Annual Review

Wednesday, 21 November 2018

Matthew Lee

Supervisors: Professor Nicholas Timpson, Dr Kaitlin Wade, Dr Laura Corbin

What lies behind the causal impact of body mass index (BMI) level and change on human health? Added value from complementary study design and deep metabolomic phenotyping

Date of commencement: 01/10/2017

Expected submission date: 03/10/2021

[1. Introduction and direction 3](#_Toc530428899)

[2. Work so far 3](#_Toc530428900)

[2.1.1. GWAS of glycosuria in ALSPAC (ALSPAC B number – B3188) 3](#_Toc530428901)

[2.1.2. Chapter 1 4](#_Toc530428902)

[2.1.3. Systematic review 5](#_Toc530428903)

[2.1.4. Measures of adiposity and metabolites – MR analysis 6](#_Toc530428904)

[2.1.5. MR visualisation tool 10](#_Toc530428905)

[2.1.6. Collaboration with Tim Frayling, Exeter 10](#_Toc530428906)

[2.1.7. Flemish Gut Flora Project observational analysis 12](#_Toc530428907)

[2.1.8. Three measures of adiposity - metabolites to disease 13](#_Toc530428908)

[3. Training so far 13](#_Toc530428909)

[3.1.1. Training courses 13](#_Toc530428910)

[3.1.2. Conferences, symposia, presentations 14](#_Toc530428911)

[3.1.3. Public engagement 14](#_Toc530428912)

[3.1.4. GW4 training 16](#_Toc530428913)

[3.1.5. Grants 16](#_Toc530428914)

[4. Next 12 months 17](#_Toc530428915)

[4.1. Immediate work 17](#_Toc530428916)

[4.2. PhD 18](#_Toc530428917)

[4.3. Training 18](#_Toc530428918)

[5. Thesis outline/plan 19](#_Toc530428919)

[5.1. Chapter 1 19](#_Toc530428920)

[5.2. Chapter 2 19](#_Toc530428921)

[5.3. Chapter 3 20](#_Toc530428922)

[5.4. Chapter 4 20](#_Toc530428923)

[5.5. Chapter 5 20](#_Toc530428924)

[5.6. Chapter 6 21](#_Toc530428925)

# Introduction and direction

I am on a 3.5-year GW4 BioMed MRC DTP PhD. This review aims to reflect my first years’ work and training progress so far. I will also outline the next 12 months and provide a thesis plan.

My PhD is looking at the impact of BMI on health. The focus is on how BMI affects metabolites and the metabolic profile of different populations, and how changes to these profiles impact health.

My overarching interest is whether metabolic profiles differ based on different measures of adiposity and how these mechanisms may be working. It is important to investigate whether the metabolic profile can provide insights into pathways that lead to disease. Primarily my work is focused on using summary statistics for Mendelian randomization (MR) analysis, but I am particularly interested in how cohort studies can be used to investigate life-course effects of adiposity and metabolic profiles. I am keen to explore the longitudinal effects of changing metabolic profiles.

# Work so far

### GWAS of glycosuria in ALSPAC (ALSPAC B number – B3188)

I have conducted a genome-wide association study (GWAS) of glycosuria in pregnancy in the Avon Longitudinal Study of Parents and Children (ALSPAC). This has been written up as a manuscript, to be submitted to Wellcome Open Research. I followed the work of George McMahon and collaborated with David Hughes and Laura Corbin to develop scripts and an analysis plan to complete the project. The draft manuscript is attached.

This project has been an excellent training opportunity, and although at times challenging, it has both increased my understanding and enjoyment of genetic epidemiology. The main benefit from the project is that it has enabled me to put into practice much of what I have learned over the course of my first year, namely the Linear and Logisitic Regression short course I attended.

Whilst conducting this project, I have handled, managed, cleaned, and prepared genetic and observational data, and learnt to store it and use scripts in line with best practice that will enable replication of my work. I have conducted a GWAS and performed follow-up to ensure associations are reliable; I have performed a meta-analysis of GWASs and prepared data for publication; I have performed additional analysis to examine the association of the GWAS signal and better understand its biological relevance to the phenotype, involving linear and logistic regression as well as use of software platforms such as LD Hub to investigate genetic correlations. Finally, working through feedback and comments from supervisors has developed my writing skills and ability to draw on multiple pieces of information to form a conclusion.

Although this project will not sit directly within my thesis, I have developed my understanding of genetics, epidemiology, and data management, which will be vital for my PhD and other projects going forward.

### Chapter 1

I have begun work on the introduction to my thesis, in the form of a systematic review (discussed in 2.1.3) and an introduction to MR.

Within the introduction I think it is important to provide a clear and definitive explanation of MR and the different methods and applications I will/have used throughout the PhD. I have been working on a document with information on new methods and applications of MR which can form the foundations of my introduction – the document is very much a work in progress, and I have attached it. From this, I have developed a better understanding of MR, its assumptions, and applications. However, I still find it a challenge to know what method is most applicable to different situations and how those methods work mathematically – I hope to build on this over course of the next year.

### Systematic review

I am working with Luke McGuinness, a research associate with Julian Higgins and Jonathan Sterne, to conduct a systematic review. The working title is: A systematic review of MR studies using adiposity as an exposure. The review question is: What has the application of MR informed us about the causal relevance of adiposity?

The review will sit within the first chapter of my thesis, and provide the foundation of my introduction, as well as being submitted for publication. The review has been pre-registered with the International Prospective Registration of Systematic Reviews (PROSPERO)1 and follows the recommended Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines2,3. Pre-registration document is attached.

There is currently no risk-of-bias tool for MR studies. This systematic review will inform the development for a risk-of-bias tool, and will be considered a formal systematic review – a number of systematic reviews on applications of MR have been published4-6 and pre-registered on PROSPERO7-15. The reviews mostly miss assessing the MR studies themselves and instead focus mainly on the results.

In the last week a systematic review of MR with obesity as an exposure and a number of cardiovascular related outcomes has been published6. This systematic review, though closely aligned to ours, is not as comprehensive and has limitations16. The study includes 7 articles (5 included in the meta-analysis) focussing on a single measure of adiposity (obesity) and a small number of outcomes providing a narrow overview of obesity and traits such as type 2 diabetes and coronary artery disease. Additionally, it does not include an assessment of many of the aspects of an MR study that one needs to consider, including MR methods. Our systematic review, incorporating many measures of adiposity and all outcomes and multiple aspects of MR studies, is therefore still warranted and will provide a comprehensive picture of the associations between adiposity and outcomes in an MR context as well as an assessment of the study’s methodology.

Through this project I have: established a collaboration, conceived of and designed with collaborators the study, completed a literature search strategy, developed an analysis plan and methodology, and written and published pre-registration documentation following guidelines. This project is currently on hold while my collaborator is away from Bristol. In January we will perform our search and begin abstract screening with an aim of completing this by the end of January. We aim to complete full paper screening by the end of March, analysis by the end of May, and write up by the end of June, with submission aimed for July.

### Measures of adiposity and metabolites – MR analysis

I have begun performing MR analysis of different measures of adiposity and metabolites. My aim is to build a comprehensive map of the metabolic profile for multiple adiposity phenotypes and profiling types (nuclear magnetic resonance (NMR), Metabolon mass spectrometry (MS)). We will investigate links with disease, which will form Chapter 3 of my thesis. Figures 1 and 2 show results for BMI, waist hip ratio, and body fat % for Shin (MS)17 and Kettunen (NMR)18 metabolites respectively. Both figures are ordered by the effect size of BMI on metabolites from highest to lowest to enable easier comparison of adiposity measures.



Figure 1. MR analysis of BMI (1st track; blue), waist hip ratio (2nd track; green), and body fat % (3rd track; orange) to Shin17 metabolites – beta for IVW (multiplicative random effects) is shown by coloured points with red points representing a p value < 0.05. Metabolites are grouped into classes (1. Amino Acid; 2. Carbohydrates; 3. Cofactors and Vitamins; 4. Energy; 5. Fatty Acid; 6. Lipid; 7. Nucleotide; 8. Peptide; 9. Unknown Metabolite; 10. Xenobiotics) and ordered within those classes by the effect size of BMI on metabolites from highest to lowest.



Figure 2. MR analysis of BMI (1st track; blue), waist hip ratio (2nd track; green), and body fat % (3rd track; orange) to Kettunen18 metabolites – beta for IVW (multiplicative random effects) is shown by coloured points with red points representing a p value < 0.05. Metabolites are grouped into classes (1. Amino Acid; 2. Carbohydrates; 3. Energy; 4. Fatty Acid; 5. Keto Acid; 6. Lipid; 7. Metabolite Salt; 8. Metabolites ratio; 9. Protein) and ordered within those classes by the effect size of BMI on metabolites from highest to lowest.

My next step is to produce a data library of adiposity exposures, and for different populations (EU, all ancestry, male, female, different ages etc.). Once this data library is established, performing the MR analysis will be relatively straightforward as I have already developed the pipeline using BMI, waist hip ratio, and body fat %. MR Base has many adiposity measures available within its catalogue of GWASs and I have begun using this as a starting point for the data library. I will also consult the GWAS catalogue and collaborators for measures not present in MR Base – for example favourable adiposity instruments have been acquired from Tim Frayling. GWAS data on scan measurements may be a challenge with few studies currently available.

I will also use MR Base for many of the outcomes when looking at the second stage of my PhD. My preliminary work, looking solely at BMI as an exposure has found a number lack the necessary instruments for MR analysis. MR Base also lacks many disease outcomes, specifically cancer outcomes, and thus other sources may be required. This is a challenge for cancer data where studies are reluctant to share.

A comprehensive list of available metabolic GWASs is already publicly available19 (<http://www.metabolomix.com/list-of-all-published-gwas-with-metabolomics/>). I will use this as the template to build my data library for metabolite outcomes – an issue here is that a number of the studies have been included in larger meta-analysis and identifying which of these have and haven’t will take time. In addition to this I take part in the CHARGE (Cohorts for Heart Aging Research in Genomic Epidemiology) Metabolomics Working Group calls. A paper is currently being circulated in the group on a meta-GWAS of metabolomic platforms (Dianna/Carolina/Debbie are providing ALSPAC data) which will be helpful in identifying more metabolite sources.

A major challenge with metabolite, and omics data in general, is how to present, perform sensitivity analysis, and select exposure-outcome associations of interest. In terms of presenting results, we are developing a visualisation tool (discussed in 2.1.5) which will efficiently produce meaningful visualisations of the exposure-outcome associations.

Another challenge in metabolomics is in selecting exposure-outcome associations since metabolites are highly correlated. This is a particular issue when looking at groups of metabolites, or all metabolites as a metabolic profile, and how one deals with this is important. I have discussed this with Kate Tilling who has given me some direction on different ways to approach this.

### MR visualisation tool

One of the difficulties with omics data is the size. The two large metabolite studies from Shin17 and Kettunen18 provide GWAS summary level data on 452 and 123 metabolites respectively. When Claudia Langenberg and colleagues publish their metabolite GWAS the number of individual metabolites is likely to rise further. This size presents a unique challenge in how one presents analysis and uses visualisations to go beyond the initial stages of analysis.

As a group we have discussed the most appropriate way to visualise these data and have begun using Circos plots20,21 (Figure 2 and 3). Working with Osama Mahmoud I have begun to develop a platform that automates the creation of these plots, much like MR Base automates MR analysis. The platform is currently in development stage and can be accessed here - <http://bristol-medical-stat.bristol.ac.uk:3838/MR-Vis/>. One extension of visualising our data, particularly for two-step MR analysis and investigating SNP effects, is the Hive Plot22 which would enable visualisation of mediation analysis and networks from exposure to multiple outcomes – this is a long term aim for our visualisation platform.

In addition to providing a tool to visualise data we aim to introduce components of sensitivity analysis for MR analysis and to enable visualisation of this within plots. The platform will either form a standalone publication or be an addition to one of the papers that comes from my work, most likely that investigating three measures of adiposity to metabolites and to disease.

One of the main components of the platform is the ability to visualise a single step MR analysis and to extract instruments for outcomes of interest for two-step MR analysis. The work I have discussed with Kate Tilling in terms of how one groups metabolites together will be particularly relevant in informing how to provide instruments to individuals.

### Collaboration with Tim Frayling, Exeter

I am collaborating with Professor Tim Fraying and his group in Exeter on their recent work identifying ‘favourable adiposity’ SNPs. We will investigate the effect these SNPs have, in the context of a genetic risk score (GRS), on the metabolite profile compared with BMI as instrumented by a GRS derived from the most recent GIANT BMI GWAS23. For this work we are also exploring whether to apply for data from the INTERVAL study if power is a limiting factor within ALSPAC.

**My ALSPAC proposal for this collaboration**

Yaghootkar et al (2016)24 have identified SNPs where the allele associated with body fat percentage is associated with a lower risk of type 2 diabetes and favourable biomarker profile; fat allele goes with higher HDLC and lower triglycerides and lower insulin. Most of these SNPs are associated with lower waist-hip ratio in women, but not men, and are associated with similar effects on body fat percentage. A genetic risk score for these SNPs is associated with more subcutaneous and less liver fat – the effect on liver fat is potentially stronger in women. Commonly, individuals with such a profile are described as having ‘favourable adiposity’ or being ‘metabolically healthy obese’. Currently no-one has looked at these variants’ effects on the metabolite profile, we intend to investigate the effects of these SNPs on the metabolite profile of individuals within ALSPAC.

We will perform MR for our GRS and a standard BMI GRS (from the most recent GIANT GWAS) for association with metabolites18 across ALSPAC individuals. We will undertake hypothesis 1 and 2 within all ALSPAC individuals with available metabolite data. We will follow this up by looking at whether the effect is present in ALSPAC child metabolite data pre- and post-puberty.

Hypothesis

1. The metabolite profile from our GRS will be opposite to that of a standard BMI GRS, and we expect it to be in line with a favourable lipid fraction profile.
2. Favourable lipid fraction profile will be similar in men and women despite differences in body shape effects – alleles make men and women have more subcutaneous fat and less liver and visceral fat. In men this fat goes all over the body, so men end up fatter around the waist but metabolically healthier. In women this fat goes to the lower body

I have worked with Kaitlin to create a GRS for BMI and favourable adiposity. I am now working through the MR course Kaitlin delivered in Baltimore this year to understand how to perform MR in a one-sample setting. I have also spoken with Sean Harrison on how to perform sensitivity analysis in the one-sample setting and the analysis plan he uses for MR with GRSs, and after speaking with Kaitlin have a better idea of how to achieve this with ALSPAC genetic data.

My aim is to complete the first stage of analysis, performing the MR for both GRSs, before December, with a goal of having results and an idea for a paper to discuss with Tim in January.

### Flemish Gut Flora Project observational analysis

The Flemish Gut Flora Project25 (FGFP) is a large-scale cross-sectional faecal microbiome study with extensive phenotyping and available genetic data. I will look at the observational associations of BMI and other traits and metabolites to define the FGFP dataset and examine relationships with BMI and metabolites that can then be explored in further studies. Additionally I am going to use GRSs of BMI as prediction tools for metabolites, specifically I will use different genome wide thresholds to extract BMI SNPs for the GRSs, as well as using a GRS tool (LDpred) recently used by Sekar Kathiresan26. Once the observational and prediction work has been completed I will look to focus on using MR in the one-sample setting and investigate confounder associations with BMI and metabolites in this sample.

The observational and prediction work here, as well as use of one-sample MR to investigate confounders, will fit well with my PhD and provide a triangulation approach to investigating the association of BMI with metabolites and disease outcomes. I have so far worked with Laura and Caroline Bull to develop and test the scripts to produce cleaned data sets for release and further analysis. My next step is to look at a small number of token phenotypes and look at relationships with confounders. I will then begin looking at the dataset as a whole and investigating it using linear and logistic regression methods. At the same time, I will begin working on the prediction aspect of the project.

### Three measures of adiposity - metabolites to disease

In line with the work with Tim Frayling and the bulk of my PhD looking at measures of adiposity and associations with metabolites and disease, we are interested in how measures of adiposity associate with different metabolite profiles or groupings. Using BMI, waist hip ratio, and body fat % as a starting point there is a distinction between the three in terms of the pattern of association with metabolites (Figure 1 and 2). For this work we are particularly interested in looking at the association pattern for a scan measure.

This work is not an immediate focus and instead will develop as the rest of my PhD progresses, specifically it requires instruments from scan data for MR analysis and a better understanding from myself on how best to group and look at patterns of association across the metabolite measures available.

# Training so far

The following are activities and projects which have provided training opportunities. I have also been a part of the Epidemiology and Statistics book clubs in the MRC IEU; regularly attend the MR clinic meetings, seminars within the IEU, and the statistics, data science, and data visualisation discussion groups. The groups are excellent opportunities to discuss ideas and learn about new approaches to problems that may benefit my work - they also expose me to researchers from different disciplines and outside of the IEU.

### Training courses

I have attended the following training courses:

|  |  |  |
| --- | --- | --- |
| Name | Date | Course |
| Introduction to Statistics | 11/17 | Short course |
| Introduction to R | 22/11/17 | Short course |
| Introduction to STATA | 20/11/17 | Short course |
| Mendelian Randomization | 09/07/18 | Short course |
| Introduction to Linear and Logistic Regression | 16/04/18 | Short course |
| Advanced Epigenetic Epidemiology | 14/03/18 | Short course testing |
| Version control using Git | 02/02/18 | ACRC course |
| How to run an EWAS | 20/10/17 | Gemma Sharp |
| Introduction to Linux | 30/01/18 | ACRC course |
| Python 1: Beginning Python | 09/02/18 | ACRC course |
| Introduction to HPC | 05/02/18 | ACRC course |
| Introduction to programming | 11/06/18 | GW4 course – Bath |
| Best programming practices for open science | 25/05/18 | GW4 course – Cardiff |

### Conferences, symposia, presentations

I have attended the following conferences/symposia:

|  |  |  |
| --- | --- | --- |
| Event | Date | Activity |
| UK Biobank Conference | 21/06/18 | Attendee |
| GW4 Congress | 4-5/06/18 | Poster – MR analysis of BMI and metabolites  Talk – overview of poster |
| GW4 Population Health symposium | 23/04/18 | Organising committee |
| Population Health Science Institute Symposium | 07/06/18 | Poster – MR analysis of BMI and metabolites |
| Faculty of Health Sciences Research Day | 19/09/18 | Poster – MR analysis of BMI and metabolites |
| GW4 Orientation | 16/10/18 | Talk – overview of PhD |
| TARG lab meeting | 13/07/18 | Presentation to TARG on MR Vis |

### Public engagement

I have been involved in a number of public engagement activities since joining the University in 2016, and these have continued into the first year of my PhD. The most prominent of these was Creative Reactions Bristol 2018, a science art exhibition where we paired scientists and artists together to produce pieces of work that then go on display in a gallery. For this project I secured an MRC grant for £5,000 along with £750 from the MRC IEU.

As my second year commences, I will spend less time doing public engagement activities. Going forward there are two engagement events I am focusing on, the MRC IEU stand at Greenman Festival and Creative Reactions Bristol 2019.

|  |  |
| --- | --- |
| Activity | Summary |
| School workshops | Developed and delivered workshops for schools on genetics, MR, epidemiology, my research and PhD |
| ALSPAC | Delivered engagement activities at festivals and events for ALSPAC – communicating ALSAPC and its contribution to research, re-/engaging with participants |
| Bristol Neuroscience Festival | Organising committee – responsible for organising research stands and activities  Poster – neuroscience and epidemiology |
| BBC Bristol Phil Hammond Show | With Nancy McBride discussing importance of STEM, women/representation in STEM, science engagement, and our PhDs |
| BCFM Radio | With Charlie Hatcher discussing importance of STEM, women/representation in STEM, science engagement, and our PhDs |
| Creative Reactions Bristol 2018 | Organising committee – responsible for funding and collaborations |
| Creative Reactions Bristol 2019 | Lead organiser – committee of 10 |
| Bristol 24/7 | Interview about impact of community engagement and representation in sciences |
| Greenman Festival | Research stand with MRC IEU |

### GW4 training

There are training opportunities for GW4 PhD students to be involved in. Each year students are required to attend ‘Advanced Training Element’ relevant to their PhD and can attend as many others as they wish. The ATE for my first year was a symposium (GW4 Population Health Symposium) open to all population health and related fields which I helped to organise. I am attending the ATEs on 5th December in ‘Advanced R and Statistics’ and a ‘Genetics Day’ in Exeter on March 21st organised by Kate Tilling and Tim Frayling.

Online training for GW4 students comprises 30 weekly units spread over three modules and is intended as a discussion forum with 8-10 other students participating. Topics included: research philosophy, experimental design, writing, software packages, statistics, different types of data, wet-lab work, dry-lab work. The most useful units were the dry-lab specific ones, where the principles of epidemiology/population health and cohort studies were discussed.

### Grants

I have written and provided comments on a number of grants. This has enabled me to understand better the grant writing process, the requirements, and how to effectively meet those requirements, improving my writing.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Grant | Date | Amount | Role | Decision |
| The MRC | 2017 | £5,000 | Lead applicant - Creative Reactions Bristol 2018 | Awarded |
| The Biochemical Society | 2017 | £500 | Lead applicant - for working with youth centres and under privileged children. This project will now form part of Creative Reactions Bristol | Awarded |
| The Genetics Society | 2018 | £500 | Lead applicant - Creative Reactions Bristol 2018 | Unsuccessful |
| PGR Ventures Fund | 2018 | £500 | Lead applicant - Creative Reactions Bristol 2019 | Awaiting decision |
| Alumni Foundation | 2018 | £1,000 | Lead applicant - Creative Reactions Bristol 2019 | Awaiting decision |
| Institute of Physics | 2018 | £700 | Lead applicant - Creative Reactions Bristol 2019 | Awaiting decision |
| EPSRC Impact Acceleration Account | 2018 | £14,987 | Sole co-applicant - Creative Reactions Bristol 2019 | Awaiting decision |
| University of Bristol:  Widening Participation  Engineering  Life Sciences  Health Sciences | 2018 | £500  £500  £500  £500 | Lead applicant - Creative Reactions Bristol 2019 - shared application system so only one written grant is required | Writing |

# Next 12 months

## Immediate work

My immediate focus is submission of the GWAS on glycosuria, for which I am currently addressing comments from supervisors before sending to authors. Alongside this, I am working on the project with Tim Frayling, creating GRSs for BMI and favourable adiposity and using these in an MR context to investigate metabolite profile. I will shortly start the first stage of the work on defining the FGFP dataset that we have. In January I will continue work with Luke McGunniess on the systematic review – we aim to have a draft ready to circulate in June. Over the course of the next year I am to make consistent progress with the main focus of my PhD, establishing a library of adiposity exposures and metabolite outcome MR results.

There are a number of conferences next year which I hope to have work to present at, including: the MR conference in Bristol, ASHG, Society for Epidemiologic Research, and Metabolomics 2019.

## PhD

I will continue work on the second chapter of my thesis, looking at different measures of adiposity and metabolite profile. I hope to collate this work into a paper focussing on three different measures of adiposity and the different metabolite profiles they associate with, and diseases these profiles associate with – this may be an opportunity to incorporate the MR Vis platform into a paper.

## Training

There are a number of training opportunities I wish to attend this year. I am unsure of the first two Wellcome Genome Campus courses listed below but feel the course on Genetic analysis of population-based association studies would be of most benefit to me and my PhD going forward. I am also unsure whether attending Stephen Burgess MR course in Cambridge is worthwhile in terms of keeping up with advances in MR.

|  |  |  |
| --- | --- | --- |
| Name | Date | Course |
| Causal inference in epidemiology: concepts and methods - waiting list | 01/07/19 | Short course |
| Statistical methods for mediation analysis | 04/07/19 | Short course |
| Genetic epidemiology | 03/06/19 | Short course |
| Multiple imputation for missing data | 14/03/19 | Short course |
| Analysis of repeated measures | 30/01/19 | Short course |
| Introduction to data visualisation and web applications using R - waiting list | 10/01/19 | Short course |
| **Genetic analysis of Mendelian and Complex disorders** - provide broad overview of genetic analysis with in-depth focus on family studies, linear mixed models, risk prediction and other analysis methods | 18/07/19 | Wellcome Genome Campus |
| **Summer School in Bioinformatics** - may be beneficial for metabolite analysis (though this might be a limited focus of the course) | 25/06/19 | Wellcome Genome Campus |
| **Genetic analysis of population-based association studies** - from GWAS all the way through analysis of traits through different methods (GRS, MR, fine mapping etc.) | 09/2019 | Wellcome Genome Campus |

# Thesis outline/plan

My PhD is looking at the impact of BMI on health. The primary focus is on how BMI affects the metabolic profile of different populations (e.g. men, women, young, old), and how changes to these profiles impact health. Below is the thesis outline.

## Chapter 1

* Background on PhD
* Systematic review
* Data/repositories and MR (ALSPAC, FGFP, UK Biobank, MR-Base etc.)
* Aims/objectives

The first chapter will be an introduction to the PhD, the bulk of which will be a systematic review of ‘Mendelian randomization studies using BMI and other measures of adiposity as an exposure’. The systematic review will provide an overview of the landscape of current work examining how BMI and other measures of adiposity measures have a causal effect on health and disease. Overviews of data and repositories used will follow along with an overview of MR.

## Chapter 2

* Background – Peter Wurtz
* Measures of adiposity -> metabolites
* Build a causal metabolic profile

This chapter will aim to build a causal metabolic profile of the effects of BMI and other adiposity measures. This expands considerably on Peter Wurtz27 work from 2014 investigating BMI and metabolites in an MR framework. I will frame this chapter in light of this work and what we have gained and is still to gain in this area.

This work will focus on the effects of different measures of adiposity (BMI, WHR, BF%, etc.) on metabolites and the metabolic profile using MR. The collaboration with Tim Frayling, investigating BMI and favourable adiposity GRSs on metabolite profiles of ALSPAC individuals, will contribute to this chapter. The observational work using FGFP data will be used to introduce the chapter and provide an aspect of triangulation to support the MR analysis.

## Chapter 3

* Metabolites and the metabolic profile -> disease

Chapter 3 will focus on the second stage of investigating the landscape of BMI and association with metabolic intermediates and disease. Taking the metabolites and metabolic profiles identified in Chapter 2, Chapter 3 will look at the association with disease outcomes.

## Chapter 4

* Change in metabolic profile over the life course

Chapter 2 and 3 will look at the association of BMI -> metabolites -> disease at a single point in time. Chapter 4 will look at how this relationship changes over time. By using multiple cohort studies, we will build a picture of the relationship between BMI and disease across multiple time points building a life course picture of this association.

## Chapter 5

* Wet lab

We hope to identify a number of metabolic intermediates in the BMI -> metabolite -> disease pathway through chapters 2 and 3 that we can take into the wet lab to investigate specific mechanisms of action. This work will most likely focus on cancer as a disease outcome, and there are labs in which our group collaborates already (Ann Williams, Emma Vincent).

It may not be possible to undertake wet lab work during the PhD, however there are opportunities to extend this work beyond the PhD by applying for fellowships and this is something I have discussed with supervisors. The Elizabeth Blackwell Institute have a discipline hopping/bridging fellowship for recent PhD candidates to move disciplines and take their skill sets and apply them in a new context. This is particularly important with the increasing focus on triangulation, both within and outside of the IEU.

## Chapter 6

* Discussion
  + Introduction
  + Summary of thesis
  + Strengths and limitations
  + Future research
  + Conclusions

References

1 Booth, A. PROSPERO’s progress and activities 2012/13. *Systematic Reviews* **2**, 111, doi:10.1186/2046-4053-2-111 (2013).

2 Liberati, A. *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS medicine* **6**, e1000100-e1000100, doi:10.1371/journal.pmed.1000100 (2009).

3 Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. & Group, P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine* **6**, e1000097-e1000097, doi:10.1371/journal.pmed.1000097 (2009).

4 Kuźma, E. *et al.* Which Risk Factors Causally Influence Dementia? A Systematic Review of Mendelian Randomization Studies. *Journal of Alzheimer's disease : JAD* **64**, 181-193, doi:10.3233/JAD-180013 (2018).

5 Pingault, J.-B., Cecil, C. A. M., Murray, J., Munafò, M. R. & Viding, E. Causal Inference in Psychopathology: A Systematic Review of Mendelian Randomisation Studies Aiming to Identify Environmental Risk Factors for Psychopathology. *Psychopathology Review* **a4**, 4-25, doi:10.5127/pr.038115 (2016).

6 Riaz, H., Khan, M., Siddiqi, T. & et al. Association between obesity and cardiovascular outcomes: A systematic review and meta-analysis of mendelian randomization studies. *JAMA Network Open* **1**, e183788, doi:10.1001/jamanetworkopen.2018.3788 (2018).

7 Camille Lassale, M. T. S. F. HDL-cholesterol functionality and CVD outcomes: a systematic review and meta-analysis.

8 Elis Haan, L. S. K. W. H. S. G. T. M. M. Associations between alcohol, tobacco and caffeine consumption in pregnancy and externalising disorders in offspring: a systematic review and meta-analysis <br />.

9 Fariha Elsakloul, J. D. The genetic determinants of serum uric acid and their association with cardiovascular diseases: a systematic review.

10 Loubaba Mamluk, L. Z. T. M. A. R. Systematic review of the effects of low-moderate prenatal alcohol exposure on pregnancy and childhood outcomes.

11 Marco Solmi, B. S. A. M. F. B. S. M. M. d. T. L. B. T. T. I. T. J. F. V. R. G. C. F. M. M. F. P. F.-. Cannabinoids and cannabis: useful or harmful? An umbrella review of meta-analyses of randomized controlled trials and observational studies

12 Marco Solmi, D. P. C. B. G. O. E. D. J. I. P. F.-P. B. S. T. T. J. F. A. C. Associations of psychosis and non-communicable general medical conditions: an umbrella review of observational and mendelian randomization studies.

13 Matthew Romo, J. V. H. S. L. A. Y. J. Z. C. M. S. The conduct and reporting of Mendelian randomization studies: a systematic review.

14 Sandeep Grover, F. D. G. M. I. K. Identification of causal risk factors for neurodegenerative diseases using a mendelian randomization approach.

15 Shan Luo, J. Z. Q. Y. M. S. Mendelian randomization studies compared to randomized controlled trials: a systematic review and meta-analysis.

16 Wade, K. H. & Davey Smith, G. Adiposity and cardiometabolic outcomes: What can meta-analyses of mendelian randomization studies contribute? *JAMA Network Open* **1**, e183778, doi:10.1001/jamanetworkopen.2018.3778 (2018).

17 Shin, S.-Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nature Genetics* **46**, 543, doi:10.1038/ng.2982

<https://www.nature.com/articles/ng.2982#supplementary-information> (2014).

18 Kettunen, J. *et al.* Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nature communications* **7**, 11122-11122, doi:10.1038/ncomms11122 (2016).

19 Kastenmüller, G., Raffler, J., Gieger, C. & Suhre, K. Genetics of human metabolism: an update. *Human Molecular Genetics* **24**, R93-R101, doi:10.1093/hmg/ddv263 (2015).

20 Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome research* **19**, 1639-1645, doi:10.1101/gr.092759.109 (2009).

21 Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811-2812, doi:10.1093/bioinformatics/btu393 (2014).

22 Krzywinski, M., Birol, I., Jones, S. J. M. & Marra, M. A. Hive plots—rational approach to visualizing networks. *Briefings in Bioinformatics* **13**, 627-644, doi:10.1093/bib/bbr069 (2012).

23 Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206, doi:10.1038/nature14177 (2015).

24 Yaghootkar, H. *et al.* Genetic evidence for a link between favorable adiposity and lower risk of type 2 diabetes, hypertension and heart disease. *Diabetes* **65**, 2448-2460, doi:10.2337/db15-1671 (2016).

25 Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science* **352**, 560 (2016).

26 Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics* **50**, 1219-1224, doi:10.1038/s41588-018-0183-z (2018).

27 Wurtz, P. *et al.* Metabolic Signatures of Adiposity in Young Adults: Mendelian Randomization Analysis and Effects of Weight Change. *PLoS Medicine* **11**, doi:<http://dx.doi.org/10.1371/journal.pmed.1001765> (2014).