health Surveys'. In: *Lancet* 370.9590, pp. 851–858. ISSN: 01406736. DOI: 10.1016/S0140-6736(07)61415-9.

Mühlig, Y. et al. (2016). 'Are bidirectional associations of obesity and depression already apparent in childhood and adolescence as based on high-quality studies? A systematic review'. In: *Obesity Reviews* 17.3, pp. 235–249. ISSN: 1467789X. DOI: 10.1111/obr.12357.

Mulder, J. D. and Hamaker, E. L. (2021). 'Three Extensions of the Random Intercept Cross-Lagged Panel Model'. In: *Structural Equation Modeling: A Multidisciplinary Journal* 28.4, pp. 638–648. ISSN: 1070-5511. DOI: 10.1080/10705511.2020.1784738. URL: https://www.tandfonline.com/doi/full/10.1080/10705511.2020.1784738.

Muthen, B. and Muthen, L. (2022). *Chi-Square Difference Testing Using the Satorra-Bentler Scaled Chi-Square*. URL: https://www.statmodel.com/chidiff.shtml (visited on 18/03/2022).

Muthen, L. and Muthen, B. (2017). *Mplus User Guide*. DOI: DOI. arXiv: ArXivID. URL: https://www.mendeley.com/library/.

Mykletun, A. et al. (2007). 'Anxiety, Depression, and Cause-Specific Mortality: The HUNT Study'. In: *Psychosomatic Medicine* 69.4, pp. 323–331. ISSN: 0033-3174. DOI: 10.1097/PSY.0b013e31803cb862. URL: https://journals.lww.com/00006842-200705000-00006.

Mykletun, A. et al. (2009). 'Levels of anxiety and depression as predictors of mortality: the HUNT study'. In: *British Journal of Psychiatry* 195.2, pp. 118–125. ISSN: 0007-1250. DOI: 10.1192/bjp.bp.108.054866. URL: https://www.cambridge.org/core/product/identifier/S000712500000773X/type/journal{_}article.

Naska, A., Lagiou, A. and Lagiou, P. (2017). 'Dietary assessment methods in epidemiological research: Current state of the art and future prospects'. In: *F1000Research* 6.0, pp. 1–8. ISSN: 1759796X. DOI: 10.12688/f1000research.10703.1.

Naylor, C. et al. (2012). 'Long-term conditions and mental health: The cost of co-morbidities'. In: *The King's Fund and Centre for Mental Health*, pp. 1–32. ISSN: 9781857176339 1857176332. DOI: ISBN9781857176339. URL: http://www.kingsfund.org.uk/sites/files/kf/field/field{_}publication{_}file/long-term-conditions-mental-health-cost-comorbidities-naylor-feb12.pdf.

Netemeyer, R. G. (2003). *Scaling procedures issues and applications* /. Thousand Oaks, Calif.

Newton, S., Braithwaite, D. and Akinyemiju, T. F. (2017). 'Socio-economic status over the life course and obesity: Systematic review and meta-analysis'. In: *PLOS ONE* 12.5. Ed. by G. Xiao, e0177151. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0177151. URL: https://dx.plos.org/10.1371/journal.pone.0177151.

NHS (2019). *The NHS Long Term Plan*. Tech. rep. URL: https://www.longtermplan.nhs.uk/.

NHS Digital (2017). *Statistics on Obesity, Physical Activity and Diet*. March. ISBN: 978-1-78386-982-4. DOI: 978-1-84636-274-3.

NHS Digital (2020a). *Fruit and Vegetables*. URL: http://healthsurvey.hscic.gov.uk/data-visualisation/data-visualisation/explore-the-trends/fruit-vegetables.aspx (visited on 12/03/2020).

NHS Digital (2020b). *Weight*. URL: http://healthsurvey.hscic.gov.uk/data-visualisation/data-visualisation/explore-the-trends/weight.aspx (visited on 12/03/2020).

Niazi, M. A., Hussain, A. and Kolberg, M. (2009). 'Verification & Validation of Agent Based Simulations using the VOMAS (Virtual Overlay Multi-agent System) approach'. In: *Proceedings of the Third Workshop on Multi-Agent Systems and Simulation'09 (MASS '09), as part of MALLOW 09, Sep 7-11, 2009, Torino, Italy* 494, pp. 340–346. arXiv: 1708.02361. URL: http://arxiv.org/abs/1708.02361.

Ning, Y. et al. (2013). 'Results Differ by Applying Distinctive Multiple Imputation Approaches on the Longitudinal Cardiovascular Health Study Data'. In: *Experimental Aging Research* 39.1, pp. 27–43. ISSN: 0361-073X. DOI: 10.1080/0361073X.2013.741968. URL: http://www.tandfonline.com/doi/abs/10.1080/0361073X.2013.741968.

Ohrnberger, J., Fichera, E. and Sutton, M. (2017). 'The relationship between physical and mental health: A mediation analysis'. In: *Social Science and Medicine* 195.October, pp. 42–49. ISSN: 18735347. DOI: 10.1016/j.socscimed.2017.11.008. URL: https://doi.org/10.1016/j.socscimed.2017.11.008.

Olinsky, A., Chen, S. and Harlow, L. (2003). 'The comparative efficacy of imputation methods for missing data in structural equation modeling'. In: *European Journal of Operational Research* 151.1, pp. 53–79. ISSN: 03772217. DOI: 10.1016/S0377-2217(02)00578-7.

Organisation of Economic Co-operation and Development (2019). *Health Risks*. DOI: https://doi.org/10.1787/1c4df204-en. URL: https://www.oecd-ilibrary.org/social-issues-migration-health/health-risks/indicator-group/english{_}1c4df204-en{\#}:{~}:text=Riskfactorsareanyattribute,andbeingoverweightorobese. (visited on 01/11/2021).

Pampel, F. C., Krueger, P. M. and Denney, J. T. (2010). 'Socioeconomic Disparities in Health Behaviours'. In: *Annual Review of Sociology* 36, pp. 349–70. ISSN: 00034819. DOI: 10.1146/annurev.soc.012809.102529.

Paolicelli, C. W. (2016). *Weight management and obesity*. First edit. New York: New York : Momentum Press Health, 2016. ISBN: 1-60650-764-8.

Papadopoulos, S. and Brennan, L. (2015). 'Correlates of weight stigma in adults with overweight and obesity: A systematic literature review'. In: *Obesity* 23.9, pp. 1743–1760. ISSN: 1930739X. DOI: 10.1002/oby.21187. URL: https://onlinelibrary.wiley.com/doi/10.1002/oby.21187.

Paul, L. A. and Hall, N. (2013). *Causation: A User's Guide*. eng. Oxford: Oxford University Press, p. 304. ISBN: 9780199673445. DOI: 10.1093/acprof:oso/9780199673445.001.0001. URL: https://oxford.universitypressscholarship.com/10.1093/acprof:oso/9780199673445.001.0001/acprof-9780199673445.

Pearl, J. (2001). 'Causal inference in the health sciences: A conceptual introduction'. In: *Health Services and Outcomes Research Methodology* 2.3-4, pp. 189–220. ISSN: 13873741. DOI: 10.1023/A:1020315127304.

Pearl, R. L. and Puhl, R. M. (2018). *Weight bias internalization and health: a systematic review*. DOI: 10.1111/obr.12701. URL: https://onlinelibrary.wiley.com/doi/10.1111/obr.12701.

Pearl, R. L. and Puhl, R. M. (2016). 'The distinct effects of internalizing weight bias: An experimental study'. In: *Body Image* 17, pp. 38–42. ISSN: 17401445. DOI: 10.1016/j.bodyim.2016.02.002. URL: http://dx.doi.org/10.1016/j.bodyim.2016.02.002.

Pereira-Miranda, E. et al. (2017). 'Overweight and Obesity Associated with Higher Depression Prevalence in Adults: A systematic Review and Meta-Analysis'. In: *Journal of the American College of Nutrition*. URL: https://www.researchgate.net/profile/M{_}Pereira-santos/publication/315875029{_}Overweight{_}and{_}Obesity{_}Associated{_}with{_}Higher{_}Depression{_}Prevalence{_}in{_}Adults{_}A{_}Systematic{_}Review{_}and{_}Meta-Analysis{_}Overweight{_}and{_}Obesity{_}Associated{_}with{_}Higher{_}Depression{_}Preval.

Perry, J. L. et al. (2015). 'Assessing Model Fit: Caveats and Recommendations for Confirmatory Factor Analysis and Exploratory Structural Equation Modeling'. In: *Measurement in Physical Education and Exercise Science* 19.1, pp. 12–21. ISSN: 1091-367X. DOI: 10.1080/1091367X.2014.952370. URL: http://dx.doi.org/10.1080/1091367X.2014.952370https://doi.org/10.1080/1091367X.2014.952370http://www.tandfonline.com/doi/abs/10.1080/1091367X.2014.952370.

Pescosolido, B. A. and Martin, J. K. (2015). 'The Stigma Complex'. In: *Annual Review of Sociology* 41.1, pp. 87–116. DOI: 10.1146/annurev-soc-071312-145702.

Petraitis, P. S., Dunham, A. E. and Niewiarowski, P. H. (1996). 'Inferring Multiple Causality: The Limitations of Path Analysis'. In: *Functional Ecology* 10.4, p. 421. ISSN: 02698463. DOI: 10.2307/2389934. URL: https://www.jstor.org/stable/2389934?origin=crossref.

Preiss, K., Brennan, L. and Clarke, D. (2013). 'A systematic review of variables associated with the relationship between obesity and depression'. In:

Obesity Reviews 14.11, pp. 906–918. ISSN: 14677881. DOI: 10.1111/obr.12052.

Public Health England (2017). *Health matters: obesity and the food environment*. URL: https://www.gov.uk/government/publications/health-matters-obesity-and-the-food-environment/health-matters-obesity-and-the-food-environment--2 (visited on 19/10/2021).

Puhl, R. and Brownell, K. D. (2003). 'Ways of coping with obesity stigma: review and conceptual analysis'. In: *Eating Behaviors* 4.1, pp. 53–78. ISSN: 14710153. DOI: 10.1016/S1471-0153(02)00096-X. URL: https://linkinghub.elsevier.com/retrieve/pii/S147101530200096X.

Puhl, R. M. and King, K. M. (2013). 'Weight discrimination and bullying'. In: *Best Practice and Research: Clinical Endocrinology and Metabolism* 27.2, pp. 117–127. ISSN: 1521690X. DOI: 10.1016/j.beem.2012.12.002. URL: http://dx.doi.org/10.1016/j.beem.2012.12.002.

Puhl, R. M., Moss-Racusin, C. A. and Schwartz, M. B. (2007). 'Internalization of Weight Bias: Implications for Binge Eating and Emotional Well-being'. In: *Obesity* 15, pp. 19–23.

Pyne, J. M. et al. (1997). 'Assessment of the quality of life of patients with major depression.' In: *Psychiatric services (Washington, D.C.)* 48.2, pp. 224–30. ISSN: 1075-2730. DOI: 10.1176/ps.48.2.224. URL: http://www.ncbi.nlm.nih.gov/pubmed/9021855.

Quek, Y. et al. (2017). 'Exploring the association between childhood and adolescent obesity and depression: a meta-analysis'. In: *Obesity Reviews* 18.7, pp. 742–754. ISSN: 1467-789X. DOI: 10.1111/OBR.12535. URL: http://onlinelibrary.wiley.com.libproxy.lib.unc.edu/wol1/doi/10.1111/obr.12535/full.

Quirk, S. E. et al. (2013). 'The association between diet quality, dietary patterns and depression in adults: a systematic review'. In: *BMC Psychiatry* 13.1, p. 175. ISSN: 1471-244X. DOI: 10.1186/1471-244X-13-175. URL: https://bmcpsychiatry.biomedcentral.com/articles/10.1186/1471-244X-13-175.

Rahe, C., Unrath, M. and Berger, K. (2014). 'Dietary patterns and the risk of depression in adults: A systematic review of observational studies'. In:

European Journal of Nutrition 53.4, pp. 997–1013. ISSN: 14366215. DOI: 10.1007/s00394-014-0652-9.

Rebar, A. L. et al. (2015). 'A meta-meta-analysis of the effect of physical activity on depression and anxiety in non-clinical adult populations'. In: *Health Psychology Review* 9.3, pp. 366–378. ISSN: 1743-7199. DOI: 10.1080/17437199.2015.1022901. URL: https://www.tandfonline.com/doi/full/10.1080/17437199.2015.1022901.

Robinson, E., Sutin, A. and Daly, M. (2017). 'Perceived weight discrimination mediates the prospective relation between obesity and depressive symptoms in U.S. and U.K. adults.' In: *Health Psychology* 36.2, pp. 112–121. ISSN: 1930-7810. DOI: 10.1037/hea0000426. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/hea0000426.

Robinson, S. et al., eds. (2010). *Conceptual Modeling for Discrete-Event Simulation*. CRC Press. ISBN: 9781439810385. DOI: 10.1201/9781439810385. URL: https://www.taylorfrancis.com/books/9781439810385.

Rooke, S. and Thorsteinsson, E. (2008). 'Examining the temporal relationship between depression and obesity: meta-analyses of prospective research'. In: *Health Psychology Review* 2.1, pp. 94–109. ISSN: 1743-7199. DOI: 10.1080/17437190802295689. URL: http://www.tandfonline.com/doi/abs/10.1080/17437190802295689.

Rouhani, M. H. et al. (2014). 'Is there a relationship between red or processed meat intake and obesity? A systematic review and meta-analysis of observational studies'. In: *Obesity Reviews* 15.9, pp. 740–748. ISSN: 14677881. DOI: 10.1111/obr.12172. URL: http://doi.wiley.com/10.1111/obr.12172.

Rubin, D. B. (1976). 'Inference and missing data'. In: *Biometrika* 63.3, pp. 581–592. ISSN: 00063444. DOI: 10.1093/biomet/63.3.581.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Ed. by D. B. Rubin. Wiley Series in Probability and Statistics. Hoboken, NJ, USA: John Wiley & Sons, Inc. ISBN: 9780470316696. DOI: 10.1002/9780470316696. URL: http://doi.wiley.com/10.1002/9780470316696.

Russo, A. et al. (2007). 'Depression and Physical Function: Results From the Aging and Longevity Study in the Sirente Geographic Area (ilSIRENTE Study)'. In: *Journal of Geriatric Psychiatry and Neurology* 20.3, pp. 131–137.

ISSN: 0891-9887. DOI: 10.1177/0891988707301865. URL: http://journals.sagepub.com/doi/10.1177/0891988707301865.

Ruxton, C. H. S., Gardner, E. J. and McNulty, H. M. (2009). 'Is Sugar Consumption Detrimental to Health? A Review of the Evidence 1995—2006'. In: *Critical Reviews in Food Science and Nutrition* 50.1, pp. 1–19. ISSN: 1040-8398. DOI: 10.1080/10408390802248569. URL: http://www.tandfonline.com/doi/abs/10.1080/10408390802248569.

Sabia, S. et al. (2012). 'Effect of intensity and type of physical activity on mortality: Results from the whitehall II cohort study'. In: *American Journal of Public Health* 102.4, pp. 698–704. ISSN: 00900036. DOI: 10.2105/AJPH.2011.300257.

Salecker, J. et al. (2019). 'The <scp>nlrx r</scp> package: A next-generation framework for reproducible NetLogo model analyses'. In: *Methods in Ecology and Evolution* 10.11. Ed. by L. Graham, pp. 1854–1863. ISSN: 2041-210X. DOI: 10.1111/2041-210X.13286. URL: https://onlinelibrary.wiley.com/doi/10.1111/2041-210X.13286.

Salgado, M. and Gilbert, N. (2013). 'Agent Based Modelling'. In: *Handbook of Quantitative Methods for Educational Research*. Rotterdam: SensePublishers, pp. 247–265. ISBN: 978-94-6209-404-8 (Online). DOI: 10.1007/978-94-6209-404-8_12. URL: http://link.springer.com/10.1007/978-94-6209-404-8{_}12.

Sanhueza, C., Ryan, L. and Foxcroft, D. R. (2013). 'Diet and the risk of unipolar depression in adults: systematic review of cohort studies'. In: *Journal of Human Nutrition and Dietetics* 26.1, pp. 56–70. ISSN: 09523871. DOI: 10.1111/j.1365-277X.2012.01283.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1365-277X.2012.01283.x.

Santos, I. et al. (2017). 'Prevalence of personal weight control attempts in adults: a systematic review and meta-analysis'. In: *Obesity Reviews* 18.1, pp. 32–50. ISSN: 1467789X. DOI: 10.1111/obr.12466.

Satorra, A. and Bentler, P. (2001). 'A Scaled Difference Chi-Square Test Statisitic for Moment Structure Analysis'. In: *Psychometrika* 66.4, pp. 507–514. URL: https://link.springer.com/content/pdf/10.1007{\%}2FBF02296192.pdf.

Sattler, K. M. et al. (2018). 'Gender differences in the relationship of weight-based stigmatisation with motivation to exercise and physical activity in overweight individuals'. In: *Health Psychology Open* 5.1. ISSN: 20551029. DOI: 10.1177/2055102918759691.

Savalei, V. and Falk, C. F. (2014). 'Robust Two-Stage Approach Outperforms Robust Full Information Maximum Likelihood With Incomplete Nonnormal Data'. In: *Structural Equation Modeling* 21.2, pp. 280–302. ISSN: 10705511. DOI: 10.1080/10705511.2014.882692. URL: http://dx.doi.org/10.1080/10705511.2014.882692https://doi.org/10.1080/10705511.2014.882692.

Schafer, J. L. and Graham, J. W. (2002). 'Missing data: Our view of the state of the art'. In: *Psychological Methods* 7.2, pp. 147–177. ISSN: 1082989X. DOI: 10.1037/1082-989X.7.2.147.

Schwingshackl, L. et al. (2015). 'Fruit and vegetable consumption and changes in anthropometric variables in adult populations: A systematic review and meta-analysis of prospective cohort studies'. In: *PLoS ONE* 10.10, pp. 1–19. ISSN: 19326203. DOI: 10.1371/journal.pone.0140846.

Selig, J. P. and Little, T. D. (2012). 'Autoregressive and Cross-Lagged Panel Analysis for Longitudinal Data'. In: *Handbook of Developmental Research Methods*. Chapter 12, pp. 265–278.

Selig, J. P. and Preacher, K. J. (2009). 'Mediation Models for Longitudinal Data in Developmental Research'. In: *Research in Human Development* 6.2-3, pp. 144–164. ISSN: 1542-7609. DOI: 10.1080/15427600902911247. URL: http://www.tandfonline.com/doi/abs/10.1080/15427600902911247https://www.tandfonline.com/doi/full/10.1080/15427600902911247.

Sheline, Y. I., Gado, M. H. and Kraemer, H. C. (2003). 'Untreated depression and hippocampal volume loss'. In: *American Journal of Psychiatry* 160.8, pp. 1516–1518. ISSN: 0002953X. DOI: 10.1176/appi.ajp.160.8.1516.

Shrier, I. and Platt, R. W. (2008). 'Reducing bias through directed acyclic graphs'. In: *BMC Medical Research Methodology* 8.1, p. 70. ISSN: 1471-2288. DOI: 10.1186/1471-2288-8-70. URL: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-8-70.

Sikorski, C. et al. (2011). 'The stigma of obesity in the general public and its implications for public health - a systematic review'. In: *BMC Public Health*

11.1, p. 661. ISSN: 1471-2458. DOI: 10.1186/1471-2458-11-661. URL: http://www.biomedcentral.com/1471-2458/11/661http://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-11-661.

Sobal, J. and Stunkard, A. J. (1989). *Socioeconomic status and obesity: A review of the literature.* US. DOI: 10.1037/0033-2909.105.2.260.

Spahlholz, J. et al. (2016). 'Obesity and discrimination - a systematic review and meta-analysis of observational studies'. In: *Obesity Reviews* 17.1, pp. 43–55. ISSN: 14677881. DOI: 10.1111/obr.12343. URL: https://onlinelibrary.wiley.com/doi/10.1111/obr.12343.

Stanford, J. N. and Mccabe, M. P. (2002). 'Body Image Ideal among Males and Females: Sociocultural Influences and Focus on Different Body Parts'. In: *Journal of Health Psychology* 7.6, pp. 675–684. ISSN: 1359-1053. DOI: 10.1177/1359105302007006871. URL: http://journals.sagepub.com/doi/10.1177/1359105302007006871.

Stansfeld, S. A. et al. (2003). 'Social inequalities in depressive symptoms and physical functioning in the Whitehall II study: exploring a common cause explanation'. In: *Journal of Epidemiology & Community Health* 57.5, pp. 361–367. ISSN: 0143005X. DOI: 10.1136/jech.57.5.361. URL: https://jech.bmj.com/lookup/doi/10.1136/jech.57.5.361.

Stegenga, B. T. et al. (2012). 'Depression, anxiety and physical function: exploring the strength of causality'. In: *Journal of Epidemiology and Community Health* 66.7, e25–e25. ISSN: 0143-005X. DOI: 10.1136/jech.2010.128371. URL: https://jech.bmj.com/lookup/doi/10.1136/jech.2010.128371.

Stevens, J. (1992). *Applied multivariate statistics for the social sciences, 2nd ed.* Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc, pp. xvii, 629–xvii, 629. ISBN: 0-8058-1154-0 (Hardcover); 0-8058-1197-4 (Paperback).

Stevens, S. D. et al. (2017). 'Adult and childhood weight influence body image and depression through weight stigmatization'. In: *Journal of Health Psychology* 22.8, pp. 1084–1093. ISSN: 1359-1053. DOI: 10.1177/1359105315624749. URL: http://journals.sagepub.com/doi/10.1177/1359105315624749.

Stunkard, A. J., Faith, M. S. and Allison, K. C. (2003). 'Depression and Obesity'. In: *Biological Psychiatry* 54.330-7. DOI: 10.1016/S0006-3223(03)00608-5. URL: https://ac.els-cdn.com/S0006322303006085/1-s2.0-S0006322303006085-

main.pdf?{_}tid=44fe605e-ace0-11e7-850e-00000aab0f6b{\&}acdnat=1507546669{_}5fd426a339f5d2b13f40e91b653629eb.

Suk, S.-H. et al. (2003). 'Abdominal Obesity and Risk of Ischemic Stroke: The Northern Manhattan Stroke Study'. In: *Stroke* 34.7, pp. 1586–1592. ISSN: 0039-2499. DOI: 10.1161/01.STR.0000075294.98582.2F. URL: http://stroke.ahajournals.org/cgi/doi/10.1161/01.STR.0000075294.98582.2F.

Sutin, A. R. and Terracciano, A. (2013). 'Perceived Weight Discrimination and Obesity'. In: *PLoS ONE* 8.7. Ed. by R. L. Newton, e70048. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0070048. URL: https://dx.plos.org/10.1371/journal.pone.0070048.

Svedberg, P. et al. (2016). 'The validity of socioeconomic status measures among adolescents based on self-reported information about parents occupations, FAS and perceived SES; Implication for health related quality of life studies'. In: *BMC Medical Research Methodology* 16.1. ISSN: 14712288. DOI: 10.1186/s12874-016-0148-9. URL: http://dx.doi.org/10.1186/s12874-016-0148-9.

Tanumihardjo, S. A. et al. (2009). 'Strategies to Increase Vegetable or Reduce Energy and Fat Intake Induce Weight Loss in Adults'. In: *Experimental Biology and Medicine* 234.5, pp. 542–552. ISSN: 1535-3702. DOI: 10.3181/0810-RM-293. URL: https://doi.org/10.3181/0810-RM-293.

Tapsell, L. C. et al. (2014). 'Weight loss effects from vegetable intake: a 12-month randomised controlled trial'. In: *European Journal of Clinical Nutrition* 68.7, pp. 778–785. ISSN: 0954-3007. DOI: 10.1038/ejcn.2014.39. URL: http://www.nature.com/articles/ejcn201439.

Tarlov, A. R. et al. (1989). 'The Medical Outcomes Study. An application of methods for monitoring the results of medical care'. In: *JAMA: The Journal of the American Medical Association* 262.7, pp. 925–930. ISSN: 00987484. DOI: 10.1001/jama.262.7.925. URL: http://jama.ama-assn.org/cgi/doi/10.1001/jama.262.7.925.

Taylor, H. L. et al. (1978). 'A questionnaire for the assessment of leisure time physical activities'. In: *Journal of Chronic Diseases* 31.12, pp. 741–755. ISSN: 00219681. DOI: 10.1016/0021-9681(78)90058-9. URL: https://linkinghub.elsevier.com/retrieve/pii/0021968178900589.

Taylor, P., Funk, C. and Craighill, P. (2006). 'Americans See Weight Problems Everywhere But In the Mirror'. In: *Pew Research Center*.

Teychenne, M., Ball, K. and Salmon, J. (2008). 'Physical activity and likelihood of depression in adults: A review'. In: *Preventive Medicine* 46.5, pp. 397–411. ISSN: 00917435. DOI: 10.1016/j.ypmed.2008.01.009.

Thiele, J. C., Kurth, W. and Grimm, V. (2014). 'Jan C . Thiele , Winfried Kurth and Volker Grimm (2014) Facilitating Parameter Estimation and Sensitivity Analysis of Agent-Based Models : A Cookbook Using NetLogo and R'. In: *Journal of Artificial Societies and Social Simulation* 17.3, pp. 1–45. ISSN: 1460-7425.

Thompson, F. E. and Subar, A. F. (2013). 'Dietary Assessment Methodology'. In: *Nutrition in the Prevention and Treatment of Disease*. Elsevier, pp. 5–46. ISBN: 9780123918840. DOI: 10.1016/B978-0-12-391884-0.00001-9. URL: https://linkinghub.elsevier.com/retrieve/pii/B9780123918840000019.

Tiggemann, M (2012). 'Sociocultural Perspectives on Body Image'. In: *Encyclopedia of Body Image and Human Appearance*. San Diego, CA, US: Elsevier, pp. 758–765. ISBN: 978-0-12-386451-2 (Hardcover). DOI: 10.1016/B978-0-12-384925-0.00120-6. URL: https://linkinghub.elsevier.com/retrieve/pii/B9780123849250001206.

Tobias, D. K. et al. (2015). 'Effect of low-fat diet interventions versus other diet interventions on long-term weight change in adults: a systematic review and meta-analysis'. In: *The Lancet Diabetes & Endocrinology* 3.12, pp. 968–979. ISSN: 22138587. DOI: 10.1016/S2213-8587(15)00367-8. URL: http://www.ncbi.nlm.nih.gov/pubmed/26527511https://linkinghub.elsevier.com/retrieve/pii/S2213858715003678.

Tomiyama, A. J. (2014). 'Weight stigma is stressful. A review of evidence for the Cyclic Obesity/Weight-Based Stigma model'. In: *Appetite* 82, pp. 8–15. ISSN: 01956663. DOI: 10.1016/j.appet.2014.06.108. URL: http://dx.doi.org/10.1016/j.appet.2014.06.108https://linkinghub.elsevier.com/retrieve/pii/S0195666314003560.

Truesdale, K. P. and Stevens, J. (2009). 'Do the obese know they are obese?' In: *North Carolina medical journal* 69.3, pp. 188–94. ISSN: 0029-2559. arXiv: NIHMS150003. URL: http://www.ncbi.nlm.nih.gov/pubmed/18751350http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3234679.

Trumbo, P. R. and Rivers, C. R. (2014). 'Systematic review of the evidence for an association between sugar-sweetened beverage consumption and risk of obesity'. In: *Nutrition Reviews* 72.9, pp. 566–574. ISSN: 00296643. DOI: 10.1111/nure.12128. URL: https://academic.oup.com/nutritionreviews/article-lookup/doi/10.1111/nure.12128.

Tu, Y.-K. (2012). 'Directed acyclic graphs and structural equation modelling'. In: *Modern Methods for Epidemiology*. Springer, pp. 191–203.

Usami, S., Murayama, K. and Hamaker, E. L. (2019). 'A unified framework of longitudinal models to examine reciprocal relations.' In: *Psychological Methods* 24.5, pp. 637–657. ISSN: 1939-1463. DOI: 10.1037/met0000210. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/met0000210.

Van Buuren, S. et al. (2006). 'Fully conditional specification in multivariate imputation'. In: *Journal of Statistical Computation and Simulation* 76.12, pp. 1049–1064. ISSN: 00949655. DOI: 10.1080/10629360600810434.

Van Buuren, S. (2012). *Flexible Imputation of Missing Data*. 1st Editio. New York: Chapman and Hall/CRC, p. 342. ISBN: 9780429065408. DOI: https://doi.org/10.1201/b11826.

Van Buuren, S. and Groothuis-Oudshoorn, K. (2011). 'mice: Multivariate imputation by chained equations in R'. In: *Journal of Statistical Software* 45.3, pp. 1–67. ISSN: 15487660.

Vandenbroeck, P., Goossens, J. and Clemens, M. (2007). *Tackling Obesities: Future Choices - Building the Obesity System Map*. Tech. rep. s1. Government Office for Science, Foresight. DOI: 10.1111/j.1467-789X.2007.00344.x. URL: www.foresight.gov.ukhttp://www.ncbi.nlm.nih.gov/pubmed/17316292https://onlinelibrary.wiley.com/doi/10.1111/j.1467-789X.2007.00344.x.

Vartanian, L. R., Pinkus, R. T. and Smyth, J. M. (2014). 'The phenomenology of weight stigma in everyday life'. In: *Journal of Contextual Behavioral Science* 3.3, pp. 196–202. ISSN: 22121447. DOI: 10.1016/j.jcbs.2014.01.003. URL: http://dx.doi.org/10.1016/j.jcbs.2014.01.003https://linkinghub.elsevier.com/retrieve/pii/S2212144714000040.

Vartanian, L. R., Pinkus, R. T. and Smyth, J. M. (2018). 'Experiences of weight stigma in everyday life: Implications for health motivation.' In: *Stigma and*

Health 3.2, pp. 85–92. ISSN: 2376-6964. DOI: 10.1037/sah0000077. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/sah0000077.

Ware, J. et al. (1993). 'SF36 Health Survey: Manual and Interpretation Guide'. In: *Lincoln, RI: Quality Metric, Inc, 1993* 30.

Ware, J. E. and Sherbourne, C. D. (1992). 'The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection.' In: *Medical care* 30.6, pp. 473–83. ISSN: 0025-7079. URL: https://pubmed.ncbi.nlm.nih.gov/1593914/http://www.ncbi.nlm.nih.gov/pubmed/1593914.

Wareham, N. (2007). 'Physical activity and obesity prevention'. In: *Obesity Reviews* 8.s1, pp. 109–114. ISSN: 1467-7881. DOI: 10.1111/j.1467-789X.2007.00328.x. URL: http://doi.wiley.com/10.1111/j.1467-789X.2007.00328.x.

Wen, L., Terrera, G. M. and Seaman, S. R. (2018). 'Methods for handling longitudinal outcome processes truncated by dropout and death'. In: *Biostatistics* 19.4, pp. 407–425. ISSN: 14684357. DOI: 10.1093/biostatistics/kxx045.

Westfall, P. H. (2014). 'Kurtosis as Peakedness, 1905 – 2014. R.I.P.' In: *American Statisticians* 68.3, pp. 191–195. DOI: 10.1080/00031305.2014.917055.Kurtosis.

Whigham, L. D. et al. (2012). 'Increased vegetable and fruit consumption during weight loss effort correlates with increased weight and fat loss'. In: *Nutrition & Diabetes* 2.10, e48–e48. ISSN: 2044-4052. DOI: 10.1038/nutd.2012.22. URL: http://dx.doi.org/10.1038/nutd.2012.22http://www.nature.com/articles/nutd201222.

Wilenski, U. (1999). *NetLogo*. URL: http://ccl.northwestern.edu/netlogo/.

Wilenski, U. (2021). *NetLogo User Manual*. URL: http://ccl.northwestern.edu/netlogo/docs/ (visited on 03/11/2021).

Wilensky, U. and Reisman, K. (2006). 'Thinking Like a Wolf, a Sheep, or a Firefly: Learning Biology Through Constructing and Testing Computational Theories—An Embodied Modeling Approach'. In: *Cognition and Instruction* 24.2, pp. 171–209. ISSN: 0737-0008. DOI: 10.1207/s1532690xci2402_1. URL: http://www.tandfonline.com/doi/abs/10.1207/s1532690xci2402{_}1.

Wing, R. R. and Phelan, S. (2005). 'Long-term weight loss maintenance'. In: *American Journal of Clinical Nutrition* 82.May, pp. 222–225.

Woodward, J. (2002). 'What Is a Mechanism? A Counterfactual Account'. In: *Philosophy of Science* 69.S3. Ed. by R. A. Jackson, S366–S377. ISSN: 0031-8248. DOI: 10.1086/341859. URL: http://ebook.rsc.org/?DOI=10.1039/9781847551337-00001https://www.journals.uchicago.edu/doi/10.1086/341859.

World Health Organization (2021). *Obesity and Overweight*. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.

Wright, S. (1960). 'Path Coefficients and Path Regressions: Alternative or Complementary Concepts?' In: *Biometrics* 16.2, p. 189. ISSN: 0006341X. DOI: 10.2307/2527551. URL: https://www.jstor.org/stable/2527551?origin=crossref.

Yanovski, S. (1993). 'Binge eating disorder: current knowledge and future directions.' In: *Obesity Research* 1.4, pp. 306–324.

Yuan, K.-H. and Bentler, P. M. (2000). '5. Three Likelihood-Based Methods for Mean and Covariance Structure Analysis with Nonnormal Missing Data'. In: *Sociological Methodology* 30.1, pp. 165–200. DOI: 10.1111/0081-1750.00078. URL: http://journals.sagepub.com/doi/10.1111/0081-1750.00078.

Zyphur, M. J. et al. (2020). 'From Data to Causes I: Building A General Cross-Lagged Panel Model (GCLM)'. In: *Organizational Research Methods* 23.4, pp. 651–687. ISSN: 1094-4281. DOI: 10.1177/1094428119847278. URL: http://journals.sagepub.com/doi/10.1177/1094428119847278.

Appendix A

Example code from SEM Models

Below is an example of the model code used for the SEM Model presented in Chapter 5. For the sake of brevity, not all the separate models have been included, as much of the code remains the same across each model.

A.1 The Strong Invariant Model

```
TITLE: Strong model for gender-sex multigroup analysis
DATA: FILE = "Wh_Model_final_measures_no_diet_lat.dat";
VARIABLE:
NAMES = ID stdygrp tage_c statusx vdepres xhlthyr grlump educ xbmi xghq24 xghq25
xghq29 xghq30 vtpfsc vtbpsc vtrlppsc xfamprb2 tbmi tghq24 tghq25 tghq29 tghq30
ttpfsc ttbpsc ttrlppsc tcalssc tcarbssc tfatsc ttmetsc tfamprb5 tfamprb6 mbmi
mghq24 mghq25 mghq29 mghq30 mtpfsc mtbpsc mtrlppsc mcalssc mcarbssc mfatsc
mtmetsc mfamprb5 mfamprb6 jbmi jghq24 jghq25 jghq29 jghq30 jtpfsc jtbpsc
jtrlppsc jcalssc jcarbssc jfatsc jtmetsc jfamprb5 jfamprb6;

USEVARIABLES = ID stdygrp tage_c statusx vdepres xhlthyr grlump educ xbmi
xghq24 xghq25 xghq29 xghq30 vtpfsc vtbpsc vtrlppsc xfamprb2
tbmi tghq24 tghq25 tghq29 tghq30 ttpfsc ttbpsc ttrlppsc
tcalssc ttmetsc tfamprb5 tfamprb6
mbmi mghq24 mghq25 mghq29 mghq30 mtpfsc mtbpsc mtrlppsc mcalssc
mtmetsc mfamprb5 mfamprb6 jbmi jghq24 jghq25 jghq29 jghq30 jtpfsc jtbpsc
jtrlppsc jcalssc jtmetsc jfamprb5 jfamprb6;

IDVARIABLE = ID;
GROUPING IS stdygrp (1 = non-white-female
2 = non-white-male
3 = white-female
```

```
4 = white-male);

AUXILIARY = statusx vdepres xhlthyr grlump educ xbmi
xghq24 xghq25 xghq29 xghq30 vtpfsc vtbpsc vtrlppsc xfamprb2;

MISSING=.;

ANALYSIS:
ESTIMATOR = MLR; !Use Robust Maximum likelihood estimation

MODEL:
!------ Depression construct
Dep1 BY tghq24* tghq25 tghq29 tghq30 (L1-L4);
Dep2 BY mghq24* mghq25 mghq29 mghq30 (L1-L4);
Dep3 BY jghq24* jghq25 jghq29 jghq30 (L1-L4);
!Set scale using fixed factor method, free at t2 and t3
Dep1@1 Dep2* Dep3*;
!Fix factor means to zero, free at t2 and t3
[Dep1@0 Dep2* Dep3*];

!------ Physical Function
pf1 BY ttpfsc* ttrlppsc ttbpsc (L5-L7);
pf2 BY mtpfsc* mtrlppsc mtbpsc (L5-L7);
pf3 BY jtpfsc* jtrlppsc jtbpsc (L5-L7);
!Set scale using fixed factor method, free at t2 and t3
PF1@1 PF2* PF3*;
!Fix factor means to zero, free at t2 and t3
[PF1@0 PF2* PF3*];

!Socioeconomic Position
SEP1 BY tfamprb5* (L8)
tfamprb6 (L8);
SEP2 BY mfamprb5* (L8)
mfamprb6 (L8);
SEP3 BY jfamprb5* (L8)
jfamprb6 (L8);
!Set scale using fixed factor method, free at t2 and t3
SEP1@1 SEP2* SEP3*;
```

```
!Fix factor means to zero, free at t2 and t3
[SEP1@0 SEP2* SEP3*];

!------- covariance between observed and latent vars

Dep1 WITH DEP2 DEP3 PF1 PF2 PF3 SEP1 SEP2 SEP3 (C1-C8)
tage_c  (C9)
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C10-C18);

Dep2 WITH DEP3 PF1 PF2 PF3 SEP1 SEP2 SEP3 (C19-C25)
tage_c  (C26)
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C27-C35);

DEP3 WITH PF1 PF2 PF3 SEP1 SEP2 SEP3 (C36-C41)
tage_c  (C42)
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C43-C51);

PF1 WITH PF2 PF3 SEP1 SEP2 SEP3 (C52-C56)
tage_c  (C57)
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C58-C66);

PF2 WITH PF3 SEP1 SEP2 SEP3 (C67-C70)
tage_c  (C71)
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C72-C80);

PF3 WITH SEP1 SEP2 SEP3 (C81-C83)
tage_c  (C84)
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C85-C93);

SEP1 WITH SEP2 SEP3 (C94-C95)
tage_c  (C96)
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C97-C105);

SEP2 WITH SEP3 (C106)
tage_c  (C107)
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C108-C116);

SEP3 WITH tage_c  (C117)
```

```
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C118-C126);

!BMI
tbmi WITH tage_c  (C127)
mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C128-C135);
mbmi WITH tage_c  (C136)
jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C137-C143);
jbmi WITH tage_c  (C144)
tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C145-C150);

!Activity vars
ttmetsc WITH tage_c  tcalssc mcalssc jcalssc mtmetsc jtmetsc (C151-C156);
mtmetsc WITH tage_c  tcalssc mcalssc jcalssc jtmetsc (C157-C161);
jtmetsc WITH tage_c  tcalssc mcalssc jcalssc (C162-C165);

!Diet variables
tcalssc WITH tage_c  mcalssc jcalssc (C166-C168);
mcalssc WITH tage_c  jcalssc (C169-C170);
jcalssc WITH tage_c  (C171);


!------ Correlated Residuals --------!

!Depression vars
tghq24 WITH mghq24 jghq24;
mghq24 WITH jghq24;

tghq25 WITH mghq25 jghq25;
mghq25 WITH jghq25;

tghq29 WITH mghq29 jghq29;
mghq29 WITH jghq29;

tghq30 WITH mghq30 jghq30;
mghq30 WITH jghq30;

!PF vars
ttpfsc WITH mtpfsc jtpfsc;
```

```
mtpfsc WITH jtpfsc;

ttrlppsc WITH mtrlppsc jtrlppsc;
mtrlppsc WITH jtrlppsc;

ttbpsc WITH mtbpsc jtbpsc;
mtbpsc WITH jtbpsc;

!SEP vars
tfamprb5 WITH mfamprb5 jfamprb5;
mfamprb5 WITH jfamprb5;

tfamprb6 WITH mfamprb6 jfamprb6;
mfamprb6 WITH jfamprb6;

!Item intercepts
[tghq24 tghq25 tghq29 tghq30] (I1-I4);
[ttpfsc ttrlppsc ttbpsc] (I5-I7);
[tfamprb5 tfamprb6] (I8-I9);

[mghq24 mghq25 mghq29 mghq30] (I1-I4);
[mtpfsc mtrlppsc mtbpsc] (I5-I7);
[mfamprb5 mfamprb6] (I8-I9);

[jghq24 jghq25 jghq29 jghq30] (I1-I4);
[jtpfsc jtrlppsc jtbpsc] (I5-I7);
[jfamprb5 jfamprb6] (I8-I9);

!Manifest variances
tbmi mbmi jbmi (V10-V12);
ttmetsc mtmetsc jtmetsc (V13-V15);
tcalssc mcalssc jcalssc (V16-V18);
tage_c  (V19);

!Factor means
[tbmi mbmi jbmi];
[ttmetsc mtmetsc jtmetsc];
[tcalssc mcalssc jcalssc];
```

```
MODEL non-white-female:
!------ Depression construct
Dep1 BY tghq24* tghq25 tghq29 tghq30 (L1-L4);
Dep2 BY mghq24* mghq25 mghq29 mghq30 (L1-L4);
Dep3 BY jghq24* jghq25 jghq29 jghq30 (L1-L4);
!Set scale using fixed factor method, free at all time points
Dep1@1 Dep2* Dep3*;
!Factor means freed
[Dep1@0 Dep2* Dep3*];

!------ Physical Function
pf1 BY ttpfsc* ttrlppsc ttbpsc (L5-L7);
pf2 BY mtpfsc* mtrlppsc mtbpsc (L5-L7);
pf3 BY jtpfsc* jtrlppsc jtbpsc (L5-L7);
!Set scale using fixed factor method, free at all time points
PF1@1 PF2* PF3*;
!Factor means freed
[PF1@0 PF2* PF3*];

!Socioeconomic Position
SEP1 BY tfamprb5* (L8)
tfamprb6 (L8);
SEP2 BY mfamprb5* (L8)
mfamprb6 (L8);
SEP3 BY jfamprb5* (L8)
jfamprb6 (L8);
!Set scale using fixed factor method, free at all time points
SEP1@1 SEP2* SEP3*;
!Factor means freed
[SEP1@0 SEP2* SEP3*];

!------- covariance between observed and latent vars

Dep1 WITH DEP2 DEP3 PF1 PF2 PF3 SEP1 SEP2 SEP3 (C1-C8)
tage_c  (C9)
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C10-C18);
```

```
Dep2 WITH DEP3 PF1 PF2 PF3 SEP1 SEP2 SEP3 (C19-C25)
tage_c  (C26)
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C27-C35);

DEP3 WITH PF1 PF2 PF3 SEP1 SEP2 SEP3 (C36-C41)
tage_c  (C42)
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C43-C51);

PF1 WITH PF2 PF3 SEP1 SEP2 SEP3 (C52-C56)
tage_c  (C57)
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C58-C66);

PF2 WITH PF3 SEP1 SEP2 SEP3 (C67-C70)
tage_c  (C71)
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C72-C80);

PF3 WITH SEP1 SEP2 SEP3 (C81-C83)
tage_c  (C84)
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C85-C93);

SEP1 WITH SEP2 SEP3 (C94-C95)
tage_c  (C96)
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C97-C105);

SEP2 WITH SEP3 (C106)
tage_c  (C107)
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C108-C116);

SEP3 WITH tage_c  (C117)
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C118-C126);

!BMI
tbmi WITH tage_c  (C127)
mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C128-C135);
mbmi WITH tage_c  (C136)
jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C137-C143);
jbmi WITH tage_c  (C144)
tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc (C145-C150);
```

```
!Activity vars
ttmetsc WITH tage_c  tcalssc mcalssc jcalssc mtmetsc jtmetsc (C151-C156);
mtmetsc WITH tage_c  tcalssc mcalssc jcalssc jtmetsc (C157-C161);
jtmetsc WITH tage_c  tcalssc mcalssc jcalssc (C162-C165);

!Diet variables
tcalssc WITH tage_c  mcalssc jcalssc (C166-C168);
mcalssc WITH tage_c  jcalssc (C169-C170);
jcalssc WITH tage_c  (C171);


!------ Correlated Residuals --------!

!Depression vars
tghq24 WITH mghq24 jghq24;
mghq24 WITH jghq24;

tghq25 WITH mghq25 jghq25;
mghq25 WITH jghq25;

tghq29 WITH mghq29 jghq29;
mghq29 WITH jghq29;

tghq30 WITH mghq30 jghq30;
mghq30 WITH jghq30;

!PF vars
ttpfsc WITH mtpfsc jtpfsc;
mtpfsc WITH jtpfsc;

ttrlppsc WITH mtrlppsc jtrlppsc;
mtrlppsc WITH jtrlppsc;

ttbpsc WITH mtbpsc jtbpsc;
mtbpsc WITH jtbpsc;

!SEP vars
```

```
tfamprb5 WITH mfamprb5 jfamprb5;
mfamprb5 WITH jfamprb5;

tfamprb6 WITH mfamprb6 jfamprb6;
mfamprb6 WITH jfamprb6;

!Item intercepts
[tghq24 tghq25 tghq29 tghq30] (I1-I4);
[ttpfsc ttrlppsc ttbpsc] (I5-I7);
[tfamprb5 tfamprb6] (I8-I9);

[mghq24 mghq25 mghq29 mghq30] (I1-I4);
[mtpfsc mtrlppsc mtbpsc] (I5-I7);
[mfamprb5 mfamprb6] (I8-I9);

[jghq24 jghq25 jghq29 jghq30] (I1-I4);
[jtpfsc jtrlppsc jtbpsc] (I5-I7);
[jfamprb5 jfamprb6] (I8-I9);

!Manifest variances
tbmi mbmi jbmi (V10-V12);
ttmetsc mtmetsc jtmetsc (V13-V15);
tcalssc mcalssc jcalssc (V16-V18);
tage_c  (V19);

!Factor means
[tbmi mbmi jbmi];
[ttmetsc mtmetsc jtmetsc];
[tcalssc mcalssc jcalssc];

MODEL non-white-male:
!------ Depression construct
Dep1 BY tghq24* tghq25 tghq29 tghq30 (L1-L4);
Dep2 BY mghq24* mghq25 mghq29 mghq30 (L1-L4);
Dep3 BY jghq24* jghq25 jghq29 jghq30 (L1-L4);
!Set scale using fixed factor method, free at all time points
Dep1* Dep2* Dep3*;
!Factor means freed
```

```
[Dep1* Dep2* Dep3*];

!------ Physical Function
pf1 BY ttpfsc* ttrlppsc ttbpsc (L5-L7);
pf2 BY mtpfsc* mtrlppsc mtbpsc (L5-L7);
pf3 BY jtpfsc* jtrlppsc jtbpsc (L5-L7);
PF1* PF2* PF3*; !Set scale using fixed factor method, free at t2 and t3
!Factor means freed
[PF1* PF2* PF3*];

!Socioeconomic Position
SEP1 BY tfamprb5* (L8)
tfamprb6 (L8);
SEP2 BY mfamprb5* (L8)
mfamprb6 (L8);
SEP3 BY jfamprb5* (L8)
jfamprb6 (L8);
!Set scale using fixed factor method, free at all time points
SEP1* SEP2* SEP3*;
!Factor means freed
[SEP1* SEP2* SEP3*];


!------- covariance between observed and latent vars

Dep1 WITH DEP2 DEP3 PF1 PF2 PF3 SEP1 SEP2 SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

Dep2 WITH DEP3 PF1 PF2 PF3 SEP1 SEP2 SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

DEP3 WITH PF1 PF2 PF3 SEP1 SEP2 SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

PF1 WITH PF2 PF3 SEP1 SEP2 SEP3
```

```
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

PF2 WITH PF3 SEP1 SEP2 SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

PF3 WITH SEP1 SEP2 SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc ;

SEP1 WITH SEP2 SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

SEP2 WITH SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

SEP3 WITH tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

!BMI
tbmi WITH tage_c
mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;
mbmi WITH tage_c
jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;
jbmi WITH tage_c
tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

!Activity vars
ttmetsc WITH tage_c  tcalssc mcalssc jcalssc mtmetsc jtmetsc;
mtmetsc WITH tage_c  tcalssc mcalssc jcalssc jtmetsc;
jtmetsc WITH tage_c  tcalssc mcalssc jcalssc;

!Diet variables
tcalssc WITH tage_c  mcalssc jcalssc;
mcalssc WITH tage_c  jcalssc;
```

```
jcalssc WITH tage_c ;

!------ Correlated Residuals --------!

!Depression vars
tghq24 WITH mghq24 jghq24;
mghq24 WITH jghq24;

tghq25 WITH mghq25 jghq25;
mghq25 WITH jghq25;

tghq29 WITH mghq29 jghq29;
mghq29 WITH jghq29;

tghq30 WITH mghq30 jghq30;
mghq30 WITH jghq30;

!PF vars
ttpfsc WITH mtpfsc jtpfsc;
mtpfsc WITH jtpfsc;

ttrlppsc WITH mtrlppsc jtrlppsc;
mtrlppsc WITH jtrlppsc;

ttbpsc WITH mtbpsc jtbpsc;
mtbpsc WITH jtbpsc;

!SEP vars
tfamprb5 WITH mfamprb5 jfamprb5;
mfamprb5 WITH jfamprb5;

tfamprb6 WITH mfamprb6 jfamprb6;
mfamprb6 WITH jfamprb6;

!Item intercepts
[tghq24 tghq25 tghq29 tghq30] (I1-I4);
[ttpfsc ttrlppsc ttbpsc] (I5-I7);
[tfamprb5 tfamprb6] (I8-I9);
```

```
[mghq24 mghq25 mghq29 mghq30] (I1-I4);
[mtpfsc mtrlppsc mtbpsc] (I5-I7);
[mfamprb5 mfamprb6] (I8-I9);

[jghq24 jghq25 jghq29 jghq30] (I1-I4);
[jtpfsc jtrlppsc jtbpsc] (I5-I7);
[jfamprb5 jfamprb6] (I8-I9);

!Manifest variances
tbmi mbmi jbmi;
ttmetsc mtmetsc jtmetsc;
tcalssc mcalssc jcalssc;
tage_c ;

!Manifest means
[tbmi mbmi jbmi];
[ttmetsc mtmetsc jtmetsc];
[tcalssc mcalssc jcalssc];


MODEL white-female:
!------ Depression construct
Dep1 BY tghq24* tghq25 tghq29 tghq30 (L1-L4);
Dep2 BY mghq24* mghq25 mghq29 mghq30 (L1-L4);
Dep3 BY jghq24* jghq25 jghq29 jghq30 (L1-L4);
!Set scale using fixed factor method, free at all time points
Dep1* Dep2* Dep3*;
!Factor means freed
[Dep1* Dep2* Dep3*];

!------ Physical Function
pf1 BY ttpfsc* ttrlppsc ttbpsc (L5-L7);
pf2 BY mtpfsc* mtrlppsc mtbpsc (L5-L7);
pf3 BY jtpfsc* jtrlppsc jtbpsc (L5-L7);
!Set scale using fixed factor method, free at all time points
PF1* PF2* PF3*;
!Factor means freed
```

```
[PF1* PF2* PF3*];

!Socioeconomic Position
SEP1 BY tfamprb5* (L8)
tfamprb6 (L8);
SEP2 BY mfamprb5* (L8)
mfamprb6 (L8);
SEP3 BY jfamprb5* (L8)
jfamprb6 (L8);
!Set scale using fixed factor method, free at all time points
SEP1* SEP2* SEP3*;
!Factor means freed
[SEP1* SEP2* SEP3*];

!------- covariance between observed and latent vars

Dep1 WITH DEP2 DEP3 PF1 PF2 PF3 SEP1 SEP2 SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

Dep2 WITH DEP3 PF1 PF2 PF3 SEP1 SEP2 SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

DEP3 WITH PF1 PF2 PF3 SEP1 SEP2 SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

PF1 WITH PF2 PF3 SEP1 SEP2 SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

PF2 WITH PF3 SEP1 SEP2 SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

PF3 WITH SEP1 SEP2 SEP3
tage_c
```

```
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc ;

SEP1 WITH SEP2 SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

SEP2 WITH SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

SEP3 WITH tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

!BMI
tbmi WITH tage_c
mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;
mbmi WITH tage_c
jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;
jbmi WITH tage_c
tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

!Activity vars
ttmetsc WITH tage_c  tcalssc mcalssc jcalssc mtmetsc jtmetsc;
mtmetsc WITH tage_c  tcalssc mcalssc jcalssc jtmetsc;
jtmetsc WITH tage_c  tcalssc mcalssc jcalssc;

!Diet variables
tcalssc WITH tage_c  mcalssc jcalssc;
mcalssc WITH tage_c  jcalssc;
jcalssc WITH tage_c ;


!------ Correlated Residuals --------!

!Depression vars
tghq24 WITH mghq24 jghq24;
mghq24 WITH jghq24;
```

```
tghq25 WITH mghq25 jghq25;
mghq25 WITH jghq25;

tghq29 WITH mghq29 jghq29;
mghq29 WITH jghq29;

tghq30 WITH mghq30 jghq30;
mghq30 WITH jghq30;

!PF vars
ttpfsc WITH mtpfsc jtpfsc;
mtpfsc WITH jtpfsc;

ttrlppsc WITH mtrlppsc jtrlppsc;
mtrlppsc WITH jtrlppsc;

ttbpsc WITH mtbpsc jtbpsc;
mtbpsc WITH jtbpsc;

!SEP vars
tfamprb5 WITH mfamprb5 jfamprb5;
mfamprb5 WITH jfamprb5;

tfamprb6 WITH mfamprb6 jfamprb6;
mfamprb6 WITH jfamprb6;

!Item intercepts
[tghq24 tghq25 tghq29 tghq30] (I1-I4);
[ttpfsc ttrlppsc ttbpsc] (I5-I7);
[tfamprb5 tfamprb6] (I8-I9);

[mghq24 mghq25 mghq29 mghq30] (I1-I4);
[mtpfsc mtrlppsc mtbpsc] (I5-I7);
[mfamprb5 mfamprb6] (I8-I9);

[jghq24 jghq25 jghq29 jghq30] (I1-I4);
[jtpfsc jtrlppsc jtbpsc] (I5-I7);
[jfamprb5 jfamprb6] (I8-I9);
```

```
!Manifest variances
tbmi mbmi jbmi;
ttmetsc mtmetsc jtmetsc;
tcalssc mcalssc jcalssc;
tage_c ;

!Manifest means
[tbmi mbmi jbmi];
[ttmetsc mtmetsc jtmetsc];
[tcalssc mcalssc jcalssc];


MODEL white-male:
!------ Depression construct
Dep1 BY tghq24* tghq25 tghq29 tghq30 (L1-L4);
Dep2 BY mghq24* mghq25 mghq29 mghq30 (L1-L4);
Dep3 BY jghq24* jghq25 jghq29 jghq30 (L1-L4);
!Set scale using fixed factor method, free at all time points
Dep1* Dep2* Dep3*;
!Factor means free
[Dep1* Dep2* Dep3*];

!------ Physical Function
pf1 BY ttpfsc* ttrlppsc ttbpsc (L5-L7);
pf2 BY mtpfsc* mtrlppsc mtbpsc (L5-L7);
pf3 BY jtpfsc* jtrlppsc jtbpsc (L5-L7);
!Set scale using fixed factor method, free at all time points
PF1* PF2* PF3*;
!Free factor means
[PF1* PF2* PF3*];

!Socioeconomic Position
SEP1 BY tfamprb5* (L8)
tfamprb6 (L8);
SEP2 BY mfamprb5* (L8)
mfamprb6 (L8);
SEP3 BY jfamprb5* (L8)
```

```
jfamprb6 (L8);
!Set scale using fixed factor method, free at all time points
SEP1* SEP2* SEP3*;
!Free factor means
[SEP1* SEP2* SEP3*];

!------- covariance between observed and latent vars

Dep1 WITH DEP2 DEP3 PF1 PF2 PF3 SEP1 SEP2 SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

Dep2 WITH DEP3 PF1 PF2 PF3 SEP1 SEP2 SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

DEP3 WITH PF1 PF2 PF3 SEP1 SEP2 SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

PF1 WITH PF2 PF3 SEP1 SEP2 SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

PF2 WITH PF3 SEP1 SEP2 SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

PF3 WITH SEP1 SEP2 SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc ;

SEP1 WITH SEP2 SEP3
tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

SEP2 WITH SEP3
tage_c
```

```
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

SEP3 WITH tage_c
tbmi mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

!BMI
tbmi WITH tage_c
mbmi jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;
mbmi WITH tage_c
jbmi tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;
jbmi WITH tage_c
tcalssc mcalssc jcalssc ttmetsc mtmetsc jtmetsc;

!Activity vars
ttmetsc WITH tage_c  tcalssc mcalssc jcalssc mtmetsc jtmetsc;
mtmetsc WITH tage_c  tcalssc mcalssc jcalssc jtmetsc;
jtmetsc WITH tage_c  tcalssc mcalssc jcalssc;

!Diet variables
tcalssc WITH tage_c  mcalssc jcalssc;
mcalssc WITH tage_c  jcalssc;
jcalssc WITH tage_c ;



!------ Correlated Residuals --------!

!Depression vars
tghq24 WITH mghq24 jghq24;
mghq24 WITH jghq24;

tghq25 WITH mghq25 jghq25;
mghq25 WITH jghq25;

tghq29 WITH mghq29 jghq29;
mghq29 WITH jghq29;

tghq30 WITH mghq30 jghq30;
```

```
mghq30 WITH jghq30;

!PF vars
ttpfsc WITH mtpfsc jtpfsc;
mtpfsc WITH jtpfsc;

ttrlppsc WITH mtrlppsc jtrlppsc;
mtrlppsc WITH jtrlppsc;

ttbpsc WITH mtbpsc jtbpsc;
mtbpsc WITH jtbpsc;

!SEP vars
tfamprb5 WITH mfamprb5 jfamprb5;
mfamprb5 WITH jfamprb5;

tfamprb6 WITH mfamprb6 jfamprb6;
mfamprb6 WITH jfamprb6;

!Item intercepts
[tghq24 tghq25 tghq29 tghq30] (I1-I4);
[ttpfsc ttrlppsc ttbpsc] (I5-I7);
[tfamprb5 tfamprb6] (I8-I9);

[mghq24 mghq25 mghq29 mghq30] (I1-I4);
[mtpfsc mtrlppsc mtbpsc] (I5-I7);
[mfamprb5 mfamprb6] (I8-I9);

[jghq24 jghq25 jghq29 jghq30] (I1-I4);
[jtpfsc jtrlppsc jtbpsc] (I5-I7);
[jfamprb5 jfamprb6] (I8-I9);

!Manifest variances
tbmi mbmi jbmi;
ttmetsc mtmetsc jtmetsc;
tcalssc mcalssc jcalssc;
tage_c ;
```

```
!Manifest means
[tbmi mbmi jbmi];
[ttmetsc mtmetsc jtmetsc];
[tcalssc mcalssc jcalssc];

OUTPUT:
MOD STAND;
```

A.2 Code for Model 3 - Full Mediation with AR2 Paths

```
TITLE: Full mediation structural model with AR2 paths
DATA: FILE = "Wh_Model_final_measures_no_diet_lat.dat";
VARIABLE:
NAMES = ID stdygrp tage_c statusx vdepres xhlthyr grlump educ xbmi xghq24 xghq25
xghq29 xghq30 vtpfsc vtbpsc vtrlppsc xfamprb2 tbmi tghq24 tghq25 tghq29 tghq30
ttpfsc ttbpsc ttrlppsc tcalssc tcarbssc tfatsc ttmetsc tfamprb5 tfamprb6 mbmi
mghq24 mghq25 mghq29 mghq30 mtpfsc mtbpsc mtrlppsc mcalssc mcarbssc mfatsc
mtmetsc mfamprb5 mfamprb6 jbmi jghq24 jghq25 jghq29 jghq30 jtpfsc jtbpsc
jtrlppsc jcalssc jcarbssc jfatsc jtmetsc jfamprb5 jfamprb6;

USEVARIABLES = ID stdygrp tage_c statusx vdepres xhlthyr grlump educ xbmi
xghq24 xghq25 xghq29 xghq30 vtpfsc vtbpsc vtrlppsc xfamprb2
tbmi tghq24 tghq25 tghq29 tghq30 ttpfsc ttbpsc ttrlppsc
tcalssc ttmetsc tfamprb5 tfamprb6
mbmi mghq24 mghq25 mghq29 mghq30 mtpfsc mtbpsc mtrlppsc mcalssc
mtmetsc mfamprb5 mfamprb6 jbmi jghq24 jghq25 jghq29 jghq30 jtpfsc jtbpsc
jtrlppsc jcalssc jtmetsc jfamprb5 jfamprb6;

IDVARIABLE = ID;
GROUPING IS stdygrp (1 = non-white-female
2 = non-white-male
3 = white-female
4 = white-male);

AUXILIARY = statusx vdepres xhlthyr grlump educ xbmi
```

```
xghq24 xghq25 xghq29 xghq30 vtpfsc vtbpsc vtrlppsc xfamprb2;

MISSING=.;

ANALYSIS:
ESTIMATOR = MLR; !Use Robust Maximum likelihood estimation

MODEL:
!------ Depression construct
Dep1 BY tghq24* tghq25 tghq29 tghq30 (L1-L4);
Dep2 BY mghq24* mghq25 mghq29 mghq30 (L1-L4);
Dep3 BY jghq24* jghq25 jghq29 jghq30 (L1-L4);
!Set scale using fixed factor method, free at t2 and t3
Dep1@1 Dep2* Dep3*;
!Fix factor means to zero, free at t2 and t3
[Dep1@0 Dep2* Dep3*];

!------ Physical Function
pf1 BY ttpfsc* ttrlppsc ttbpsc (L5-L7);
pf2 BY mtpfsc* mtrlppsc mtbpsc (L5-L7);
pf3 BY jtpfsc* jtrlppsc jtbpsc (L5-L7);
!Set scale using fixed factor method, free at t2 and t3
PF1@1 PF2* PF3*;
!Fix factor means to zero, free at t2 and t3
[PF1@0 PF2* PF3*];

!Socioeconomic Position
SEP1 BY tfamprb5* (L8)
tfamprb6 (L8);
SEP2 BY mfamprb5* (L8)
mfamprb6 (L8);
SEP3 BY jfamprb5* (L8)
jfamprb6 (L8);
!Set scale using fixed factor method, free at t2 and t3
SEP1@1 SEP2* SEP3*;
!Fix factor means to zero, free at t2 and t3
[SEP1@0 SEP2* SEP3*];
```

```
!------- Structural Associations ---------!
!--- Time 3 regressed on time 2
Dep3 ON DEP2 PF2 mtmetsc mcalssc mbmi SEP2 tage_c;
PF3 ON PF2 mbmi SEP2 tage_c;
jtmetsc ON DEP2 mtmetsc mbmi SEP2 tage_c;
jcalssc ON DEP2 mcalssc mbmi SEP2 tage_c;
jbmi ON DEP2 mtmetsc mcalssc mbmi SEP2 tage_c;
SEP3 ON SEP2 tage_c;

!--- Time 2 regressed on time 1
Dep2 ON DEP1 PF1 ttmetsc tcalssc tbmi SEP1 tage_c;
PF2 ON PF1 tbmi SEP1 tage_c;
mtmetsc ON DEP1 ttmetsc tbmi SEP1 tage_c;
mcalssc ON DEP1 tcalssc tbmi SEP1 tage_c;
mbmi ON DEP1 ttmetsc tcalssc tbmi SEP1 tage_c;
SEP2 ON SEP1 tage_c;

!--- Time 3 on time 1 (AR2 paths)
Dep3 ON Dep1;
Pf3 on PF1;
jtmetsc on ttmetsc;
jcalssc on tcalssc;
jbmi on tbmi;
SEP3 on SEP1;

!--- Correlations within time points
!--- Time 2
Dep2 WITH mbmi Pf2 mtmetsc mcalssc SEP2;
mbmi WITH Pf2 mtmetsc mcalssc SEP2;
Pf2 WITH mtmetsc mcalssc Sep2;
mtmetsc WITH mcalssc Sep2;
mcalssc WITH Sep2;

!--- Time3
Dep3 WITH jbmi Pf3 jtmetsc jcalssc SEP3;
jbmi WITH Pf3 jtmetsc jcalssc SEP3;
Pf3 WITH jtmetsc jcalssc Sep3;
jtmetsc WITH jcalssc Sep3;
```

```
jcalssc WITH Sep3;

!------ Correlated Residuals --------!

!Depression vars
tghq24 WITH mghq24 jghq24;
mghq24 WITH jghq24;

tghq25 WITH mghq25 jghq25;
mghq25 WITH jghq25;

tghq29 WITH mghq29 jghq29;
mghq29 WITH jghq29;

tghq30 WITH mghq30 jghq30;
mghq30 WITH jghq30;

!PF vars
ttpfsc WITH mtpfsc jtpfsc;
mtpfsc WITH jtpfsc;

ttrlppsc WITH mtrlppsc jtrlppsc;
mtrlppsc WITH jtrlppsc;

ttbpsc WITH mtbpsc jtbpsc;
mtbpsc WITH jtbpsc;

!SEP vars
tfamprb5 WITH mfamprb5 jfamprb5;
mfamprb5 WITH jfamprb5;

tfamprb6 WITH mfamprb6 jfamprb6;
mfamprb6 WITH jfamprb6;

!Item intercepts
[tghq24 tghq25 tghq29 tghq30] (I1-I4);
[ttpfsc ttrlppsc ttbpsc] (I5-I7);
[tfamprb5 tfamprb6] (I8-I9);
```

```
[mghq24 mghq25 mghq29 mghq30] (I1-I4);
[mtpfsc mtrlppsc mtbpsc] (I5-I7);
[mfamprb5 mfamprb6] (I8-I9);

[jghq24 jghq25 jghq29 jghq30] (I1-I4);
[jtpfsc jtrlppsc jtbpsc] (I5-I7);
[jfamprb5 jfamprb6] (I8-I9);

!Manifest variances
tbmi mbmi jbmi (V10-V12);
ttmetsc mtmetsc jtmetsc (V13-V15);
tcalssc mcalssc jcalssc (V16-V18);
tage_c  (V19);

!Factor means
[tbmi mbmi jbmi];
[ttmetsc mtmetsc jtmetsc];
[tcalssc mcalssc jcalssc];


MODEL non-white-female:
!------ Depression construct
Dep1 BY tghq24* tghq25 tghq29 tghq30 (L1-L4);
Dep2 BY mghq24* mghq25 mghq29 mghq30 (L1-L4);
Dep3 BY jghq24* jghq25 jghq29 jghq30 (L1-L4);
!Set scale using fixed factor method, free at t2 and t3
Dep1@1 Dep2* Dep3*;
!Fix factor means to zero, free at t2 and t3
[Dep1@0 Dep2* Dep3*];

!------ Physical Function
pf1 BY ttpfsc* ttrlppsc ttbpsc (L5-L7);
pf2 BY mtpfsc* mtrlppsc mtbpsc (L5-L7);
pf3 BY jtpfsc* jtrlppsc jtbpsc (L5-L7);
!Set scale using fixed factor method, free at t2 and t3
PF1@1 PF2* PF3*;
!Fix factor means to zero, free at t2 and t3
```

```
[PF1@0 PF2* PF3*];

!Socioeconomic Position
SEP1 BY tfamprb5* (L8)
tfamprb6 (L8);
SEP2 BY mfamprb5* (L8)
mfamprb6 (L8);
SEP3 BY jfamprb5* (L8)
jfamprb6 (L8);
!Set scale using fixed factor method, free at t2 and t3
SEP1@1 SEP2* SEP3*;
!Fix factor means to zero, free at t2 and t3
[SEP1@0 SEP2* SEP3*];

!------- Structural Associations ---------!
!--- Time 3 regressed on time 2
Dep3 ON DEP2 PF2 mtmetsc mcalssc mbmi SEP2 tage_c;
PF3 ON PF2 mbmi SEP2 tage_c;
jtmetsc ON DEP2 mtmetsc mbmi SEP2 tage_c;
jcalssc ON DEP2 mcalssc mbmi SEP2 tage_c;
jbmi ON DEP2 mtmetsc mcalssc mbmi SEP2 tage_c;
SEP3 ON SEP2 tage_c;

!--- Time 2 regressed on time 1
Dep2 ON DEP1 PF1 ttmetsc tcalssc tbmi SEP1 tage_c;
PF2 ON PF1 tbmi SEP1 tage_c;
mtmetsc ON DEP1 ttmetsc tbmi SEP1 tage_c;
mcalssc ON DEP1 tcalssc tbmi SEP1 tage_c;
mbmi ON DEP1 ttmetsc tcalssc tbmi SEP1 tage_c;
SEP2 ON SEP1 tage_c;

!--- Time 3 on time 1 (AR2 paths)
Dep3 ON Dep1;
Pf3 on PF1;
jtmetsc on ttmetsc;
jcalssc on tcalssc;
jbmi on tbmi;
SEP3 on SEP1;
```

```
!--- Correlations within time points
!--- Time 2
Dep2 WITH mbmi Pf2 mtmetsc mcalssc SEP2;
mbmi WITH Pf2 mtmetsc mcalssc SEP2;
Pf2 WITH mtmetsc mcalssc Sep2;
mtmetsc WITH mcalssc Sep2;
mcalssc WITH Sep2;

!--- Time3
Dep3 WITH jbmi Pf3 jtmetsc jcalssc SEP3;
jbmi WITH Pf3 jtmetsc jcalssc SEP3;
Pf3 WITH jtmetsc jcalssc Sep3;
jtmetsc WITH jcalssc Sep3;
jcalssc WITH Sep3;


!------ Correlated Residuals --------!

!Depression vars
tghq24 WITH mghq24 jghq24;
mghq24 WITH jghq24;

tghq25 WITH mghq25 jghq25;
mghq25 WITH jghq25;

tghq29 WITH mghq29 jghq29;
mghq29 WITH jghq29;

tghq30 WITH mghq30 jghq30;
mghq30 WITH jghq30;

!PF vars
ttpfsc WITH mtpfsc jtpfsc;
mtpfsc WITH jtpfsc;

ttrlppsc WITH mtrlppsc jtrlppsc;
mtrlppsc WITH jtrlppsc;
```

```
ttbpsc WITH mtbpsc jtbpsc;
mtbpsc WITH jtbpsc;

!SEP vars
tfamprb5 WITH mfamprb5 jfamprb5;
mfamprb5 WITH jfamprb5;

tfamprb6 WITH mfamprb6 jfamprb6;
mfamprb6 WITH jfamprb6;

!Item intercepts
[tghq24 tghq25 tghq29 tghq30] (I1-I4);
[ttpfsc ttrlppsc ttbpsc] (I5-I7);
[tfamprb5 tfamprb6] (I8-I9);

[mghq24 mghq25 mghq29 mghq30] (I1-I4);
[mtpfsc mtrlppsc mtbpsc] (I5-I7);
[mfamprb5 mfamprb6] (I8-I9);

[jghq24 jghq25 jghq29 jghq30] (I1-I4);
[jtpfsc jtrlppsc jtbpsc] (I5-I7);
[jfamprb5 jfamprb6] (I8-I9);

!Manifest variances
tbmi mbmi jbmi (V10-V12);
ttmetsc mtmetsc jtmetsc (V13-V15);
tcalssc mcalssc jcalssc (V16-V18);
tage_c  (V19);

!Factor means
[tbmi mbmi jbmi];
[ttmetsc mtmetsc jtmetsc];
[tcalssc mcalssc jcalssc];

MODEL non-white-male:
!------ Depression construct
Dep1 BY tghq24* tghq25 tghq29 tghq30 (L1-L4);
```

```
Dep2 BY mghq24* mghq25 mghq29 mghq30 (L1-L4);
Dep3 BY jghq24* jghq25 jghq29 jghq30 (L1-L4);
!Set scale using fixed factor method, free at all time points
Dep1* Dep2* Dep3*;
!Factor means freed
[Dep1* Dep2* Dep3*];

!------ Physical Function
pf1 BY ttpfsc* ttrlppsc ttbpsc (L5-L7);
pf2 BY mtpfsc* mtrlppsc mtbpsc (L5-L7);
pf3 BY jtpfsc* jtrlppsc jtbpsc (L5-L7);
!Set scale using fixed factor method, free at all time points
PF1* PF2* PF3*;
!Factor means freed
[PF1* PF2* PF3*];

!Socioeconomic Position
SEP1 BY tfamprb5* (L8)
tfamprb6 (L8);
SEP2 BY mfamprb5* (L8)
mfamprb6 (L8);
SEP3 BY jfamprb5* (L8)
jfamprb6 (L8);
!Set scale using fixed factor method, free at all time points
SEP1* SEP2* SEP3*;
!Factor means freed
[SEP1* SEP2* SEP3*];


!------- Structural Associations ---------!
!--- Time 3 regressed on time 2
Dep3 ON DEP2 PF2 mtmetsc mcalssc mbmi SEP2 tage_c;
PF3 ON PF2 mbmi SEP2 tage_c;
jtmetsc ON DEP2 mtmetsc mbmi SEP2 tage_c;
jcalssc ON DEP2 mcalssc mbmi SEP2 tage_c;
jbmi ON DEP2 mtmetsc mcalssc mbmi SEP2 tage_c;
SEP3 ON SEP2 tage_c;
```

```
!--- Time 2 regressed on time 1
Dep2 ON DEP1 PF1 ttmetsc tcalssc tbmi SEP1 tage_c;
PF2 ON PF1 tbmi SEP1 tage_c;
mtmetsc ON DEP1 ttmetsc tbmi SEP1 tage_c;
mcalssc ON DEP1 tcalssc tbmi SEP1 tage_c;
mbmi ON DEP1 ttmetsc tcalssc tbmi SEP1 tage_c;
SEP2 ON SEP1 tage_c;

!--- Time 3 on time 1 (AR2 paths)
Dep3 ON Dep1;
Pf3 on PF1;
jtmetsc on ttmetsc;
jcalssc on tcalssc;
jbmi on tbmi;
SEP3 on SEP1;

!--- Correlations within time points
!--- Time 2
Dep2 WITH mbmi Pf2 mtmetsc mcalssc SEP2;
mbmi WITH Pf2 mtmetsc mcalssc SEP2;
Pf2 WITH mtmetsc mcalssc Sep2;
mtmetsc WITH mcalssc Sep2;
mcalssc WITH Sep2;

!--- Time3
Dep3 WITH jbmi Pf3 jtmetsc jcalssc SEP3;
jbmi WITH Pf3 jtmetsc jcalssc SEP3;
Pf3 WITH jtmetsc jcalssc Sep3;
jtmetsc WITH jcalssc Sep3;
jcalssc WITH Sep3;

!------ Correlated Residuals --------!

!Depression vars
tghq24 WITH mghq24 jghq24;
mghq24 WITH jghq24;

tghq25 WITH mghq25 jghq25;
```

```
mghq25 WITH jghq25;

tghq29 WITH mghq29 jghq29;
mghq29 WITH jghq29;

tghq30 WITH mghq30 jghq30;
mghq30 WITH jghq30;

!PF vars
ttpfsc WITH mtpfsc jtpfsc;
mtpfsc WITH jtpfsc;

ttrlppsc WITH mtrlppsc jtrlppsc;
mtrlppsc WITH jtrlppsc;

ttbpsc WITH mtbpsc jtbpsc;
mtbpsc WITH jtbpsc;

!SEP vars
tfamprb5 WITH mfamprb5 jfamprb5;
mfamprb5 WITH jfamprb5;

tfamprb6 WITH mfamprb6 jfamprb6;
mfamprb6 WITH jfamprb6;

!Item intercepts
[tghq24 tghq25 tghq29 tghq30] (I1-I4);
[ttpfsc ttrlppsc ttbpsc] (I5-I7);
[tfamprb5 tfamprb6] (I8-I9);

[mghq24 mghq25 mghq29 mghq30] (I1-I4);
[mtpfsc mtrlppsc mtbpsc] (I5-I7);
[mfamprb5 mfamprb6] (I8-I9);

[jghq24 jghq25 jghq29 jghq30] (I1-I4);
[jtpfsc jtrlppsc jtbpsc] (I5-I7);
[jfamprb5 jfamprb6] (I8-I9);
```

```
!Manifest variances
tbmi mbmi jbmi;
ttmetsc mtmetsc jtmetsc;
tcalssc mcalssc jcalssc;
tage_c ;

!Manifest means
[tbmi mbmi jbmi];
[ttmetsc mtmetsc jtmetsc];
[tcalssc mcalssc jcalssc];


MODEL white-female:
!------ Depression construct
Dep1 BY tghq24* tghq25 tghq29 tghq30 (L1-L4);
Dep2 BY mghq24* mghq25 mghq29 mghq30 (L1-L4);
Dep3 BY jghq24* jghq25 jghq29 jghq30 (L1-L4);
!Set scale using fixed factor method, free at all time points
Dep1* Dep2* Dep3*;
!Factor means freed
[Dep1* Dep2* Dep3*];

!------ Physical Function
pf1 BY ttpfsc* ttrlppsc ttbpsc (L5-L7);
pf2 BY mtpfsc* mtrlppsc mtbpsc (L5-L7);
pf3 BY jtpfsc* jtrlppsc jtbpsc (L5-L7);
!Set scale using fixed factor method, free at all time points
PF1* PF2* PF3*;
!Factor means freed
[PF1* PF2* PF3*];

!Socioeconomic Position
SEP1 BY tfamprb5* (L8)
tfamprb6 (L8);
SEP2 BY mfamprb5* (L8)
mfamprb6 (L8);
SEP3 BY jfamprb5* (L8)
jfamprb6 (L8);
```

```
!Set scale using fixed factor method, free at all time points
SEP1* SEP2* SEP3*;
!Factor means freed
[SEP1* SEP2* SEP3*];

!------- Structural Associations ---------!
!--- Time 3 regressed on time 2
Dep3 ON DEP2 PF2 mtmetsc mcalssc mbmi SEP2 tage_c;
PF3 ON PF2 mbmi SEP2 tage_c;
jtmetsc ON DEP2 mtmetsc mbmi SEP2 tage_c;
jcalssc ON DEP2 mcalssc mbmi SEP2 tage_c;
jbmi ON DEP2 mtmetsc mcalssc mbmi SEP2 tage_c;
SEP3 ON SEP2 tage_c;

!--- Time 2 regressed on time 1
Dep2 ON DEP1 PF1 ttmetsc tcalssc tbmi SEP1 tage_c;
PF2 ON PF1 tbmi SEP1 tage_c;
mtmetsc ON DEP1 ttmetsc tbmi SEP1 tage_c;
mcalssc ON DEP1 tcalssc tbmi SEP1 tage_c;
mbmi ON DEP1 ttmetsc tcalssc tbmi SEP1 tage_c;
SEP2 ON SEP1 tage_c;

!--- Time 3 on time 1 (AR2 paths)
Dep3 ON Dep1;
Pf3 on PF1;
jtmetsc on ttmetsc;
jcalssc on tcalssc;
jbmi on tbmi;
SEP3 on SEP1;

!--- Correlations within time points
!--- Time 2
Dep2 WITH mbmi Pf2 mtmetsc mcalssc SEP2;
mbmi WITH Pf2 mtmetsc mcalssc SEP2;
Pf2 WITH mtmetsc mcalssc Sep2;
mtmetsc WITH mcalssc Sep2;
mcalssc WITH Sep2;
```

```
!--- Time3
Dep3 WITH jbmi Pf3 jtmetsc jcalssc SEP3;
jbmi WITH Pf3 jtmetsc jcalssc SEP3;
Pf3 WITH jtmetsc jcalssc Sep3;
jtmetsc WITH jcalssc Sep3;
jcalssc WITH Sep3;


!------ Correlated Residuals --------!

!Depression vars
tghq24 WITH mghq24 jghq24;
mghq24 WITH jghq24;

tghq25 WITH mghq25 jghq25;
mghq25 WITH jghq25;

tghq29 WITH mghq29 jghq29;
mghq29 WITH jghq29;

tghq30 WITH mghq30 jghq30;
mghq30 WITH jghq30;

!PF vars
ttpfsc WITH mtpfsc jtpfsc;
mtpfsc WITH jtpfsc;

ttrlppsc WITH mtrlppsc jtrlppsc;
mtrlppsc WITH jtrlppsc;

ttbpsc WITH mtbpsc jtbpsc;
mtbpsc WITH jtbpsc;

!SEP vars
tfamprb5 WITH mfamprb5 jfamprb5;
mfamprb5 WITH jfamprb5;

tfamprb6 WITH mfamprb6 jfamprb6;
```

```
mfamprb6 WITH jfamprb6;

!Item intercepts
[tghq24 tghq25 tghq29 tghq30] (I1-I4);
[ttpfsc ttrlppsc ttbpsc] (I5-I7);
[tfamprb5 tfamprb6] (I8-I9);

[mghq24 mghq25 mghq29 mghq30] (I1-I4);
[mtpfsc mtrlppsc mtbpsc] (I5-I7);
[mfamprb5 mfamprb6] (I8-I9);

[jghq24 jghq25 jghq29 jghq30] (I1-I4);
[jtpfsc jtrlppsc jtbpsc] (I5-I7);
[jfamprb5 jfamprb6] (I8-I9);

!Manifest variances
tbmi mbmi jbmi;
ttmetsc mtmetsc jtmetsc;
tcalssc mcalssc jcalssc;
tage_c ;

!Manifest means
[tbmi mbmi jbmi];
[ttmetsc mtmetsc jtmetsc];
[tcalssc mcalssc jcalssc];


MODEL white-male:
!------ Depression construct
Dep1 BY tghq24* tghq25 tghq29 tghq30 (L1-L4);
Dep2 BY mghq24* mghq25 mghq29 mghq30 (L1-L4);
Dep3 BY jghq24* jghq25 jghq29 jghq30 (L1-L4);
!Set scale using fixed factor method, free at all time points
Dep1* Dep2* Dep3*;
!Factor means freed
[Dep1* Dep2* Dep3*];

!------ Physical Function
```

```
pf1 BY ttpfsc* ttrlppsc ttbpsc (L5-L7);
pf2 BY mtpfsc* mtrlppsc mtbpsc (L5-L7);
pf3 BY jtpfsc* jtrlppsc jtbpsc (L5-L7);
!Set scale using fixed factor method, free at all time points
PF1* PF2* PF3*;
!Factor means freed
[PF1* PF2* PF3*];

!Socioeconomic Position
SEP1 BY tfamprb5* (L8)
tfamprb6 (L8);
SEP2 BY mfamprb5* (L8)
mfamprb6 (L8);
SEP3 BY jfamprb5* (L8)
jfamprb6 (L8);
!Set scale using fixed factor method, free at all time points
SEP1* SEP2* SEP3*;
!Factor means freed
[SEP1* SEP2* SEP3*];

!------- Structural Associations ---------!
!--- Time 3 regressed on time 2
Dep3 ON DEP2 PF2 mtmetsc mcalssc mbmi SEP2 tage_c;
PF3 ON PF2 mbmi SEP2 tage_c;
jtmetsc ON DEP2 mtmetsc mbmi SEP2 tage_c;
jcalssc ON DEP2 mcalssc mbmi SEP2 tage_c;
jbmi ON DEP2 mtmetsc mcalssc mbmi SEP2 tage_c;
SEP3 ON SEP2 tage_c;

!--- Time 2 regressed on time 1
Dep2 ON DEP1 PF1 ttmetsc tcalssc tbmi SEP1 tage_c;
PF2 ON PF1 tbmi SEP1 tage_c;
mtmetsc ON DEP1 ttmetsc tbmi SEP1 tage_c;
mcalssc ON DEP1 tcalssc tbmi SEP1 tage_c;
mbmi ON DEP1 ttmetsc tcalssc tbmi SEP1 tage_c;
SEP2 ON SEP1 tage_c;

!--- Time 3 on time 1 (AR2 paths)
```

```
Dep3 ON Dep1;
Pf3 on PF1;
jtmetsc on ttmetsc;
jcalssc on tcalssc;
jbmi on tbmi;
SEP3 on SEP1;

!--- Correlations within time points
!--- Time 2
Dep2 WITH mbmi Pf2 mtmetsc mcalssc SEP2;
mbmi WITH Pf2 mtmetsc mcalssc SEP2;
Pf2 WITH mtmetsc mcalssc Sep2;
mtmetsc WITH mcalssc Sep2;
mcalssc WITH Sep2;

!--- Time3
Dep3 WITH jbmi Pf3 jtmetsc jcalssc SEP3;
jbmi WITH Pf3 jtmetsc jcalssc SEP3;
Pf3 WITH jtmetsc jcalssc Sep3;
jtmetsc WITH jcalssc Sep3;
jcalssc WITH Sep3;



!------ Correlated Residuals --------!

!Depression vars
tghq24 WITH mghq24 jghq24;
mghq24 WITH jghq24;

tghq25 WITH mghq25 jghq25;
mghq25 WITH jghq25;

tghq29 WITH mghq29 jghq29;
mghq29 WITH jghq29;

tghq30 WITH mghq30 jghq30;
mghq30 WITH jghq30;
```

```
!PF vars
ttpfsc WITH mtpfsc jtpfsc;
mtpfsc WITH jtpfsc;

ttrlppsc WITH mtrlppsc jtrlppsc;
mtrlppsc WITH jtrlppsc;

ttbpsc WITH mtbpsc jtbpsc;
mtbpsc WITH jtbpsc;

!SEP vars
tfamprb5 WITH mfamprb5 jfamprb5;
mfamprb5 WITH jfamprb5;

tfamprb6 WITH mfamprb6 jfamprb6;
mfamprb6 WITH jfamprb6;

!Item intercepts
[tghq24 tghq25 tghq29 tghq30] (I1-I4);
[ttpfsc ttrlppsc ttbpsc] (I5-I7);
[tfamprb5 tfamprb6] (I8-I9);

[mghq24 mghq25 mghq29 mghq30] (I1-I4);
[mtpfsc mtrlppsc mtbpsc] (I5-I7);
[mfamprb5 mfamprb6] (I8-I9);

[jghq24 jghq25 jghq29 jghq30] (I1-I4);
[jtpfsc jtrlppsc jtbpsc] (I5-I7);
[jfamprb5 jfamprb6] (I8-I9);

!Manifest variances
tbmi mbmi jbmi;
ttmetsc mtmetsc jtmetsc;
tcalssc mcalssc jcalssc;
tage_c ;

!Manifest means
```

```
[tbmi mbmi jbmi];
[ttmetsc mtmetsc jtmetsc];
[tcalssc mcalssc jcalssc];

OUTPUT:
MOD STAND TECH4;
```

Appendix B

Model code for Netlogo Simulations

B.1 Model code for Chapter 6

B.1.1 Model Setup

```
to setup
clear-all
read-turtles-from-csv
make-network
reset-ticks
end
```

read-turtles-from-csv

```
to read-turtles-from-csv
file-close-all ; close all open files

if not file-exists? "ABM␣Initial␣Characteristics.csv" [
user-message "No␣file␣'ABM␣Initial␣Characteristics.csv'␣
   exists!␣Check␣for␣file␣in␣directory."
stop
]

file-open "ABM␣Initial␣Characteristics_v3.csv" ; open the
   file with the turtle data

; We'll read all the data in a single loop
while [ not file-at-end? ] [
```

```
; here the CSV extension grabs a single line and puts the
   read data in a list
let data csv:from-row file-read-line
; now we can use that list to create a turtle with the saved
   properties
create-people 1 [
set sex item 0 data
set bmi item 1 data
set depression item 2 data / 15
set income item 3 data
set time-since-promoted random 6 ; promotion occurred
   sometime in last 6 months.
set stigmatised? FALSE
set stigmatising? FALSE
set stigma-count 0
set weight-gain-susceptibility 0 ; everyones weight gain
   susceptibility is between 0 and 0.1
;above line currently means that everyone ends up obese
   eventually
set depression-susceptibility 0 ; normally distruted with
   mean 0 s.d 0.01
set size bmi / 25 ; make agent size proportional to bmi
set xcor random-xcor;
set ycor random-ycor;
]
]

file-close ; make sure to close the file
end
```

make-network

```
to make-network ; Generates a network in which everyone has
   as least 2 link neighbours
ask people [
let current-links turtle-set [other-end] of my-links
   ;prevents people from creating 2 links between same agents
```

```
create-links-with n-of 2 other people with [not member? self
   current-links] ;ask to make link with agents they don't
   currently have a link with
]
ask links[ set meeting? FALSE ]
end
```

B.1.2 Model 'go' procedure

```
to go
update-body-deviance
reset-meetings
meet-up
stigmatise
check-meetings
apply-promotion
update-weight
update-mood
tick
end
```

B.1.3 Submodels

update-body-deviance

```
to update-body-deviance
ask people with [sex = "2: female"]
[ set body-deviance max (list 0 (bmi - female-ideal-bmi)) ]
ask people with [sex = "1: male"]
[ set body-deviance max (list 0 (bmi - male-ideal-bmi)) ]
end
```

reset-meetings

```
to reset-meetings
ask people [ask my-links [set meeting? FALSE] ;reset meeting
   links from previous time point
set stigma-count 0] ;reset amount of stigma received for
   last time-point
end
```

B.1.3.1 Stigma submodel

meet-up

```
;describes part of network that is interacting socially in
   each time step
to meet-up
ask people [
ask n-of 2 my-links [set meeting? TRUE] ;choose 2 links to
   be meetings in that time point
]
end
```

Stigmatise

```
to stigmatise
ask people
[ let meetings my-links with [meeting?] ;set stigma targets as
   individuals at other end of meeting links
let targets turtle-set [other-end] of meetings
;show targets ;for debugging
let victims targets with [body-deviance - body-tolerance > 0]
set stigmatising? TRUE
ask victims
[ if body-deviance / 25 >= random-float 1 ; probability of
   stigma increases linearly up to max at 25 over ideal-bmi
[set stigmatised? TRUE
set stigma-count stigma-count + 1
set depression depression + stigma-to-depression ;slight
   increase in depression with each stigmatising encounter
set bmi bmi + stigma-to-obesity] ;slight increase in bmi in each
   stigmatising encounter.
]
]
end
```

check-meetings

```
to check-meetings
ask people with [stigma-count = 0]
```

```
[ set depression depression -
   meeting-to-depression];individuals benefit from
   non-stigmatising interactions
end
```

B.1.3.2 Promotion submodel

```
to apply-promotion
ask people [
ifelse time-since-promoted <= 6
[
set time-since-promoted time-since-promoted + 1
]
[
let promotion-penalty max (list 0 (body-deviance -
   body-tolerance))
ifelse random-float 1 < promotion-prob * (1 -
   promotion-penalty * obesity-promotion-penalty)
[
set income income * promotion-value
set time-since-promoted 0
]
[
set time-since-promoted time-since-promoted + 1
]
]
]
end
```

B.1.3.3 Update submodel

update-weight

```
to update-weight
ask people
[set bmi bmi + ( random-normal 0 0.1 ) ;makes sure that bmi
   changes
set size bmi / 25
]
end
```

update-mood

```
to update-mood
ask people
[set depression depression + random-normal 0 0.05
set depression min (list depression 1) ; makes sure that
   depression can't increase above 1
set depression max (list depression 0) ; makes sure depression
   can't go below 0
]
end
```

B.2 Model code for Chapter 7

B.2.1 Model Setup

```
to setup
clear-all
set wage-inflation 1.003859
set pension-inflation 1.004889
read-turtles-from-csv
set mean-income-init mean [income] of people
make-network
reset-ticks
end
```

Note: The `read-turtles-from-csv` and `make-network` procedures are unchanged from the model from chapter 6, although the csv file used for the initial conditions was updated to include the extra characteristics.

B.2.2 Model 'go' procedure

```
to go
update-globals
update-body-deviance
reset-meetings
meet-up
stigmatise
receive-stigma
apply-retirements
```

```
apply-promotion
apply-inflation
attempt-new-diet
check-diet-status
update-weight
update-mood
update-age
tick
end
```

B.2.3 Submodels

update-globals

```
to update-globals
set mean-wage mean-income-init * (wage-inflation ^ ticks)
   ;with [retired? = FALSE]
if ticks mod 12 = 0 [
; global means
set bmi-mean mean [bmi] of people
set income-mean mean [income] of people
set depression-mean mean [depression] of people
; Global correlations
ask people [
set bmi-diff (bmi - bmi-mean)
set income-diff (income - income-mean)
set depression-diff (depression - depression-mean)
]
let bmi-diff-sqr-sum sum [(bmi-diff) ^ 2] of people
let inc-diff-sqr-sum sum [income-diff ^ 2] of people
let dep-diff-sqr-sum sum [depression-diff ^ 2] of people
ifelse (bmi-diff-sqr-sum > 0) and (inc-diff-sqr-sum > 0)
[set bmi-inc-cor (sum [bmi-diff * income-diff] of people) /
   ((( bmi-diff-sqr-sum ) * (inc-diff-sqr-sum)) ^ (1 / 2))]
[set bmi-inc-cor 0]
ifelse (bmi-diff-sqr-sum > 0) and (dep-diff-sqr-sum > 0)
[set bmi-dep-cor (sum [bmi-diff * depression-diff] of
   people) / ((( bmi-diff-sqr-sum ) * (dep-diff-sqr-sum)) ^
   (1 / 2))]
```

```
[set bmi-dep-cor 0]

;;;;;;;;;;;; Relationships by sex ;;;;;;;;;;;
;;;;;; males ;;;;;;;
;means
let males people with [sex = "1: male"]
set bmi-mean-m mean [bmi] of males ;people with [sex = "1:
   male"]
set income-mean-m mean [income] of males ; people with [sex
   = "1: male"]
set depression-mean-m mean [depression] of males ; people
   with [sex = "1: male"]

;correlations
ask males [
set bmi-diff (bmi - bmi-mean-m)
set income-diff (income - income-mean-m)
set depression-diff (depression - depression-mean-m)
]
let bmi-diff-sqr-sum-m sum [(bmi-diff) ^ 2] of males
let inc-diff-sqr-sum-m sum [income-diff ^ 2] of males
let dep-diff-sqr-sum-m sum [depression-diff ^ 2] of males
ifelse (bmi-diff-sqr-sum-m > 0) and (inc-diff-sqr-sum-m > 0)
[set bmi-inc-cor-m (sum [bmi-diff * income-diff] of males) /
   ((( bmi-diff-sqr-sum-m ) * (inc-diff-sqr-sum-m)) ^ (1 /
   2))]
[set bmi-inc-cor-m 0]
ifelse (bmi-diff-sqr-sum-m > 0) and (dep-diff-sqr-sum-m > 0)
[set bmi-dep-cor-m (sum [bmi-diff * depression-diff] of
   males) / ((( bmi-diff-sqr-sum-m ) * (dep-diff-sqr-sum-m))
   ^ (1 / 2))]
[set bmi-dep-cor-m 0]

;;;;;;;; females ;;;;;;;
;means
let females people with [sex = "2: female"]
set bmi-mean-f mean [bmi] of females ;people with [sex = "1:
   male"]
```

```
set income-mean-f mean [income] of females ; people with
   [sex = "1: male"]
set depression-mean-f mean [depression] of females ; people
   with [sex = "1: male"]

;correlations
ask females [
set bmi-diff (bmi - bmi-mean-f)
set income-diff (income - income-mean-f)
set depression-diff (depression - depression-mean-f)
]
let bmi-diff-sqr-sum-f sum [(bmi-diff) ^ 2] of females
let inc-diff-sqr-sum-f sum [income-diff ^ 2] of females
let dep-diff-sqr-sum-f sum [depression-diff ^ 2] of females
ifelse (bmi-diff-sqr-sum-f > 0) and (inc-diff-sqr-sum-f > 0)
[set bmi-inc-cor-f (sum [bmi-diff * income-diff] of males) /
   ((( bmi-diff-sqr-sum-f ) * (inc-diff-sqr-sum-f)) ^ (1 /
   2))]
[set bmi-inc-cor-f 0]
ifelse (bmi-diff-sqr-sum-f > 0) and (dep-diff-sqr-sum-f > 0)
[set bmi-dep-cor-f (sum [bmi-diff * depression-diff] of
   females) / ((( bmi-diff-sqr-sum-f ) *
   (dep-diff-sqr-sum-f)) ^ (1 / 2))]
[set bmi-dep-cor-f 0]

]
end
```

update-body-deviance

```
to update-body-deviance ;Only stigmatises those who are
   above the ideal bmi
ask people with [sex = "2: female"]
[ set body-deviance max (list 0 (bmi - female-ideal-bmi)) ]
ask people with [sex = "1: male"]
[ set body-deviance max (list 0 (bmi - male-ideal-bmi)) ]
end
```

B.2.3.1 Social stigma Submodel

reset-meetings

```
to reset-meetings
ask people [ask my-links [set meeting? FALSE]] ;reset
   meeting links from previous time point
end
```

meet-up

```
to meet-up
ask people [
ask n-of 2 my-links[ set meeting? TRUE] ;choose 2 links to
   be meetings in that time point
]
end
```

stigmatise

```
to stigmatise
ask people
[ let meetings my-links with [meeting?] ;set stigma targets
   as individuals at other end of meeting links
let targets turtle-set [other-end] of meetings
;show targets ;for debugging
let victims targets with [body-deviance - body-tolerance >
   0] ; might be no victims so need to guard against this
ask victims
[ ifelse body-deviance / 25 >= random-float 1 ; probability
   of stigma increases linearly up to max at 25 over
   ideal-bmi
[set stigmatised? TRUE ;status here is only to check
   consistency of the method
set stigma-count stigma-count + 1]
[set stigmatised? FALSE]
]
ifelse any? victims with [stigmatised?]
[set stigmatising? TRUE]
[set stigmatising? FALSE]
]
end
```

receive-stigma

```
to receive-stigma
;set stigma-total sum [stigma-count] of people
ask people with [stigma-count > 0]
[ set depression-risk-stigma stigma-to-depression *
   stigma-count ;increased risk of depressive symptoms from
   stigma
set weight-risk-stigma stigma-to-obesity *
   stigma-count;increased risk of bmi increase from being
   stigmatised.
set stigma-count 0 ;reset stigma count for the next
   time-point
]
end
```

B.2.3.2 The employment submodel

apply-retirements

```
to apply-retirements
ask people with [age > 55 and retired? = FALSE]
[ ;set retired? TRUE
;set income income * pension-weight

ifelse age > 65
[
set retired? TRUE
set income income * pension-weight
]
[
let retirement-prob (income - mean-wage) /
   (income-retirement-threshold * mean-wage)
if retirement-prob > random-float 1 [
set retired? TRUE
set income income * pension-weight]
]
]
end
```

apply-promotion

```
to apply-promotion
ask people with [retired? = FALSE] [
ifelse time-since-promoted <= promotion-wait
[
set time-since-promoted time-since-promoted + 1
]
[
let body-excess max (list 0 (body-deviance -
   body-tolerance-work))
ifelse random-float 1 < promotion-prob * (max (list 0 (1 -
   body-excess * obesity-promotion-penalty)))
[
set income income * promotion-value
set time-since-promoted 0
]
[
set time-since-promoted time-since-promoted + 1
]
]
]
end
```

apply-inflation

```
to apply-inflation
ask people with [retired? = TRUE]
[
set income income * pension-inflation
]
ask people with [retired? = FALSE]
[
set income income * wage-inflation
]
end
```