INVESTIGATING FACTORS INFLUENCING FOOD CHOICE BEHAVIOURS INCORPORATING PSYCHOLOGICAL DISTRESS

Agent Based Model Approach

Ruairi McElhatton

Student Reg No: 190134101

May 2024

Supervisors:
Robin Purshouse and Charlotte Buckley

University of Sheffield | Automatic and Control Systems Engineering

A dissertation submitted in partial fulfilment of the requirements for the degree of Robotic and Mechatronic Engineering MEng

Abstract

Introduction: Unhealthy eating, including consumption of foods with high fat, salt and sugar (HFSS), is a risk factor for non-communicable diseases. Studies linking psychological distress (PD) and consumption of HFSS are limited, especially in an agent based model (ABM) context. The aim of this project is to build on an existing ABM of food choice behaviours by incorporating PD and then to use this model to analyse the impact of potential government policies related to distress.

Method: PD and body mass index (BMI) were added to the existing ABM to introduce a feedback loop representing widely accepted theory regarding PD's impact on food choice. This pathway was initialised by simulating individuals past and desired eating habits which were calculated by combining, through imputation, two large surveys in the UK and England.

Results: The model was a good fit to population HFSS consumption and PD levels for both men and women. It was not as accurate for BMI levels. The trends for BMI were accurate in women but not in men. Simulating an increase in the availability of NHS mental health services led to a reduction in population PD levels, which led to better dietary choices, which in turn led to improved BMI levels.

Conclusion: This project provides more evidence that a significant relationship between PD and HFSS consumption exists, as well as relationships between PD and BMI, and BMI and HFSS consumption. The simulation model is a useful tool for estimating changes in population health following hypothetical population wide interventions.

Individual contribution

Existing work and resources

In this project, I was building upon an existing ABM. This meant I was given some previous work, mostly coding scripts, to add to and tailor to my needs.

These files can be found in a repository [1]. This repository contains the orignal ABM and the calibration R files I was given. The only file that may have been changed in here is "rscripts/0_NDNS_read.R". I had to make minor changes to this and then couldn't find the original file when I was putting this repository together.

Other resources included helpful guidance from my supervisors including useful information on ABMs and psychological theory relevant to the study area. I was pointed in the right direction in my literature review. Robin gave me feedback on the structure of my report which I have tried to implement.

Individual work undertaken

In terms of programming and plotting results, everything that is new or different in the updated model repository [2] to the original repository [1] is my own work. Aside from that, I conducted the literature review by myself bar being sent several relevant articles. I wrote the report by myself and decided what to include in it. Figure 21 shows the work breakdown structure which should help illustrate the work I carried out for this project. This is in appendix A.

Acknowledgements

I would like to express my gratitude to the following people, without whom it would have been impossible to complete this project.

Thank you to to my supervisor Robin Purshouse for being a great help with guiding me in the right direction at the beginning of my project. Thank you also to Charlotte Buckley, who was able to help me greatly with her extensive knowledge on the subject matter in Robin's absence.

I would also like to give special thanks to my family, friends and girlfriend for putting up with me when I was stressed and giving me good general advice not just throughout my project, but through my whole higher education.

Contents

1 Introduction and background

1.1 Aims and objectives

1 - Literature review of food choice behaviours and psychological distress

The first objective was to complete a literature review outlining the theory that the ABM was based on. This can be seen in 2.

2 - Generate descriptive statistics from literature review to represent the link between psychological distress and consumption of foods with HFSS content.

There are a few relationships which will be added to the existing ABM. The idea behind this objective was to highlight the need for these relationships to be given a numerical value.

3 - Impute psychological distress into existing agents

This objective highlights the need for adding PD into the agents which are generated from the national diet and nutrition survey (NDNS) [3] using data from the health survey for England (HSE) [4].

4 - Redesign the existing HFSS ABM software architecture to incorporate psychological distress pathways

New pathways had to be created in the ABM to accomodate PD, these are shown in a few diagrams for clarity. UML class diagram enables one to see how different parts of the model interact with eachother.

5 - Implement the ABM Model in Python using the repast HPC toolkit

The changes from objective 4 had to be coded into the model using various equations which try to replicate real life behaviour of a population, with an element of randomness.

6 - Advanced objective - Demonstrate an application of this model to test the impact of a hypothetical scenario where policies designed to reduce psychological distress are introduced to the population

The scenario that I chose to introduce was to have NHS mental health services improve by different amounts, then depict how different levels of improvement would be reflected in the populations PD, BMI and diet.

1.2 Background

Unhealthy eating

Unhealthy eating is a significant concern for population health, it is estimated that "in 2017, 11 million deaths and 255 million Disease-Adjusted Life-Years (DALYs) were attributable to dietary risk factors" [5]. These deaths are predominantly due to a bad diet being a leading cause of non-communicable diseases. The alarming rise of diet-related disease is also putting great strain on health systems. It is important to understand unhealthy eating at a population level because it will increase the possibility of finding better ways of reducing unhealthy eating than methods used up to this point, particularly in England [6]. Unhealthy eating is quite a broad term so in this paper it was decided to specify unhealthy eating in terms of consumption of foods high in fat, salt, and sugar (HFSS).

Psychological distress

Psychological distress (PD) is also a concern for public health. PD is a term used to describe the symptoms of depression and anxiety, not necessarily both together [7]. "PD is broadly defined as a state of emotional suffering characterized by symptoms of depression (e.g., loss of interest; unhappiness; desperateness) and anxiety (e.g., restlessness; feeling tense)" [8]. Recent studies have found a dose-response relationship between PD and risk of mortality from all causes. There is evidence that "clinically significant levels of distress (i.e. a GHQ-12 [9] score of four or greater) have been found to increase risk of incident arthritis, cardiovascular disease (CVD), chronic pulmonary disease and diabetes mellitus" [7]. The effects of less severe PD symptoms are poorly understood.

Agent-based models

"Agent Based Modelling is a computational method that stimulates individuals into making decisions according to programmable rules. Those rules are set by the modeller to represent key elements of real-world decisions, including the individuals' own characteristics and their social and physical environment" [10]. This paper will look at how PD is related to HFSS consumption and model this relationship in an Agent-Based Model (ABM). This will be done by building upon an existing ABM which looks at how food advertising is related to consumption of HFSS [11]. A thorough search of the relevant literature yielded nothing on any studies to date relating PD to HFSS consumption in an ABM, but there were many related studies.

ABM could be very useful for the purpose of this investigation due to it being able to model complex dynamics in "artificial societies" on computers. Every individual actor (agent) in the system is explicitly represented in computer code [12]. Due to the individualistic nature of the effect of PD on diet [13], and the causes of PD, this approach of modelling is perfect.

1.3 Project management

Overall I am happy with how I approached the project and I am proud of the final report I have produced along with the results it explains. I have produced a work breakdown structure (21) and a Gantt chart to help illustrate the work I have carried out and the time period this was done over. These can both be seen in A.

Explaining progress against plan

As can be seen in the Gantt chart in appendix A, my progress was behind my plan for the whole year. This was because my initial plan was quite ambitious and I struggled quite a bit with the programming side of the project having not programmed outside the demands of my university course until this year which meant Python was a new language to me. This struggle is reflected in the technical activities progress in the Gantt chart. I was tweaking the software until a week before the deadline and analysing results of the intervention until later than I would have liked too. The software was not as easy as it may have been had the original model I was building upon been structured according to the MBSSM architecture standard [14]. This is designed to make it easier for additions to be made as will be mentioned throughout the report.

However, before the programming was the literature review where it can be clearly seen that I fell behind here and this was due to a big underestimate of what a literature review is. There was so much literature around the subject that I found very interesting but may not have been completely relevant.

Another reason I fell behind is because the project was not my priority for the whole year, I had other modules with more pressing deadlines throughout the year as well as job-seeking engagements. In the end I caught up after my time management skills were forced to improve and I have enjoyed the dissertation thoroughly.

Progress

Aside from the obvious of trying my best and working hard, I had been trying to make the most of my meetings with my supervisors from the start of the year. I did this by following an effective management technique called progress, plan, problems (PPP) which had been used by my manager on my year in industry. I found that this was a great way to make progress and I would come away from meetings with a clear idea of what to do next.

Although not following this technique staunchly, having it in the back of my mind was enough to ensure I was able to make steady progress throughout the year. It also has to be said that my supervisors were very knowledgeable on the subject area and every meeting was useful.

2 Literature review

As the ABM in this project is designed to model real world social behaviour, it was necessary to research relevant psychological theory. This would ensure that additions to the existing ABM would stem from reality, thus making the model more accurate. It was also necessary to research ABMs in related fields to get a better idea on how to build these.

2.1 Context for psychological distress

Due to the existing ABM being designed without PD being accounted for, it was necessary to research PD to understand how it should be implemented into the model.

Review of the GHQ-12 as a measure of psychological distress

There are many different measures of PD, but upon consulting the literature, a measure which appeared persistently was the General Health Questionnaire (GHQ). The GHQ is the measure of PD used in the dataset [4] that is used in this project.

The GHQ-12 (12 item short-screening instrument, shorter than original 60 item version) is, upon research, the most common and most accepted way of measuring PD, it is also the primary method of measuring well-being and mental health in the Health Survey for England (HSE) [4]. In a World Health Organisation (WHO) study of mental illness in general health care [15], the GHQ-12 was described as "remarkably robust". Despite being widely used there are a few papers that question the reliability of the questionnaire. For example, [16] systematically scores the model and finds that there is response bias on the negative items, which limits the validity of the findings of GHQ-12 when screening for mental illness. Despite this, after reviewing all of the literature it should be said with confidence that the GHQ-12 is reliable.

Common causes of psychological distress

There has been extensive research into the causes of high PD. In [17], family-to-work conflict (where family interferes with work) is strongly associated with PD, but this conflicts with other literature, such as [18], which says the opposite (work-to-family conflict) is true so not much can be discerned here. [17] also mentioned that job-dissatisfaction was strongly associated with PD, which is a reoccurring association within the literature. Another associated factor which appeared in the literature was loneliness, this was seen to be just as big a cause of high PD as job-dissatisfaction and was equal in both genders. Supporting this, participation in social networks and the presence of social support were associated with having lower levels of PD. Other factors associated with increased PD include financial difficulties [19], smoking [20] and at-risk drinking [21]. However, these factors had a less prominent effect on PD than that of social and work-related factors [17].

Psychological distress in the UK

Levels of mental health in the UK are currently very bad when compared with other developed countries, in fact the UK is not in the top 10 of a mental health well-being table produced by comparethemarket in 2023 [22]. "Current levels of good mental health" in the UK are "disturbingly low" according to the UK mental health foundation [23]. The foundation commissioned a survey in March 2017 to understand mental health patterns in the UK. This survey also found that young adults are reporting having experienced mental health problems in their lives when compared with older adults, despite having had fewer years to experience this. In recent years, this issue is widely agreed to have worsened due to the COVID-19 pandemic, where the proportion of mental health disorders among 17-19 year olds increased from 10% to 26% [24].

According to the British medical association, the main reason for the low levels of good mental health in England is because "demand for mental health services is outpacing the resources afforded to them" [25]. One in seven full time roles within NHS mental health services is currently vacant, with this likely being an underestimate due to many hospitals not being able to afford to fill posts and not advertising them as a result. This part of the project was made to show what fixing this gap in services could do for the country.

Effectiveness of mental health treatment

There are a number of studies on the efficacy of various mental health treatments. One meta-analysis is very comprehensive in this matter [26]. This paper states, drawing from several detailed studies, that on average 75% of therapy clients are better off than untreated individuals. This makes a strong case for introducing more widespread mental health care in the UK.

The paper also states that there is "virtually no difference in effectiveness between the class of all behavioural therapies"[26]. This suggests that any improvement in availability of mental health services will lead to great strides in the UK's mental health.

2.2 Psychological distress and unhealthy eating

Before diving into the literature surrounding the relationship under investigation itself, it is important to understand what the two elements of the relationship are. In section 1, "unhealthy eating" is already specified as the consumption of HFSS, this is not what unhealthy eating is all the time, it is just the measure decided upon in this project due to the nature of the surveys used. PD is also defined in section 1 and GHQ-12 is to be used as the measure for this.

The effect of psychological distress on HFSS intake

When consulting the literature of this project's investigation specifically, there are several studies which find a strong correlation between stress and unhealthy eating in all members of the population, regardless of sex or prevalence of an eating disorder [27], [28]. Although these studies often use a smaller population (e.g., first year Australian university students) their findings may be applicable to the outside world. There were also many papers looking at the link in the opposite direction [29]–[31]. On top of these papers which only showed that a link exists, there are many papers which look at this bi-directional relationship whilst also discussing theories of mechanisms explaining it.

One such mechanism which is used to explain the link is *comfort eating* [32], p. 1. In the academic literature comfort eating is defined as eating to relieve negative emotions. This is synonymous with *emotional eating* and *stress eating* [32]. The limitation of comfort eating as an explanation for the relationship under investigation is that the *comfort foods* consumed when looking to provide relief vary so much from one individual to another, with a difference between sexes [27]. There are also studies which have associated mild stress with increased eating and more severe stress with reduced food intake [13, p. 1]. According to some of the literature *emotional eaters* are typically in the minority amongst non-clinical populations.

Another mechanism explaining the link is *food insecurity* [33]. Food insecurity is defined as a lack of stable access to nutritious food, and it has often been associated with obesity and a less nutritious diet through an increase in stress and the use of food as a coping mechanism. To go a step further, higher availability of low-cost, high-energy foods in neighbourhoods with High Household Food Insecurity (HHFI) is believed to play a key role in the association between food insecurity and BMI [33, p. 2]. HHFI is directly associated with distress and poorer diet quality but not does not directly contribute to eating to cope or BMI [33]. This is backed up further by [34] which states that food groups with more favourable nutrient profiles were associated with higher cost per unit of energy, going on to say that this could be a barrier to low-income households adopting food-based dietary advice.

Bi-directionality of the relationship between psychological distress and diet

Throughout the literature, a relationship which appeared in several studies was the direct effect of diet quality on PD [29]–[31]. This relationship broadly stated that a bad diet would lead to bad mental health, and by extension, higher levels of psychological distress.

As well as studies focussing on a direct impact, there is also strong evidence of an indirect effect. The pathway of this being that a bad diet leads to weight gain [35], [36], this weight gain leads to obesity which then leads to a higher susceptibility to depression and psychological distress [37], [38]. Despite widespread acceptance that this relationship exists, "there is increasing recognition that obesity is strikingly heterogeneous with respect to depression psychopathology" [39].

This means that the link differs among the population, with gender being a big factor in this difference. The "female gender has been consistently associated with an increased risk of depression among obese individuals" [39], [40]. Markowitz goes on to explain the possible pathways of this relationship. One pathway mentioned is the "social stigma" pathway, which we have seen elsewhere in the literature [38]. Many studies accept the theory that there is a greater correlation between BMI and depression amongst women due to this social stigma pathway. This has been put down to the "exaggerated pressure on women to be thin" [40].

Another theory mentioned in the literature for differing psychological effects of obesity is that of the severity of obesity. One study stating that "the most severely obese individuals (BMI > 40) were significantly more likely to be depressed than those with BMI between 20 and 34.9 (12.51% versus 3.55%)" as discussed by Markowitz in [41]. The pathway which could be said to explain this this is named as *poor adherence pathway*, depressed patients have a tendency to amplify physical symptoms of chronic medical conditions [42]. This could mean these patients may be suffering from severe pain which would make it more difficult to follow weight loss exercise programs, therefore making it harder to become a healthy weight.

Path analysis

Path analysis is a way of assessing the effects of a set of variables acting on a specified outcome via different pathways [43]. In other words, it is a way of putting a number on relationships between variables in our model. *Path analysis* assists with forming the equations which enable researchers to put a numerical value on the strength of real life relationships between behaviours and traits, or anything for that matter. This numerical value is often referred to as a path coefficient. Relevant examples are seen in [44]–[46] among other articles.

Varying results in the existing literature

Within the literature, there are varied results of the strength and direction of the correlation under investigation. This can be explained by several things. One possible reason is the difficulty of obtaining valid and reliable information on food intake if the clinical population are not under constant observation. All methods of obtaining dietary information rely on the population recording it themselves or recalling what they have eaten in a prior period. Even if a subject records their intake perfectly, it still may be difficult to find the exact nutritional content within some consumed foods.

Another explanation which is mentioned in the literature is the issue of individual differences in dietary response to stress [13], [47]. This theory is supported a bit more when looking at studies conducted over a longer period so it can be seen who is consistently hyperphagic (eating more) and who is consistently hypophagic (eating less) in response to stress. This theory can be developed further into restrained and non-restrained eaters [13], [48]. This is the idea that a restrained eater's diet becomes worse when under stress, whereas non-restrained eaters are less reactive. It is important to consider this when looking at the relationship under investigation.

2.3 Agent-based models

The definition for agent based models (ABMs) can be found in section 1.2. In this part of the literature review, the utility of ABMs is explored, specifically within the area of public health.

Folk psychology

Folk psychology is often used in ABMs to try and accurately predict agent behaviour. "Folk psychology is a name traditionally used to denote our everyday way of understanding, or rationalizing, intentional actions in mentalistic terms"[49]. A common and relatively simple way of implementing folk psychology in ABMs is through logit models, an example of which is seen in [50]. This is how agent behaviour will be modelled in this project.

ABMs exploring public health problems

The next area of literature which was examined consisted of ABMs in related fields to the one under investigation. It soon became apparent that there have been calls for more ABMs to be used to understand public health issues due to their realistic nature and ability to account for feedback [51], [52]. An ABM built using the mechanism-based social systems modelling (MBSSM) architecture detailed in [14] will give this project a platform to identify which factors affect PD and HFSS consumption. The ability of an ABM to model feedback is very useful in the case of HFSS consumption and PD due to the bi-directional nature of the relationship.

MBSSM architecture for ABM

The MBSSM architecture includes situational, action and transformational mechanisms. Situational mechanisms encode interaction, where an individual's attributes are altered based on the observable attributes and actions of other entities. Action mechanisms encode a process an individual decides to take based on its own attributes, individuals will have different attributes which will lead to different courses of action being taken. Transformational mechanisms are significantly more complex and are dissected in [14, p. 6].

Another reason the MBSSM architecture is to be used is because it can model how individual factors cause dependent variables to change. The mechanisms described in the literature will be used to create a series of dynamical equations to generate HFSS consumption due to PD. This will enable the modelling of individual-level factors and social context which contribute to dietary behaviour. The ability of an ABM to model individual factors is very important for the purposes of this investigation and it is what sets it apart from other modelling techniques such as System Dynamics. System Dynamics uses a more *top-down* approach and key objects are usually represented at macro level, which limits insight into the individual level and means it can't show agent diversity or adaptation [12, p. 5].

The predominant reason for the use of the MBSSM architecture standard is that it simplifies adding new theories and pathways to existing models. This will in theory simplify this project and any future work. The overarching idea is that researchers can use ABMs to simulate any theories they would like to in any combination.

ABMs on mental health and well-being

As far as the research for this paper goes, there has been no prior ABM created for the specific purpose of relating dietary habits in any capacity to PD. However, there have been several ABMs created in similar areas of social science. In the literature, there was one paper using an ABM in the hope of finding ways to help healthcare administrators discover interventions that increase population wellness [53]. This research has a very similar goal to the advanced objective outlined in 1.1 and is a paper which could be very useful in this investigation. It is also one of the few models within the literature to have considered relationships, social networks, and cultural norms. This consideration of various theories and mechanisms makes for an architecture that is similar to the MBSSM. However, a limiting factor from this model is that it looks at a population which is not very representative of the larger, general population. It focuses on Philadelphia's Medicaid [54] population. This means that low-income families and individuals are over-represented in the referenced study. The model in this project will be representative of a more general population.

Another paper in the literature was an Agent-Based Simulation to examine the relationship between relative income and depression in Canadian mothers [55]. This paper is relevant not only because it analyses a relationship between two variables but also because they are two variables which are both going to feature to some degree in the model of this project. This model has its limitations though, being a model that only looks at the relationship stated in mothers. This is a very specific sample and is certainly not representative of the whole population with prevalence of depression in mothers in Canada being higher than the average of the population [56]. Although the model includes factors of income and social networks, it doesn't include anything around the characteristics of individual agents, which have been shown to be important and are represented in MBSSM.

ABMs on population dietary habits

Along with ABMs on mental health, there were a number found within the literature which looked at how ABMs could be used to model and/or influence dietary actions of agents. One was modelled around the subject of obesity and depression [38], this is closely related with the matter of this investigation although the model focuses heavily on social norms and obesity causing depression as opposed to the other way round like in our investigation, it also chooses not to include other factors which play a part in obesity. There are several ABMs based on obesity interventions, such as [57], which relates to the advanced objective in this project in the fact that it focuses on interventions, even though it is to do with influencing diet as opposed to PD. In fact, [57] does not include PD, ignoring it as a contributing factor to dietary habits.

Another area looked at by ABMs in the literature is food accessibility, with a particular look at food insecurity [58] which has been linked with stress [33]. Although [58] raises an important issue, it is quite a specialised model and only has limited relevance to the model in this investigation. This issue of being a model which doesn't look at many factors puts it in the *Broad-Shallow* category of agents [51, p. 63]. According to this definition of type of ABM, the one produced in this project will be *mixed modelling* because it includes both detail of individual agents and considers numerous theories. A model which this project will be almost identical to in structure is in [11], which is obvious due to it being the starting point of the work in this project. It will also be similar to [59]. Although this model is surrounding alcohol use behaviours as opposed to dietary intake, there are still many parallels to be drawn.

2.4 Gaps in the literature that this project will attempt to fill

One of the main motives of this project is to highlight how reducing PD across a population could indirectly lead to a reduction in unhealthy eating. This may be possible due to the ability of ABMs to evaluate different targeting strategies and then finally point to an optimal solution. An ABM's "principal strength is the ability to model and capture emergent collective behaviour arising from dynamic adaptation of knowledgeable actors who seek strategic solutions in the face of environmental constraints and whose complex interactions create emergent patterns that cannot be predicted or understood using conventional methods that do not permit non-linear dynamics" [57, p. 1].

Having reviewed the relationship between PD and consumption of HFSS, it can be said to be a multi-faceted, bi-directional relationship which means there are many factors to consider whilst building an ABM to represent a population that has changing characteristics to show how this relationship may change. However, although the model will be complex, there is no doubt that an ABM is a great method of modelling for this application. It is hoped that as well as finding a way to reduce PD in the population, a successful initiative with the aim of reducing unhealthy eating can also be found.

As mentioned in section 2.3, there are existing ABMs in similar subject areas to the on under investigation here. These are more specialised studies focussing on particular sub-groups within the general population [53], [55]. In this study, the aim is to expand upon these and make the model representative of a general population. Another aspect this study will incorporate is to have more detailed characteristics of individual agents than that seen in the ABM representing Canadian mothers [55].

3 Agent-based modelling

3.1 Conceptual modelling

After conducting research on the theory behind PDs link with diet, it was time to start building on the HFSS model to include PD. An overview of what this would look like can be seen in figure 1.

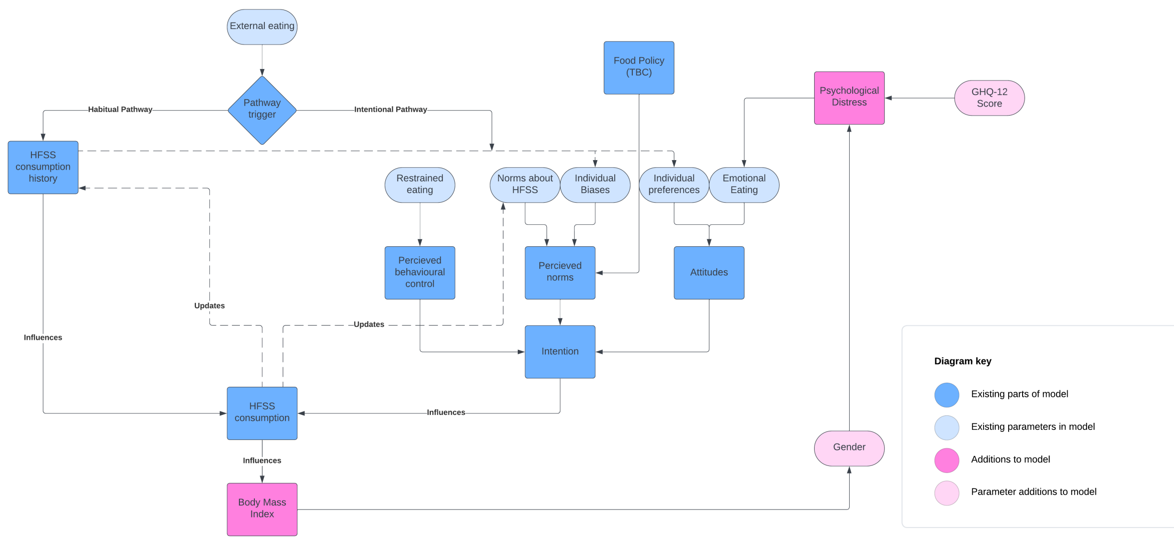


Figure 1: New schematic diagram showing theory behind what the model will incorporate

As seen in figure 1, there are some additions to the theoretical model which is to be implemented in Repast4Py. The main goal of the project is to put PD into the ABM and have this influence other characteristics. Here, PD is initially found using GHQ-12 scores from the HSE and imputing this into the agents which go into the model.

Recap of relevant theories explaining link from PD to HFSS

The relationship between PD and diet quality is mentioned throughout the literature and different mechanisms are used to explain it. It is discussed in section 2.2. One mechanism which appeared frequently in the literature was the *emotional eating* theory [13], [32]. At a high level this theory states that generally, individuals with high levels of PD will eat worse foods to make themselves feel better.

Another theme in the literature was that of a bi-directional relationship between quality of diet and PD [29]–[31]. This is a theory that extends to obesity being a cause of depression [38], which is a big factor of PD [8]. This presented an opportunity to implement a feedback loop into the model. Our agents in our data all had BMI values. BMI is a popular method used globally to determine if a person is underweight, a healthy weight, overweight, or obese. Therefore, this theory would be implemented so that agents with a high BMI, defined in table 1 from the NHS guidance [60], would have their PD increase as a result.

Recap of relationship between BMI and psychological distress

As seen in figure 1, there is a feedback loop introduced where BMI, which is influenced by overall calorie consumption and by HFSS calorie consumption, influences PD. Discussed in the literature review in section 2.2, it is widely accepted that there is correlation between obesity and depression, particularly amongst women and the severely obese. It was now necessary to put a number on this relationship.

Table 1: Table showing ranges of BMI scores and their corresponding categories

BMI score (x)	**Category**
$x < 18.5$	This is described as underweight
$18.5 \geq x < 24.9$	This is described as "healthy" or "normal"
$25 \geq x < 29.9$	This is described as overweight
$30 \geq x$	This is described as obese or severly obese

Recap of relationship between diet quality and BMI

This relationship is briefly discussed in 2.2. In summary, the literature agrees that bad diet quality will lead to an increase in BMI. This infers a link to obesity.

3.2 Agent-based model design

The model was designed with an aim to matching the MBSSM architecture [14] mentioned in section 2.3. This approach was limited due to the structure of the model that this was building upon not being very close to that structure.

After reviewing the existing model, which had been converted from the alcohol model made in RepastHPC [59] to be in Repast4Py and to be modelling a population's intake of HFSS, a way to get PD involved had to be found.

The model architecture can be seen in figure 2. The functions and variables that were added in this project can be seen in red, the functions changed are in blue. What can also be seen is a class in orange, which represents a proposed intervention from the original model that was added to in this project. This was included in the diagram to show that there is potential for this to be made active in the future.

One of the main motivations behind the MBSSM architecture is to ensure that multiple theories can be represented in a single model and it should be straightforward adding new theories. This is mentioned in section 2.3 of the literature review. In this project, the theory from the original model was being built upon and therefore some of the existing classes had to be changed.

The *NormTheory* class includes most of what is represented by 1. The theory is passed through the mediator classes to the agent class where most of it is called at each tick, some is called less frequently, at an interval which is unique for each agent called a habit update interval. This is discussed further in section 3.3.

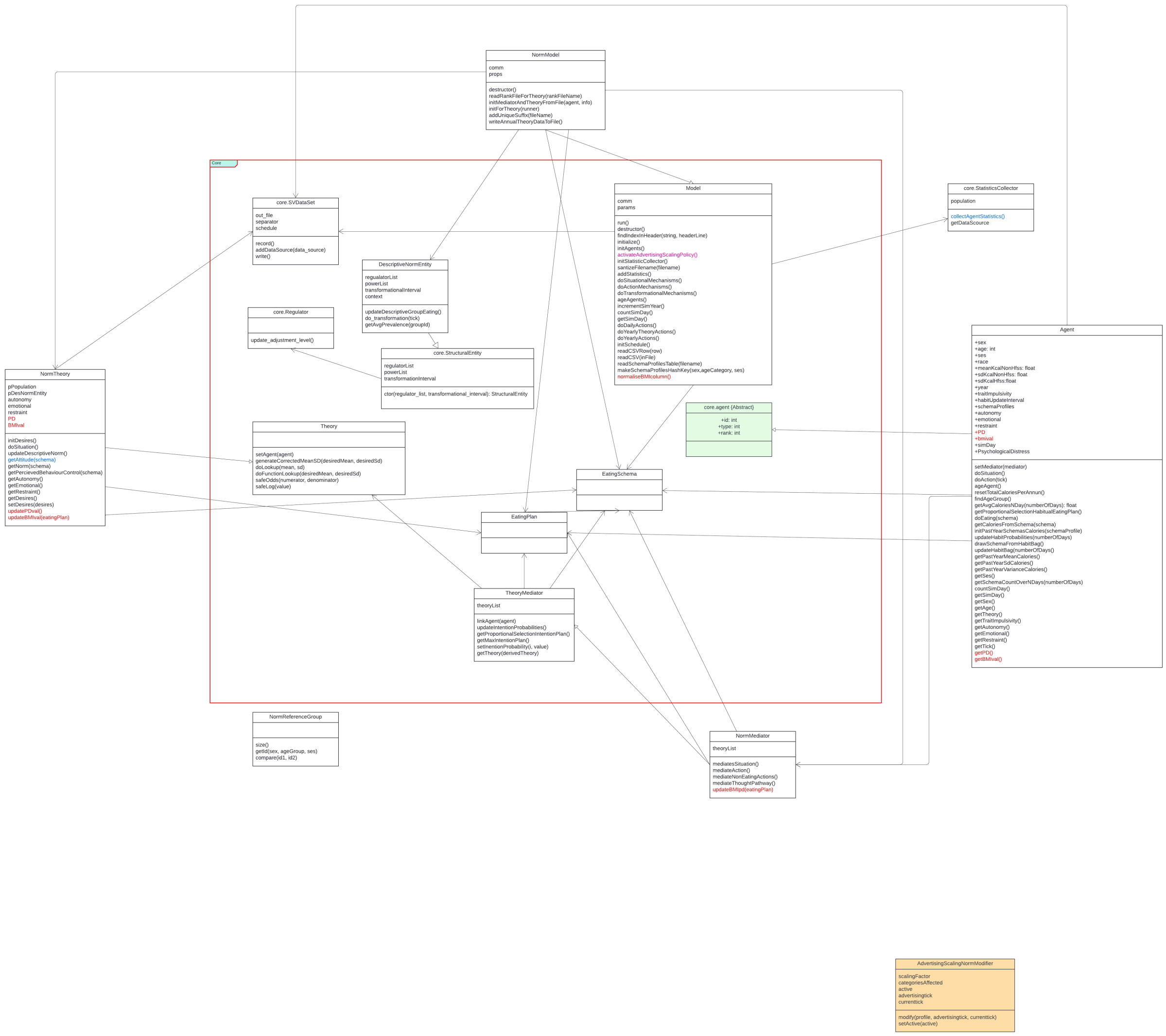


Figure 2: UML class diagram showing model architecture

3.3 Agent-based model implementation

The architecture seen in figure 2 is reflected in the Repast4Py model which is on Bitbucket [2]. This section will go into detail explaining what is new in the model such as the algorithms and equations that have been introduced.

Logit models are known to be an effective way of modelling agent behaviour as mentioned in 2.3. These are present in the functions seen in this section. Therefore, before running the model, we can be reasonably confident that this implementation will produce the desired results.

Updating Emotional Eating with PD

The *NormTheory* class seen in figure 2 was updated appropriately to include the variables needed for the updated emotional eating function *getAttitude*. The equation 1 was then implemented to create thresholds for how an agent's behaviour would change.

$$logit \quad SchemaProb(t^{+}) \leq M * (\alpha + E_{\beta} * E_{\text{agent}} + PD_{\beta} * PD_{\text{agent}} + BMI_{\beta} * BMI_{\text{agent}}) \tag{1}$$

Table 2: Variables in equation to find emotional eating

Symbol	**Variable**
$SchemaProb$	value which is fed into a logit model to get a probabilty
M	Multiplier dependent on if $PD < 4$ or not, where $PD > 4$ is high
α	$Attitude_{\alpha}$
E_{β}	Weighting of importance of historic emotional eating tendencies
E_{agent}	Historical emotional eating tendencies
PD_{β}	Weighting of importance of agent PD
PD_{agent}	Agent's PD
BMI_{β}	Weighting of importance of agent BMI
BMI_{agent}	Agent's BMI

The value for PD_{β} is an estimate of the strength of the relationship between PD and emotional eating. The value for BMI_{β} is the equivalent for the relationship between BMI and emotional eating. This equation can be found in the *getAttitude()* function within the *NormTheory* class seen in figure 2. It can be seen in listing 2 as it appears within the model.

The array that is created from the logit model mentioned in the first row of table 2 is fed by the function logitFunctionSchemaChoice. In lines 16-18, it can be seen that it is more likely that the agent will choose a lower HFSS schema diet if they have low PD. It can then be seen that they will choose a high HFSS schema diet if they have high PD (lines 19-21).

This array of probabilities made based off equation 1 is then added to the agents existing array for choosing a certain HFSS schema. This summed array is then rescaled so all 5 values in it add to one, this is then passed into another function in line 48 of 2 which updates an agents intention probabilities using a combination of three theories including this one, the other two of which are updated in similar ways. An element of randomisation is then introduced to simulate the uncertainty of predicting real-life people's decisions when it comes to food.

When initially producing results from the model, it became clear that the initial assumption that a linear relationship between PD and HFSS consumption was not going to produce results to match the target data. Therefore, some non-linearity had to be introduced. One benefit of an ABM is that it is capable of introducing non-linearity [57], as mentioned in section 2.4.

Updating agent's BMI

After consulting the relevant literature, there were several factors which could be implemented accurately into our model. Firstly, and most obviously, HFSS consumtion will directly impact weight gain. Each agent has a diet history and this data is used to put each agent into categories/schema. Depending on which schema the agent is in, the rate and direction of BMI change will change.

Another factor which was implemented into the model was age. Within the literature there was acceptance that age and weight gain are strongly related [61]. The general pattern of weight gain over a life time is said to be that weight increased up to about 60 years of age and decreases after this point. This was straightforward to implement into the model. The age and HFSS eating schema can be found in the agent class *updateBMIval*.

There were also factors mentioned in the literature which could not be included in the model, one of which was psychological state of an individual. This was something that just fed into how much people eat so it was not necessary to include. Another factor which wasn't possible to include in the model was genetics, some people have a higher susceptibility to becoming obese. However, due to the dataset not having information on this factor, it wasn't possible to implement.

Once a sufficient amount of research had been carried out, it was time to put this into a function *updateBMIval*, this can be found in the *NormTheory* class. It can be seen in listing 3. The equation 2 can be seen in line 12 and line 18 and 21. It applied differently dependent on the current BMI of the agent. This makes the chances that the agent will lose, gain, or stay the same weight vary in different ways.

Between line 33 and 58, we see that the agent will lose and gain weight by different amounts depending on different characteristics which attempt to reflect the theory on the matter.

We can also see in line 35, a random number is generated. The purpose of this number was to make it possible for each agent in our model to gain or lose weight at different rates. This is necessary because everyone puts on weight at different rates in real life due to there being many contributing factors [62]. The changing weight would be represented in the model by a changing BMI.

$$logit \quad BMIchangeProb(t^{+}) \leq \alpha + HFSSBMIVAL_{\beta} * EatingSchema \tag{2}$$

Table 3: Variables to find chances of BMI change and in which direction

Symbol	**Variable**
$BMIchangeProb$	value which is fed into a logit model to get a probabilty
α	BMI_{α}
$HFSSBMIVAL_{\beta}$	Weighting of how important HFSS schema is on the BMI of an agent
$EatingSchema$	Current eating schema of agent

Updating PD based off BMI

As discussed in section 2.2, there is a link between obesity and depression. This is something the ABM had to be redesigned to account for. This was put into the *NormTheory* class within *updatePDval* which can be seen in figure 2. By putting this function in, it meant that the PD of an individual would change, which would lead to changes in their future behaviour such as their emotional eating tendencies. This would complete the feedback loop mentioned in section 3.1. The function can be seen in listing 4, it is based off equation 3.

$$logit \quad PDchangeProb(t^+) \leq \alpha + BMIVALPD_\beta * BMI \tag{3}$$

As seen in equation 3, which is in lines 11, 14 and 18 of listing 3, the chances of PD changing depend on what BMI the agent has. The amount that the PD changes based off BMI is also varied based on the gender and current BMI of the agent. This is seen between lines 30 and 44 in listing 3.

Intervention

An intervention was introduced which is designed to represent an improvement in NHS mental health services, which are widely known to be in poor condition as discussed in 2.1. Mentioned throughout this project, bad mental health and high levels of PD could be having a knock-on effect on the population's dietary habits. This in turn creates more issues with the health of the country. It is desirable for the UK to improve public health including the population's mental health.

The intervention can be seen between lines 47 and 61 in listing 4. It is called on line 64 when the intervention is turned on. Within the function of the intervention, the array "probAppointmentEffective" represents the chances how an individual recieving treatment will respond. The values assigned to each element in this array are selected based off findings from a meta-analysis [26] which looked at the efficacy of various psychotherapy treatments. This was mentioned in 2.1. Also mentioned was the idea that there is very little difference in effectiveness between types of treatment which means there is no need for further detail in this function.

In 4, once the intervention is turned on, the model input will be changed. The variable that is to be changes is $InterventionScale_\beta$. This will be varied from 0.1 to 1. The value it takes represents the size of improvement in NHS mental health services. E.g. a value of 0.1 means a 10% improvement, 0.2 represents a 20% improvement.

4 Model inputs and parameterisation

The work in this section built upon all the work that had been done for the imputation that had only used data from the NDNS [3]. The prior work had then related demographic characteristics with dietary trends it had then finished by creating a representative sample of people in the NDNS with various characteristics such as age, sex, ethnicity and desires for different HFSS schemas included. All of which were relevant to the building of the ABM.

The work that needed to be done here was to include PD, this would be done by imputing PD from the HSE [4] into the original NDNS dataset using various common demographic information. It also meant including BMI, which had not been considered in the original model.

4.1 Datasets

The NDNS is a "continuous, cross-sectional survey. It is designed to collect detailed, quantitative information on the food consumption, nutrient intake and nutritional status of the general population living in private households in the UK" [3]. With regards to the project, it is the sole source of dietary data to be used.

The HSE "monitors trends in the nation's health and care, providing information about adults aged 16 and over, and children aged 0 to 15, living in private households in England" [4]. It is the sole source of psychological data used in this project.

As well as comparing the two datasets, relationships between variables in the surveys were compared using Chi squared tests [63] and correlation. Chi squared tests are able to tell us if the correlation between two variables is a significant one.

Relationships between BMI and emotional eating, and BMI and PD are analysed in this section. Ideally, there would have been a direct analysis of the relationship between PD and emotional eating, but due to the limited information in each survey, this was not possible until imputation had taken place 4.2.

Comparing the two datasets

Due to the NDNS and HSE measuring things in different ways, variables had to be recoded. For example, the age of the respondent variable in NDNS is given as the actual age rather than in categories like in the HSE. A table made to help with this process can be seen in figure 22 in appendix B.

To validate the processed data, it was necessary to perform some descriptive analysis and ensure that this had been done correctly so that the two datasets were compatible. In figure 3 we can see that the distribution of ages in both datasets is similar but that the NDNS has a higher proportion of young people in it. This served as enough evidence that the recoding had been done correctly.

Another variable which had been recoded was ethnicity. In the NDNS this had a good rate of response in a convenient format 4. The HSE responses were recoded to match this and in figure 4 we can see how the datasets compared. Once again the survey response

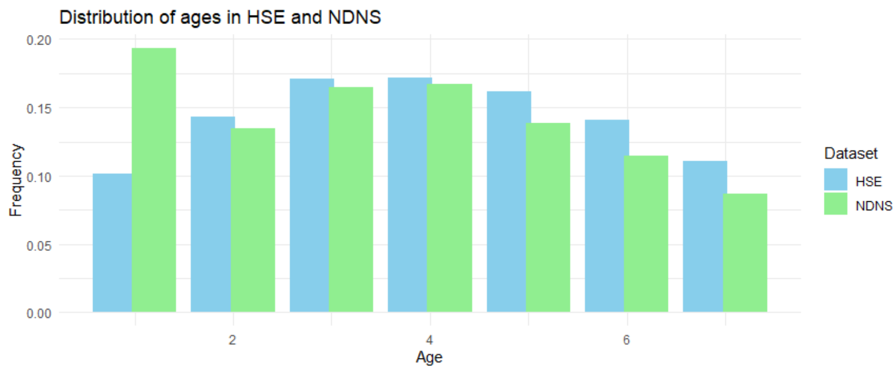


Figure 3: Showing the distribution of ages in both datasets

had a similar distribution which was a good indicator it had been correctly done. This population was noticeably a white majority which meant that the results from this could be slightly bias.

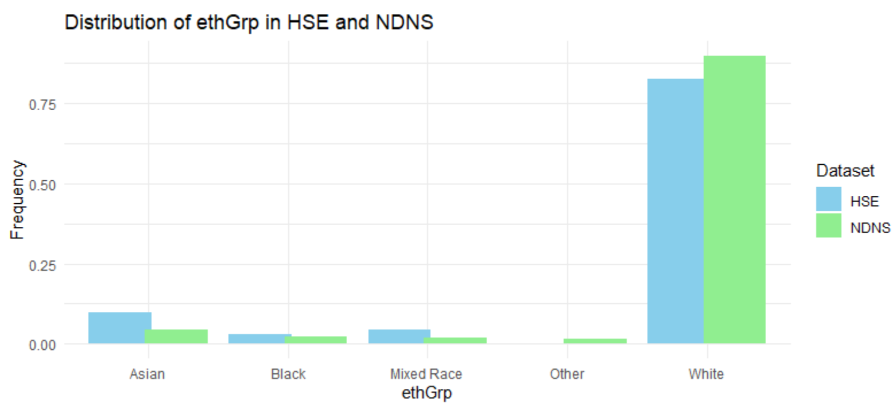


Figure 4: Showing the distribution of ethnicity in both datasets

A third important variable which had been recoded was BMI. The distribution of response rates can be seen in 5. As we can see, both surveys have a similar distribution. Some values within the HSE dataset are outliers like the person with a BMI of 81. Outliers were removed from the data to improve imputation accuracy.

Analysing relationship between BMI and emotional eating

One relationship which was tested was that between BMI and emotional eating. Due to the limited data in NDNS alone, emotional eating was inferred by HFSS consumption. HFSS consumption was found using a normalised value for a combination of total fat, salt and sugar consumed by an agent. The relationship between this and BMI was then analysed. After conducting a Chi squared test, the $p-value$ was 0.24. Even at a 5% level of significance, this would not hold up as a significant relationship despite a correlation of 0.1.

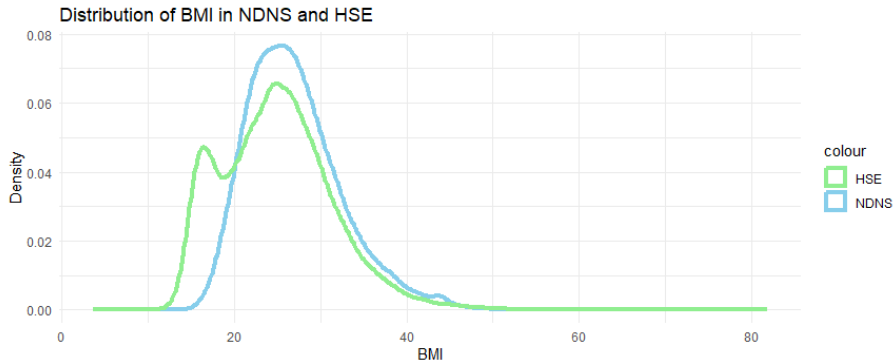


Figure 5: Showing the distribution of BMI in both datasets

This finding is contrary to what was found in the literature in section 2.2. It also goes against one article [46] which puts a path coefficient of strength 0.059 on this relationship for women, tested at $p < 0.01$. The $p - value$ in our test does not change when we exlude men. Path coefficients were introduced as a concept in 2.2.

Analysing relationship between BMI and PD

This relationship had to be tested using data from the HSE due to the NDNS not having any information regarding PD. The raw data was used to find the strength and significance of the relationship.

The Chi squared test yielded a $p - value$ of $2.715e^{-5}$ which is significant and a correlation strength of 0.07. This supports the literature in 2.2 where there was limited research done to put a numerical value on the relationship between obesity and PD. In Vittengl's mediation [46], a number was put on the strength of the link between obesity and depression over time. The strength was 0.264 tested at $p < 0.01$. Despite this being about depression, which is a factor of PD, as opposed to PD as a whole there is a lot than can be drawn from his results. These have been tested for men and women, with a significant relationship only being found for women.

The relationship found in the HSE data supports this claim. The caveat to this is that Vittengl's results [46] are from surveys taken over a long period of time (18 years), which is on a much longer scale than our model. The relationship with strength 0.264 is between obesity and emotional eating 9 years after obesity was recorded. Our model uses a theory that each individual takes a different period of time for their habits to change, with most people's *habit update interval* being less than half a year.

The relationship can be seen grapically in 6. The line of best fit shows a very small correlation (0.07).

4.2 Imputation

To impute PD into the NDNS it was necessary to use *hot-deck imputation*. This involves "each missing value being replaced with an observed response from a 'similar' unit"[64]. In this case, the missing data was the NDNS data for PD because the NDNS does not

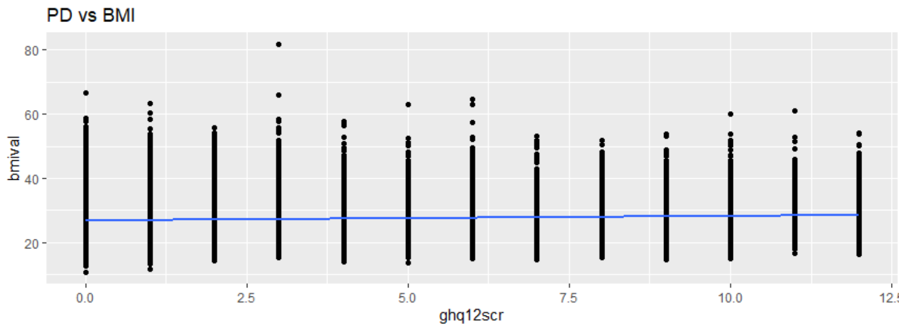


Figure 6: Plot of BMI against PD from the HSE

have a satisfactory measure of PD. The similar units were constructed from the HSE data and replaced the missing data. In other words, the NDNS [3] acted as the *base* data, the HSE [4] as the *donor* dataset.

One part of processing was to replace negative values with *NA*. Negative values are inserted when data is missing as common practice. These nagative values are different because they correspond to different reasons, e.g. "the participant refused to answer". However, they had to removed because the negative numbers would make the imputation algorithm inaccurate.

The variables that were used to impute PD into the NDNS agents and how they were measured can be seen in table 4.

Table 4: Variables used in imputation of NDNS agents

Variable	**Detail**
Year	The year of the survey in which person responded
Age	Age of respondent. This had to be done in categories
Sex	Sex of respondent
Ethnic Group	Ethnic group of respondent, white, black, asian, mixed, or other
Equivalised income tertiles	Each survey year the income of respondents was split into tertiles
GHQ-12-Score	Score given after completion of GHQ-12 by respondent in HSE
BMI value	BMI was measured in both surveys
Restrained eater	HSE asked respondent, "Are you trying to lose or gain weight or neither?"
Weight Conscious	HSE asked respondent, "Would you describe yourself as too light, too heavy or the right weight?"
HFSS intake quintile	This used data from the NDNS HFSS dataset to find what level of HFSS the agent consumed

Dummy variables

A common method to improve imputation accuracy is to create dummy variables [65] on categorical data. An example of this in the project was recoding the *ethnic group* variable. For each ethnic group, represented by a number in the data before the creation of dummy variables, a column was created and it would be filled with either ones or zeros. With a one if the person was of this ethnic group, and a zero if not.

The reason dummy variables are necessary, particularly for categorical values, is because categorical values cannot be represented by numbers, if they were, this would introduce unintended meaning or order to the data. For example if the number 2 represented black ethnicity, and 3 represented asians, this would imply that black ethnicity is in some way less than asian with regards to the variable being imputed. This would skew the imputation making it less accurate. Ordinal values such as age were processed in the same way for the same reasons. No order or meaning was assumed unless the literature had given reason otherwise.

No dummy variables were created for BMI value, GHQ-12-Score, or HFSS intake quintile. This was because the literature was clear in saying that these were related to eachother and it was decided that this was a good way of ensuring these relationships were not lost and overall information was kept to a maximum.

Choosing imputation method

An R package called MICE (multivariate imputation via chained equations) was made for the purpose of imputation [66]. This package has many features which are very adaptable and easy to use.

Imputation is used in research and industry to counter the problem of missing data, MICE is considered a robust method. It is "flexible, reduces bias, preserves relationships between variables, and is particularly useful for large datasets with missing data" [67]. It is also known for dealing with lots of missing data. Based off research of the technique, it was decided that this would be the method used going forward.

There was another question of what method within MICE should be used for each column of data. As we only wanted to impute PD, it was not necessary to waste computational power on imputing the other missing variables. There were a number of options available in MICE for choosing the actual method for the PD column. One of the most popular methods is predictive mean matching (PMM). PMM is particularly useful when your variable is skewed as PD is, this is seen in figure 7, with PD being skewed to the left. This is a pattern seen in other studies too [68].

Validating imputation

As seen in figure 7, a good early sign is that the distribution of PD remained almost identical after imputation. Comparing imputed data with original data is the most popular way of checking whether imputation is accurate or not.

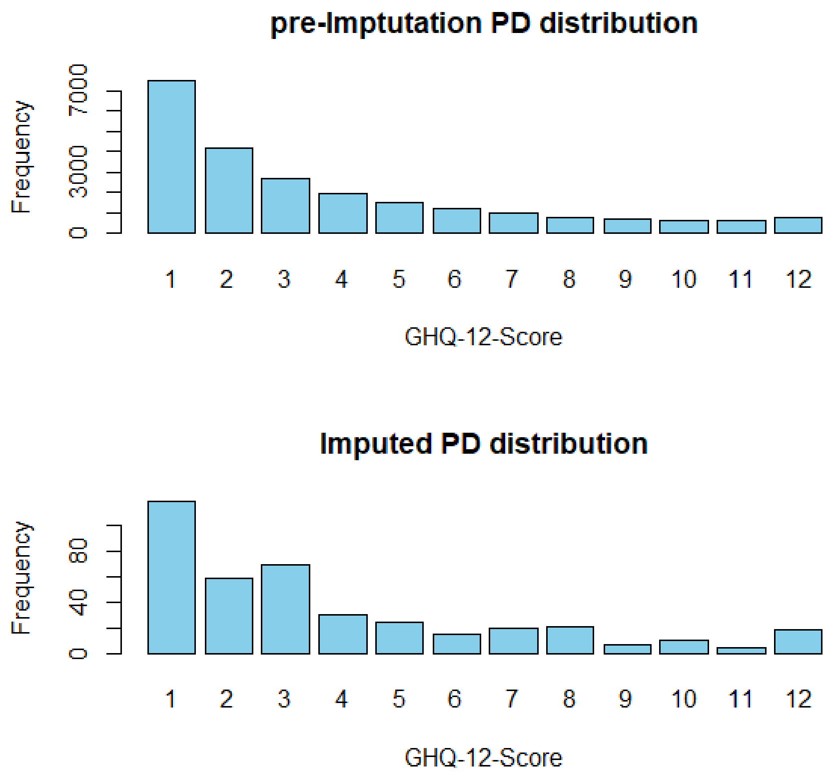


Figure 7: Pre and post imputation distribution of PD by GHQ-12-Score

Post imputation analysis of relationship between PD and diet quality seen in data

Having wanted to analyse the relationship between PD and diet quality in section 4.1 but not being able to, it was now possible post imputation. Disappointingly, the data did not reflect what the literature had suggested would be the case.

As seen in figure 8, a historical tendency to consume a high amount of HFSS foods is related with a lower PD. In figure 9, a high PD is associated with less desire for HFSS foods. Both of these relationships were tested using Chi squared tests and had $p-value$ scores of $1.417e^{-0.5}$ and $2.2e^{-16}$ respectively. The HFSS data was generated using a weighted sum of the probability that an individual would choose to eat in one of the schemas. These schemas are seen in table 5 and the equation to find the weighted sum is in 4. The data used is the model input data.

Table 5: A table showing HFSS schemas and how they were split

Schema	**1**	**2**	**3**	**4**	**5**
Calories from HFSS food	0 - 200	201 - 500	501 - 750	751 - 1000	1001+

$$weightedHFSS = s1 * 0 + s2 * 200 + s3 * 500 + s4 * 750 + s5 * 1000 \tag{4}$$

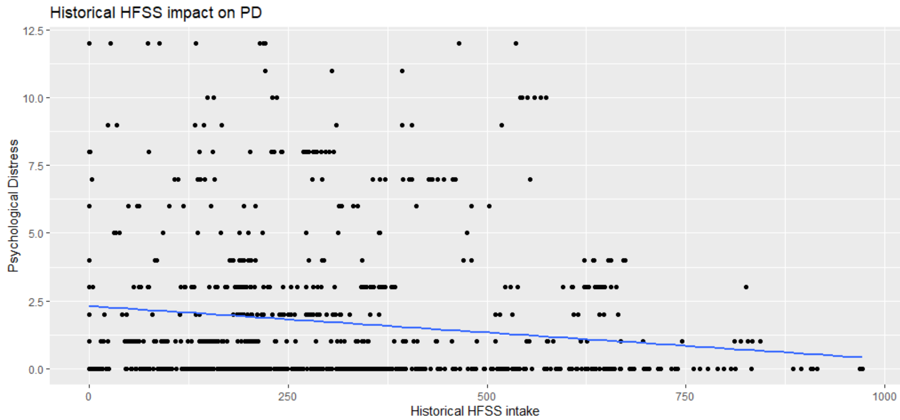


Figure 8: Scatter plot showing impact of historical consumption of HFSS on psychological distress

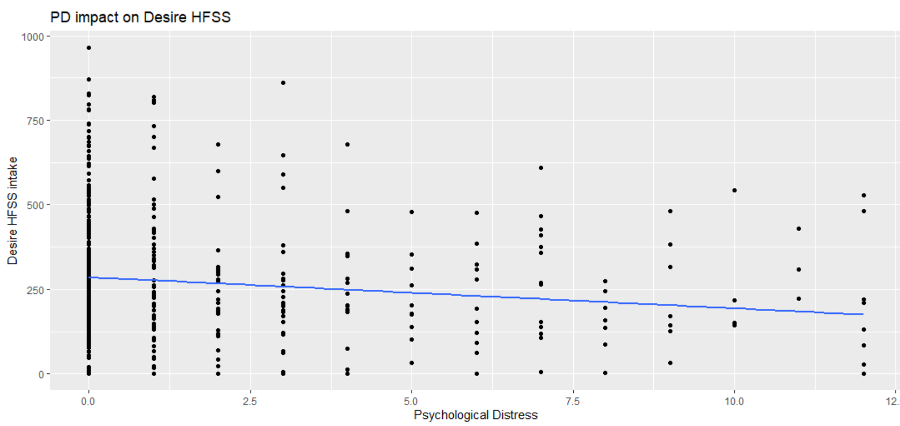


Figure 9: Scatter plot showing impact of PD on HFSS desire

This conflicts with a path analysis on the relationship between PD and emotional eating had been done in [33]. This article was mainly focussed on the effect of HHFI on PD and other things, but had also characterised this relationship. It had assigned a 0.472 strength correlation, or path coefficient, at the statistically highly significant level ($p < 0.001$), which means this strength has less than a one in a thousand chance of being wrong.

Following an expansion of the literature review, other articles with a relevant path analysis were found. These articles assigned slightly weaker path coefficients to the relationship in question with [44] giving the relationship a correlation of 0.015, again at the statistically highly significant level ($p < 0.001$). In [45], a slightly different but still very relevant relationship was being recorded. This was the relationship between anxiety and binge eating. Anxiety is a large part of the definition for PD [8], and binge eating is known to have a lot overlap with emotional eating [32], [47].

A possible reason for the data not reflecting the widely agreed upon theories in the literature is that the literature didn't look at diet quality in the same way as this project. Total HFSS consumption and desire for HFSS foods is not the same as emotional eating or binge eating. A relationship was inferred by generalising what these types of eating are using research from [32], [47].

Post imputation analysis of relationship between BMI and HFSS intake

Although this was already analysed in 4.1, it was decided to try again with a more cohesive measure of HFSS consumption. This produced the results seen in 10 and 11.

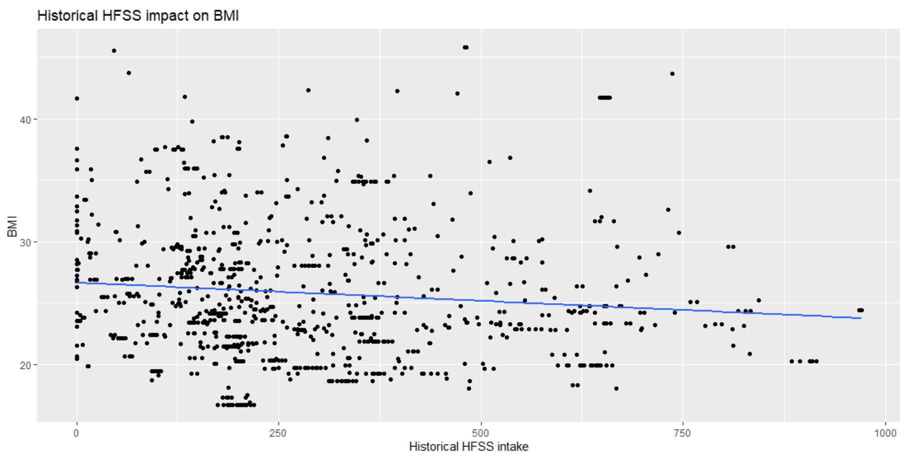


Figure 10: Scatter plot showing impact of historical consumption of HFSS on BMI

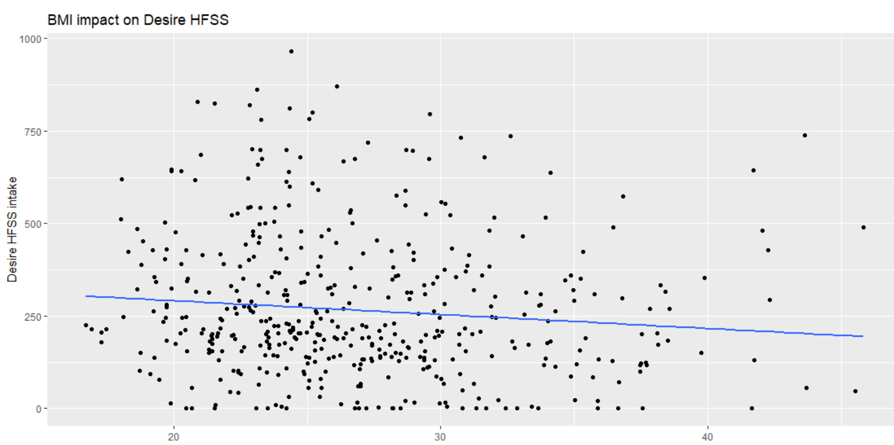


Figure 11: Scatter plot showing impact of BMI on HFSS desire

Similar to the equivalent analysis for impact of/on PD in section 4.2, the relationship was the opposite of what was expected. Once again the correlation was tested to a very high significance and was proven to be a significant relationship. Based off the findings in the literature, the model would try and fix this.

4.3 Model inputs and parameters

The final steps of preparation before the model was to be ran was to get the model inputs. As seen at the start of this section 4, the ABM is set up to recieve a representative population. It also requires a YAML file which contains parameters which would be calibrated to fit real-life historical data.

Due to the involvement of PD and BMI in the emotional eating equation seen in section 3.3, it was necessary to normalise the values of these variables so they ranged from 0 to 1, just like the other factors involved with each agent's emotional eating habits. This would ensure that they would not be more heavily weighted than the other contributing factors. The process of doing this is explained in appendix B.

Parameters

Due to the nature of ABM, with various algorithms representing relationships between characteristics, it is necessary to characterise these relationships and put a numerical figure on them. These parameters are altered to fit the model to real life historical data. The final parameters can be seen in table 6.

Path analysis was introduced in 2.2 and has been mentioned in section 4.2 regarding the relationship between PD and food choices, it is a way of putting a number on relationships between variables in our model. Although useful to see that research has been conducted on path coefficients in the relevant subject area, due to models being simplified versions of reality, the path coefficients found by previous study will not give much of an indication as to what to expect from our model.

These parameters are specific to the model and have been calibrated using Latin hypercube sampling (LHS) [69]. This is a relatively computationally cheap way of parameter optimisation when building a model. In this project, there was a limitation on how many parameter combinations could be tested due to a lack of comptational power. Therefore, it is likely that better model input parameters exist and could be found with more resources.

The optimal parameters were those which led to the output of the model matching the real-world data. This data was the average HFSS calories consumed by men and women each year between 2008 and 2018.

In table 6 the parameters represent the importance of agent characteristics when calculating other characteristics. These parameters appear in the model in the parts represented by the equations in section 3.3.

5 Results

Once the model had been completed, the parameters were calibrated using LHS as mentioned in section 4.3. These optimised parameters can be seen in table 6.

5.1 HFSS intake trends and parameters

The real-life HFSS intake data can be seen in figure 12 as the thick black line. Over the period 2008-2018, there is a decrease in intake which is more prevalent in men. The output from the calibrated model in red is more stable than that of the real life data apart from an early decrease in the first year. The model is able to replicate the patterns seen from women more closely, perhaps due to women's historical intake levels being more stable. All of the model outputs throughout calibration can be seen in figure 23.

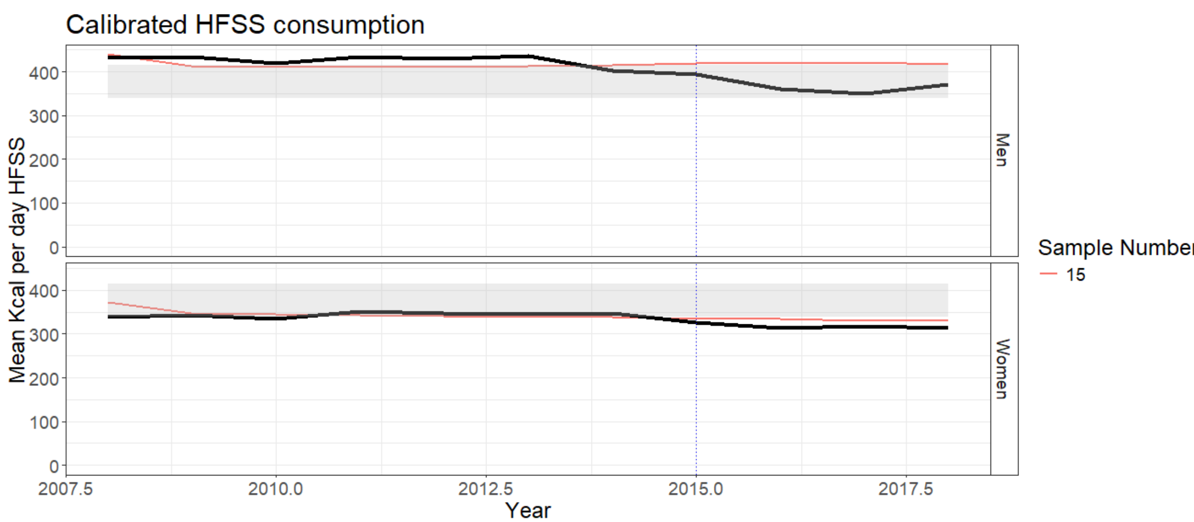


Figure 12: Plot showing optimal model output. The black line is the real-life target data. Calibration only took place up to 2015.

Table 6: A description of unobserved parameters in the model and their allocated values following the calibration process

Model parameter	**Calibrated value**	**Description**
$Attitude_\beta$	0.42	The weighting given to attitudes when calculating intentions
$Norms_\beta$	0.48	The weighting given to norms when calculating intentions
PBC_β	0.50	The weighting given to PBC when calculating intentions
$Restraint_\beta$	0.47	The weighting given to restrained eating in the model when calculating perceived behavioural control
$Emotional_\beta$	0.63	The weighting given to emotional eating in the model when calculating attitudes
PD_β	0.24	The weighting given to psychological distress in the model when calculating attitudes
$BMIvalEmo_\beta$	0.75	The weighting given to BMI in the model when calculating attitudes
$HFSSbmival_\beta$	0.71	The weighting given to HFSS schema choice in the model when updating BMI
$bmivalPD_\beta$	0.33	The weighting given to BMI in the model when updating PD
$attitude_\alpha$	0.46	The threshold in the model when calculating attitude
PD_α	0.67	The threshold in the model when calculating PD
BMI_α	0.27	The threshold in the model when calculating BMI

The parameters for finding an agents tendency toward emotional eating are seen in table 6. As can be seen there, $Emotional_{\beta} = 0.63$ and $BMIvalEmo_{\beta} = 0.75$ are given significantly more weight than $PD_{\beta} = 0.24$. This could be to do with what was mentioned in section 3.3. The relationship between PD and HFSS seems to be non-linear, which meant that the equation 1 was multiplied by different factors dependent on whether an individual was psychologically distressed or not. When distressed, the agent's probabilty of eating poorly inrease by a factor of 10, when not distressed, this was only increased by a factor of 2.

So although the weighting for PD_{β} is relatively low, it may be because PD is already taken into account before this equation 1 is applied.

Psychological distress trends and parameters

For PD, the most effective way of doing this was by looking at the proportion of the population that was classified as being psychologically distressed in the model compared to the real-life data gathered. This can be seen, split by sex, in figure 13.

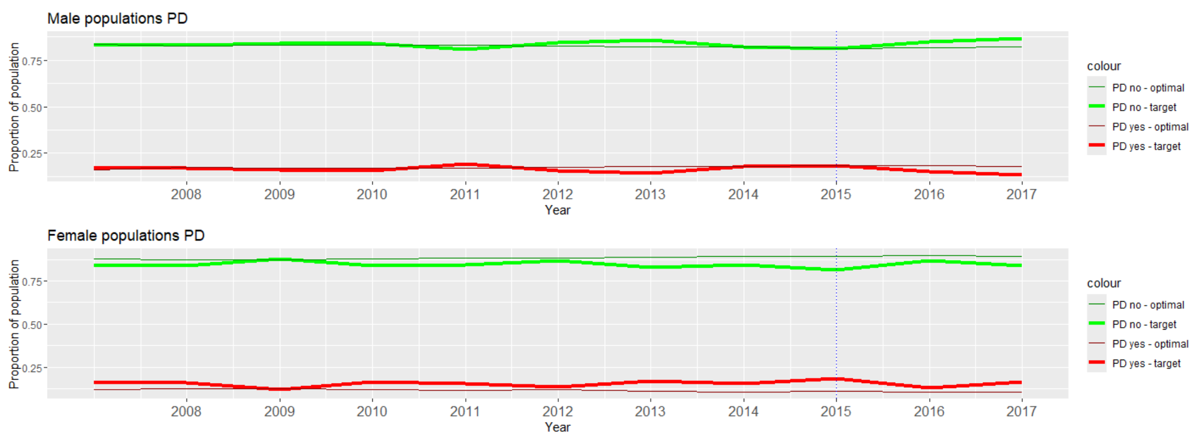


Figure 13: Plot showing best calibrated model output for PD in men and women according to NDNS and HSE datasets. The thicker lines represent the real-life "target data". The thin lines represent the model output. Note this was only calibrated up to 2015.

As we can see in figure 13, the target data fluctuates more than the model output data. Despite this, the model data matches this very closely, which shows that the additions to the model have led to a realistic mechanism for PD.

Although very small, there is a difference between how the simulation data predicts female PD compared with how it predicts male PD. The model predcits a higher proportion of women to be distressed than men from the start. PD amongst men also decreases slightly whereas this does not happen with the women. This could be because of how BMI effects women compared to men in the model, with women's PD being more reactive to BMI than men.

The two parameters used to calculate probabilty in how an agents PD would update were $bmivalPD_{\beta}$ and PD_{α}, as seen in table 6. $bmivalPD_{\beta} = 0.33$ is quite low and suggests that in order to keep PD close to the target data, an agents BMI should not have too much of an effect on PD. $PD_\alpha = 0.67$ suggests that BMI categories have a larger effect on an agents mental health updating than their exact BMI. As both are to do with the agents BMI, there is clear evidence here that BMI is an important factor for PD levels.

BMI trends and parameters

Similar to recording PD, the best way to measure this was to look at the proportion of the population in each category of BMI. These categories are seen in 1. The trends of the population's BMI, split by sex, can be seen in figure 14.

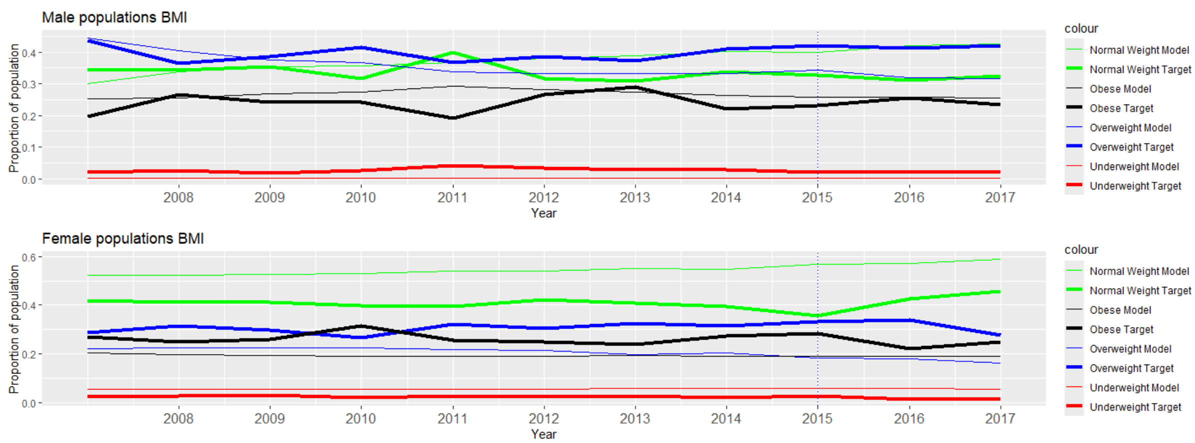


Figure 14: Plot showing best calibrated model output for BMI in men and women according to NDNS and HSE datasets. The thicker lines represent the real-life "target data". The thin lines represent the model output. Note this was only calibrated up to 2015.

As can be seen in 14, the model output for women is much closer to the target data. The model predicts the male population to become a healthier weight. A similar theme occurs in the female data. The model overestimates the proportion of people with a healthy BMI, this could be down to limited calibration.

The model accurately reflects the historical data in showing that the proportion of obese women is close to the proportion of overweight women. Whereas there are significantly more overweight men compared to obese men.

The two parameters used to calculate the probabilty of how an agents BMI would update upon each habit interval were $HFSSbmival_\beta$ and BMI_α, seen in table 6. $HFSSbmival_\beta = 0.71$ is high and shows that an agent's diet is very important when considering how their BMI, and by extension their weight, will change. $BMI_\alpha = 0.27$ shows that a low importance is placed on an agents current BMI when looking at what effects future BMI.

5.2 Effect of government intervention on NHS mental health services

HFSS intake

The premise of the intervention is explained in section 3.3. In figure 15 we see that as the intervention size increases, HFSS consumption decreases more.

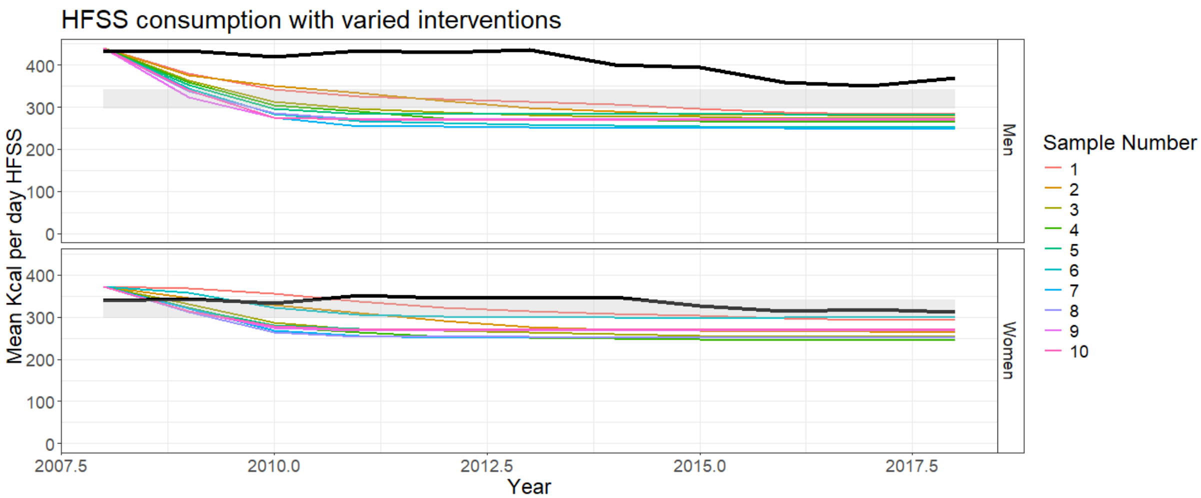


Figure 15: Plot showing how the intervention has a direct effect on HFSS intake. Numbers of the plots correspond to size of intervention, e.g. 1 = 10% improvement in NHS mental health services.

PD levels

Another way in which it was pertinent to measure the effect of the intervention was looking at the size of its impact on PD. To demonstrate this, the PD levels for men and women are plotted with results from a simulated 100% and 20% improvement in NHS mental services availability.

This can be seen in figures 16 and 17. The theme in both of these figures is that the proportion of the population who are psychologically distressed drops to zero for both men and women when the intervention is introduced. A significant decrease is also seen when the size of the intervention is at 20%, which may be a little more realistic than the 100% increase.

As seen in figure 17, the PD levels for men and women significantly improve when the intervention is active at 20%. It takes longer for the 20% increase to improve PD levels than the 100% increase, as expected. In figure 16, we see the what is in reality an unattainable goal of PD being eradicated altogher.

BMI levels

It can be seen in figure 18 that the number of overweight and obese people in the population decreased as a result of the intervention. Again, shown here is the effect of a 100% and 20% increase in availability of NHS mental health services.

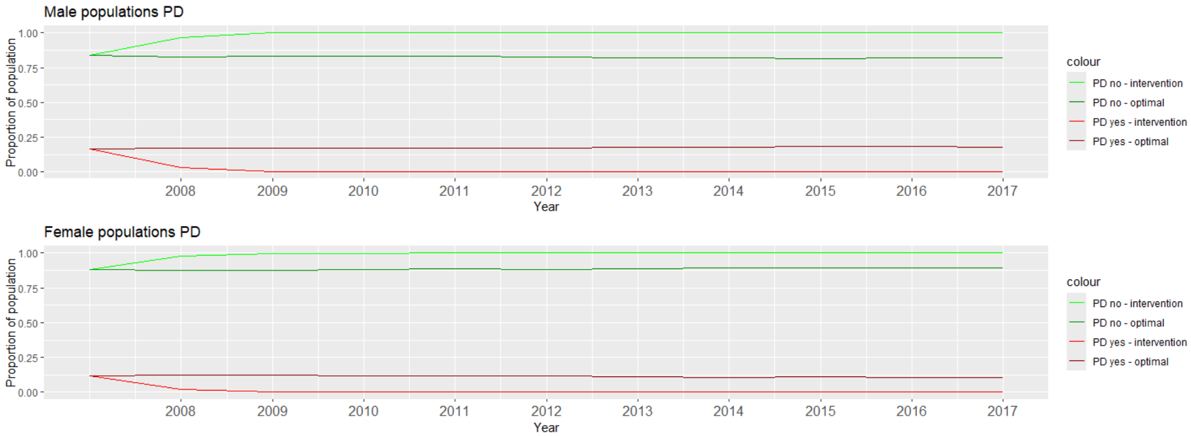


Figure 16: Plot showing how an intervention simulating a 100% increase in availability of NHS mental health services has a direct effect on PD levels. Here it is compared to the optimal model PD levels.

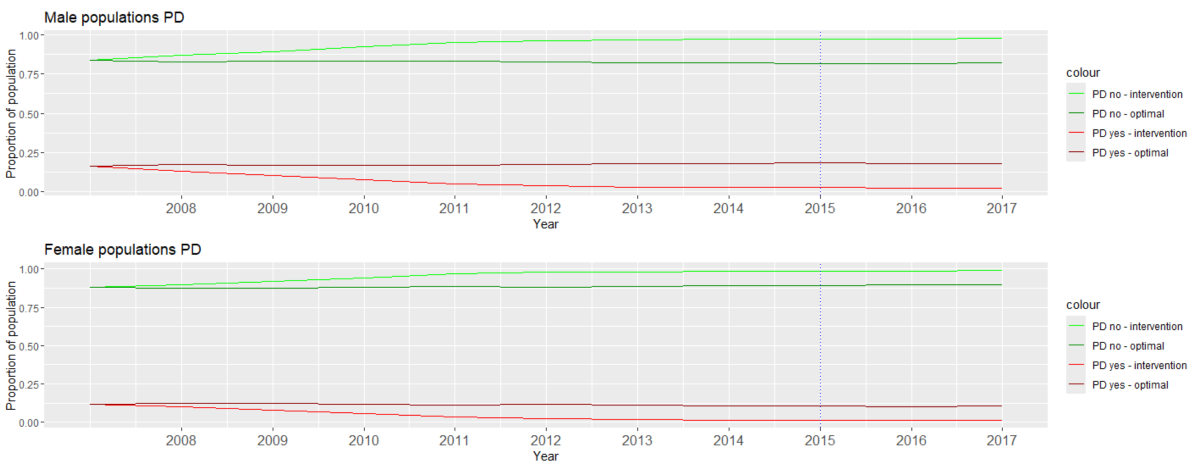


Figure 17: Plot showing how an intervention simulating a 20% increase in availability of NHS mental health services has a direct effect on PD levels. Here it is compared to the optimal model PD levels.

With both a 100% and a 20% increase in availability of NHS mental health services, there is a drastic improvement in obesity levels of men and women. With the overweight population, there is less of an impact, in fact, the optimal model seems to reduce the overweight population more than either of the intervention models. This could be because PD does not carry great weight when calculating HFSS schema probabilities, which is very important when predicting BMI. So if an overweight agent has significantly improved PD, this won't affect their desire to eat HFSS foods, which means their BMI will stay the same.

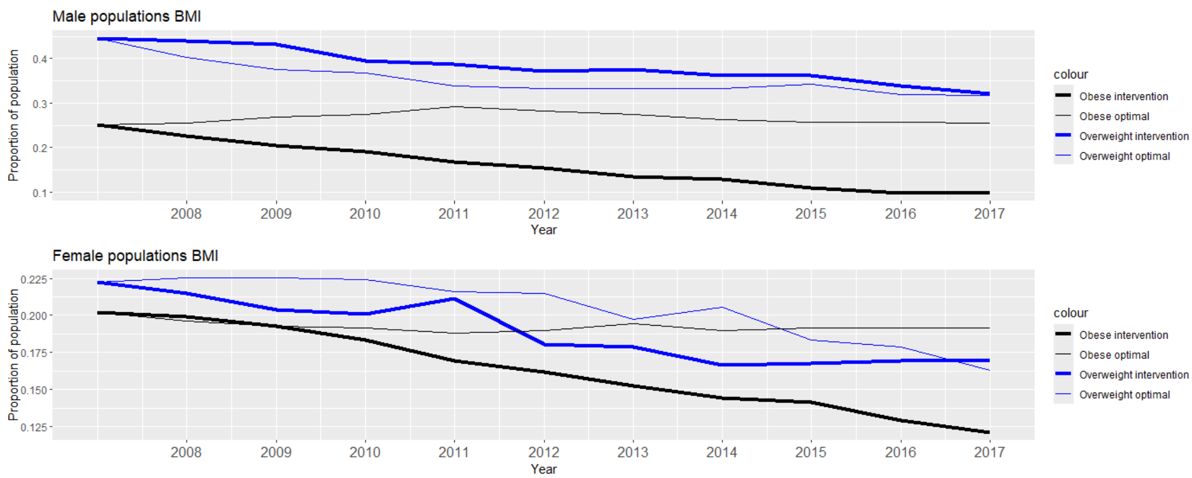


Figure 18: Plot showing how an intervention simulating a 100% increase in availability of NHS mental health services has a direct effect on BMI levels. Here it is compared to the optimal model BMI levels for the propotions relative to total population of the overweight and obese.

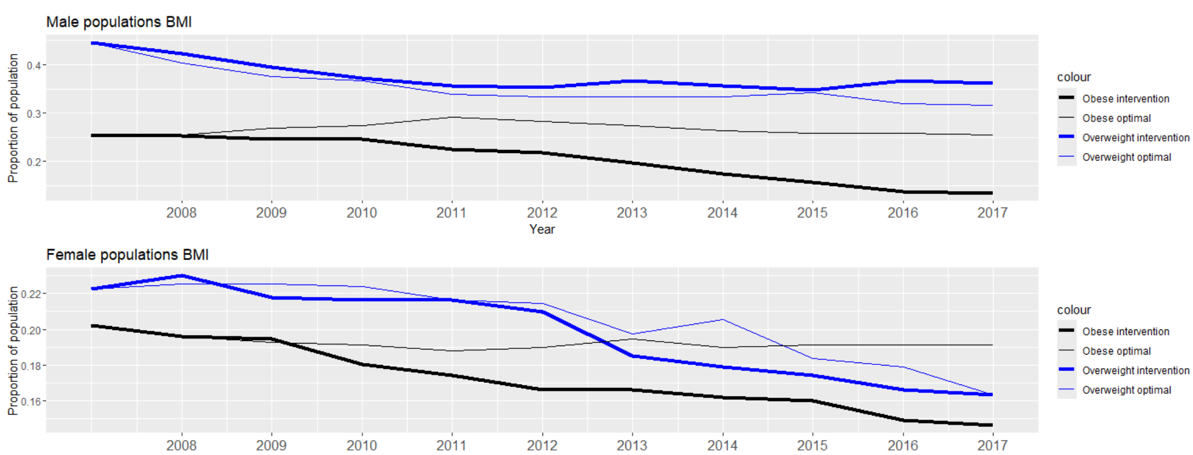


Figure 19: Plot showing how an intervention simulating a 20% increase in availability of NHS mental health services has a direct effect on BMI levels. Here it is compared to the optimal model BMI levels for the propotions relative to total population of the overweight and obese.

5.3 Habit update period

Each agent was given a habit update period in which their BMI, PD and HFSS desires would update. The distribution of the whole populations *habitUpdateInterval* is seen in 20. With regards to PD and BMI, these were updated once per *habitUpdateInterval* for each agent by a maximum amount to maintain a realistic model.

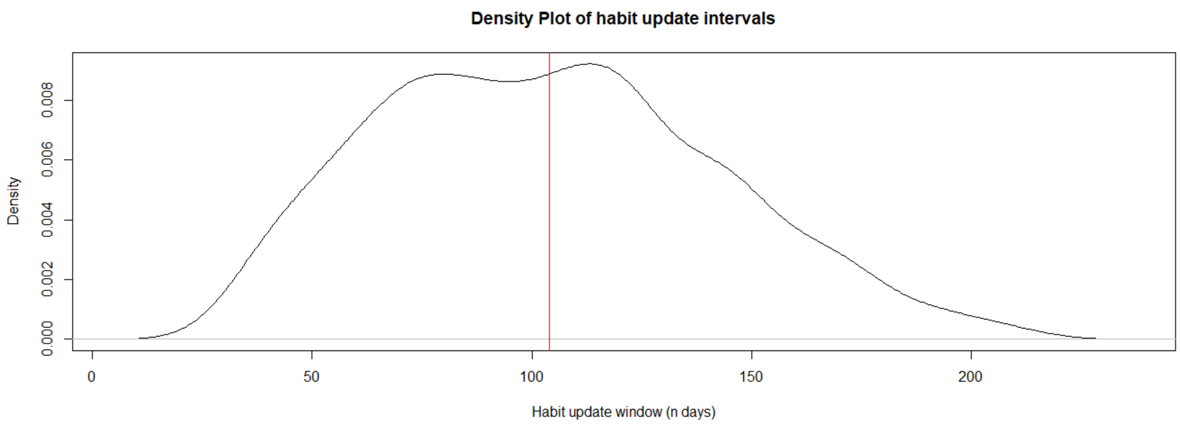


Figure 20: Density plot showing distribution of the input population's habit update interval

6 Conclusions and future work

Conclusions

As previously stated, this model was building PD into a pre-existing model analysing different factors influencing food choice behaviours. After seeing the results, it can be said that the model is accurate to real life in predicting food choice behaviours of a population. One characteristic that could have been more accurately modelled was population BMI. With more calibration, the parameters could have been perfected and the BMI trends would be more closely matched to the target data.

The model provides evidence that the theories introduced into the model are generalisable to the population as a whole. These include introducing psychological distress as a factor for emotional eating, the bi-directionality of the relationship between PD and BMI, and BMIs impact on emotional eating. Using equations representing these theories, historical trends in quantities of HFSS consumed on average by a population were recreated accurately, as were levels of PD. BMI was also accurately modelled, albeit not as accurately as PD and HFSS consumption.

The parameters generated, seen in table 6, can be used in the future to help predict an individuals tendencies for HFSS consumption. As mentioned in 3.3, the model suggests that being psychologically distressed has a large influence on HFSS consumption, more so than for those not distressed.

The model is also useful in simulating the potential impact of hypothetical scenarios. The example presented in 5.2 shows how an increase in NHS mental health services availability by 20% and 100% could affect a populations characteristics. As was expected, there was a more moderate impact for the 20% intervention, although it was still significant when compared to no intervention.

After analysing the results, it was clear to see that a correlation had been successfully forged between BMI and HFSS by the model, as was discussed in section 4.2. This supports previous research on the subject. The model was also able to replicate the relationship between BMI and PD that was dissected in section 4.1. The final relationship introduced to the model in this project was the effect of PD on HFSS desire.

There are clear trends that the agent characteristics follow. As PD improves (decrease in GHQ-12 score), HFSS consumption falls, and the proportion of overweight and obese people in the population drops. This supports the widely accepted theory surrounding the subject.

Future work

Something that has been accepted from the start of the project is that this model would in no way be a finished article and that there would be many things that could be added to it. This is one of the main ideas behind the MBSSM architecture [14]. As mentioned in section 2.3, one of the main inspirations behind this architecture was to make it easier for additions to existing models to be made. The end goal is to have an ABM which can use all relevant psychological theory to produce very accurate estimates of behaviour.

This model was built upon an existing model, adding mechanisms which make the model richer to try and fit the target data more accurately than before. Despite this, there is still huge scope for development. One way in which there could be improvement is by adding social elements to the model. Social comparison theory [70] could be implemented to take a number of things into account. Interaction between agents in ABM is possible and is certainly relevant to obesity, and by extension PD and diet.Another possible avenue for development is to add causes of PD. This could be pertinent if there is unreliable data on current PD amongst agents, which is entirely possible. Some of these causes are discussed in section 2.1. One other possible addition to the model could be based around anorexia, or more generally, possible links from PD to eating less. Additionally HHFI could be added and incorporated into the model, this was discussed in section 2.2.

One basic concept which is missing from the model is agent life and death, this is something that can also be modelled using ABM. The implementation of this could vary in complexity depending on how much detail was put into the theory behind this as there are so many different factors determining life and death.

However, no matter what the additions may be, they need to implemented with caution. Not only will additions cause the model to take longer to run, but it is very difficult to incorporate every nuance of a theory into an equation.

The simulation is more than a good foundation for future food choice modelling, it can be adapted relatively easily despite not being built perfectly to the MBSSM architecture standard, and it allows the inclusion of a broad range of factors that contribute to food choice. The model is a good fit to the real-world historical data, especially for population PD and HFSS consumption. The model suggests that in addition to habits and intentions contributing to food choices, PD and BMI are also significant factors.

7 Bibliography

References

[1] R. McElhatton, *Abmcodeoriginal*, Accessed: 2023, 2024. [Online]. Available: `https://bitbucket.org/ruairimcfinalyprojectabm/abmcodeoriginal/src/main/`.

[2] R. McElhatton, *Abmcodeupdated*, Accessed: 2023, 2023. [Online]. Available: `https://bitbucket.org/ruairimcfinalyprojectabm/abmcodeupdated/src/main/`.

[3] O. for Health Improvement and Disparities, *National diet and nutrition survey*, Accessed: 2023, 2023. [Online]. Available: `https://www.gov.uk/government/collections/national-diet-and-nutrition-survey`.

[4] N. Digital, Accessed: 2023, 2019. [Online]. Available: `https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england`.

[5] G. 2. D. Collaborators, "Health effects of dietary risks in 195 countries, 1990–2017: A systematic analysis for the global burden of disease study 2017," *The Lancet*, vol. 393, no. 10184, pp. 1958–1972, 2019. DOI: `10.1016/S0140-6736(19)30041-8`.

[6] D. R. Theis and M. White, "Is obesity policy in england fit for purpose? analysis of government strategies and policies, 1992–2020," *The Milbank Quarterly*, vol. 99, no. 1, pp. 126–170, 2021. DOI: `10.1111%2F1468-0009.12498`.

[7] K. J. McLachlan and C. R. Gale, "The effects of psychological distress and its interaction with socioeconomic position on risk of developing four chronic diseases," *Journal of Psychosomatic Research*, vol. 109, pp. 79–85, 2018. DOI: `10.1016%2Fj.jpsychores.2018.04.004`.

[8] A. S. Belay, M. M. Guangul, W. N. Asmare, and G. Mesafint, "Prevalence and associated factors of psychological distress among nurses in public hospitals, southwest, ethiopia: A cross-sectional study," *Ethiopian Journal of Health Science*, vol. 31, no. 6, pp. 1247–1256, 2021. DOI: `10.4314/ejhs.v31i6.21`.

[9] N. Digital, Accessed: November 2023, 2016. [Online]. Available: `http://healthsurvey.hscic.gov.uk/support-guidance/public-health/health-survey-for-england-2016/well-being-and-mental-health.aspx`.

[10] J. Badham, E. Chattoe-Brown, N. Gilbert, Z. Chalabi, F. Kee, and R. F. Hunter, "Developing agent-based models of complex health behaviour," *Health and Place*, vol. 54, pp. 170–177, 2018. DOI: `10.1016/j.healthplace.2018.08.022`.

[11] C. Buckley, P. Breeze, A. Brennan, R. Colosanti, M. Leach, and R. Purshouse, "Exploring the relationship between food advertising and consumption of foods high in fat, salt, and sugar in england: An agent-based modelling study," *Appetite*, vol. 189, 2023. DOI: `10.1016/j.appet.2023.106933`.

[12] R. A. Hammond, "Complex systems modeling for obesity research," *Preventing Chronic Disease - Public Health Research, Practice, and Policy*, vol. 6, no. 3, pp. 1–10, 2009. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2722404/.

[13] G. Oliver and J. Wardle, "Perceived effects of stress on food choice," *Physiology and Behaviour*, vol. 66, no. 3, pp. 511–515, 1999. DOI: 10.1016/s0031-9384(98)00322-9.

[14] T. M. Vu, C. Probst, A. Nielsen, *et al.*, "A software architecture for mechanism-based social systems modelling in agent-based simulation models," *The Journal of Artificial Societies and Social Simulation*, vol. 23, no. 3, pp. 1–25, 2019. DOI: 10.18564/jasss.4282.

[15] D. P. Goldberg, R. Gater, N. Sartoruis, *et al.*, "The validity of two versions of the ghq in the who study of mental illness in general health care," *Ethiopian Journal of Health Science*, vol. 27, no. 1, pp. 191–197, 1997. DOI: 10.1017/s0033291796004242.

[16] M. Hankins, "The reliability of the twelve-item general health questionnaire (ghq-12) under realistic assumptions," *BMC Public Health*, vol. 8, no. 335, pp. 1–7, 2008. DOI: 10.1186/1471-2458-8-355.

[17] S. Viertiö, O. Kiviruusu, M. Piirtola, *et al.*, "Factors contributing to psychological distress in the working population, with a special reference to gender difference," *BMC Public Health*, vol. 21, no. 611, pp. 1–17, 2021. DOI: 10.1186/s12889-021-10560-y.

[18] E. L. Kelly, P. Moen, J. M. Oakes, *et al.*, "Changing work and work-family conflict: Evidence from the work, family, and health network," *American Sociological Association*, vol. 79, no. 3, pp. 485–516, 2014. DOI: 10.1177/0003122414531435.

[19] M. Nagasu, K. Kogi, and I. Yamamoto, "Association of socioeconomic and lifestyle-related risk factors with mental health conditions: A cross-sectional study," *BMC Public Health*, vol. 19, no. 1759, 2019. DOI: 10.1186/s12889-019-8022-4.

[20] M. Fluharty, A. E. Taylor, M. Grabski, and M. R. Munafò, "The association of cigarette smoking with depression and anxiety: A systematic review," *Nictotine and Tobacco Research*, vol. 19, no. 1, pp. 3–13, 2017. DOI: 10.1093/ntr/ntw140.

[21] A. C. Edwards, E. Sihvola, T. Korhonen, *et al.*, "Depressive symptoms and alcohol use are genetically and environmentally correlated across adolescence," *Behaviour Genetics*, vol. 41, pp. 476–487, 2010. DOI: 10.1007/s10519-010-9400-y.

[22] H. Norton, *The best countries for mental wellbeing*, Accessed: May 2024. [Online]. Available: https://www.comparethemarket.com.au/health-insurance/features/best-countries-for-mental-wellbeing/#:~:text=Switzerland%20took%20the%20crown%20as,on%20the%20leaderboard%20(3.27)..

[23] M. H. Foundation, *Surviving or thriving? the state of the uk's mental health*, Accessed: April 2024. [Online]. Available: `https://www.mentalhealth.org.uk/explore-mental-health/publications/surviving-or-thriving-state-uks-mental-health`.

[24] N. A. Office, *Progress in improving mental health services in england*, Accessed: April 2024. [Online]. Available: `https://www.nao.org.uk/reports/progress-in-improving-mental-health-services-in-england/`.

[25] B. M. Association, *Mental health pressures in england*, Accessed: April 2024. [Online]. Available: `https://www.bma.org.uk/advice-and-support/nhs-delivery-and-workforce/pressures/mental-health-pressures-data-analysis#:~:text=Demand%20for%20mental%20health%20services%20is%20rising&text=For%20children%20and%20young%20people,more%20than%201%20in%206.`.

[26] M. L. Smith and G. V. Glass, "Meta-analysis of psychotherapy outcome studies," *American Psychologist*, vol. 32, no. 9, pp. 752–760, 1977. DOI: `10.1037//0003-066x.32.9.752`.

[27] P. Keren, F. Ahmed, P. Lee, and J. Wiseman, "Stress and dietary behaviour among first-year university students in australia: Sex differences," *The International Journal of Applied and Basic Nutritional Sciences*, vol. 31, no. 2, pp. 324–330, 2014. DOI: `10.1016/j.nut.2014.08.004`.

[28] M. C. Whatnall, A. J. Patterson, Y. Y. Siew, F. Kay-Lambkin, and M. J. Hutchesson, "Are psychological distress and resilience associated with dietary intake among australian university students?" *International Journal of Environmnental Research and Public Health*, vol. 16, no. 21, pp. 1–15, 2019. DOI: `10.3390/ijerph16214099`.

[29] S. Collins, M. Lotfalian, W. Marx, *et al.*, "Associations between indicators of diet quality and psychological distress, depression and anxiety in emerging adults: Results from a nationally representative observational sample," *Mental Health and Prevention*, vol. 24, pp. 1–8, 2021. DOI: `10.1016/j.mhp.2021.200220`.

[30] F. N. Jacka, A. Mykletun, M. Berk, I. Bjelland, and G. S. Tell, "The association between habitual diet quality and the common mental disorders in community-dwelling adults: The hordaland health study," *Psychosomatic Medicine*, vol. 73, no. 6, pp. 483–490, 2011. DOI: `10.1097/psy.0b013e318222831a`.

[31] O. Sadeghi, A. H. Keshteli, H. Afshar, A. Esmaillzadeh, and P. Adibi, "Adherence to mediterranean dietary pattern is inversely associated with depression, anxiety and psychological distress," *Nutritional Neuroscience*, vol. 24, no. 4, pp. 248–259, 2021. DOI: `10.1080/1028415x.2019.1620425`.

[32] E. L. Gibson, "The psychobiology of comfort eating: Implications for neuropharmacological interventions," *Behavioural Pharmacology*, vol. 23, no. 5 and 6, pp. 442–460, 2012. DOI: `10.1097/fbp.0b013e328357bd4e`.

[33] G. S. Keenan, P. Christansen, and C. A. Hardman, "Household food insecurity, diet quality, and obesity: An explanatory model," *Obesity*, vol. 29, no. 1, pp. 143–149, 2020. DOI: `10.1002/oby.23033`.

[34] M. Maillot, N. Darmon, M. Darmon, L. Lafay, and A. Drewnowski, "Nutrient-dense food groups have high energy costs: An econometric approach to nutrient profiling," *The Journal of Nutrition*, vol. 137, no. 7, pp. 1815–1820, 2007. DOI: 10.1093/jn/137.7.1815.

[35] J. Vaitkevičiūtė and A. Petrauskienė, "The associations between body mass index of seven and eight-year-old children, dietary behaviour and nutrition-related parenting practices.," *Medicina*, vol. 55, no. 1, 2019. DOI: 10.3390/medicina55010024.

[36] L. E. Gutiérrez-Pliego, E. d. S. Camarillo-Romero, L. P. Montenegro-Morales, and J. d. J. Garduño-García, "Dietary patterns associated with body mass index (bmi) and lifestyle in mexican adolescents," *BMC Public Health*, vol. 16, 2016. DOI: 10.1186/s12889-016-3527-6.

[37] L. Huet, I. Delgado, B. Aouizerate, N. Castanon, and L. Capuron, *Chapter 16 - Obesity and Depression: Shared Pathophysiology and Translational Implications*. Academic Press, 2019. DOI: 10.1016/B978-0-12-813333-0.00016-0.

[38] S. J. Mooney and A. M. El-Sayed, "Stigma and the etiology of depression among the obese: An agent-based exploration," *Social Science and Medicine*, vol. 148, pp. 1–7, 2016. DOI: 10.1016/j.socscimed.2015.11.020.

[39] J. Ma and L. Xiao, "Obesity and depression in us women: Results from the 2005–2006 national health and nutritional examination survey," *Obesity*, vol. 18, no. 2, pp. 347–353, 2010. DOI: 10.1038/oby.2009.213.

[40] M. A. Friedman and K. D. Brownell, "Psychological correlates of obesity: Moving to the next research generation," *Psychological Bulletin*, vol. 117, no. 1, pp. 3–20, 1995. DOI: 10.1037/0033-2909.117.1.3.

[41] S. Markowitz, M. A. Friedman, and S. M. Arent, "Understanding the relation between obesity and depression: Causal mechanisms and implications for treatment," *Clinical Psychology: Science and Practice*, vol. 15, no. 1, pp. 1–20, 2008. DOI: 10.1111/j.1468-2850.2008.00106.x.

[42] W. Katon and P. Ciechanowski, "Impact of major depression on chronic medical illness.," *Journal of Psychosomatic Research*, vol. 53, no. 4, pp. 859–863, 2002. DOI: 10.1016/s0022-3999(02)00313-6.

[43] S. Wright, "The method of path coefficients," *The Annals of Mathematical Statistics*, vol. 5, no. 3, pp. 161–215, 1934. [Online]. Available: http://www.jstor.org/stable/2957502.

[44] A. Guerrini-Usubini, R. Cattivelli, A. Scarpa, *et al.*, "The interplay between emotion dysregulation, psychological distress, emotional eating, and weight status: A path model," *International Journal of Clinical and Health Psychology*, vol. 23, no. 1, p. 100 338, 2023, ISSN: 1697-2600. DOI: 10.1016/j.ijchp.2022.100338.

[45] L. S. Duarte-Guerra, E. Kortchmar, E. C. S. Maraviglia, *et al.*, "Longitudinal patterns of comorbidity between anxiety, depression and binge eating symptoms among patients with obesity: A path analysis," *Journal of Affective Disorders*, vol. 303, pp. 255–263, 2022. DOI: 10.1016/j.jad.2022.02.030.

[46] J. R. Vittengl, "Mediation of the bidirectional relations between obesity and depression among women," *Psychiatry Research*, vol. 264, pp. 254–259, 2018. DOI: 10.1016/j.psychres.2018.03.023.

[47] J. Wardle and E. L. Gibson, *Stress: Concepts, Cognition, Emotion, and Behavior: Handbook in Stress Series, Volume 1.* Academic Press, 2016, ch. 55, pp. 435–443. DOI: 10.1016/B978-0-12-800951-2.00058-3.

[48] J. Wardle, A. Steptoe, G. Oliver, and Z. Lipsey, "Stress, dietary restraint and food intake," *Journal of Psychosomatic Research*, vol. 48, no. 2, pp. 195–202, 2000. DOI: 10.1016/s0022-3999(00)00076-3.

[49] S. E. of Philosophy, *Folk psychology as a theory*, Accessed: May 2024. [Online]. Available: https://plato.stanford.edu/entries/folkpsych-theory/.

[50] M. Lages and A. Scheel, "Logistic mixed models to investigate implicit and explicit belief tracking," *Frontiers in Psychology*, vol. 7, no. 1, pp. 40–64, 2010. DOI: 10.3389/fpsyg.2016.01681.

[51] E. Silverman, U. Gostoli, S. Picascia, *et al.*, "Situating agent-based modelling in population health research," *Emerging Themes in Epidemiology*, vol. 18, no. 10, pp. 1–15, 2021. DOI: 10.1186/s12982-021-00102-7.

[52] A. M. El-Sayed, P. Scarborough, L. Seemann, and S. Galea, "Social network analysis and agent-based modeling in social epidemiology," *Epidemiologic Perspectives and Innovations*, vol. 9, no. 1, pp. 1–9, 2012. DOI: 10.1186%2F1742-5573-9-1.

[53] B. G. Silverman, N. Hanrahan, G. Bharathy, K. Gordon, and D. Johnson, "A systems approach to healthcare: Agent-based modeling, community mental health, and population well-being," *Artificial Intelligence in Medicine*, vol. 63, no. 2, pp. 61–71, 2015. DOI: 10.1016/j.artmed.2014.08.006.

[54] C. on Budget and P. Priorities, Accessed: 2023, 2020. [Online]. Available: https://www.cbpp.org/research/policy-basics-introduction-to-medicaid.

[55] C. Benny, S. Yamamoto, S. McDonald, R. Chari, and R. Pabayo, "Modelling maternal depression: An agent-based model to examine the complex relationship between relative income and depression," *International Journal of Environmental Research and Public Health*, vol. 19, no. 7, pp. 1–12, 2022. DOI: 10.3390/ijerph19074208.

[56] S. M. Horwitz, M. J. Briggs-Gowan, A. Storfer-Isser, and A. S. Carter, "Prevalence, correlates, and persistence of maternal depression," *Journal of Women's Health*, vol. 16, no. 5, pp. 678–691, 2007. DOI: 10.1089/jwh.2006.0185.

[57] R. Beheshti, M. Jalalpour, and T. A. Glass, "Comparing methods of targeting obesity interventions in populations: An agent-based simulation," *SSM - Population Health*, vol. 3, pp. 211–218, 2017. DOI: 10.1016%2Fj.ssmph.2017.01.006.

[58] K. Koh, R. Reno, and A. Hyder, "Examining disparities in food accessibility among households in columbus, ohio: An agent-based model," *Food Security - The Science, Sociology and Economics of Food Production and Access to Food*, vol. 11, pp. 317–331, 2019. DOI: 10.1007/s12571-019-00900-7.

[59] C. Buckley, M. Field, T. M. Vu, *et al.*, "An integrated dual process simulation model of alcohol use behaviours in individuals, with application to us population-level consumption, 1984-2012," *Addictive Behaviours*, vol. 124, pp. 1–10, 2022. DOI: 10.1016/j.addbeh.2021.107094.

[60] N. inform, *Body mass index (bmi)*, Accessed: May 2024. [Online]. Available: https://www.nhsinform.scot/healthy-living/food-and-nutrition/healthy-eating-and-weight-loss/body-mass-index-bmi/.

[61] A. DE, F. L, B. M, S. SA, and H. TB, "A research agenda: The changing relationship between body weight and health in ageing," *The Journals of Gerantology*, vol. 63, no. 11, pp. 1257–1259, 2008. DOI: 10.1093/gerona/63.11.1257.

[62] E. K. S. N. I. of CHild Health and H. Development, *What causes obesity and overweight?* Accessed: April 2024. [Online]. Available: https://www.nichd.nih.gov/health/topics/obesity/conditioninfo/cause.

[63] T. BMJ, *8. the chi squared tests*, Accessed: May 2024. [Online]. Available: https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/8-chi-squared-tests.

[64] R. R. Andridge and R. J. A. Little, "A review of hot deck imputation for survey non-response," *International Statistical Review*, vol. 78, no. 1, pp. 40–64, 2010. DOI: 10.1111/j.1751-5823.2010.00103.x.

[65] S. M. Advanced Research Computing and U. Data Analytics, *Coding systems for categorical variables in regression analysis*, Accessed: December 2023. [Online]. Available: https://stats.oarc.ucla.edu/spss/faq/coding-systems-for-categorical-variables-in-regression-analysis-2/#DUMMYCODING.

[66] S. v. Buuren and K. Groothuis-Oudshoorn, "Mice: Multivariate imputation by chained equations in r," *Journal of Statisitical Software*, vol. 45, no. 3, pp. 1–67, 2011. DOI: 10.18637/jss.v045.i03.

[67] B. Soni, *Topic:9 mice or multivariate imputation with chain-equation*, Accessed: December 2023, 2023. [Online]. Available: https://medium.com/@brijesh_soni/topic-9-mice-or-multivariate-imputation-with-chain-equation-f8fd435ca91.

[68] S. Tomitaka, Y. Kawasaki, K. Ide, M. Akutagawa, Y. Ono, and T. A. Furukawa, "Distribution of psychological distress is stable in recent decades and follows an exponential pattern in the us population," *Scientific Reports*, vol. 9, no. 11982, 2019. DOI: 10.1038/s41598-019-47322-1.

[69] A. Olsson, G. Sandberg, and O. Dahlblom, "On latin hypercube sampling for structural reliability analysis," *Structural Safety*, vol. 25, no. 1, pp. 47–68, 2003. DOI: 10.1016/S0167-4730(02)00039-5.

[70] L. Festinger, "A theory of social comparison processes," *Human Relations*, vol. 7, no. 2, pp. 117–140, 1954. DOI: 10.1177/001872675400700202.

A Appendix - Project plan and progress

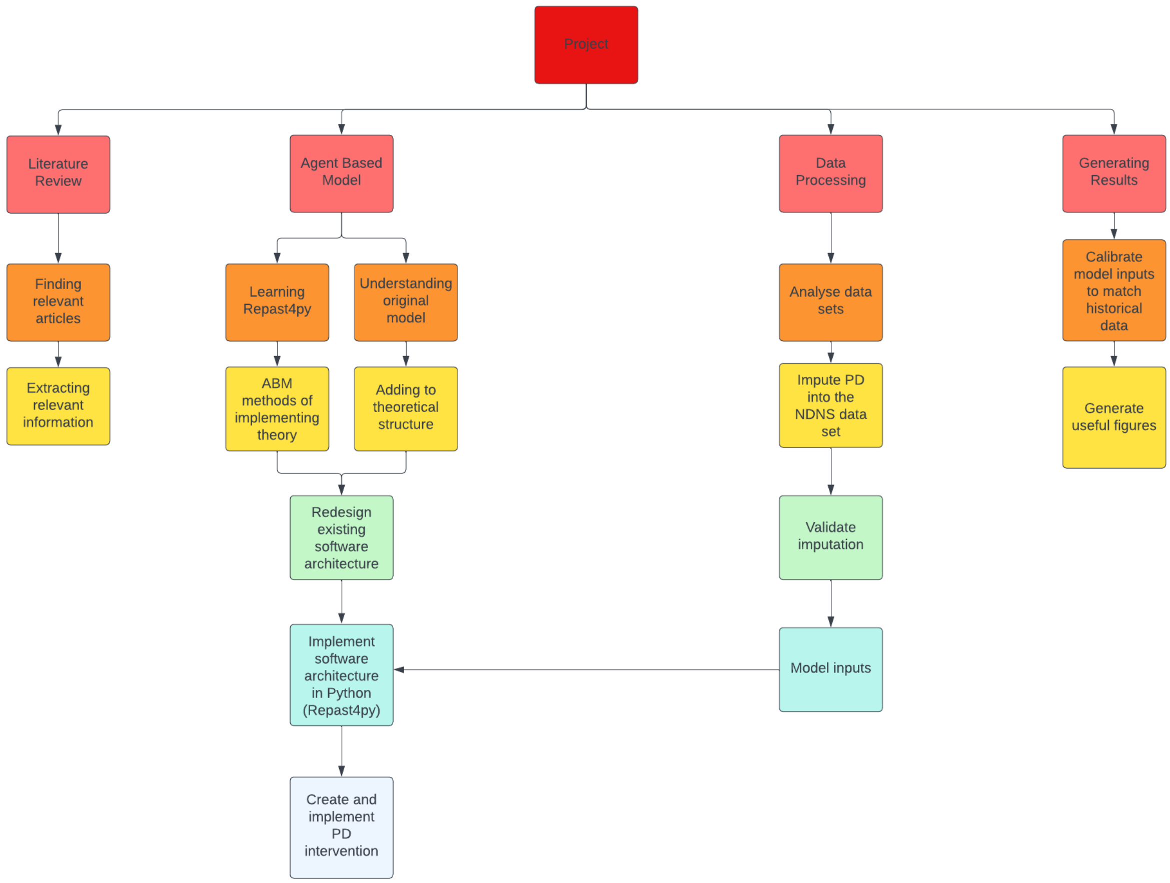


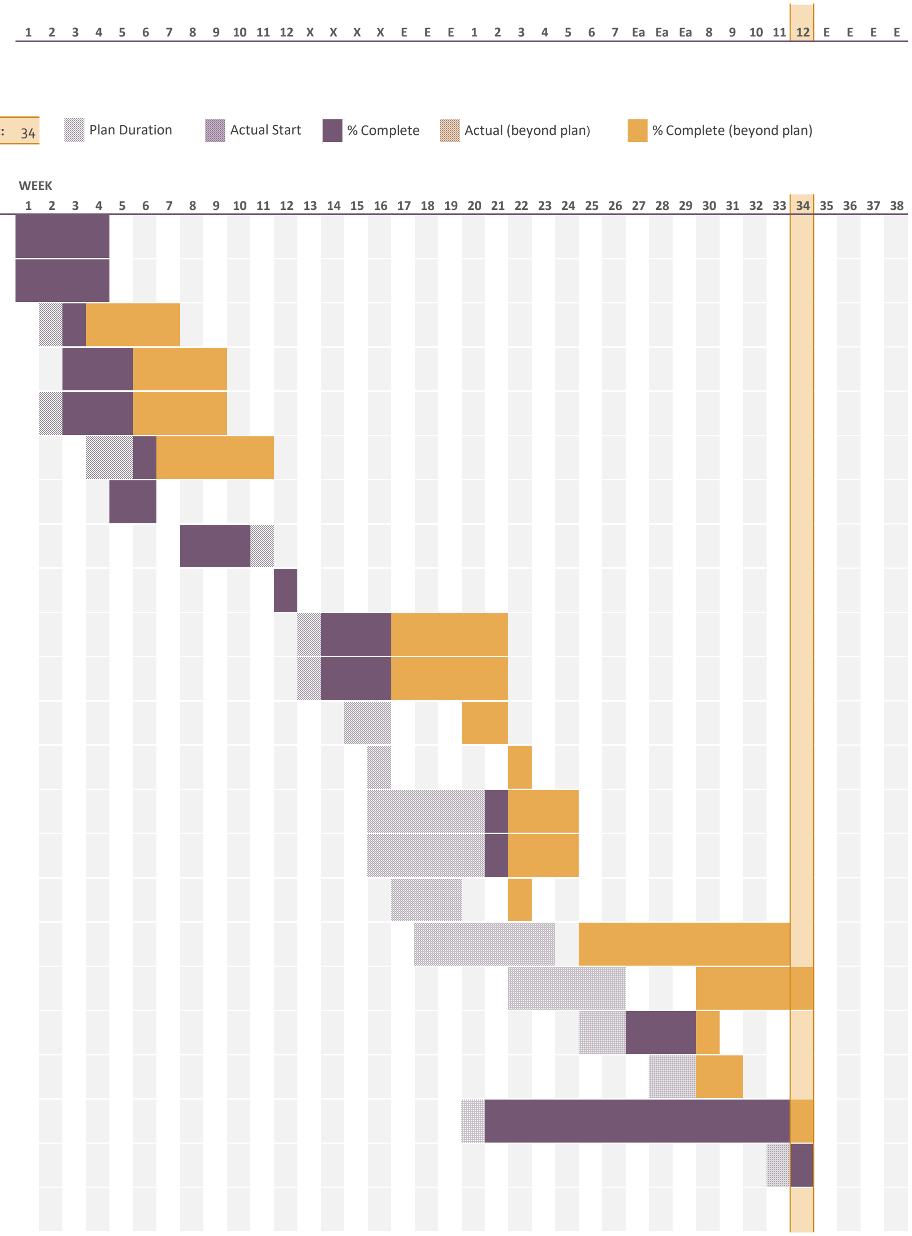
Figure 21: Diagram showing work breakdown structure for the project

Project Planner

Select a period to highlight at right. A legend describing the charting follows.

Period Highlight: 34

ACTIVITY	PLAN START	PLAN DURATION	ACTUAL START	ACTUAL DURATION	PERCENT COMPLETE
Aims and Objectives	1	4	1	4	100%
Project Planner	1	4	1	4	100%
Find relevant articles for literature review	2	2	3	5	100%
Extract Relevant information	3	3	3	7	100%
Literature Review	2	4	3	7	100%
Assess and Process Data	4	3	6	6	100%
Progress Form	5	2	5	2	100%
Interim Report	8	4	8	3	100%
Interview with Second Reader	12	1	12	1	100%
Descriptive Statistics Development for Psychological Distress	13	4	14	8	100%
Descriptive Statistics Development HFSS	13	4	14	8	100%
Generate Agent Population Including Psychological Distress	15	2	20	2	100%
Validate imputation and input data	16	1	22	1	100%
Familiarise myself with Repast4Py	16	6	21	4	100%
Understanding original model	16	6	21	4	100%
Redesign Existing Software Architecture	17	3	22	1	100%
Implement Software Architecture in Python (Repast4Py)	18	6	25	9	100%
Intervention and Analysis	22	5	30	5	100%
Update literature review	25	5	27	4	100%
Generate Results and calibrate model parameters	28	2	30	2	100%
Dissertation Final Report	20	14	21	14	100%
Oral Presentation	33	2	34	1	100%
Activity x	0	0	0	0	0%



B Appendix - Model inputs and parameterisation

Variable in NDNS	In 2008?	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Age of respondent 16+, grouped into 4 cats	Has age 16+ in 10 year bands and in 20 year bands	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)
Age of respondent 16+, grouped into 5 cats	Has age 16+ in 10 year bands and in 20 year bands	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)
Age of respondent 16+, grouped into 3 cats	Has age 16+ in 10 year bands and in 20 year bands	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)	Has age 16+ in 10 year bands and in 20 year bands (ag16g10)
Sex	Sex measured in a very similar way (1=man, 2=woman)	Sex measured in a very similar way (1=man, 2=woman) 'sex'	same as 2009	Changes variable name to 'Sex' (captial s)	Changes variable name to 'Sex' (captial s)	Changes variable name to 'Sex' (captial s)	Changes variable name to 'Sex' (captial s)	Changes variable name to 'Sex' (captial s)	Changes variable name to 'Sex' (captial s)	Changes variable name to 'Sex' (captial s)	Changes variable name to 'Sex' (captial s)	Changes variable name to 'Sex' (captial s)
Ethnic Group, 5 Groups, White, Mixed, Black, Asian, other	origin' has more categories than origin2 from later years	origin' has more categories than origin2 from later years	origin' has more categories than origin2 from later years	Origin' has more categories than origin2 from later years	Origin' has more categories than origin2 from later years	Origin' has more categories than origin2 from later years	origin2', same cats as 2015	origin2', same cats as 2016	Origin2' grouped ethnic categories White, Black, Asian, Mixed, Other	Origin2' grouped ethnic categories White, Black, Asian, Mixed, Other	origin2' grouped ethnic categories White, Black, Asian, Mixed, Other	origin2' grouped ethnic categories White, Black, Asian, Mixed, Other
Qual7	Very similar scale called topqual2	Very similar scale called topqual2	Very similar scale called topqual2	Very similar scale called topqual2	Very similar scale called topqual2	Very similar scale called topqual2	Very similar scale called topqual2	Very similar scale called topqual2	Very similar scale called topqual2	Very similar scale called topqual2	Very similar scale called topqual2	Very similar scale called topqual2
Economic Status	Similar called econact	Similar called econact	Similar called econact	Similar called econact	Similar called econact	Similar called econact	Similar called econact	econact' not here, most similar variable seems to be 'Activb2'	econact' not here, most similar variable seems to be 'Activb2'	econact' not here, most similar variable seems to be 'Activb2'	econact' not here, most similar variable seems to be 'Activb2'	econact' not here, most similar variable seems to be 'Activb2'
NTypHours												
NS-SEC8	Yes	yes	yes	Yes	yes	yes	Yes	yes	yes	Yes	yes	yes
Total Household income in last 12 months	Yes, but more categories	yes, but more categories. 'totinc'	yes, but more categories. 'totinc'	yes, but more categories. 'totinc'	yes, but more categories. 'totinc'	yes, but more categories. 'totinc'	yes, but more categories. 'totinc'	totinc replaced by 'JNTINC2'	totinc replaced by 'JNTINC2'	totinc replaced by 'JNTINC2'	totinc replaced by 'JNTINC2'	totinc replaced by 'JNTINC2'
GHQ-12 score	ghq12scr'	ghq12scr'	ghq12scr'	no ghq variables?	ghq12scr'	no ghq variables?	ghq12scr'	no ghq variables?	ghq12scr'	no ghq variables?	ghq12scr'	no ghq variables?
Lost much sleep	ghqsleep'	ghqsleep'	ghqsleep'	no ghq variables?	ghq12scr'	no ghq variables?	ghq12scr'	no ghq variables?	ghq12scr'	no ghq variables?	ghq12scr'	no ghq variables?
Under stress/strain	ghqstrain'	ghqstrain'	ghqstrain'	no ghq variables?	ghq12scr'	no ghq variables?	ghq12scr'	no ghq variables?	ghq12scr'	no ghq variables?	ghq12scr'	no ghq variables?
Could not overcome dificulties	ghqover'	ghqover'	ghqover'	no ghq variables?	ghq12scr'	no ghq variables?	ghq12scr'	no ghq variables?	ghq12scr'	no ghq variables?	ghq12scr'	no ghq variables?
WEMWBS score	none	none	wemwbs'	wemwbs'	wemwbs'	wemwbs'	wemwbs'	wemwbs'	wemwbs'	none	none	wemwbs'

Figure 22: Table to document names of variables in HSE over the years compared to NDNS

Normlalising PD and BMI

Due to PD having a fixed maximum on the GHQ-12 [16]questionnaire of 12, it was easy to normalise these and to regain the original range of values on the output from the model. The .csv file produced in R of the base population, with the 1000 agents in the model, was adjusted at the end of its construction within R, by simply dividing all values in the PD column by 12. The output of the model in python was then multiplied by 12.

For BMI, it was a bit more tricky. The base population produced in R has different agents each time, these agents have varying BMIs, with the maximum BMI changing each time. Therefore, it was necessary to make a new python function which would find the maximum BMI in the population, then pass this value to another function which would divide all BMI values by this maximum, normlalising all the agents BMIs. As with the PD, all BMI values would be multiplied by this maximum so the model ouput is easier to interpret.

Listing 1: Function made to normalise BMI within the model

```
def normaliseBMIColumn(self): # this is to normalise BMI column and use max value so that we can change output file accordingly
    maxValue = None
    normalized_column = []

    headerLine = self.infoTable[0]

    self.indexBmival = self.findIndexInHeader("bmival", headerLine)

    # Step 1: Find the max value in the column
    for row in self.infoTable[1:]:  # Skip header row
        value = float(row[self.indexBmival])
        if maxValue is None or value > maxValue:
            maxValue = value

    # Step 2: Normalize all values in the "bmival" column
    for row in self.infoTable[1:]:  # Skip header row
        value = float(row[self.indexBmival])
        # print(value)
        normalizedValue = value / maxValue
        row[self.indexBmival] = normalizedValue  # Update the original value with the normalized value
```

C Appendix - Implement the model in python

Listing 2: getAttitude function

```
def getAttitude(self, schema):

    schemaId = schema
    # schemaId = EatingSchema
    # desire = 0.0
    if self.desires:
        def softmaxSchemaChoice(x):
            expX = np.exp(x - np.max(x))
            return expX / np.sum(expX, axis = 0)

        def logitFunctionSchemaChoice(params, schemaId, PD, emotional, BMI):
            alpha, betaEmotional, betaPD, betaBMI = params
            HealthyPDmultiplier = 2
            UnhealthyPDmultiplier = 10
            # [VERY_LOW, LOW, MED, HIGH, VERY_HIGH]
            if (self.PD < 4/12): # if low/good PD, higher chance of choosing better diet
                lowerSchema = HealthyPDmultiplier*(alpha + (betaEmotional * emotional) + (betaPD * PD) + (betaBMI * BMI))
                return softmaxSchemaChoice([lowerSchema, lowerSchema, 0, 0, 0])
            elif (self.PD >= 4/12):
                higherSchema = UnhealthyPDmultiplier*(alpha + (betaEmotional * emotional) + (betaPD * PD) + (betaBMI * BMI))
                return softmaxSchemaChoice([0, 0, 0, higherSchema, higherSchema])

        alpha = Globals.ALPHA_ATTITUDE
        betaBMI = Globals.BETA_BMIVALEMO
        betaPD = Globals.BETA_PD
        betaEmotional = Globals.BETA_EMOTIONAL
        params = [alpha, betaEmotional, betaPD, betaBMI]

        if (self.PD >= 4/12): # this is needed to make sure strength of PD in the above equation is consistent
            PD = ((self.PD*12) - 4)/8 # if it wasn't here, importance/weight of high PD would be greater than low PD
        elif (self.PD < 4/12):
            PD = 1 - (self.PD*12)/4

        newDesires = np.array(logitFunctionSchemaChoice(params, schemaId, PD, self.emotional, self.bmival))

    # Rescale the rest of the schemas to sum to 1
    self.desires = self.desires + newDesires
```

```
    desireSum = sum(self.desires)

    if desireSum != 0:
        self.desires = [d/desireSum for d in self.desires]
        desire = self.safeOdds(self.desires[schemaId], self.desires[0])
    else:
        print("WARNING: Agent has no desires")
    if desire == 0:
        desire = 1e-30 # small number for log

    return self.safeLog(desire)
```

Listing 3: updateBMIval function

```
def updateBMIval(self, eatingPlan):
    schemaId = eatingPlan.schema

    def softmax(x):
        expX = np.exp(x - np.max(x))
        return expX / np.sum(expX, axis = 0)

    def logit_function(params, schemaId, BMI): # now goes too high
        # print(schemaId)
        alpha, beta = params

        if schemaId == EatingSchema.VERY_HIGH or schemaId == EatingSchema.HIGH:
            increaseBMI = alpha + beta * schemaId
            return softmax([increaseBMI, 0, 0])
        elif schemaId == EatingSchema.VERY_LOW or schemaId == EatingSchema.LOW:
            # print([schemaId, "is very low"])
            if ((BMI < 25/maxValue)): # if healthy weight or underweight, it is completely random whether BMI changes and which way
                decreaseBMI = alpha + beta * schemaId
                return softmax([0, 0, decreaseBMI])
            else:
                decreaseBMI = alpha + beta * schemaId
                return softmax([0, decreaseBMI, 0])
                # return softmax([0, 0, 0])
        else:
            sameBMI = alpha + beta * schemaId # if med or low, will most likely stay the same weight
            return softmax([0, 0, sameBMI])

    # alpha = 0.1
    alpha = Globals.ALPHA_BMI
    beta = Globals.BETA_HFSSBMIVAL
    params = [alpha, beta]

    probPDchange = np.array(logit_function(params, schemaId, self.bmival)) # will output something like [0.2, 0.7, 0.1] for BMI = 22
```

```
where its [gain, loss, same]

    randGainOrLoss = repast4py.random.default_rng.random() # once in a bmi category, this decides whether weight is gained or lost

    if self.age>60:
        if (self.bmival < 25/maxValue):
            randGainAmount = repast4py.random.default_rng.random() * 0.01 * 0.5 # For healthy or underweight, 2% change at a time is fine
            if (randGainOrLoss < probPDchange[0]):
                self.bmival = self.bmival + self.bmival * randGainAmount # Weight gain
            elif (randGainOrLoss > probPDchange[0] and (randGainOrLoss < (probPDchange[1] + probPDchange[0]))):
                self.bmival = self.bmival - self.bmival * randGainAmount # weight loss
        else:
            randGainAmount = repast4py.random.default_rng.random() * 0.02 * 0.5 # anything larger than 5% weight gain is unrealistic over time frame this is being applied on
            if (randGainOrLoss < probPDchange[0]):
                self.bmival = self.bmival + self.bmival * randGainAmount # Weight gain
            elif (randGainOrLoss > probPDchange[0] and (randGainOrLoss < (probPDchange[1] + probPDchange[0]))):
                self.bmival = self.bmival - self.bmival * randGainAmount # weight loss
    elif self.age<=60:
        if (self.bmival < 25/maxValue):
            randGainAmount = repast4py.random.default_rng.random() * 0.01 # For healthy or underweight, 2% change at a time is fine
            if (randGainOrLoss < probPDchange[0]):
                self.bmival = self.bmival + self.bmival * randGainAmount # Weight gain
            elif (randGainOrLoss > probPDchange[0] and (randGainOrLoss < (probPDchange[1] + probPDchange[0]))):
                self.bmival = self.bmival - self.bmival * randGainAmount # weight loss
        else:
            randGainAmount = repast4py.random.default_rng.random() * 0.02  # anything larger than 5% weight gain is unrealistic over time frame this is being applied on
            if (randGainOrLoss < probPDchange[0]):
                self.bmival = self.bmival + self.bmival * randGainAmount # Weight gain
            elif (randGainOrLoss > probPDchange[0] and (randGainOrLoss < (probPDchange[1] + probPDchange[0]))):
                self.bmival = self.bmival - self.bmival * randGainAmount # weight loss

```

```
    if self.bmival < 13/maxValue or self.bmival > maxValue: # BMI of 13
        is considered fatal so have made a floor here
        self.bmival == 13/maxValue

    return self.bmival
```

Listing 4: updatePDval function

```
def updatePDval(self):

    # Making a cumulative logit equation to see whether PD will [
        increase, decrease, stay the same]
    def softmax(x):
        expX = np.exp(x - np.max(x))
        return expX / np.sum(expX, axis = 0)

    def logit_function(params, BMI):
        alpha, beta = params
        if BMI > 25/maxValue: # if overweight, you have a higher chance
            of worse PD
            increasePD = alpha + beta * BMI
            return softmax([increasePD, 0, 0])
        elif (BMI < 25/maxValue) and (BMI > 18.5/maxValue): # if healthy
            weight, higher chance of improving PD
            decreasePD = alpha + beta * BMI
            return softmax([0, decreasePD, 0])
        else:
            samePD = alpha + beta * BMI
            return softmax([0, 0, samePD])

    # alpha = 1.1
    alpha = Globals.ALPHA_PD # threshold
    beta = Globals.BETA_BMIVALPD # strength of relationship between BMI
        and PD
    params = [alpha, beta]

    probPDchange = np.array(logit_function(params, self.bmival)) # will
        output something like [0.2, 0.7, 0.1] for BMI = 22, # [gain, loss
        , stay]
    randGainOrLoss = repast4py.random.default_rng.random() # random
        number to see which path will be taken
    randGainAmount = repast4py.random.default_rng.random() * 0.1 #
        anything larger than 10% PD change is unrealistic over time frame
         this is being applied on

    if (self.sex == "f"): # BMI value affects women's mental health
        differently compared to men
        if ((randGainOrLoss < probPDchange[0]) and (self.bmival >= 30/
            maxValue)): # more extreme impact as a result of being obese
            compared to overweight
            self.PD = self.PD + self.PD * randGainAmount * 1.2 # PD
                increased by 12% and below is 3%, this represents theory
        elif (randGainOrLoss < probPDchange[0] and ((self.bmival >= 25/
```

```
            maxValue) and (self.bmival < 30/maxValue))):
            self.PD = self.PD + self.PD * randGainAmount * 0.3 # PD increased
        elif ((randGainOrLoss > probPDchange[0]) and (randGainOrLoss < probPDchange[0] + probPDchange[1])):
            self.PD = self.PD - self.PD * randGainAmount * 1.2
    elif (self.sex == "m"):
        if ((randGainOrLoss < probPDchange[0]) and (self.bmival >= 30/maxValue)):
            self.PD = self.PD + self.PD * randGainAmount * 0.6 # PD increased half as much as females
        elif (randGainOrLoss < probPDchange[0] and (self.bmival >= 25/maxValue) and (self.bmival < 30/maxValue)):
            self.PD = self.PD + self.PD * randGainAmount * 0.15 # PD increased
        elif ((randGainOrLoss > probPDchange[0]) and (randGainOrLoss < probPDchange[0] + probPDchange[1])):
            self.PD = self.PD - self.PD * randGainAmount * 0.6

    # INTERVENTION - TURN OFF FOR CALIBRATION
    def intervention():
        # [gain, loss, stay]
        probAppointmentEffective = [0.75, 0.05, 0.2] # [decrease, increase, same] # These are the probabilities based off research on how effective treatment is on average for people

        randAppointment = repast4py.random.default_rng.random() # chance that this patient gets seen to - based off level of improvement in services atm
        randEffectiveOrNot = repast4py.random.default_rng.random() # choosing whether the treatment will be effective or not
        randEffectiveness = repast4py.random.default_rng.random() # The size of the impact that treatment has on people

        if randAppointment < Globals.BETA_INTERVENTIONSCALE: # the global value here represents the level at which the UK mental health services improve
            if (randEffectiveOrNot < probAppointmentEffective[0]):
                # self.PD = self.PD - self.PD * Globals.BETA_INTERVENTIONEFFECT * randEffectiveness
                self.PD = self.PD - self.PD * randEffectiveness
            elif (randEffectiveOrNot > probAppointmentEffective[0] + probAppointmentEffective[1]):
                # self.PD = self.PD + self.PD * Globals.BETA_INTERVENTIONEFFECT * randEffectiveness # This line was changed because there wasn't time to include this
                self.PD = self.PD + self.PD * randEffectiveness

        return self.PD

    # if intervention is called:
```

```
    # self.PD = intervention() # turn this off for calibration

    if self.PD >= 1: # GHQ-12 score cannot go above 12 and it has been normalised so 1 represents 12.
        self.PD = 1

    return self.PD
```

D Appendix - Additional results

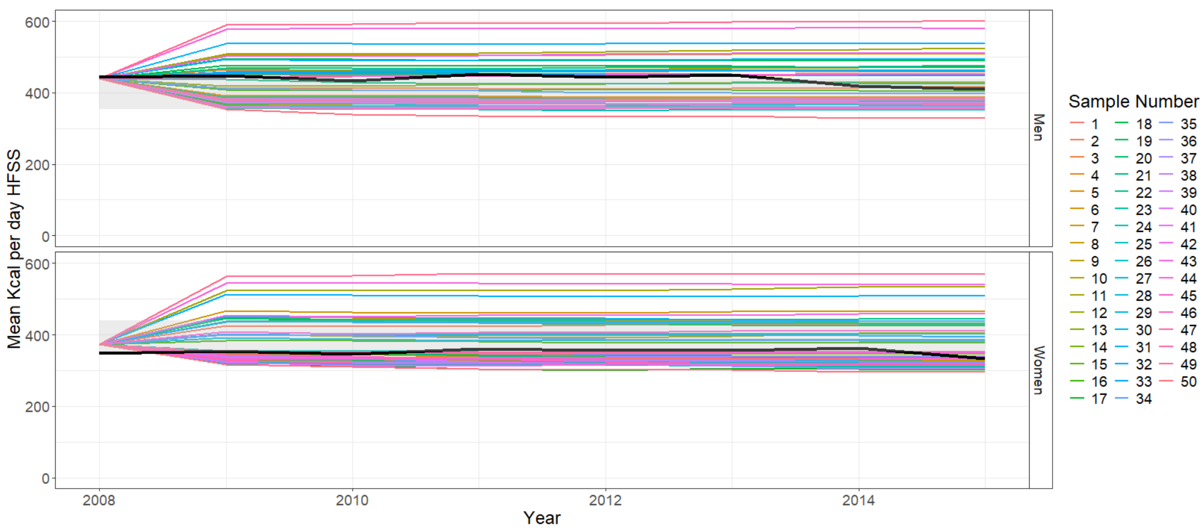


Figure 23: Plot showing all models involved in calibration process. The black line is the real-life target data

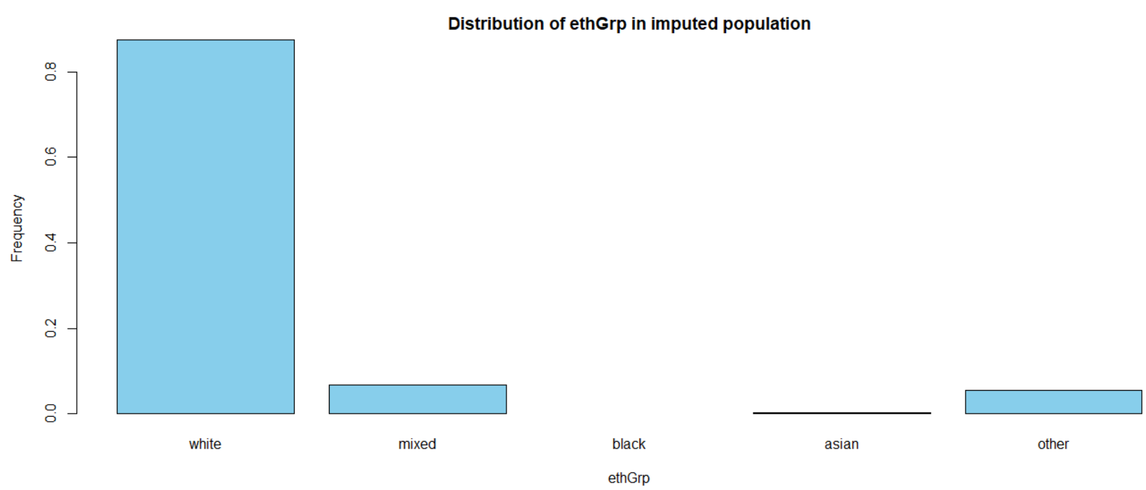


Figure 24: Density plot showing distribution of the input population's ethnicity