



1
2 What lies behind the causal impact of body mass index and change on
3 human health? Added value from complementary study design and deep
4 metabolomic phenotyping.

5 Matthew A Lee

6 MRC Integrative Epidemiology Unit, Population Health Sciences
7 Bristol Medical School
8 Faculty of Health Sciences
9 University of Bristol

10 A dissertation submitted to the University of Bristol in accordance with the
11 requirements for award of the degree of PhD Population Health Sciences in the
12 Faculty of Health Sciences

13

14 Bristol Medical School, November 2021, Word Count: 9999999

15 Abstract

16 Increased adipose tissue, adiposity, is associated with many diseases and overall
17 mortality. Studies have suggested that metabolites may play a role in these relationships.
18 Within this thesis, I aimed to identify whether metabolites play an intermediary
19 role in the relationship between adiposity and diseases using a variety of resources
20 and methods to strengthen causal inference.

21 In a comprehensive systematic review and meta-analysis, the causal effects
22 of adiposity were observed across a broad spectrum of diseases. Meta-analyses
23 highlighted an increasing effect of adiposity on many cancers, including endometrial
24 cancer, as well as metabolic and cardiovascular traits such as blood pressure. Evi-
25 dence from a narrative synthesis of over 2,000 MR analyses supported results from
26 the meta-analyses and many findings from the observational literature. There was
27 evidence from the narrative synthesis that many metabolites, predominantly lipids,
28 are associated with adiposity.

29 Within the Avon Longitudinal Study of Parents and Children (ALSPAC), evidence
30 for an effect of body mass index (BMI), waist hip ratio (WHR), and body fat percentage
31 (BF) on up-to 230 predominantly lipid based metabolites was found. There was broad
32 consistency across the measures of adiposity. In these linear models, the effect of
33 adiposity persisted after adjustment for age, sex, education, smoking status, alcohol
34 consumption, diet, and physical activity. In addition, not only did effects persist across
35 multiple time points (~9, ~18, ~24, ~50 years old), but the effect size tended to
36 increase with age.

37 Observational studies are limited by the potential for unmeasured confounding
38 and reverse causation. In two independent datasets, Mendelian randomization
39 (MR), which can overcome limitations in observational analyses, provided further
40 evidence for an effect of BMI and WHR on a large number of predominantly lipid
41 based metabolites. These effects were consistent in sensitivity analyses. However,
42 the effect of BF on metabolites in MR analyses was not clear. Broadly speaking,
43 directions of effect were opposite to those for BMI and WHR (in observational analyses

44 directions of effect were highly consistent). Additional analyses suggested this may
45 be due to the complexity in instrumenting BF. Nine metabolites, associated with BF in
46 observational analyses of adults (~50 years old) had consistent direction of effect in
47 MR analyses and were taken forward. For BMI and WHR, meta-analysis across the
48 two datasets identified 46 and 48 metabolites respectively. These metabolites had
49 consistent directions of effect in observational analysis of adults. A total of 56 unique
50 metabolites were taken forward for investigation with adiposity-associated diseases.

51 Endometrial cancer, identified in the systematic review and meta-analysis as
52 associated with adiposity, was selected to investigate the potential intermediary
53 role of metabolites. MR analyses provided evidence for an increasing effect of
54 BMI and BF on overall endometrial and endometrioid cancer. A weaker effect was
55 observed for non-endometrioid cancer. WHR was associated with an increase in non-
56 endometrioid cancer. Five of the 56 adiposity associated metabolites were associated
57 with endometrial cancer in MR analyses, two of which, triglycerides in small and very
58 small VLDL, were taken forward for multivariable MR (MVMR) analysis. In MVMR,
59 the direct effect of the exposure controlling for the intermediate is estimated. When
60 comparing MVMR and MR estimates, there was evidence for a potential intermediary
61 role of both metabolites on the effect of WHR and BF on non-endometrioid cancer.
62 Weak instruments may have biased these results however.

63 Acknowledgements

64 I will forever be grateful for the love and support that you have selflessly given me

65 This thesis is a result of the support and effort that has been given to me. It is
66 undoubtedly a better piece of work because of this. I can only hope that I am
67 able to support and give as much as has been given to me in the years to come.
68 Nic/Kaitlin/Laura, thanks for putting up with my shit and just being great fun to work
69 with! David/Vanessa/Caroline/Sam, thanks for being patient and helping me with
70 stupid questions. Emma, thanks so much for getting me out of the lab!

71 char/nance/steph/jake-ollie-etc

72 Contributions

73 All of the work in Chapter 1, 4, and 5 was performed by myself. In Chapter 2, work
74 was conducted in collaboration with Luke A McGuinness, Charlie Hatcher, Nancy
75 McBride, Thomas Battram, Si Fang, Wenxin Wan, and Kaitlin H Wade. All contributed
76 to data extraction. I also performed data extraction and, in collaboration with Charlie
77 Hatcher, checked all extracted data for all included studies. I performed all analyses in
78 Chapter 2. In Chapter 3, work was conducted in collaboration with Osama Mahmoud
79 and Luke A McGuinness. The original project was started by Osama Mahmoud under
80 the supervision of Nicholas Timpson. I worked with Osama to develop the code
81 that would form the original draft of the visualisation tool. I took over the project and
82 remade the base code for the R package and Shiny application. Luke McGuinness
83 helped with aspects of the Shiny application. The following colleagues provided
84 feedback in the development of EpiViz: Caroline Bull, Charlie Hatcher, Kurt Taylor,
85 Nancy McBride, Neil Goulding, Steph Suddell. Work from Chapter 3 is presented in
86 Bos et al. (2020)¹. In Chapter 6, all of the work was performed by myself. Vanessa
87 Tan previously performed a two-sample MR of metabolites on endometrial cancer
88 and Caroline Bull previously performed a multivariable MR analysis to understand
89 the intermediary role of metabolites in the relationship between BMI and colorectal
90 cancer, both shared their scripts with me which I used for reference in my analyses.
91 Eleanor Sanderson provided helpful guidance and feedback on the interpretation of
92 multivariable MR results in Chapter 6.

93 Work in this thesis would not have been possible without the grateful contribution
94 of all the families who took part in ALSPAC, the midwives for their help in recruiting
95 them, and the whole ALSPAC team, which includes interviewers, computer and
96 laboratory technicians, clerical workers, research scientists, volunteers, managers,
97 receptionists and nurses. Thanks to: Adam Butterowtrh for letting me use summary
98 statistics from INTERVAL (which should be publicly available anyway), Carolina for UK
99 Biobank summary statistics, and Emma Anderson for diet trajectory data in ALSPAC.

¹⁰⁰ Research Output

¹⁰¹ The following outputs are a result of this PhD and their relevance to the content of
¹⁰² this thesis are noted.

¹⁰³ **Matthew A Lee**, George McMahon, Ville Karhunen, Kaitlin H Wade, Laura J
¹⁰⁴ Corbin, David A Hughes, George Davey Smith, Debbie A Lawlor, Marjo-Riitta Jarvelin,
¹⁰⁵ Nicholas J Timpson, Common variation at 16p11.2 is associated with glycosuria in
¹⁰⁶ pregnancy: findings from a genome-wide association study in European women,
¹⁰⁷ [Human Molecular Genetics, Volume 29, Issue 12, 15 June 2020, Pages 2098–2106](#).
¹⁰⁸ *Not relevant.*

¹⁰⁹ **Matthew A Lee**, Osama Mahmoud, Luke J McGuinness, David Hughes, Kaitlin H
¹¹⁰ Wade, Laura J Corbin, Nicholas J Timpson, Epiviz: an implementation of Circos plots
¹¹¹ for epidemiologists, 2020, [github.com/mattlee821/EpiViz](#). *This R package and web*
¹¹² *application are the focus of Chapter 3*

¹¹³ David A Hughes, Kurt Taylor, Nancy McBride, **Matthew A Lee**, Dan Mason, Debo-
¹¹⁴ rah A Lawlor, Nicholas J Timpson, Laura J Corbin. metaboprep: an R package for pre-
¹¹⁵ analysis data description and processing, 2020, [github.com/MRCIEU/metaboprep](#).
¹¹⁶ *The R package discussed in this pre-print was used in Chapter 4 to process*
¹¹⁷ *metabolomics data.*

¹¹⁸ Linda M. O'Keeffe, Joshua A. Bell, Kate N. O'Neill, **Matthew A Lee**, Mark Wood-
¹¹⁹ ward, Sanne Peters, George Davey Smith, Patricia M. Kearney, Sex-specific associa-
¹²⁰ tions of adiposity with cardiometabolic traits: multi-life-stage cohort study with repeat
¹²¹ metabolomics, [medRxiv, 2020](#). *This paper is referenced within the thesis.*

¹²² Bos, M.M., Goulding, N.J., **Lee, M.A.** et al. Investigating the relationships between
¹²³ unfavourable habitual sleep and metabolomic traits: evidence from multi-cohort
¹²⁴ multivariable regression and Mendelian randomization analyses. [BMC Med 19, 69, 2021](#).
¹²⁵ *This paper is a use case for the EpiViz R package and is referenced in Chapter*
¹²⁶ *3.*

Declaration

¹²⁷

¹²⁸ I declare that the work in this dissertation was carried out in accordance with
¹²⁹ the requirements of the University's Regulations and Code of Practice for Research
¹³⁰ Degree Programmes and that it has not been submitted for any other academic award.
¹³¹ Except where indicated by specific reference in the text, the work is the candidate's
¹³² own work. Work done in collaboration with, or with the assistance of, others, is
¹³³ indicated as such. Any views expressed in the dissertation are those of the author.

¹³⁴

Signed

¹³⁵

Dated

136 Table of Contents

137	List of Tables	xv
138	List of Figures	xxi
139	Abbreviations	1
140	Chapter 1: Introduction	3
141	Chapter summary	3
142	1.1 Background	4
143	1.2 Adipose tissue	5
144	1.2.1 Energy storage	5
145	1.2.2 Insulation	6
146	1.2.3 Signalling	7
147	1.3 Adiposity: genetics	8
148	1.4 Adiposity: measures	9
149	1.5 Adiposity: morbidities	11
150	1.6 Adiposity: summary	22

151	1.7 Metabolites	23
152	1.8 Mendelian randomization	25
153	1.9 Previous work	32
154	1.10 Aims and objectives	33
155	Chapter 2: Systematic review: What has the application of Mendelian randomization informed us about the causal relevance of adiposity and health outcomes? .	35
157	Chapter Summary	35
158	2.1 Introduction	36
159	2.2 Methods	37
160	2.2.1 Data sources and search strategy	37
161	2.2.2 Study selection	37
162	2.2.3 Data extraction	38
163	2.2.4 Quality assessment	39
164	2.2.5 Meta-analysis	39
165	2.2.6 Narrative synthesis	42
166	2.3 Results	42
167	2.3.1 Literature search	42
168	2.3.2 Data extraction	44
169	2.3.3 Quality assessment	48
170	2.3.4 Meta-analysis	49
171	2.3.5 Narrative synthesis	54
172	2.4 Discussion	62

173	2.4.1 Summary of the evidence	62
174	2.4.2 Quality	64
175	2.4.3 Limitations	65
176	2.4.4 Conclusion	67
177	Chapter 3: EpiViz: a tool to visualise large association analyses	69
178	Chapter summary	69
179	3.1 Introduction	70
180	3.2 Methods	70
181	3.2.1 Circos plots for association analyses	70
182	3.2.2 Implementation	72
183	3.2.3 Operation	72
184	3.3 Discussion	76
185	Chapter 4: Associations between multiple measures of adiposity and metabolites: observational analysis	79
187	Chapter Summary	79
188	4.1 Introduction	80
189	4.2 Methods	81
190	4.2.1 Data overview	81
191	4.2.2 Exposures: adiposity	82
192	4.2.3 Outcomes: metabolites	84
193	4.2.4 Covariates	85
194	4.2.5 Statistical analysis	86

195	4.3 Results	88
196	4.3.1 Data overview	88
197	4.3.2 Statistical analysis	92
198	4.3.3 Comparison with previous work	103
199	4.4 Discussion	105
200	4.4.1 Limitations	106
201	4.4.2 Conclusion	109
202	Chapter 5: Associations between multiple measures of adiposity and metabolites: Mendelian randomization analysis	111
204	Chapter summary	111
205	5.1 Introduction	112
206	5.2 Methods	112
207	5.2.1 Instrumentation	114
208	5.2.2 Data	115
209	5.2.3 Two-sample Mendelian randomization: effect of adiposity measures on metabolites	120
211	5.2.4 Meta-analysis of two-sample Mendelian randomization results	123
212	5.2.5 Comparison of two-sample Mendelian randomization and observational results from Chapter 2	124
214	5.3 Results	124
215	5.3.1 Two-sample Mendelian randomization: effect of adiposity measures on metabolites	124
217	5.3.2 Meta-analysis of two-sample Mendelian randomization results	145

218	5.3.3 Comparison of two-sample Mendelian randomization and observational	146
219	results from Chapter 2	
220	5.4 Discussion	149
221	5.4.1 Limitations	152
222	5.4.2 Conclusion	155
223	Chapter 6: Associations between adiposity associated metabolites and endometrial	
224	cancer: Mendelian randomization analysis	157
225	Chapter summary	157
226	6.1 Introduction	158
227	6.2 Methods	159
228	6.2.1 Data	162
229	6.2.2 Two-sample Mendelian randomization	164
230	6.2.3 Two-sample multivariable Mendelian randomization	166
231	6.3 Results	168
232	6.3.1 Two-Sample Mendelian randomization: effect of adiposity measures on	
233	endometrial cancer	168
234	6.3.2 Two-sample Mendelian randomization: effect of adiposity measures on	
235	metabolites	169
236	6.3.3 Two-sample Mendelian randomization: effect of adiposity-associated	
237	metabolites on endometrial cancer	170
238	6.3.4 Two-sample multivariable Mendelian randomization: intermediate ef-	
239	fects of adiposity associated metabolites on endometrial cancer	174
240	6.4 Discussion	180
241	6.4.1 Instrumentation	183
242	6.4.2 Limitations	183

243	6.4.3 Conclusion	185
244	Chapter 7: Discussion	187
245	Chapter summary	187
246	7.1 Overview	187
247	7.1.1 Chapter 2: Adiposity is causally associated with numerous risk factors and diseases	188
248	7.1.2 Chapter 4: Adiposity is associated with individual and whole subclasses of metabolites across time points	189
251	7.1.3 Chapter 5: The effect of adiposity on metabolites is consistent across observational and Mendelian randomization analyses	190
252	7.1.4 Chapter 6: Intermediary metabolites may partly explain the association between adiposity and diseases identified in Chapter 2	191
255	7.2 Strengths and limitations	192
256	7.2.1 Replication, meta-analysis, and triangulation of evidence	192
257	7.2.2 Data, availability, and presentation	194
258	7.2.3 Methodology, instrumentation, and assumptions	196
259	7.2.4 General assumptions of MR analyses	199
260	7.2.5 Additional considerations	201
261	7.3 Future work	202
262	Conclusion	205
263	Appendix A:	207
264	A.1 Chapter 1: Introduction	208
265	A.2 Chapter 2: Systematic Review	212

266	A.3 Chapter 4: Observational	219
267	A.3.1 Methods	219
268	A.3.2 Results	224
269	A.4 Chapter 5: MR	235
270	A.4.1 Methods	235
271	A.4.2 Results	243
272	A.5 Chapter 6: Mediation	267
273	A.5.1 Methods	267
274	A.5.2 Results	269
275	References	285

List of Tables

277	2.1	Number and frequency of exposures used across all 2214 MR analyses	46
278	2.2	Number and frequency of 10 most used outcomes across all 2214 MR analyses	47
279	2.3	Number and frequency of outcomes within each outcome category across all 2214 MR analyses	47
280			
281	2.4	Number of times a study was used in meta-analyses	49
282	4.1	Metabolomics data available prior to pre-analysis processing	89
283	4.2	Metabolomics data available after pre-analysis processing	90
284	4.3	Measures of adiposity available for individuals with metabolomics data	91
285	4.4	Covariates available for individuals with metabolomics data	91
286	4.5	Total number of individuals with metabolomics data and total number of individuals included in each linear model	92
287			
288	4.6	Number of tests reaching a multiple testing threshold	93
289	5.1	Number of tests in which Cochran's Q exceeds genetic instrumental variable degrees of freedom	136
290			
291	5.2	Metabolites with a consistent direction of effect between BMI and WHR which reached a multiple testing threshold (p-value < 0.00045) in meta-analysis . . .	145
292			

293	6.1	Two-sample Mendelian randomisation results: the effect of adiposity measures on endometrial cancer	169
295	6.2	Two-sample Mendelian randomization: Effect of adiposity measures on triglycerides in small and very small VLDL	174
297	6.3	Two-sample Mendelian randomization: Effect of triglycerides in small and very small VLDL on endometrial cancer	175
299	6.4	F-statistics for adjusted exposures used in multivariable MR analyses	179
300	6.5	Multivariable MR Q statistics	180
301	A.1	MELDOI analysis: intermediates between "body mass index" and "mortality" .	209
302	A.2	Quality assessment tool	213
303	A.3	Quality assessment results for studies included in meta-analyses	214
304	A.4	Meta-analyses results from MR analyses using adiposity as an exposure . . .	216
305	A.5	Metabolites used in MR and meta-analyses	237
306	A.6	Metabolites used in MR analyses	267

³⁰⁷ List of Figures

308	1.1	Proportion of individuals who suffer from overweight or obesity globally and in 309 select locations	4
310	1.2	Adipocyte fat deposition and mobilisation	6
311	1.3	Directed acyclic graph of the Mendelian randomization principle	26
312	1.4	Directed acyclic graph of the two-step Mendelian randomization principle . . .	30
313	1.5	Directed acyclic graph of the multivariable Mendelian randomization principle 314 using two exposures	31
315	2.1	Inclusion criteria for meta-analysis: flowchart	41
316	2.2	PRISMA flowchart	43
317	2.3	Distribution of publication year per article and average exposure and outcome 318 sample sizes across included studies up to February 2019	44
319	2.4	Distribution of study design across 173 included articles	45
320	2.5	Quality assessment: distribution of quality assessment scores for studies 321 included in the meta-analyses	48
322	2.6	Meta-analysis: effect estimates and 95% confidence intervals for binary outcomes	52
323	2.7	Meta-analysis: effect estimates and 95% confidence intervals for continuous 324 outcomes	53
325	3.1	Circos plot highlighting the six elements used to plot data	72

326	3.2 Example Circos plot using EpiViz R package and example data	75
327	4.1 Directional consistency between effect estimates across three linear models 328 within exposures and age groups	94
329	4.2 Directional consistency between effect estimates for linear model 2 across 330 exposures within age groups	95
331	4.3 Effect estimates from linear regression of adiposity measures on metabolites 332 in children	97
333	4.4 Effect estimates from linear regression of adiposity measures on metabolites 334 in adolescents	98
335	4.5 Effect estimates from linear regression of adiposity measures on metabolites 336 in young adults	99
337	4.6 Effect estimates from linear regression of adiposity measures on metabolites 338 in adults	100
339	4.7 Effect estimates from linear regression of adiposity measures on metabolites 340 in all age groups for small, medium, and large LDL subclasses	102
341	4.8 Comparison of the directions of effect estimates from linear regression of 342 adiposity measures on metabolites across all age groups and analysis by 343 Wurtz et al. (2014) ¹⁵⁷	104
344	5.1 Analysis overview: Effect of adiposity on metabolites	114
345	5.2 Mendelian randomization analysis: Effect of adiposity measures on 123 NMR 346 derived metabolites using Kettunen data	126
347	5.3 Mendelian randomization analysis: Effect of adiposity measures on 230 NMR 348 derived metabolites using INTERVAL data	127
349	5.4 Directional consistency of two-sample MR effect estimates	129
350	5.5 Directional consistency of main and sensitivity analyses from two-sample 351 Mendelian randomization results using Kettunen and INTERVAL data	132
352	5.6 Directionally consistent effects across all two-sample Mendelian randomization 353 models using the Kettunen data	134

354	5.7	Directionally consistent effects across all two-sample Mendelian randomization models in the INTERVAL analysis	135
355			
356	5.8	Schematic representative figure: Single-SNP MR analysis of the effect of adiposity on a metabolite	137
357			
358	5.9	Schematic representative figure: Leave-one-out MR analysis of the effect of adiposity on a metabolite	138
359			
360	5.10	Representative figure: Funnel plot of of BMI on one metabolite using Kettunen data	139
361			
362	5.11	Representative figure: Funnel plot of of BF on one metabolite using Kettunen data	140
363			
364	5.12	Representative figure: Leave-one-out MR analysis of BF on acetoacetate using Kettunen data	141
365			
366	5.13	Metabolites associated with measures of adiposity in observational, MR, and MR meta-analysis	148
367			
368	6.1	Analysis overview	161
369	6.2	Directed acyclic graph of multivariable MR principle	167
370	6.3	Distribution of the number of SNPs individual metabolites were associated with	171
371	6.4	Distribution of the number of metabolites individual SNPs were associated with	172
372	6.5	Two-sample MR: Effect of metabolites on endometrial cancer	173
373	6.6	Multivariable MR: Effect of adiposity and metabolites on endometrial cancer	177
374	A.1	Distribution of raw impedance and calculated body fat percentage in children and adolescents	220
375			
376	A.2	Correlation between different body fat percentage measures in children and adolescents	221
377			
378	A.3	Correlation between body fat percentage and anthropometric measures in children and adolescents	222
379			

380	A.4 Correlation between body fat percentage and anthropometric measures in 381 adolescents	223
382	A.5 Distribution of adiposity measures across all age groups	228
383	A.6 Effect estimates from linear regression of adiposity measures on metabolites .	229
384	A.7 Effect estimates from linear regression of adiposity measures on derived 385 metabolites	230
386	A.8 Effect estimates from linear regression of adiposity measures on derived 387 supplementary metabolites in children	231
388	A.9 Effect estimates from linear regression of adiposity measures on derived 389 supplementary metabolites in adolescents	232
390	A.10 Effect estimates from linear regression of adiposity measures on derived 391 supplementary metabolites in young adults	233
392	A.11 Effect estimates from linear regression of adiposity measures on derived 393 supplementary metabolites in adults	234
394	A.12 F-statistics for individual genetic instrumental variables and mean F-statistic 395 for each exposure used in Mendelian randomization analyses	242
396	A.13 Global metabolic profile of BMI on 123 NMR derived metabolites from Kettunen 397 et al. (2016) ³¹⁸ : sensitivity analysis	244
398	A.14 Global metabolic profile of BMI on 230 NMR derived metabolites from INTER- 399 VAL: sensitivity analysis	245
400	A.15 Global metabolic profile of WHR on 123 NMR derived metabolites from Ket- 401 tunen et al. (2016) ³¹⁸ : sensitivity analysis	246
402	A.16 Global metabolic profile of WHR on 230 NMR derived metabolites from INTER- 403 VAL: sensitivity analysis	247
404	A.17 Global metabolic profile of BF on 123 NMR derived metabolites from Kettunen 405 et al. (2016) ³¹⁸ : sensitivity analysis	248
406	A.18 Global metabolic profile of BF on 230 NMR derived metabolites from INTERVAL: 407 sensitivity analysis	249

408	A.19 Representative figure: Single SNP MR analysis of BMI on glycerol from Ket-	
409	tunen et al. (2016) ³¹⁸	251
410	A.20 Representative figure: Funnel plot of of BMI on one metabolite from Kettunen	
411	et al. (2016) ³¹⁸	252
412	A.21 Representative figure: Leave-one-out MR analysis of BF on Acetoacetate from	
413	Kettunen et al. (2016) ³¹⁸	253
414	A.22 Representative figure: Leave-one-out MR analysis of BMI on Acetate from	
415	Kettunen et al. (2016) ³¹⁸	254
416	A.23 Single SNP MR analysis of BMI on 123 NMR derived metabolites from Kettunen	
417	et al. (2016) ³¹⁸ : additional sensitivity analysis - representative figure	256
418	A.24 Representative figure: Funnel plot of of BMI on one metabolite from INTERVAL	257
419	A.25 Representative figure: Leave-one-out MR analysis of BF on Acetoacetate from	
420	Kettunen et al. (2016) ³¹⁸	258
421	A.26 Representative figure: Leave-one-out MR analysis of BMI on Glycine from	
422	INTERVAL	259
423	A.27 Global metabolic profile of BMI on 123 NMR derived metabolites from Kettunen	
424	et al. (2016) ³¹⁸ : additional analysis	261
425	A.28 Global metabolic profile of BMI on 230 NMR derived metabolites from INTER-	
426	VAL: additional analysis	262
427	A.29 Global metabolic profile of WHR on 123 NMR derived metabolites from Ket-	
428	tunen et al. (2016) ³¹⁸ : additional analysis	263
429	A.30 Global metabolic profile of WHR on 230 NMR derived metabolites from INTER-	
430	VAL: additional analysis	264
431	A.31 Global metabolic profile of BF on 123 NMR derived metabolites from Kettunen	
432	et al. (2016) ³¹⁸ : additional analysis	265
433	A.32 Global metabolic profile of BF on 230 NMR derived metabolites from INTERVAL:	
434	additional analysis	266
435	A.33 Two-sample MR: Effect of adiposity on endometrial cancer	269

436	A.34 Representativ figure: Scatter plot of MR effect estimates from two-sample MR of adiposity on endometrial cancer	270
438	A.35 Representativ figure: Funnel plot of MR effect estimates from two-sample MR of adiposity on endometrial cancer	271
440	A.36 Representativ figure: Leave one out plot of MR effect estimates from two- sample MR of adiposity on endometrial cancer	272
442	A.37 Representativ figure: Single SNP plot of MR effect estimates from two-sample MR of adiposity on endometrial cancer	273
444	A.38 Two-sample MR: Effect of adiposity on metabolites	275
445	A.39 Representativ figure: Scatter plot of MR effect estimates from two-sample MR of adiposity on metabolites	276
447	A.40 Representativ figure: Funnel plot of MR effect estimates from two-sample MR of adiposity on metabolites	277
449	A.41 Representativ figure: Leave one out plot of MR effect estimates from two- sample MR of adiposity on metabolites	278
451	A.42 Representativ figure: Single SNP plot of MR effect estimates from two-sample MR of adiposity on metabolites	279
453	A.43 Representativ figure: Scatter plot of MR effect estimates from two-sample MR of metabolites on endometrial cancer	280
455	A.44 Representativ figure: Funnel plot of MR effect estimates from two-sample MR of metabolites on endometrial cancer	281
457	A.45 Representativ figure: Leave one out plot of MR effect estimates from two- sample MR of metabolites on endometrial cancer	282
459	A.46 Representativ figure: Single SNP plot of MR effect estimates from two-sample MR of metabolites on endometrial cancer	283

Abbreviations

Abbreviation	Term
Apo	Apolipoprotein
BF	Body fat percentage
BMI	Body mass index
CRP	C reactive protein
CVD	Cardiovascular disease
CKD	Chronic kidney disease
CI	Confidence interval
CHD	Coronary heart disease
DXA	Dual energy x-ray absorptiometry
FG	Fasting glucose
GWAS	Genome-wide association study
HDL	High-density lipoprotein
HC	Hip circumference
HCadjBMI	Hip circumference adjusted for body mass index
IV	Instrumental variable
IL	Interleukin
LD	Linkage disequilibrium
MS	Mass spectrometry
MR	Mendelian randomization
MVMR	Multivariable Mendelian randomization
NAFLD	Non-alcoholic fatty liver disease
NMR	Nuclear magnetic resonance
OR	Odds ratio
PPAR- γ	Peroxisome proliferator-activated receptor- γ
PCOS	Polycystic ovary syndrome
RCT	Randomised control trial
SNP	Single nucleotide polymorphism
SD	Standard deviation
SE	Standard error

(continued)

Abbreviation	Term
STROBE-MR	Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomization
TG	Triglycerides
TNF-alpha	Tumour necrosis factor-alpha
WC	Waist circumference
WCadjBMI	Waist circumference adjusted for body mass index
WHR	Waist hip ratio
WHRadjBMI	Waist hip ratio adjusted for body mass index

462 **Chapter 1**

463 **Introduction**

464 **Chapter summary**

465 This chapter provides the context of the thesis. It presents a summary of the literature
466 relevant to the thesis, where evidence is limited and how this thesis will contribute. Firstly, I
467 give background to the issue of adiposity, describe the biological role of adipose tissue, its
468 genetic components and different measurements. I discuss associations with diseases and
469 outline potential underlying mechanisms. I give focus to metabolites as potential intermediates
470 linking adiposity with diseases and discuss the use of Mendelian randomization to disentangle
471 these associations.

472 **1.1 Background**

473 Excess and increased adipose tissue (adiposity) is a global health concern. Globally, the
474 prevalence of overweight (body mass index (BMI) of 25–29.9 kg/m²) and obesity (BMI > 30
475 kg/m²) is 39% and 13% respectively^{2,3} (Figure 1.1). Obesity is estimated to be responsible for
476 8% of global deaths⁴ and this number is likely to rise as the prevalence of obesity increases^{5–7}.
477 Overweight and obesity are categorisations of excess adipose tissue and are associated with
478 numerous diseases⁸. These associations are ultimately a result of the underlying functions
479 adipose tissue exerts within the body, including its distribution. Identifying where these
480 functions link with disease may improve health outcomes.

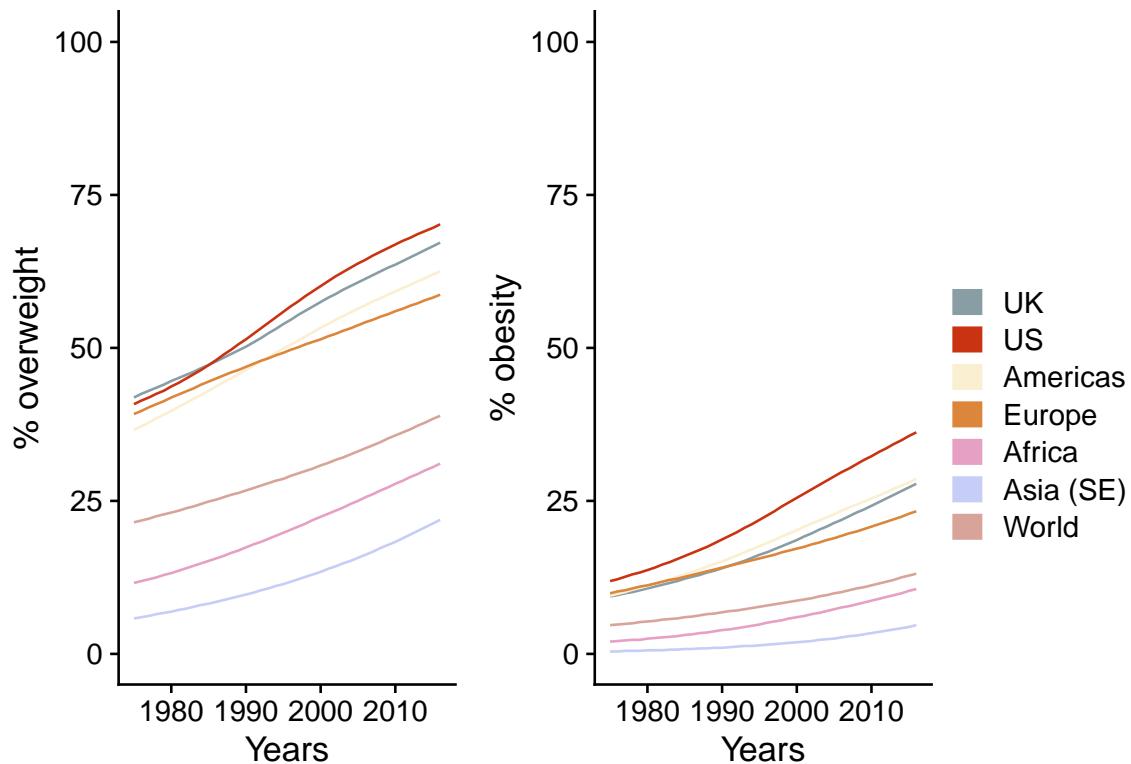


Figure 1.1: **Proportion of individuals who suffer from overweight or obesity globally and in select locations.** Data from Ritchie and Roser (2019)³ for individuals over 18 in the United Kingdom (UK), United States (US), Americas, South-East Asia (Asia (SE)), and the world.

481 **1.2 Adipose tissue**

482 Weight is made up of two components, fat free mass and fat mass. Fat free mass
483 encompasses muscle, bone and water mass. Fat mass is an all encompassing term for
484 adipose tissue. Adipose tissue is predominantly made up of adipocytes, with other tissues
485 and cells such as the stromal vascular fraction, preadipocytes and fibroblasts making up
486 smaller proportions⁹⁻¹¹. The main function of adipose tissue is energy storage in the form of
487 lipids, with a secondary function to insulate the body and maintain thermoregulation. These
488 two functions can broadly be separated into two types of adipose tissue, white and brown
489 respectively¹⁰. In addition to energy storage and insulation, adipose tissue is considered an
490 endocrine organ, responsive to afferent signalling as well as being a prolific signaller itself¹².
491 Additionally, single nucleotide polymorphisms (SNPs) and genes associated with increased
492 adipose tissue have been identified¹³.

493 **1.2.1 Energy storage**

494 Energy storage is determined by energy intake and energy expenditure. An increase
495 or decrease in one leads to a change in energy balance and thus an increase or decrease
496 in total energy storage as *Energy balance = energy in – energy out*. Energy stores are
497 comprised primarily of proteins, carbohydrates, and fats. For protein, there is little change in
498 total energy stores outside of a growth stimulus (i.e. exercise is needed to increase protein
499 stores)¹⁴. Carbohydrate stores fluctuate markedly throughout the day as a result of limited
500 storage capacity and the fact that they comprise the majority of energy production¹⁴. Fats are
501 the largest energy store. Daily fat intake is ~1% of the total available fat store¹⁴. Given the
502 tight controls over protein and limited availability of carbohydrate storage, fat storage is the
503 only expandable reservoir of excess energy intake^{9,14}. As a result, an energy imbalance will
504 be reflected in the fat stores and not elsewhere^{9,14}.

505 Excess energy is stored in adipocytes in the form of lipid droplets (triglycerides; Figure
506 1.2) via lipogenesis. The release of these fat stores, in the form of fatty acids, occurs through
507 lipolysis. As the main store of excess energy, triglycerides provide an accurate reflection of
508 energy imbalance, while adipocytes reflect the deposition and mobilisation of triglycerides⁹.
509 Deposition and mobilisation of triglycerides are a product of a complex interplay of genetic
510 and hormonal signals with leptin and insulin playing key roles¹⁵.

511 Insulin stimulates the conversion of acetyl-CoA to triglycerides by encouraging uptake
512 of glucose by adipocytes and promoting production of SREBP1 (sterol regulatory element-
513 binding protein 1), which regulates fatty acid, triglyceride, and cholesterol synthesis^{9,11}. In
514 addition, lipoprotein lipase plays a key role in hydrolysing circulating triglycerides into fatty
515 acids enabling their uptake by adipocytes⁹. Though they can expand, individual adipocytes
516 have limited storage capacity for triglycerides. Once *full*, adipocytes have the ability to

517 multiply⁹. The amount of expansion adipocytes can achieve is limited^{9,16} and thought to be
518 influential in the rate of adipogenesis, the rate of fat mobilisation around the body, and the
519 development of disease¹⁶.

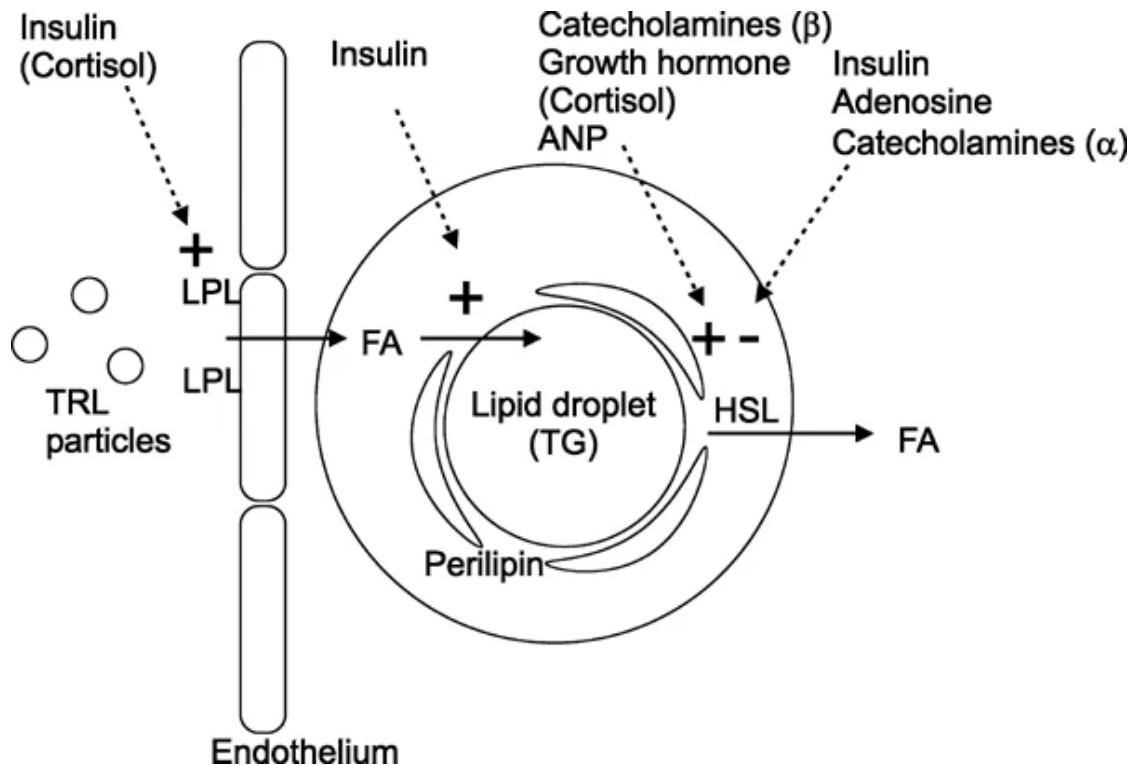


Figure 1.2: **Adipocyte fat deposition and mobilisation.** Stimuli such as insulin result in positive (+) fat deposition, while other stimuli such as cortisol may result in positive or negative (-) fat deposition depending on the location of the adipocyte within the body. ANP, atrial natriuretic peptide; FA, fatty acids; LPL, lipoprotein lipase; TG, triglyceride; TRL, TG-rich lipoproteins. Reproduced from Frayn et al. (2003)⁹.

520 1.2.2 Insulation

521 Energy storage of fats is managed predominantly by white adipose tissue. These deposits
522 are located mainly within subcutaneous tissue. During infancy, brown adipose tissue is
523 abundant; however, as humans age, these deposits *whiten* leaving few adult brown fat
524 deposits. The remaining deposits of brown adipose tissue in adults are located around
525 the neck, thoracic section of the spine, aortic body, and adrenal glands; all locations with
526 high blood flow^{11,17}. Thermogenesis by these tissues is regulated by the hypothalamus and
527 is achieved by uncoupling of the respiratory chain of oxidative phosphorylation via UCP1
528 (uncoupling protein 1). When this process is active, lipids and glucose are used as fuel¹⁷.

529 Due to the abundant vascularization of areas where brown adipose tissue are located,
530 the heat generated from this process is quickly distributed via the circulatory system. White
531 adipose tissue can undergo a *beiging* process taking on thermogenic properties of brown
532 adipose tissue. Beige adipose tissue is a half way point between white and brown adipose
533 tissue and is more widely dispersed than brown adipose tissue, being located mainly within
534 subcutaneous white adipose tissue. Beige adipose tissue, much like brown adipose tissue, is
535 cold activated but can be recruited through signalling that mimics the stressed state induced
536 by cold. The *beiging* process is not well characterized but is thought to be a result of signalling
537 changes during differentiation of preadipocytes¹¹. The *beiging* process is reversible but has
538 been suggested as a therapeutic avenue for weight loss.¹⁷

539 1.2.3 Signalling

540 It is important to consider adipose tissue as an organ in its own right. Not solely comprised
541 of adipocytes, adipose tissue includes a multitude of tissues and cells including connective
542 and nerve tissue as well as immune cells. All respond to, and secrete, signalling molecules
543 locally and systemically. It is thought this signalling is primarily to maintain appropriate energy
544 stores and includes signals influencing deposition and mobilisation of fats and differentiation
545 of new adipocytes^{9,11,12}. Functionally, signalling molecules have metabolic effects and/or are
546 involved in steroid hormone production¹².

547 Adipogenesis, the process of adipocyte formation, has been well characterized and
548 PPAR γ (peroxisome proliferator-activated receptor γ) is the master regulator^{9,11}. Over-
549 expression of PPAR γ leads to differentiation, and under-expression of PPAR γ results in
550 lipodystrophy. Other signalling molecules such as KLFs (Kruppel-like factors) and C/EBPs
551 (CCAAT-enhancer-binding proteins) influence adipogenesis through PPAR γ ¹¹. Because of the
552 master regulatory function of PPAR γ , exploring regulatory function and expression has been
553 considered as a potential therapeutic avenue for obesity^{11,18}. Though not well characterized,
554 brown adipocytes are thought to be influenced heavily by PRDM16 (PR/SET Domain 16) and
555 PGC1 α (peroxisome proliferator-activated receptor-gamma coactivator 1 α), with the latter
556 required for thermogenesis and not necessarily adipogenesis¹¹.

557 The breakdown of stored triglycerides via lipolysis results in the release of fatty acids
558 and glycerol molecules for oxidation and gluconeogenesis respectively. Fatty acids can also
559 be broken down into ketone bodies via ketogenesis. While insulin abundance activates
560 lipogenesis, the relative absence of insulin promotes lipolysis. The lipolytic pathway, which is
561 also activated by cAMP-dependent (cyclic adenosine monophosphate) PKA (protein kinase
562 A), relies on the function of ATGL (adipocyte triglyceride lipase) and HSL (hormone sensitive
563 lipase) to catalyse the hydrolysis of triglycerides to di- and mono-glyceride's respectively.
564 Inhibition of ATGL can result in impaired lipolysis and obesity¹⁹.

565 The signalling molecules adipocytes produce, known as adipokines, are numerous and

566 act on the auto- and endo-crine systems^{20,21}. There are adipose deposit specific effects on
567 expression and secretion of adipokines and the movement these adipokines can be expected
568 to undertake. Subcutaneous adipose tissue adipokines travel through the systemic system
569 while visceral adipose tissue adipokines can travel via the portal system with direct access to
570 the liver. Adipocyte receptors are also expressed differentially based on deposit location¹².
571 The main adipokines produced by adipocytes are leptin and adiponectin which function to
572 regulate metabolism and inflammation systemically. Other cells within the adipose tissue,
573 including immune and endothelial cells, produce many of the other adipokines such as TNF α
574 (tumour necrosis factor α) and IL6 (interleukin 6)¹¹.

575 Observational evidence has shown that abnormal levels of adipokines are harmful, as
576 they lead to impaired adipose tissue function and subsequent downstream effects such as
577 insulin resistance^{11,21-23}. As adipose tissue abundance increases so too does the likelihood
578 of abnormal levels of adipokines. This is of particular interest as adipokine levels can change
579 as a result of diseases such as obesity, thus introducing feed-back loops which serve to alter
580 normal processes^{21,22}. However, there are outstanding questions about how functionally
581 abnormal levels of adipokines leads to the development of disease²¹. Additionally, recent
582 causal analyses has shown inconsistent and weak evidence for an effect of adipokines on
583 adverse health outcomes^{24,25}.

584 1.3 Adiposity: genetics

585 Studies have identified numerous genes associated with adipose tissue, including for
586 lipogenesis, storage capacity, lipolysis, mobilisation of fat deposits, and energy expenditure¹³.
587 This also includes genes for adipose specific proteins such as leptin (*LEP*^{26,27}), adiponectin
588 (*ADIPOQ*²⁸), and *PPAR γ* ^{27,29}. Expression of these and other adipose associated genes^{27,30}
589 is tissue specific (subcutaneous/visceral and adipose/not-adipose)²⁷. Differential expression
590 along with the presence of adipose-specific genes is associated with obesity and related
591 diseases²⁷. For example, *SLC19A3* is an adipose-specific gene²⁷ encoding a thiamine trans-
592 porter; thiamine dependent enzymes have been associated with obesity³¹. Adipose expand-
593 ability is limited, and given that an excess of fatty acids results in increased adipogenesis^{9,16},
594 it is possible that expandability is fixed by adipose-specific genes and expression. This may
595 have an impact on the development of diseases such as type-2 diabetes where incidence
596 differs between different ancestral populations with the same BMI³².

597 In addition to adipose-specific genetics, a wealth of genome-wide association studies
598 (GWAS) have identified numerous genes and SNPs associated with adiposity^{30,33-43}. These
599 studies have focused on large population based studies, predominantly including individuals
600 of European ancestries and using specific measures of adiposity. These studies have also
601 revealed the genetic contribution to adiposity to be relatively high across many measures
602 including BMI (twin-based = 60-75%^{44,45}, family-based = 40-45%⁴⁴, population-based = 20-

603 40%^{36,40,46,47}) waist hip ratio (WHR; twin-based = 30-60%^{37,48}, family-based = 20-50%^{37,49},
604 population-based = 10%^{37,50}) and body fat percentage (BF; family-based = 64%⁵¹).

605 The first of these GWAS identified the *FTO* locus to be associated with BMI and
606 obesity^{52,53}. Subsequent studies including larger sample sizes have identified over 900
607 SNPs associated with BMI^{33,36,41,42}. This same increase in associated SNPs is true for other
608 measures of adiposity including WHR^{37,42}, and WHR adjusted for BMI (WHRadjBMI)^{37,42}. BF
609 has also been associated with increasing numbers of SNPs^{39,43}, however given BF measures
610 are not as easily obtained as BMI and WHR, sample sizes have been much smaller. Studies
611 are also increasingly focusing on non-European ancestries and identifying ancestry specific
612 genetic associations^{30,40,54,55}. Though allele frequencies and effect sizes are different, on the
613 whole, associations shared across ancestries are directionally consistent.

614 Genetic variants associated with complex traits such as adiposity are spread throughout
615 the genome in coding and non-coding regions⁵⁶. Ascribing causal links between variants and
616 genes⁵⁷ and between variants and functions⁵⁵ is challenging. However, functional analyses
617 have revealed different pathways associated with adiposity, as well as measurement specific
618 pathways. Variants associated with BMI have been linked with the central nervous system
619 (hypothalamus, pituitary gland, hippocampus, and limbic system)^{36,58} while WHRadjBMI
620 variants show links with adipogenesis and angiogenesis^{37,58}. These differences in genetics
621 highlight the underlying biology of the measurements of adiposity, it suggests that BMI is not a
622 solely physical measurement, but that it also captures a behavioral component of adiposity³⁶.

623 1.4 Adiposity: measures

624 Adiposity is generally measured, and used to categorise individuals, using BMI. Body
625 mass index is a measure of weight given an adjustment of height; $\frac{\text{Weight (kg)}}{\text{Height}^2 (\text{cm})}$. It thus gives
626 an approximation of body composition. Given this simplicity, BMI is easily and widely used
627 as a measure of adiposity. BMI can be classified into sub-types according to the range of
628 values seen in the general population. Whilst ethnicity, sex, and age specific, the international
629 standards set by the WHO^{59,60} estimate a normal weight classification at a population level to
630 be a BMI of 18.5–24.9 kg/m², with an underweight class below this. Underweight is a specific
631 condition that may be secondary to or symptomatic of an underlying disease and is not within
632 the scope of this thesis. An excess of adipose tissue classified as overweight is given as
633 a BMI of 25–29.9 kg/m². Obesity is classed as a BMI greater than or equal to 30 kg/m² –
634 additional obesity classifications are sometimes used, e.g. the NHS classifies severely obese
635 as a BMI equal to or greater than 40 kg/m².

636 A main driving force behind the use of BMI is its association with many of the most abundant
637 diseases and causes of death worldwide. This includes all cause mortality, hypertension,
638 type 2 diabetes, stroke, respiratory problems, many cancers, and more^{61,62}. The ability of

639 BMI to categorise individuals as at risk in an efficient manner is a major benefit over other
640 more costly and time consuming measures.

641 Though BMI shows similar relationships to other anthropometric measures with disease⁶³,
642 a main limitation is its inability to differentiate lean mass and fat mass. One can have a
643 high/low BMI and a low/high total fat mass⁶⁴. This is particularly evident in sex comparisons,
644 where generally men have a higher BMI than women, even though women generally have
645 greater fat mass⁶⁵. Differences are not limited to sex, they are also apparent for age⁶⁵,
646 race^{66,67}, and ethnicity^{68,69}.

647 The reliability of BMI, the ability to obtain consistent and stable results repeatedly, is
648 affected by numerous factors. These include diurnal variation in height and weight fluctuations
649 due to clothing and food consumption. Though this can be managed relatively well using
650 standard operating procedures, the question of validity remains. Specificity is reported
651 to be high, however sensitivity is much lower, therefore incorrectly classifying individuals
652 with excess body fat mass. Numerous studies have found that BMI lacks the resolution to
653 accurately measure body composition⁷⁰⁻⁷⁴. There are also questions around the relationship
654 with morbidity and mortality⁷⁵⁻⁷⁷, though these questions are likely a result of confounding
655 and other biases⁷⁸⁻⁸⁰.

656 Evidence has pointed to a more important role for fat deposition in the relationship with
657 mortality^{81,82}. WHR is commonly used to measure fat deposition, and is calculated as
658
$$\frac{\text{Waist (cm)}}{\text{Hip (cm)}}$$
. A WHR > 0.85 in women and > 0.9 in men is considered equivalent to a BMI of >
659 30 kg/m².⁸³ Unlike BMI, WHR requires two measurements and is therefore prone to more
660 measurement error; hip measurements are generally harder to perform than waist measures,
661 especially where there is excess adipose tissue. It is estimated that error can be as high
662 as 1.56 cm⁸³. Additionally, interpretation can be difficult given that an increased WHR can
663 be caused by both increased abdominal adiposity and a decrease in lean mass at the hips⁸.
664 WHR is a predictor for many diseases and correlate strongly with direct measures of body
665 fat⁸.

666 Direct measurement of body fat has been argued as key in understanding the development
667 of diseases associated with adiposity⁷⁰. Broadly, the aim of these measures is to quantify BF
668 which provides a more accurate estimation of body composition⁸⁴. Skinfold callipers can be
669 used to measure subcutaneous fat at various locations around the body. Equations derived
670 from gold standard measurements are used to convert the multiple skinfold measures to
671 BF. In general, skinfold measurement is the cheapest and easiest of direct BF assessments.
672 That being said, measurements are prone to error, especially as many skinfold measures are
673 needed. Additionally, equations are derived in specific populations and do not always translate
674 well⁸. Bioelectrical impedance devices, which send a small electrical current through the
675 body, can be used to estimate BF by measuring the time taken for the current to pass through
676 the body. Fat mass has a greater resistance than lean mass and water. Similar to skinfold
677 callipers, impedance devices are quite easy to use and are portable. However, calibration

678 can be difficult and derived equations may not be well translated to all populations; age,
679 height, weight, sex and more can be used in proprietary equations, which are commercially
680 sensitive and difficult to appraise⁸. Evidence does however show strong correlations for some
681 equations with more accurate measures of BF^{85,86}. It is important to note that although these
682 techniques can be delivered easily, they are not used as widely as BMI and WHR, as such
683 sample sizes are much lower.

684 Where as previous measurement techniques indirectly measure BF, imaging techniques
685 allow for direct measurement. Dual energy x-ray absorptiometry (DXA) uses X-rays, which
686 pass through body tissues differentially, to image and then quantify using software fat mass,
687 fat free mass, and bone mineral density. As DXA directly measures body tissues it is highly
688 accurate. However, DXA uses X-rays which are harmful to certain individuals, it is expensive,
689 and can't easily be transported. Additionally, it is not possible to distinguish subcutaneous
690 and visceral fat. Computerized tomography (CT) and magnetic resonance imaging (MRI) are
691 considered the gold standard for BF measurement. Both are able to measure specific tissues,
692 however a hard call must still be made as to whether the measured area is coded as fat-free
693 or fat mass. That being said, there is high inter-individual reproducibility of DXA measures⁸⁷.
694 Imaging techniques are expensive, static, and can not be used with certain individuals much
695 like DXA. As with other BF measures, sample sizes are low given the cost and complexity
696 of using DXA, CT, and MRI devices at scale. For all BF measurements, there is also the
697 potential for measurement error as a result of the fasted status of the individual and whether
698 they had recently performed exercise.

699 Given the limitations of BMI, WHR, and easily acquired BF measurements, a complimentary
700 assessment of adiposity, using a combination of body composition measures, may be
701 beneficial when investigating associations with disease^{83,88}. This is pertinent given genetic
702 analyses have also revealed different underlying biological pathways for BMI, WHR, and
703 BF⁵⁵.

704 1.5 Adiposity: morbidities

705 Body mass index^{75–77}, WHR^{71,89–92} and BF^{89,92–94} are associated with increased risk
706 of mortality. Literature mining to identify intermediate diseases linked with BMI and mor-
707 tality (Appendix A.1) highlights 9 broad categories: Cancer, Cardiovascular, immune, Kid-
708 ney, Liver, Neurological/behavioural, Other, Pregnancy, Respiratory – *other* includes dis-
709 eases like diabetes and the metabolic syndrome. Similar links are identified for WHR and
710 cancer⁹⁵, cardiovascular^{96,97}, kidney⁹⁸, liver⁹⁹, neurological/behavioural^{100,101}, pregnancy¹⁰²,
711 respiratory¹⁰³, and diabetes¹⁰⁴ outcomes. Fewer studies are reported for BF, however there
712 are links with cancer⁹³, cardiovascular^{93,105}, kidney¹⁰⁶, respiratory⁹³, and diabetes¹⁰⁷ out-
713 comes.

714 **Cancer**

715 In one of the largest studies of its kind, BMI was found to be associated with overall
716 cancer risk as well as site specific cancers in up to 5.24 million UK adults¹⁰⁸. For oral, lung,
717 pre-menopausal breast, and prostate cancer BMI was protective. However after exclusion
718 of ever smokers, associations with oral and lung were inconclusive. There was no change
719 with pre-menopausal breast and prostate after ever smoker exclusion. For oesophageal,
720 stomach, and pancreatic cancer associations reach a traditional significance threshold for
721 an increasing effect of BMI after exclusion of ever smokers. For the remaining site specific
722 cancers exclusion of ever smokers did not change the increased risk associated with BMI:
723 colon, rectum, liver, gallbladder, post-menopausal breast, cervix, uterus, ovarian, kidney, and
724 leukaemia. There was weaker evidence of an increased risk for: melanoma, bladder, brain,
725 thyroid, non-Hodgkin lymphoma, and melanoma.

726 The protective effect observed for BMI and some cancers has been challenged more
727 recently, with evidence showing that among Korean populations BMI increases the risk of
728 prostate cancer and this may be linked with abdominal fat deposition¹⁰⁹. Additionally, a sys-
729 tematic review and meta-analysis has highlighted that BMI was associated with an increased
730 risk of advanced prostate cancer but also with a reduced risk of localised prostate cancer¹¹⁰.
731 This may be a result of multiple factors, including selection bias, whereby individuals with a
732 higher BMI are diagnosed at a later stage¹¹¹. It may also be the case that advanced prostate
733 cancer is more likely in individuals with a higher BMI because of a beneficial environment that
734 promotes cancer development, i.e. increased production of hormones¹¹¹.

735 There is an association between breast cancer and adiposity in men¹¹² however this is
736 complicated in women by hormonal status, with pre-menopausal women at reduced risk but
737 post-menopausal women at increased risk¹⁰⁸. This picture is further complicated by receptor
738 status, with oestrogen (ER) and progesterone receptor (PR) positive post-menopausal breast
739 cancer than for ER- and PR- post-menopausal status¹¹³. There is also an association with
740 increased post-menopausal breast cancer risk and WHR and waist circumference (WC)¹¹⁴.

741 Though studies have suggested associations with post-menopausal breast cancer and
742 WHR and WC may be a result of overall adiposity as opposed to deposition^{115,116}, there is
743 wider evidence for associations between WC, hip circumference (HC), and WHR with multiple
744 cancers^{117,118}. Similarly, there is evidence for an association with weight gain¹¹⁷ and BF¹¹⁹,
745 however, limited data has led to some inconsistent results for weight gain¹¹⁷ and few studies
746 using BF. Additional considerations in the association of adiposity and cancer, include age, sex,
747 and ancestry^{117,120}. Prostate cancer for example is more strongly associated with adiposity
748 in African American ancestries than white ancestries¹²⁰. In regards to age, evidence has
749 highlighted adiposity in childhood and adolescence to be associated with later life cancers¹¹⁷.
750 For sex, the association with adiposity and cancer may be multifactorial, involving hormonal,
751 deposition, and homeostatic differences¹²¹.

752 **Cardiovascular**

753 Adiposity is associated with many cardiovascular traits^{122,123}, including risk factors (e.g. hypertension) and diseases (e.g. coronary heart disease). Generally, traits and diseases affecting the cardiovascular system are grouped as cardiovascular disease (CVD) and there is strong evidence for an increasing effect of adiposity¹²⁴. These associations include anaemia¹²⁵, thrombosis¹²⁶, myocardial infarction¹²⁷, dyslipidemia¹²⁸ and more. Importantly, the cardiovascular system undergoes adiposity induced adaptations. This includes increased cardiac output (primarily via stroke volume) to compensate for the increased blood flow required by adipose tissue. Overtime, these adaptations can lead to cardiomyopathy and in some cases adipose infiltration and replacement of cardiac cells¹²⁹.

762 Evidence for an independent effect of adiposity on CVD is strong^{130,131}. A number of 763 large studies have identified deposition as a stronger indicator of CVD risk compared to 764 overall adiposity^{132–135}, strongly suggesting that additional measures of adiposity, including 765 those that can distinguish fat types, can prove beneficial when investigating CVD and CVD 766 risk factors¹³⁵. It has also been argued that other morbidities independently associated with 767 adiposity may be influential in CVD development¹²⁹ – e.g. adiposity increases type 2 diabetes 768 risk which increases hypertension risk independent of adiposity's effect on hypertension.

769 Atherosclerosis is a major component of coronary heart disease (CHD; coronary artery 770 disease) and evidence shows atheroma build up is accelerated due to adiposity¹²². Both 771 overall and abdominal adiposity are associated with atherosclerosis and associations persist 772 after adjustment for smoking, hypertension, hyperlipidemia, and diabetes¹³⁶. Overtime, 773 atherosclerosis can lead to CHD, whereby the supply of blood to the heart muscles is reduced. 774 Adiposity is independently associated with CHD and accelerated progression¹³⁷. Alongside 775 CHD, adiposity is associated with heart failure and, in particular, intermediate risk factors 776 such as increased cardiac output and ventricular hypertrophy^{122,129}.

777 Adiposity is associated with an increased risk of sudden cardiac death, including associated 778 risk factors arrhythmia and atrial fibrillation^{138–140}. There even appears to be a dose 779 response relationship in the association with atrial fibrillation¹⁴⁰. There is evidence to suggest 780 that the relationship with atrial fibrillation is as a result of structural changes, namely atrial 781 hypertrophy¹³⁹. Of particular importance in the association with arrhythmia, is increased 782 cardiac adipose tissue which appears more predictive than overall and abdominal adipose 783 tissue^{140,141}.

784 Adiposity is associated with hypertension primarily through adiposity induced adaptations. 785 For example, increased blood flow to adipose tissue results in increased cardiac output 786 and subsequently the development of hypertension. The increase in overall adipose tissue 787 increases the risk of hypertension further due to adipose tissue increasing systemic and pul- 788 monary vascular resistance¹²⁹. Hypertension itself is associated with downstream diseases 789 such as stroke. Importantly, BMI and WHR are both independently associated with increased

790 risk of stroke (ischemic and haemorrhagic)^{142,143}. There has been evidence to suggest that
791 these independent effects are a result of inflammatory markers such as C-reactive protein
792 (CRP)¹⁴⁴. CRP has also been suggested as playing a role in the wider adiposity CVD associa-
793 tion. However, (and discussed in detail in section 1.8) these associations are believed not to
794 be causal and instead limitations of observational studies¹⁴⁵.

795 **Metabolic**

796 Metabolic disorders involve the dysfunction of metabolism, including not enough and
797 too many of a particular substance. A well studied, yet complex example of this metabolic
798 dysfunction is metabolic syndrome. A combination of three or more of adiposity, reduced
799 high-density lipoproteins (HDL), increased triglycerides (TG), increased fasting glucose
800 (FG), and hypertension. Metabolic syndrome is directly associated with the development
801 of CVD and type 2 diabetes¹⁴⁶. A large component of metabolic syndrome is adiposity
802 and insulin resistance. Dyslipidemia¹²⁸, hyperglycaemia¹⁴⁷, and hypertension¹²⁹ are all
803 independently associated with adiposity. While insulin resistance is associated with the
804 development of dyslipidemia¹⁴⁸, evidence points to hyperglycaemia¹⁴⁹ and hypertension¹⁵⁰
805 being associated with the development of insulin resistance. There is also a strong association
806 with adiposity and the development of insulin resistance^{146,151} and this may be through
807 adiposity's contribution to the components of metabolic syndrome¹⁴⁶.

808 Diabetes is a metabolic dysfunction in which blood glucose regulation is impaired as
809 a result of insufficient insulin. Type 1 diabetes is an autoimmune condition and makes up
810 about ~10% of cases. Type 2 diabetes is much more prevalent (~90%) and is a result
811 of either insufficient insulin production and/or insulin resistance¹⁵². Generally, adiposity is
812 not associated with type 1 diabetes¹⁵³. There is some evidence to support an accelerator
813 hypothesis, whereby adiposity at a young age accelerates the development of type 1 diabetes
814 via insulin resistance^{154,155}. More data is needed however to investigate the factors that
815 impact this relationship^{154,155}.

816 Adiposity is strongly associated with the development of type 2 diabetes¹⁵³. This associa-
817 tion is thought to be partly through the association between adiposity and insulin resistance,
818 however not all individuals with adiposity and insulin resistance develop hyperglycaemia¹⁵³. β -
819 cell dysfunction is an additional component; β -cells produce and secrete insulin, dysfunction
820 therefore results in aberrant insulin availability. The function of β -cells is altered as a result
821 of adipose tissue requiring more glucose, but also adipose tissue increases the abundance
822 of circulating fatty acids which in turn affects insulin secretion. Adiposity is associated with
823 increased circulating fatty acids. Over time, continual stimulation by fatty acids can reduce
824 insulin production¹⁵⁶.

825 Alongside dyslipidemia and hyperglycaemia, adiposity is associated with an increase in
826 many other circulating metabolites¹⁵⁷ as well as the metabolome in general¹⁵⁸. Changes in

827 the metabolome appear to track adiposity changes over time¹⁵⁸. Global associations, like
828 that with the metabolome, are also found for the gut-microbiome, where adiposity leads to
829 changes in both the abundance of, and types of, species that make up the microbiome^{159,160}.
830 The exact relationship between adiposity and the gut microbiome is unclear, but there
831 are association between the microbiome and several diseases, including CVD and type
832 2 diabetes¹⁶¹. Additionally, there is evidence for an association between adiposity and
833 the proteome^{162,163} which has been shown to influence CVD risk¹⁶⁴. Much of the work
834 investigating the metabolome, microbiome, and proteome has involved limited sample sizes
835 and requires further investigation.

836 **Immune**

837 Adiposity is associated with a state of overall chronic inflammation which is itself as-
838 sociated with numerous diseases, including CVD and type 2 diabetes¹⁶⁵. This chronic
839 inflammation is indicated by higher circulating levels of inflammatory markers as well as an
840 increased production of markers such as TNF- α , interleukin-6 (IL-6), and CRP at adipose
841 tissue deposits^{166–168}. There is a strong association between BMI and total body fat with
842 these inflammatory markers, however, evidence points to a stronger association with central
843 adiposity (WC and WHR)^{166,167,169,170}. These associations are found in children as well as
844 adults.^{171,172} In addition, the function of the immune system is shown to be impaired as a
845 result of adiposity, this includes diminished vaccination response and reduced T and B cell
846 counts, but increased leukocyte counts¹⁷³.

847 Adipokine abundance, naturally, increases in parallel with adiposity. Many adipokines
848 have immunomodulatory functions, not least leptin which stimulates lymphocyte activation
849 and subsequent cytokine production^{173,174}. Adiponectin, unlike leptin, is anti-inflammatory
850 and has a role in the inhibition of TNF- α as well as in the production of anti-inflammatory
851 cytokines such as IL-10¹⁷⁴. Leptin and adiponectin are the two main adipokines, there are
852 many more with similar immunomodulatory functions. Additionally, fatty acids, which increase
853 with adiposity, influence inflammation. Notably, fatty acids are associated with adiponectin
854 levels¹⁷⁵ and are thought to be involved in the production of TNF- α and IL-6¹⁷⁶.

855 As discussed previously, there is a strong link between adiposity and type 2 diabetes¹⁵³.
856 This association is thought to be somewhat via insulin resistance. In a large review of the area,
857 adipose tissue resident and infiltrating immune cells were shown to lead to the development
858 of insulin resistance¹⁷⁷. Studies highlighted the change in immune cell recruitment and
859 differentiation as a result of adiposity, particularly the increase in CD8 $^{+}$ T-cell recruitment and
860 the differentiation of macrophages from fixed states into mixed pro- and anti-inflammatory
861 states. These changes occurred in parallel with reductions in anti-inflammatory immune cells.
862 The function that these changes have on the development of insulin resistance is unclear¹⁷⁷,
863 however studies highlighted the increases in TNF- α and IL-6. These studies were carried out
864 in mice and are therefore not wholly translateable to humans.

865 Adiposity is strongly associated with CVD as discussed previously. This association
866 may be a result of inflammation and changes to immune cell recruitment and differentiation
867 similar to that observed for type 2 diabetes¹⁷⁸. Chronic inflammation develops as a result of
868 the parallel increase in adipose tissue, adipokine production, immune cell recruitment, and
869 cytokine production. There is evidence that this chronic inflammation results in endothelial
870 dysfunction which is a precursor to atherosclerosis. For example, leptin has been shown
871 to increase atheroma formation via cholesterol uptake by macrophages, while TNF- α , IL-6,
872 and IL-1 promote inflammation of blood vessels¹⁷⁹. In addition, CRP, which causes cell
873 death through the complement cascade¹⁸⁰, is an immune marker associated with adiposity¹⁸¹
874 and CVD¹⁸². CRP has been directly associated with the development of atherosclerosis¹⁸³.
875 Numerous other inflammatory markers also show links with CVD, including adiponectin¹⁸³.

876 Alongside chronic inflammation, type 2 diabetes, and CVD, a number of other immune
877 related conditions have been associated with adiposity. This includes the development of
878 gallstones¹⁸⁴ and pancreatitis¹⁸⁵. Adiposity may influence gallstone development through
879 diet¹⁸⁶ or through physically impeding detection of gallstones¹⁸⁷. Gallstones are themselves
880 associated with the development of pancreatitis. Additionally, adiposity may result in pancre-
881atitis through increased triglycerides and type 2 diabetes¹⁸⁵, both of which are associated
882 with adiposity.

883 Kidney

884 Adiposity is associated with numerous kidney related disorders¹⁸⁸, including risk factors
885 for the development of chronic kidney disease (CKD) such as hypertension and type 2
886 diabetes. These associations with BMI also include the development of proteinuria¹⁸⁹,
887 reduced estimated glomerular filtration rate¹⁹⁰⁻¹⁹², and end-stage renal failure^{193,194}. Similar
888 associations are found for WHR and WC independent of BMI^{190,195,196}. The association
889 between adiposity and CKD remains after adjustment for potential mediators such as type 2
890 diabetes and hypertension¹⁸⁸.

891 A possible reason for these associations is that hyperfiltration is needed to accommodate
892 the increased metabolic demand of adipose tissue, which over time places a permanent
893 stress on the kidneys¹⁸⁸. However, as a majority of individuals with obesity do not develop
894 CKD¹⁸⁸, the effects of adiposity on CKD development may be mediated by other factors such
895 as adiponectin¹⁹⁷ and leptin¹⁹⁸. The consequences of which include inflammation and insulin
896 resistance as described previously, as well as aberrant lipid metabolism which leads to a toxic
897 renal environment¹⁹⁹.

898 **Liver**

899 There are numerous types of liver disease; non-alcoholic fatty liver disease (NAFLD)
900 categorises a number of conditions related to liver disease that includes a build up of adipose
901 cells within the liver but does not involve alcohol. There is a strong association between
902 adiposity and NAFLD which is becoming more common as adiposity increases globally^{200,201}
903 – ancestry influences this association^{202–204}. NAFLD is also associated with other adiposity
904 associated disorders such as type 2 diabetes, hypertension, dyslipidemia, and insulin
905 resistance¹⁴⁶.

906 Steatosis, the abnormal retention of lipids within cells/organs, is the main feature of NAFLD.
907 Hepatic steatosis begins when fatty acid uptake and synthesis is greater than oxidation and
908 export; clinically hepatic steatosis is defined as a fatty acid content > 5% of total liver weight²⁰¹.
909 This increase in fatty acids has been linked with increased lipolysis as a result of adiposity as
910 well as an increase in *de novo* lipogenesis at the liver²⁰⁵. The increased fatty acid content
911 in the liver is associated with changes in glucose abundance, lipoprotein metabolism, and
912 inflammation, however whether this is a consequence of, a cause of, or concomitant with fatty
913 acid accumulation is unclear²⁰⁰.

914 **Neurological/behavioural**

915 There are a number of neurological consequences of adiposity which can broadly be
916 defined as structural, psychological and behavioral, and physiological²⁰⁶. There is some
917 conflicting evidence for an association between adiposity and mild cognitive impairment^{207,208}.
918 Mild cognitive impairment is a precursor to Alzheimer's disease which is associated with
919 adiposity^{209,210}. The adiposity Alzheimer's association includes elevated levels of β -amyloid
920 compared to normal weight individuals²¹¹. These associations between adiposity and mild
921 cognitive impairment and Alzheimer's disease may be a result of structural changes. Cerebral
922 atrophy is associated with both mild cognitive impairment and Alzheimer's disease²¹² and
923 there is evidence of reduced hippocampal volume with adiposity²¹³.

924 Structurally, evidence suggests that adiposity not only impacts brain volume
925 negatively^{214,215} but also increases the risk of abnormal neuronal activity²¹⁶, though
926 the latter may be a result of accelerated aging as a result of adiposity. There are also
927 associations between adiposity associated disorders such as type 2 diabetes with structural
928 changes in the brain²¹⁷. These structural changes are present in individuals without cognitive
929 decline²¹⁷ as well as children^{218,219}.

930 Animal models have shown that high fat diet induced adiposity is associated with cognitive
931 decline^{220,221}. This cognitive decline may result from inflammation within the hypothalamus,
932 with elevated levels of TNF- α present in high fat diet mice compared with controls²²². There is

933 evidence to suggest that inflammation in the brain resulting from adiposity goes on to impair
934 the control of satiety²²³, with leptin resistance shown to build²²⁴ – leptin binds to receptors in
935 the brain to make us feel *full*.

936 In regards to other conditions, there is some evidence for an association between adiposity
937 and migraines²²⁵ which includes links with adipokines²²⁶. Additionally, there is a strong
938 relationship between adiposity and depressive symptoms and depression²²⁷, specifically
939 there appears a *U* shaped distribution with underweight and obese individuals exhibiting
940 depression more often than their normal weight counterparts²²⁸. However, this association
941 appears reciprocal with depression also shown to increase the risk of developing obesity²²⁷.
942 In addition, the relationship between adiposity and depression/depressive symptoms is
943 complicated due to the stigmatization of adiposity²²⁹ and the detrimental effect adiposity has
944 on an individuals quality of life²³⁰.

945 **Respiratory**

946 Adiposity is associated with a number of respiratory disorders²³¹. One of the first respira-
947 tory consequences of adiposity is dyspnea and wheezing both at rest and during exercise²³².
948 This may be a result of increased oxygen demand from adipose tissue and the resulting
949 increased respiratory workload²³³. These associations may be further impacted by the asso-
950 ciation between adiposity and reduced respiratory muscle function, though this association
951 may result from a reduction in fat free mass in parallel to the increase in fat mass²³⁴. That
952 being said, not all individuals with overweight or obesity develop dyspnea or wheezing²³²,
953 potential confounding factors such as smoking may impact on the discussed relationships.

954 Alongside wheezing and dyspnea, asthma is also found to be more frequent among
955 individuals with overweight or obesity²³⁵. Asthma severity, as well as use of medication, and
956 hospital admissions are also increased as a result of adiposity²³⁶. The underlying mechanism
957 of the adiposity-asthma relationship is unclear and, increased asthma frequency has not
958 been related to over- or mis-diagnosis²³⁷. There is some evidence to suggest that sys-
959 temic inflammation as a result of adiposity is associated with glucocorticoid insensitivity in
960 asthma²³⁸.

961 Obstructive sleep apnoea is caused by a collapsing of the upper airway and results in
962 oxygen desaturation and poor sleep²³⁹. There is a strong association between adiposity
963 and obstructive sleep apnoea, with fat deposition around the neck a major factor due to the
964 resulting increased pressure on the airway²³⁹. This is reflected in neck circumference being
965 strongly associated with obstructive sleep apnoea²⁴⁰.

966 Chronic obstructive pulmonary disorder (COPD) is primarily a smoking related disorder
967 characterized by obstructions to airflow. There is some evidence that adiposity is more preva-
968 lent in individuals with early stage COPD. However, at later COPD stages this relationship

969 is reversed²⁴¹. It is likely that this association is confounded and the association between
970 adiposity and COPD is instead related to changes in lung function²⁴².

971 The driving force behind the associations between adiposity and many respiratory dis-
972 orders is adiposities effects on lung volume. Studies have shown that adiposity results in
973 reduced forced expiratory volume, forced vital capacity, functional residual capacity, and
974 expiratory reserve volume²⁴³⁻²⁴⁵. For example, reduced lung volume results in the upper
975 airway being under lower tension and making it more susceptible to collapse²⁴⁶. These
976 associations are predominantly found for abdominal adiposity^{231,243,247}.

977 **Reproductive**

978 There are a number of reproductive outcomes associated with adiposity many of which
979 are specific to women²⁴⁸. This also includes the developmental origins of health and disease
980 (DOHaD) which states that exposures during formative periods of development and growth
981 may have lasting consequences of health²⁴⁹. It should also be noted that the prevalence
982 of adiposity among mothers varies by ancestry²⁴⁸. Reproductive research is primarily on
983 women given it is easier to see the impact their decisions have on the offspring. However,
984 recent work has highlighted paternal effects on offspring health and proposes more should
985 be done in studying paternal effects going forward²⁵⁰.

986 Adiposity is negatively associated with fertility, both overall and in women who gain weight
987 in adolescence²⁵¹. There is also evidence to suggest that there is a dose response relationship
988 between BMI and infertility²⁵² as well as variations in fertility rates alongside variations in
989 adiposity among ancestries²⁵³. One of the leading causes of infertility is polycystic ovary
990 syndrome (PCOS) for which there is an increased risk with adiposity^{248,254}. The severity of
991 PCOS is exacerbated by the effects adiposity has on other factors such as insulin resistance
992 which in turn worsen PCOS symptoms²⁵⁵. There is also overlap between PCOS and metabolic
993 syndrome; many women who have PCOS also have metabolic syndrome²⁵⁶. This means
994 women with PCOS are at greater risk of developing other adiposity related conditions such
995 as type 2 diabetes²⁵⁷ and CVD²⁵⁸. Adiposity in men is also associated with infertility, with
996 associations found between overall and abdominal adiposity and sperm abnormalities such
997 as motility, morphology, and count²⁵⁹. In men, the metabolic syndrome is associated with
998 hypogonadism²⁶⁰ and erectile dysfunction²⁶¹.

999 The effect of adiposity on maternal and foetal outcomes includes increased risk of
1000 hypertension, pre-eclampsia, gestational diabetes, preterm birth, macrosomia, still birth,
1001 and miscarriage²⁶²⁻²⁶⁵. One way women are at increased risk of hypertensive disorders
1002 of pregnancy is via pre-pregnancy hypertension²⁶⁴. Hypertension is strongly linked with
1003 adiposity as discussed previously¹²⁹ and increases in maternal hypertension are associated
1004 with the rise in pre-pregnancy adiposity²⁶³. There is also an association between hypertension
1005 and pre-eclampsia during pregnancy and insulin resistance independent of current diabetes

1006 status. Additionally, hypertension and pre-eclampsia during pregnancy are strongly associated
1007 with the development of maternal type 2 diabetes later on. This association persists after
1008 accounting for gestational diabetes²⁶⁶.

1009 Type 2 diabetes is strongly associated with adiposity as discussed previously¹⁵³. Women
1010 entering pregnancy with pre-existing type 2 diabetes are at risk of birth complications associ-
1011 ated with gestational diabetes such as large for gestational age²⁴⁸. There is also evidence
1012 for an increased risk of gestational diabetes in women entering pregnancy with overweight
1013 or obesity absent pre-existing type 2 diabetes²⁶³. There is a dose response relationship be-
1014 tween adiposity and risk of gestational diabetes²⁶². This has long term health consequences,
1015 including subsequent development of type 2 diabetes²⁶⁶.

1016 Foetal outcomes are directly impacted by maternal exposures. This includes preterm
1017 complications, birth weight, congenital anomalies, and mortality²⁴⁸. Maternal adiposity is
1018 associated with preterm birth²⁶⁵ with increasing adiposity associated with increased risk
1019 of preterm birth²⁶². Birth weight is also strongly associated with maternal adiposity in a
1020 dose dependent manner^{262,267}. Extreme birth weight, macrosomia, and large for gestational
1021 age are much more prevalent for women with obesity^{262,265}. Low birth weight and small for
1022 gestational age are also associated with maternal adiposity^{268,269}. In a large umbrella review,
1023 birth weight was shown to be associated with later life adiposity and there was suggestive
1024 evidence for an association with a number of later life health outcomes including type 2
1025 diabetes and CVD²⁷⁰.

1026 Maternal adiposity is associated with a number of congenital anomalies including spinabi-
1027 fida and cleft lip and palate^{271,272}. Additionally, maternal adiposity is associated with neonatal
1028 and infant mortality, with increasing adiposity associated with increasing risk of mortality^{262,273}.
1029 A similar association is found for stillbirth²⁷³.

1030 **Underlying aetiology**

1031 One of the key points from the literature is the broad array of outcomes adiposity is
1032 associated with, or has an effect on. However, the underlying aetiology of these relationships
1033 is not always clear. In some instances, associations between adiposity and disease are
1034 explained by other factors. For example, associations with reduced quality of life is mostly
1035 explained by the presence of co-morbidities which increases the likelihood of poor outcomes,
1036 including stigmatisation^{274,275}. Similarly, the relationship with many sleep complications is
1037 likely a result of chronic pulmonary diseases²⁷⁴⁻²⁷⁶.

1038 Type 2 diabetes development is likely to follow a process of impaired glucose clearance as a
1039 result of adiposity, which leads to increased insulin resistance. There are likely wider metabolic
1040 changes that influence this process which are also a consequence of adiposity^{88,274,275}.
1041 Respiratory diseases are likely a result of reductions in forced expiratory volume, forced vital

1042 capacity, lung and residual capacity, and expiratory reserve. Each of these is a consequence
1043 of weakened muscles and reduced compliance of the chest which can be caused by the
1044 physical burden of adiposity around the chest and lungs^{275,276}. With respiratory disease there
1045 is also the prospect of confounding as a result of smoking status, which is associated with
1046 adiposity²⁷⁷.

1047 In the case of CVD, hypertension may be related to changes in sympathetic activity,
1048 blood flow and viscosity, and dietary intake as a result of adiposity^{274,275,278}. Some of these
1049 changes might similarly be a result of metabolic, inflammatory, and/or hormonal changes.
1050 Both dyslipidemia and reductions in HDL result from adiposity and these changes may be
1051 important in the development of heart disease. Adiposity induced adaptations, including
1052 structural, functional, and hormonal changes may influence CVD development by altering
1053 homeostasis and inducing, for example, inflammatory responses. Additionally, co-morbidities
1054 which are also associated with adiposity (e.g., type 2 diabetes) may have concomitant effects
1055 on CVD development¹²⁹. There is also evidence of altered myocardial metabolism as a result
1056 of adiposity, this includes increased fatty acid oxidation and decreased glucose oxidation²⁷⁹.

1057 Unlike diabetes and respiratory diseases, many other diseases have a less well under-
1058 stood process of development as a result of adiposity. For cancer, hypotheses for development
1059 differs based on the type of cancer. Metabolic, inflammatory, and hormonal changes as a
1060 result of adiposity are proposed as leading to the development of a number of different
1061 cancers^{88,108,274,275}. The protective effect of adiposity on pre-menopausal breast cancer sta-
1062 tus is thought to be a result of oestrogen production. Oestrogen is upregulated in individuals
1063 with overweight or obesity; in pre-menopausal women oestrogen is primarily produced at the
1064 ovaries, while in post-menopausal women oestrogen is produced via androgen conversion
1065 and aromatase. Aromatase can be produced by adipose tissue and is upregulated in breast
1066 tissue of women with adiposity²⁸⁰⁻²⁸². Additionally, the association between adiposity and
1067 breast cancer is strongest for ER+ cases compared with ER- cases^{116,283}. Evidence also
1068 shows that hormone therapy reduces the risk of post-menopausal breast cancer²⁸⁴ and
1069 that adiposity associated factors such as leptin²⁸¹ may also be involved in breast cancer
1070 development.

1071 The location of fat deposits may also be important; deposition of adipose tissue around
1072 the heart may result in inflammation of the myocardium, but this might also be subsequent
1073 to dyslipidemia and reduction in HDL^{88,274,275,285}. Similarly, distribution of fat mass around
1074 the neck has been associated with sleep complications²⁷⁴⁻²⁷⁶. Osteoarthritis, though likely
1075 a result of the physical burden of adiposity, may also be a product of changes to cartilage
1076 and bone metabolism^{274,275}. Similar metabolic changes may play a role in a number of
1077 other diseases. An increased risk of gallstones is associated with increased cholesterol²⁷⁴,
1078 and increased salt intake has been suggested as a potential link between adiposity and
1079 stroke^{88,275}.

1080 **1.6 Adiposity: summary**

1081 The body of work discussed here highlights the wide array of effects that adiposity has
1082 on health and disease. Many proposed mechanisms of disease development involve the
1083 physical burden of fat mass and/or changes to different pathways, particularly the roles of
1084 adipokines and metabolic changes. A key point that many studies surmised was that losing
1085 weight would be beneficial in reducing the burden of adiposity.

1086 There is evidence to show that weight loss interventions reduce the risk of mortality^{286,287}
1087 as well as the risk of developing conditions such as CVD and cancer²⁸⁸⁻²⁹². There are however
1088 many barriers to behavioral weight loss interventions²⁹³ and their long term effectiveness is
1089 unclear given many studies do not follow-up for longer than two years. As such, bariatric
1090 surgery²⁹⁴ is thought to be an effective long term intervention^{295,296}. This type of surgery has
1091 a number of disadvantages²⁹⁵, not least the fact it is invasive. Additionally, and true for all
1092 interventions, identifying individuals who will benefit most from a specific intervention strategy
1093 is challenging.

1094 Identifying mechanisms by which adiposity exerts its effect on the development of diseases
1095 may help to prioritise interventions. It may also help to identify targets for interventions.
1096 Studies focusing on targeting adiposity driven pathways have been successful. In humans,
1097 pharmacological interventions have been effective in aiding weight loss²⁹⁷, while acting on
1098 the leptin pathway reduces adiposity in individuals with obesity²⁹⁸ and reduces the effects
1099 of lipodystrophy²⁹⁸. In mice, Withaferin A, a leptin sensitizing agent, has shown promise in
1100 aiding weight loss and improving glucose metabolism independent of the leptin pathway²⁹⁹.
1101 Similarly, agents that target the adiponectin pathway in mice have been effective in improving
1102 insulin resistance and thus diabetic outcomes³⁰⁰.

1103 The studies discussed in the previous section focused primarily on observational method-
1104 ologies using varying sized samples. Observational analyses are subject to a number of
1105 limitations which must be considered when investigating mechanisms of disease development.
1106 This includes confounding, reverse causation, and various forms of bias^{301,302}, which, even
1107 with careful study design, can not all wholly be addressed. Importantly, observational studies
1108 struggle to identify causality.

1109 Studies aim to assess the association between an exposure and an outcome. Measured or
1110 unmeasured factors that are associated with the exposure and/or the outcome, confounders,
1111 can bias results. In order to account for confounding bias, studies can include measured
1112 confounders within their models or stratify analyses based on the confounder, for example
1113 sex. If confounders are poorly measured, or not measured at all, this can introduce residual
1114 confounding bias. If a measured factor is a mediator of the exposure outcome effect instead
1115 of a confounder this can introduce collider bias if subsequent confounders of the association
1116 are not properly adjusted for^{303,304}. An additional limitation of observational studies is the
1117 difficulty in obtaining the causal direction of effect. A central tenet of causality is temporality,

1118 the exposure must come before the outcome. When the temporal sequence is unknown the
1119 outcome may come before the exposure. It is difficult to obtain this information in observational
1120 settings³⁰¹.

1121 Observational studies can be ranked based on their design and the evidence they pro-
1122 vide for causality. Ecological studies are ranked at the bottom given the data they utilise
1123 is aggregate and is particularly subject to the ecological fallacy³⁰¹. Cross-sectional and
1124 longitudinal studies are ranked above ecological studies, with case control and cohort studies
1125 ranked above these. Randomised control trials (RCT) sit at the top of the hierarchy as they
1126 experimentally control for, and test, the effects of an exposure on an outcome whereas the
1127 other designs only look for the presence of an outcome. Many of the studies discussed
1128 in this section are not RCTs and are therefore unable to provide direct causal evidence.
1129 Understanding what the causal literature says about the effects of adiposity on diseases will
1130 be important for this thesis. As will the use of causal methodology, discussed in detail in
1131 Section 1.8.

1132 1.7 Metabolites

1133 Many of the diseases discussed have hypothesised development processes involving
1134 metabolic, inflammatory and hormonal changes. As a complex signalling organ, with both
1135 local and systemic effects, adipose tissue is likely to influence all three of these processes at
1136 both local and systemic levels. These pathways and processes can be targeted in order to
1137 reduce the burden of adiposity^{305–307}. It is not within the scope of this thesis to investigate
1138 all three, but recent advances in measurement methodologies and the availability of large,
1139 and deeply phenotyped population based studies may now provide the data necessary to
1140 investigate metabolic effects.

1141 The metabolome, the total abundance of small-molecules, is a reflection of genetic
1142 and non-genetic factors and sits between the proteome and the phenotype^{308–311}. The
1143 metabolome can be separated into endogenous (internally produced) and exogenous (ex-
1144 ternally produced) metabolites, whereby the majority of metabolites are the result of cellular
1145 processes, with multiple functions including as energy, signalling, transportation, and struc-
1146 tural components. Metabolic effects can be far reaching and also include post-translational
1147 modifications^{311,312}. During homeostasis metabolic effects are tightly controlled, however the
1148 many functions they play mean that imbalances can be detrimental^{308,311,312}.

1149 Measurement of individual metabolites, at scale, is achieved predominantly through mass
1150 spectrometry (MS) and nuclear magnetic resonance (NMR). Both MS and NMR have differing
1151 limitations with full coverage of the metabolome not achieved by either. Complimentary usage
1152 of the two methods is desirable³¹³; however, as MS is destructive and both methods are
1153 costly this is not always possible. Many population-based studies have metabolomics data

1154 from only one measurement method limiting the number of metabolites available for analysis.

1155 The number, and type, of metabolites identified by MS and NMR methods is depen-
1156 dent upon whether a targeted, semi-targeted, or un-targeted approach is taken. Targeted
1157 metabolomics analysis uses an internal standard to characterize individual metabolites^{314,315}
1158 whereas un-targeted metabolomics analysis measures all metabolites within a specified
1159 range^{315,316}. Semi-targeted approaches use internal standards to quantify groups of metabo-
1160 lites with similar chemical structure³¹⁵. Targeted studies are able to identify a handful of
1161 metabolites whereas semi-targeted and un-targeted can identify hundreds to thousands.
1162 As targeted and semi-targeted methods use internal standards absolute quantification of
1163 metabolite abundance is possible. In un-targeted methods only relative quantification, the
1164 peak area of each metabolite in comparison to other samples, is possible³¹⁵.

1165 The availability of well powered population studies with metabolomics data from targeted,
1166 semi-targeted, and un-targeted methods as well as matched genome-wide data has enabled
1167 a growth in metabolite GWAS^{310,313}. These studies have revealed large variations in the heri-
1168 tability of metabolites and numerous loci associated with their abundances³¹⁷⁻³²¹. The public
1169 availability of these GWAS provides a unique opportunity to perform genetic epidemiology
1170 studies which can compliment the existing literature from observational association studies.

1171 Metabolites reflect the current condition and activity of an organism and vary in abundance
1172 depending on the state of the individual, this includes age and sex³²². This is particularly
1173 evident in fasted and non-fasted measurements³²³⁻³²⁵ but also in case control studies such
1174 as those focusing on diabetes³²⁶ and cancer^{312,327}. Differences are also apparent when
1175 studying complex traits such as BMI^{158,328} as well as many more³²⁹.

1176 These studies provide an overall assessment of the changes metabolites undergo as a
1177 result of different conditions but the relationship is not clear. Whether metabolites change as
1178 a result of a condition or lead to its development is an important question with potential clinical
1179 importance. Mutable, both from a genetic and non-genetic perspective, the metabolome
1180 can, with caution, be used to investigate the development of diseases^{309,310,313}. Particular
1181 consideration should be given to the metabolomics approach (targeted, semi-targeted, un-
1182 targeted) and whether individuals were fasted. Consideration should also be given to the fact
1183 that metabolomics analysis provides a snapshot of an individuals current state. Though few
1184 studies have investigated metabolomic stability in large populations, variability in metabolite
1185 measures is apparent^{323,330,331}.

1186 A key aspect of future work investigating relationships between metabolites and diseases
1187 are the interactions metabolites have with one another. The metabolome is a complex
1188 system involving feedback and feed-forward loops, this complexity means many metabo-
1189 lites are intercorrelated³³², have high genetic correlation³²⁰, and share a common genetic
1190 architecture³¹⁷⁻³²¹. As such, a perturbation in a single metabolite rarely occurs in isolation.
1191 Investigating metabolites as grouped entities that represent the underlying complexity, rather
1192 than individual metabolites, may help to elucidate relationships with disease.

1193 1.8 Mendelian randomization

1194 Studies investigating the associations between adiposity and metabolites and metabolites
1195 and disease are important and, when conducted in optimal conditions, provide information
1196 on the potential causes and consequences of altered metabolic states. Even with optimal
1197 conditions observational studies hold a number of limitations that can not easily be over-
1198 come. These limitations, such as confounding and reverse causation, can lead to biased
1199 results^{145,333–335}. Simply put, though a study may identify an association between two traits
1200 does not mean that one causes the other; they may be correlated because of shared causes
1201 for instance.

1202 In observational epidemiology, ideally we want to compare individuals based on the
1203 exposure and so attempt to control the experiment by accounting for confounders. In this
1204 regard we attempt to replicate an RCT, which is the gold standard for testing causality.
1205 However, the large costs and time required to develop, implement, and analyse results limits
1206 their use. More importantly, randomizing individuals to conditions known to be associated with
1207 harmful outcomes is ethically wrong. An alternative approach is to utilise the large amounts of
1208 data that are publicly available or that can be accessed through institutions. Causal inference
1209 methodologies have been established to exploit the availability of these data sets.

1210 Mendelian randomization (MR), described^{336–338} and reviewed^{339,340} elsewhere, and
1211 accompanied by a **dictionary** of terms³⁴¹, is a statistical methodology that uses genetic
1212 variants as IVs to investigate the causal relationship between an exposure and outcome^{336,337}.
1213 The reassessment of many observational associations has provided strong evidence for the
1214 relationships between risk factors and diseases, but has also highlighted the biases and
1215 limitations of observational research^{145,334,335}.

1216 Briefly, individuals inherit alleles largely at random from their mother and father. Across
1217 a large population this leads to the even distribution of confounders between the effect and
1218 non-effect alleles. As such, individuals differ because of the expressed allele rather than
1219 their environmental circumstances. This random allocation of genetic variants, which may
1220 ultimately be related to a health outcome, is analogous to an RCT where genotype groups
1221 act as the intervention and non-intervention arms of the trial.

1222 There are two main approaches to MR: one-sample and two-sample. In a one-sample MR,
1223 data on the exposure and data on the outcome are obtained from the same population sample.
1224 One-sample MR requires individual level data, however instruments and weights can be
1225 obtained externally. In a two-sample MR, summary level data is used³⁴². Summary statistics
1226 for the exposure instruments are obtained from one population sample and those same
1227 instruments are obtained from a GWAS of the outcome conducted in a second population
1228 sample. There are a number of considerations with one- and two-sample MR such as
1229 overfitting, weak instrument bias, and the underlying population differing across the two
1230 samples in two-sample.

1231 Inference derived from MR analyses relies up-on three assumptions (Figure 1.3): (i) no
1232 the instrumental variable (Z) is robustly associated with the exposure (X), (ii) there is no
1233 independent association of the instrumental variable with the outcome (Y) other than through
1234 the exposure, (iii) the instrumental variable is independent of measured or un-measured
1235 confounders (U).

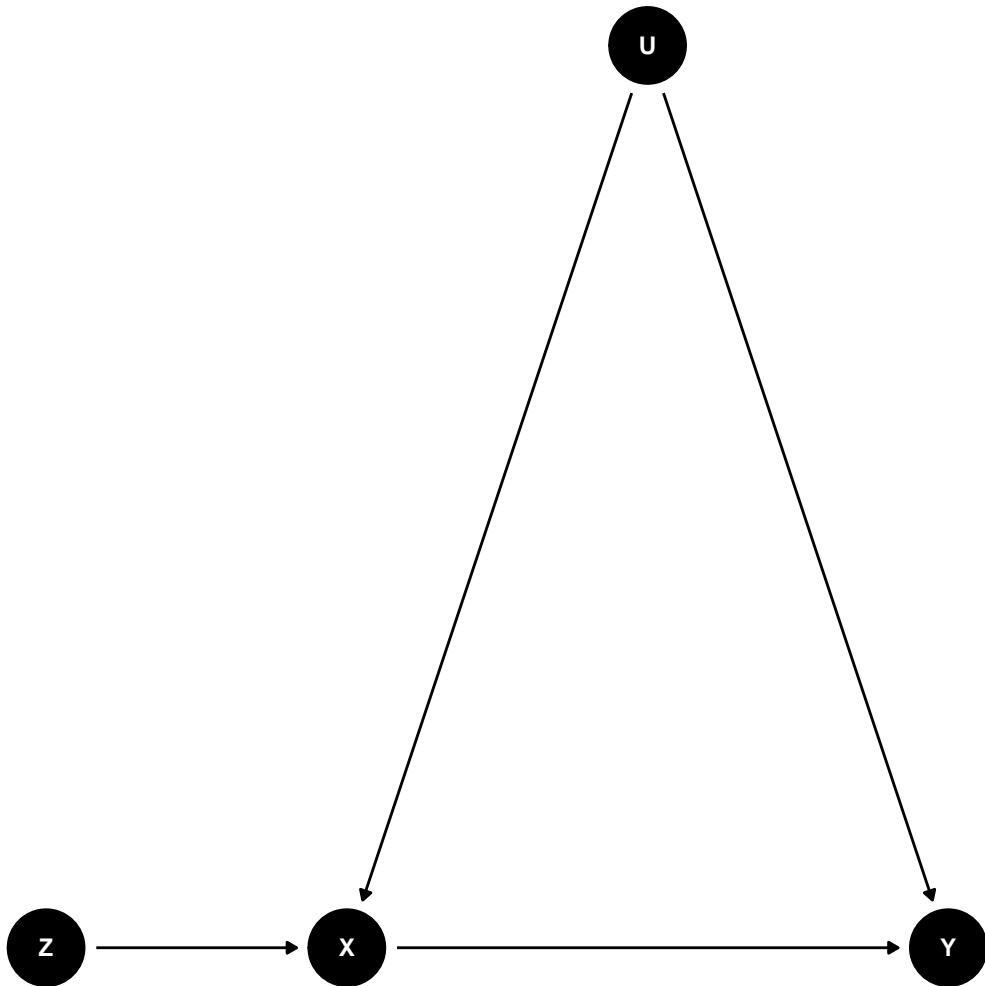


Figure 1.3: **Directed acyclic graph of the Mendelian randomization principle.** Z = instrumental variable; X = exposure; Y = outcome; U = confounders.

1236 Additional assumptions based on homogeneity, monotonicity, and effect modification
1237 are also present. The homogeneity assumption assumes the association between the

1238 instrumental variable (IV) and exposure or the effect of the exposure on the outcome is
1239 homogeneous. That is, the association or the effect is the same for all individuals in the
1240 population. Monotonicity can be deterministic or stochastic. Deterministic monotonicity
1241 assumes that the effect of the IV is consistent in all individuals of the population. That is,
1242 the effect of the IV does not increase the exposure in one group and decrease it in another.
1243 Stochastic monotonicity assumes deterministic monotonicity conditional on confounders.

1244 Based up-on Mendel's laws of inheritance, MR relies on the assumption that genetic
1245 variants are unlikely to be associated with one another (outside of linkage disequilibrium (LD))
1246 or with environmental factors. Deviation from which would mean an uneven distribution of
1247 alleles across a population. Consideration in MR analyses should therefore also be given to
1248 dynamic effects, population structure, and assortative mating. Within family MR can be used
1249 to obtain the true causal effect in these situations^{343,344}.

1250 Dynamic effects, a form of confounding, are a consequence of traits transmitted across
1251 generations which then influence the causal effect estimate^{344,345}. That is, the parental
1252 genotype directly effects the offspring phenotype. For example, the effect of BMI on CVD may
1253 be biased by the IVs for BMI being correlated across parent and offspring and the effect of
1254 maternal BMI on offspring development, which has an effect on future CVD. In this instance,
1255 the second MR assumption would be violated. Within family studies are proposed, and
1256 simulations have shown, to overcome some of the consequences of dynamic effects^{344,345}.

1257 Population structure is a result of allele frequencies differing across geographic regions.
1258 This would violate the assumption that IVs are independent of confounding factors. In MR
1259 analyses it is assumed that latent structure is accounted for in the GWAS in which the IVs are
1260 discovered³⁴⁶. As the sample sizes of GWAS has increased the potential for subtle effects of
1261 population structure have been observed^{346,347}.

1262 Assortative mating is the principle by which partners select one another based on a
1263 particular phenotype. This is either cross-trait (one trait selecting for another trait) or single-
1264 trait (one trait selecting for the same trait). MR results can be biased by both types of
1265 assortative mating, even when the phenotypes of interest are not those which influenced the
1266 mating³⁴³.

1267 Canalization, whereby what would otherwise be developmentally deleterious genetic
1268 effects are nullified by compensatory mechanisms, is broadly equivalent to non-adherence
1269 in an RCT. Any effects of canalization would attenuate effect sizes³³⁷, however there are
1270 currently no methods to detect its presence in an MR context. The effects of canalization
1271 are unlikely to be present in MR studies which utilise maternal genotypes for environmental
1272 exposures of the offspring such as during gestation³⁴⁸. For complex traits it is possible that
1273 canalization occurs at the level of the system rather than at the gene level³⁴⁹. As such, any
1274 outcome of a genetic mutation in regards to its role in the canal would likely be unpredictable.

1275 In both one-sample and two-sample MR, IVs are often obtained from external GWAS.

1276 Increasingly, these are large and well powered GWAS able to identify ever increasing numbers
1277 of SNPs associated with complex traits such as with BMI^{33,36,37,41,42}. As power has increased,
1278 the ability to detect SNPs with smaller effects which explain ever smaller proportions of
1279 variance in BMI has increased⁴¹. This holds potential considerations in regards to population
1280 structure and the effects of an omnigenic model³⁵⁰. As discussed, population structure was
1281 thought to have been an issue in smaller studies and could be accounted for by adjustment.
1282 However, well powered studies have shown both latent structure^{346,347} and an inability to
1283 perform adequate adjustment³⁵¹. This has potential implications, not only for the effect sizes
1284 of associated SNPs but also for the identification of SNPs associated with the trait³⁵¹. For
1285 example, a poorly or un-adjusted GWAS could identify SNPs associated with population
1286 differences rather than the trait of interest.

1287 In an omnigenic model, variance in a trait of interest is not solely a result of directly
1288 related genes (core-genes). Rather, all genes expressed in relevant cell types have an effect,
1289 however small, on the trait of interest³⁵⁰. These peripheral-genes, which have no obvious
1290 direct link to the trait of interest, are mostly in non-coding regions with regulatory functions⁵⁶.
1291 Given that variants associated with complex traits are dispersed widely across the genome⁵⁶
1292 and that assigning a link between any particular SNP and an individual gene is difficult⁵⁷,
1293 variants associated with complex traits likely implicate many genes with the trait. Because
1294 many of these will be peripheral-genes they will ultimately have functions on other traits,
1295 which in an MR context may include the outcome and thus violate the exclusion restriction
1296 assumption.

1297 Additional considerations include random measurement error (random measurement in
1298 the exposure will bias towards the null, and increase the standard error if in the outcome),
1299 Winners curse (whereby discovery studies identify larger effects than those in replication
1300 studies), collider bias (conditioning on a variable by adjustment, restriction, or sampling can
1301 induce an association between X and Y biasing the estimate both away and towards the
1302 null), non-overlapping samples (specific to two-sample MR, where the exposure and outcome
1303 data are obtained from samples with shared individuals), horizontal pleiotropy (the IV has an
1304 affect on the outcome independent of the exposure), and vertical pleiotropy (the IV does not
1305 have an effect on the exposure directly but on traits that have an effect on the exposure).

1306 Among the considerations and limitations of MR, population stratification, horizontal
1307 pleiotropy and canalization are the most challenging to account for. Though one can restrict
1308 analyses to homogeneous groups, use principal components, and perform within family
1309 studies to examine and mitigate the effects of population stratification, biases (e.g. sampling
1310 bias) may still remain. Additionally, methods for assessing potential horizontal pleiotropy exist
1311 but formal assessment of the exclusion restriction assumption is not possible. Accounting for
1312 canalization is much harder and, though being aware of the underlying biology can inform
1313 analyses, methods for assessment do not exist. Unlike the other considerations, vertical
1314 pleiotropy does not necessarily bias MR results rather it highlights potential intermediates.

1315 Both one-sample and two-sample MR can be extended to investigate intermediates that

1316 sit on the causal pathway. Mediation analysis in MR is discussed in detail elsewhere³⁵² and
1317 can be achieved using two-step³⁵³/network MR³³⁹ and multivariable MR³⁵⁴ (MVMR). Briefly,
1318 mediation analysis is interested in identifying the total effect, the direct effect, and the indirect
1319 effect; where all act in the same direction the proportion of the total effect explained by the
1320 mediator (proportion mediated) can be calculated³⁵⁵. The total effect is the effect of the
1321 exposure on the outcome through all mediated pathways, the direct effect is the effect of
1322 the exposure on the outcome through all mediated pathways that are not the pathway of
1323 interest, the indirect effect is the effect of the exposure on the outcome through the mediator
1324 of interest. These analyses are predicated on the following assumptions: (i) that there is
1325 a causal effect of the exposure on the outcome and mediator and of the mediator on the
1326 outcome; (ii) that there is no confounding between exposure, mediator, and outcome; (iii) that
1327 there are no intermediate confounders; (iv) that there is no interaction between the exposure
1328 and mediator³⁵⁵.

1329 In two-step MR (Figure 1.4), the indirect effect is calculated by multiplying the effect of the
1330 exposure on the intermediate and the effect of the intermediate on the outcome. The three
1331 core MR assumptions (and all previous considerations) must still be met and also extended:
1332 (i) the IV (Z & $Z2$) must be robustly associated with the exposure or intermediate only (X and
1333 M), (ii) the IV for the exposure (Z) must not be associated with the intermediate (M) or the
1334 outcome (Z) other than through the exposure (X), and the intermediate IV ($Z2$) must not be
1335 associated with the exposure, and only with the outcome (Y) through the intermediate, (iii) the
1336 IV for the exposure and intermediate must not be associated with measured or unmeasured
1337 confounders. No interaction between exposure and intermediate is also assumed. Two-step
1338 MR has been used^{356–358} and combined with MVMR³⁵⁹ to gain better insight into disease
1339 aetiology, but is not strictly speaking mediation analysis.

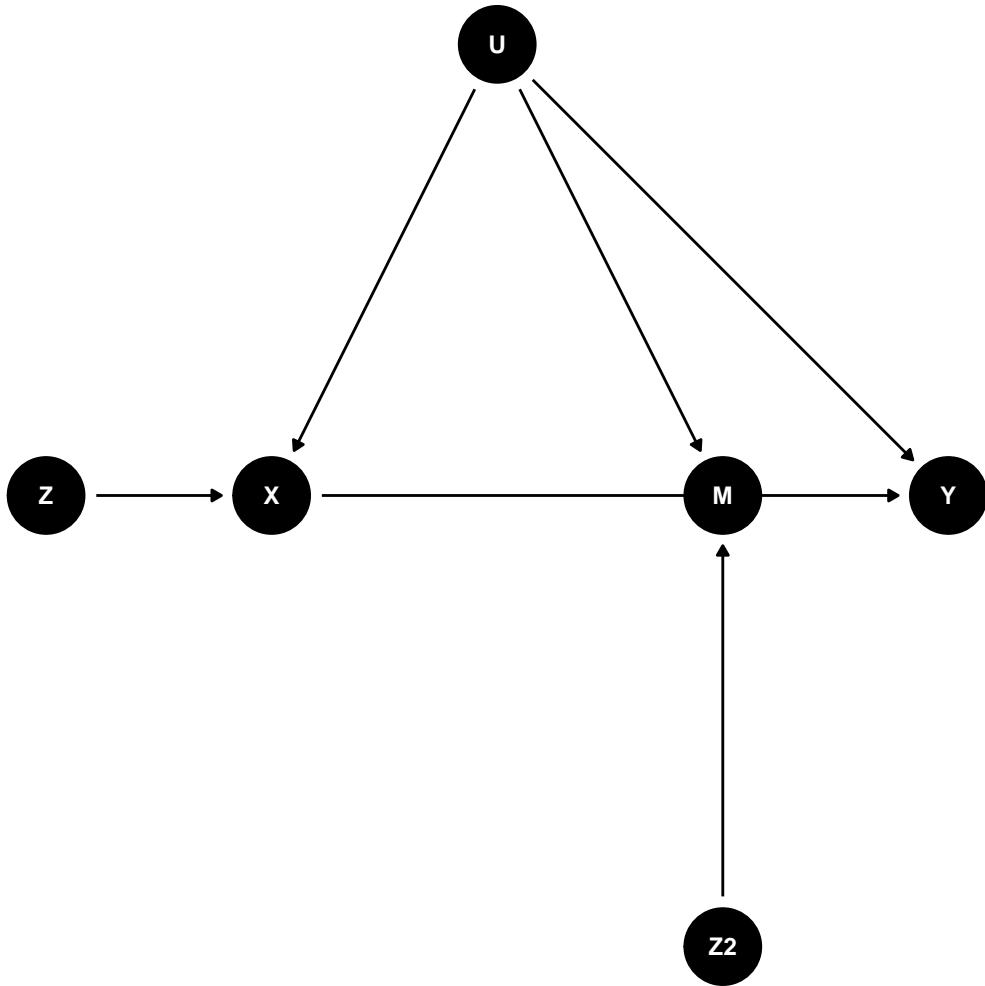


Figure 1.4: **Directed acyclic graph of the two-step Mendelian randomization principle.** Z = instrumental variable; X = exposure; M = intermediate; U = confounders; Z2 = instrumental variable for M; Y = outcome

1340 Multivariable MR is a form of mediation analysis which allows for the causal effects
 1341 of multiple exposures on an outcome to be estimated³⁵⁴ (Figure 1.5). The effect of each
 1342 exposure is estimated conditional on the other exposures and thus provides a direct estimate
 1343 of the effect. Figure 1.5 shows a simplified MVMR model with two exposures (X and X_2);
 1344 the bidirectional line between exposure one and exposure two does not make an assumption
 1345 about the exposure relationships. The indirect effect is estimated by subtraction of the direct
 1346 effect from the total effect. The total effect is calculated using univariable MR. As with two-step

¹³⁴⁷ MR, no interaction between exposure and intermediate is assumed. Though a new approach,
¹³⁴⁸ and still subject to the same assumptions as with two-step and univariable MR, MVMR has
¹³⁴⁹ shown promise in elucidating underlying aetiology of complex traits^{359–362}.

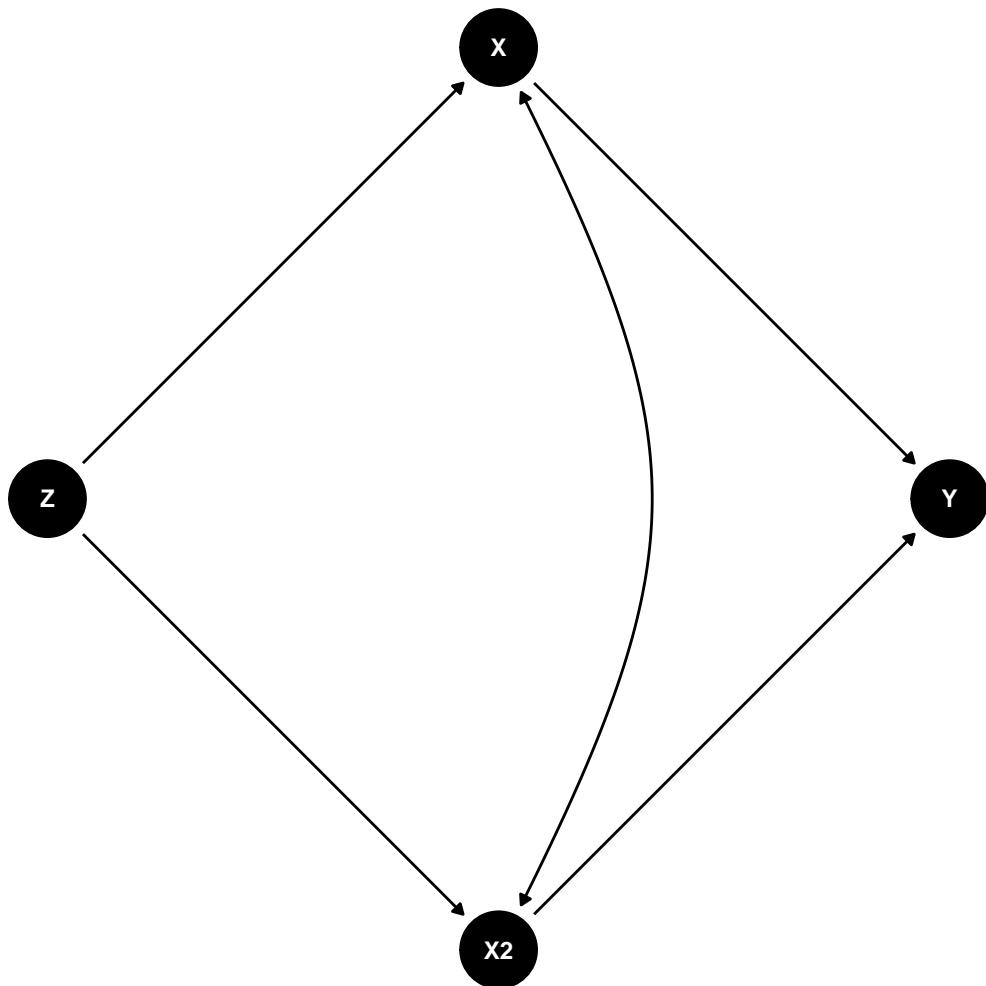


Figure 1.5: **Directed acyclic graph of the multivariable Mendelian randomization principle using two exposures.** Z = instrumental variables associated with one or more of the exposures; X = exposure; X2 = second exposure; Y = outcome

¹³⁵⁰ Though two-step MR was devised with epigenetic mechanisms in mind³⁵³ and MVMR has
¹³⁵¹ shown promise investigating metabolic intermediates³⁶¹, their application to large omic data
¹³⁵² sets is yet to be shown. An alternative approach, which instead of estimating mediated effects,

1353 is to look for overlapping signals[XXX: INSERT REF TO RCC PAPER]. In this regard, the
1354 effect of the exposure on the candidate intermediate and the effect of candidate intermediate
1355 on the outcome are ranked in terms of their effects. A candidate intermediate is considered if
1356 it ranks highly in both analyses.

1357 1.9 Previous work

1358 Metabolites have been a focus of research for some time, with many population based
1359 studies collecting data on traditional biomarkers (e.g., glucose and cholesterol). As tech-
1360 nologies such as MS and NMR have progressed, the number of metabolites studied at
1361 scale has increased. There is broad evidence for an association between adiposity and
1362 metabolites^{157,158,328,363–372}. These studies show associations between BMI and metabolites
1363 varies between sexes and over time^{157,367}; associations are stronger^{157,367} and appear earlier
1364 in men³⁶⁷. There is also evidence for associations with metabolites and BF³⁶⁷, WC^{366,371},
1365 WHR³⁶⁸, and visceral adipose tissue^{370,373}. These large scale metabolomics studies have
1366 also highlighted numerous associations between metabolites and other outcomes, from type
1367 2 diabetes³⁷⁴ and CHD³⁷⁵, to depression³⁷⁶, hypertension³⁷⁷, and more^{1,378–385}. Evidence
1368 also shows that metabolites can distinguish between cancers and adenomas¹⁸⁸.

1369 Few studies have investigated the explicit link between adiposity, metabolites, and disease
1370 outcomes^{158,386–388}. Evidence shows that BMI is linked to changes in specific metabolites and
1371 metabolite classes, and that several of these are subsequently associated with, or predictors
1372 of, later health outcomes such as insulin resistance³⁸⁷. However, these studies have three
1373 main limitations: (i) they involved a small number of individuals (N = 140-1,761), (ii) used
1374 a single measure of adiposity (BMI), and (iii) used methods that are unable to establish
1375 causality.

1376 Evidence from MR analyses shows numerous associations between adiposity and
1377 metabolites^{157,389–392}. Although many studies have focused on curated lists of metabolites,
1378 in larger analyses fewer associations are found than with observational analyses^{157,391,393}.
1379 MR studies have revealed this relationship to be a complex network of cause and effect, with
1380 metabolites being causes of, or effects of, adiposity³⁹³. Work by Hsu et al. (2020)³⁹³ found
1381 that associations with adiposity were mechanistically different based on whether a metabolite
1382 was identified as a cause or as an effect of adiposity. The causal relationship between
1383 metabolites and outcomes is much less well studied, likely due to instrumentation – few
1384 metabolites have been associated with robust and strong instruments. Associations between
1385 some metabolites have been identified for several outcomes, including type 2 diabetes^{391,394},
1386 fasting glucose^{391,394}, colorectal cancer³⁹², CHD³⁹⁵, and more^{396–399}.

1387 Few studies have looked causally at the adiposity-metabolites-outcome pathway as a
1388 whole. Xu et al. (2017)³⁵⁷ performed a two-step two-sample MR of BMI and HDL, LDL, and

1389 triacylglycerides on CHD, but did not find evidence for a pathway effect. Recently, Bull et
1390 al. (2020)³⁹² used MR to investigate the effects of BMI and WHR with 123 metabolites on
1391 colorectal cancer. Metabolites associated with BMI and/or WHR (N = 104) were used in
1392 univariable and MVMR to establish associations with colorectal cancer. BMI and WHR were
1393 associated with colorectal cancer risk, but many metabolite associations were inconsistent
1394 with this. For example, most lipids were decreased by an increase in BMI/WHR, but increased
1395 lipids increased CRC risk. Intermediate density lipoproteins and very large density lipoproteins
1396 however showed consistent directions, both increased by BMI/WHR and increasing
1397 CRC (distal colon cancer) risk. In MVMR analysis, associations for BMI and WHR were
1398 not attenuated after adjusting for intermediate density lipoproteins and very large density
1399 lipoproteins, suggesting these metabolites do not play an intermediate role in the relationship
1400 between adiposity and colorectal cancer.

1401 There is evidence of causal associations between adiposity and metabolites, between
1402 metabolites and diseases, and between adiposity and diseases. However, only the study
1403 by Bull et al³⁹² has investigated a combined adiposity-metabolite-disease pathway causally.
1404 Their analyses were likely subject to weak instrument bias however. Future analyses require
1405 more detailed metabolomic measures, with large sample sizes able to identify strong and
1406 robust instruments in order to fully address this question.

1407 1.10 Aims and objectives

1408 Adiposity is a global health concern. Many of the consequence of adiposity are known but
1409 the underlying aetiology is not well understood. Adipose tissue is a prolific signalling organ
1410 with systemic effects some of which are likely to affect the metabolome. Individual metabolites
1411 have been associated with many diseases but the complexity of the network makes these
1412 analyses difficult. MR studies provide an opportunity to investigate and disentangle the
1413 complex relationship between exposure, intermediate, and outcome. These studies must be
1414 approached carefully given the interrelatedness of metabolites. In light of these considerations
1415 this thesis aims to:

1416 ***Identify metabolites that sit on the causal pathway from adiposity to disease***

1417 In order to achieve this aim, this thesis will investigate the following objectives:

- 1418 1. Perform a systematic review and meta-analysis (Chapter 2) of all MR studies in which
1419 a measure of adiposity was used as an exposure. Identified diseases will guide later
1420 analyses (Chapter 6).
- 1421 2. Produce a visualisation tool that enables global overview of metabolite analyses (Chap-
1422 ter 3).

- 1423 3. Identify metabolites associated with adiposity using individual level data (Chapter 4).
- 1424 4. Identify metabolites associated with adiposity using MR (Chapter 5).
- 1425 5. Describe current and appropriate instrumentation in MR analyses (Chapter 5.2.1) using
1426 data from Chapter 2 and 5
- 1427 6. Identify associations between adiposity associated metabolites (Chapter 4 and 4) and
1428 outcomes (Chapter 2) using multivariable MR (Chapter 6)

¹⁴²⁹ **Chapter 2**

¹⁴³⁰ **Systematic review: What has the** ¹⁴³¹ **application of Mendelian** ¹⁴³² **randomization informed us about the** ¹⁴³³ **causal relevance of adiposity and** ¹⁴³⁴ **health outcomes?**

¹⁴³⁵ **Chapter Summary**

¹⁴³⁶ This chapter details a systematic review and meta-analysis of 173 studies investigating
¹⁴³⁷ the effects of adiposity on over 300 different outcomes. In Chapter 1, focus was given to
¹⁴³⁸ the underlying literature around adiposity, what adipose tissue is, why in excess this can
¹⁴³⁹ be detrimental to health, what observational studies have taught us about adiposity, and
¹⁴⁴⁰ potential mechanisms for adiposity-disease relationships. Here, the focus is on the causal
¹⁴⁴¹ effects of adiposity and what Mendelian randomization studies have informed us about
¹⁴⁴² adiposity-disease relationships. Results reveal the broad effect of adiposity on many health
¹⁴⁴³ outcomes, with meta-analyses highlighting a number of outcomes (e.g., endometrial cancer,
¹⁴⁴⁴ colorectal cancer, cardiovascular disease) for follow-up analysis in Chapter 6. This work was
¹⁴⁴⁵ pre-registered on [PROSPERO](#).

1446 2.1 Introduction

1447 Observational studies have indicated that adiposity is strongly associated with
1448 all-cause and cause specific mortality^{71,75–77,89–94}, and numerous risk factors and
1449 diseases^{93,95–129,136–140,142–144,146,147,151,153–163,166–176,188–196,200–219,223,224,231–248,250–256,258–267,270–273}.
1450 This includes the most common diseases, such as cardiovascular disease^{93,96,97,105,122–140,142–144}
1451 and many cancers^{93,95,108–121}, along with common risk factors such as hypertension^{129,157}.
1452 See Chapter 1 Section 1.5 for more detail.

1453 As discussed in Chapter 1 Section 1.8, observational studies hold a number of limitations
1454 that can not easily be overcome, e.g., confounding and reverse causation. These limitations
1455 can lead to biased results^{145,301,302,333–335} and, although an observational study may identify
1456 an association between two traits, this does not mean that one trait causes the other; they
1457 may be correlated because of shared causes, for instance. Furthermore, in observational
1458 studies, it is difficult to obtain the causal direction of effect as the temporal sequence is
1459 generally unknown³⁰¹. Ideally randomised controlled trials would be conducted to aid our
1460 understanding and identify causal effects, however these are costly, time consuming, and can
1461 be unethical given the assumption that adiposity is detrimental to health.

1462 Mendelian randomization (MR) analyses provide a method of obtaining causal effects
1463 outside of randomised controlled trials³³⁶. By using genetic variants, which are randomly as-
1464 signed and fixed at conception, the randomization and temporality analogous of a randomized
1465 controlled trial can be achieved (See Chapter 1 Section 1.8). There has been a rise in MR
1466 studies published in the years since it was first widely reported on in 2003³³⁶. Systematic
1467 reviews enable global overview of the literature and provide avenues for hypothesis generation.
1468 In combination with meta-analyses, systematic reviews can be used as a method for improved
1469 causal inference as pooled estimates can be more precise than estimates from individual
1470 studies⁴⁰⁰.

1471 The MR literature has not been systematically appraised in regards to the causal relevance
1472 of adiposity's effects on health outcomes. Here, a systematic review and meta-analysis are
1473 presented and will be used to inform downstream analyses within this thesis, namely selecting
1474 outcomes for which adiposity is relevant and to test associations with adiposity-related
1475 metabolic intermediates (Chapter 6).

1476 **2.2 Methods**

1477 **2.2.1 Data sources and search strategy**

1478 EMBASE and MEDLINE were searched from inception (EMBASE = 1974; MEDLINE
1479 = 1946) until February 18th 2019 using detailed search strategies including free text and
1480 controlled vocabulary terms ([GitHub](#)). The pre-print service bioRxiv was searched from
1481 inception (November 2013) until February 18th 2019. However, due to the limited search func-
1482 tionality and inability to include Boolean operators ('AND', 'OR', 'NOT') in bioRxiv searches,
1483 a restricted search strategy using four free text terms in four independent searches was
1484 used: 'Mendelian randomization', 'Mendelian randomisation', 'causal inference', and 'causal
1485 analysis'.

1486 The search strategy included synonyms for both adiposity and MR terms. For adiposity
1487 measures, this was to ensure searches returned all possible instances in which a measure of
1488 adiposity was used. For MR, synonyms were used as the term 'Mendelian randomization' has
1489 only been formalised recently and many early studies would have either been unaware they
1490 were performing an instrumental variable analysis or would have called the method something
1491 else. The search strategy is available on [GitHub](#).

1492 **2.2.2 Study selection**

1493 Articles returned through the searches of EMBASE and MEDLINE were downloaded
1494 as .ris files and imported into EndNote (version X8.2; Clarivate Analytics). De-duplication
1495 of articles identified in the EMBASE and MEDLINE searches was based on pagination
1496 identifiers described in detail elsewhere⁴⁰¹. Articles returned from bioRxiv were imported
1497 into Mendeley using the Mendeley Google Chrome extension and de-duplication performed
1498 using the Mendeley de-duplication function. After de-duplication, the titles and abstracts of
1499 all remaining articles from EMBASE, MEDLINE, and bioRxiv had their titles and abstracts
1500 screened by two independent reviewers (Matthew A Lee and Luke A McGuinness) using
1501 Rayyan⁴⁰². Each reviewer screened all articles and discrepancies at this stage were resolved
1502 through discussion between the two reviewers. Studies from EMBASE, MEDLINE, and bioRxiv
1503 meeting the pre-defined inclusion criteria (see below) were combined and, in instances where
1504 the bioRxiv study had been published and this was identified in either the EMBASE or
1505 MEDLINE search, the bioRxiv version of the study was excluded. The full texts of the
1506 combined study dataset had their full text screened by the two reviewers.

1507 For title and abstract screening and for full text screening, articles must have met the
1508 following pre-defined inclusion criteria:

- 1509 1. Be written in English
- 1510 2. Be available in full text (or in the case of conference abstracts, the authors must be
- 1511 contactable to obtain the relevant data)
- 1512 3. Be published in a peer-reviewed journal or bioRxiv
- 1513 4. Use MR methodology to investigate the causal effect of adiposity on any outcome
 - 1514 a. Adiposity was considered to be any measure which aimed to assess the amount
 - 1515 of adipose tissue an individual possessed
 - 1516 b. If a study focused on adiposity alongside other exposures, the effect of each
 - 1517 adiposity measure will be reported separately, if available, and report the joint
 - 1518 effect with these other exposures, if not available.
 - 1519 c. articles in which an MR approach is used but not explicitly called 'Mendelian
 - 1520 randomization' will be included. More specifically, any study in which genetic
 - 1521 variants are used as instrumental variables or the direct association between a
 - 1522 genetic variant and outcome is employed will be eligible, provided it meets the
 - 1523 other inclusion criteria.

1524 2.2.3 Data extraction

1525 In the first instance, data extraction was performed by nine reviewers (See ??), with
1526 articles split evenly between them, using a data extracton form ([GitHub](#)) and data extraction
1527 manual ([GitHub](#)). Once all articles had been reviewed, two reviewers (See ??) extracted data
1528 on all articles they did not review in the first instance. The same two reviewers then checked
1529 all extracted data for discrepancies which were resolved through a third review of individual
1530 articles.

1531 Articles included in data extraction may contain more than one relevant MR analysis.
1532 As such, study/studies refers to the MR analysis/analyses within an article. The following
1533 data were extracted from each articles studies: exposure(s), outcome(s), study design and
1534 sample characteristics, genetic variant and instrumental variable selection, MR methodology,
1535 sensitivity analysis, and causal estimates. Where relevant data was not reported by the article,
1536 "Not discussed" was entered into the data extraction form.

1537 Once data extraction was completed, three additional columns were added to summarise
1538 the type of outcome being studied: column 1 ("outcome") was used as a general categorisation
1539 of all outcomes across articles (e.g., the outcome "ER- breast cancer" would have the value
1540 "breast cancer"); column 2 ("outcome info") reported the outcome-specific information that
1541 distinguished outcomes within categories defined in column 1 (e.g., column 2 would contain
1542 the value "ER-" for the same breast cancer example); and column 3 ("outcome group")
1543 categorised outcomes more generally than values defined in column 1 (e.g., the breast cancer
1544 example would be categorised as "cancer"). Outcome categories were assigned based on
1545 prior biological knowledge and aimed to collapse the large number of outcomes. This could
1546 be achieved differently for some outcomes, for example smoking could go in a *respiratory*

1547 category or a *behavioural* category. Where there were few outcomes to make a category, they
1548 were grouped into an *other* category. This will include outcomes such as mortality, disease
1549 counts, epigenetic marker etc.

1550 2.2.4 Quality assessment

1551 There is currently no risk of bias tool to assess the quality of MR articles. Because of this,
1552 some reviews have not reported on the quality of MR studies^{403,404}. However, more recent
1553 studies have begun to investigate quality, either by using pre-existing risk of bias tools specific
1554 to other areas or by creating their own assessment tool^{405–416}. Some of these tools have
1555 been influenced by the recent publication of MR reporting guidelines⁴¹⁷ (STROBE-MR). The
1556 STROBE-MR guidelines allow readers to evaluate the quality of the presented evidence.

1557 For this systematic review, the tool used by Mamluk et al (2020)⁴⁰⁶ was adapted and used
1558 for quality assessment of studies included in the meta-analyses. Study quality was assessed
1559 on a 3-point scale (low = 3, medium = 2, high = 1; Appendix Table A.2) across 12 questions.
1560 These 12 questions included the five used by Mamluk et al., (2020)⁴⁰⁶. One of these five
1561 questions, relating to bias due to selection of participants, was split into two questions for
1562 exposures and outcomes to accommodate two-sample MR analyses. In addition, questions
1563 for instrumental variable (IV) association, sample overlap, whether the study performed
1564 sensitivity analyses and whether these were biased, descriptive data, data availability (data
1565 missingness), and statistical parameters were included. Given no formal risk of bias tool
1566 exists, quality assessment here was not used as a prerequisite for inclusion/exclusion in the
1567 meta-analyses. Rather, it was used to supplement the meta-analyses and aid interpretation.

1568 2.2.5 Meta-analysis

1569 To identify studies which could be meta-analysed, a set of rules were used (Figure 2.1).
1570 These rules ensured that the exposure and outcome were consistent across studies, but also
1571 that there was no population overlap between the outcomes of different studies or between
1572 the outcomes and exposure of different studies. Sample overlap can induce bias in MR
1573 studies⁴¹⁸. Where there was overlap between the outcome of one study and the outcome
1574 of another study, or where there was overlap between the exposure of one study and the
1575 outcome of another study, the study with the larger sample size was retained. Excluding
1576 studies with overlapping outcomes or overlapping exposures and outcomes would involve
1577 including non-independent data and result in overly precise estimates⁴¹⁸. Finally, studies were
1578 excluded based on whether the MR method was comparable and then on whether the units
1579 were compatible with one another (e.g., where both studies reported a standard deviation
1580 increase in BMI). Studies which had overlapping exposure populations were included as the
1581 risk of bias is low⁴¹⁸. For completeness, studies were not excluded based on the quality

1582 assessment score, but are discussed later in this chapter when interpreting the meta-analysis
1583 findings.

1584 Meta-analysis was performed using the `meta`⁴¹⁹ package in R and the function `metagen()`.
1585 In an inverse variance weighted fixed-effects model, a weighted average is calculated as:

$$\text{weighted average} = \frac{\sum Y_i(1/SE_i^2)}{\sum(1/SE_i^2)} \quad (2.1)$$

1586 Where, Y_i is the intervention effect estimates in the i^{th} study, SE_i is the standard error of that
1587 estimate, and the summation (\sum) is across all studies. The assumption here is that all effect
1588 estimates estimate the same effect; in MR analyses, we assume that studies using the same
1589 exposure and outcome will be estimating the same effect, but that the exposure and outcome
1590 is subtly different among different populations given instrumentation and measurement error
1591 and therefore consider these to be related effects. In a random-effects model the assumption
1592 is that the studies estimate related effects^{420,421}. In a random effects model, SE_i is adjusted to
1593 incorporate heterogeneity among study effects, τ^2 . In this, a random-effects model will weight
1594 smaller studies more than a fixed-effects model would, as they provide more information on
1595 the distribution of effects as opposed to more information on the overall effect. This does
1596 not mean that random-effects models account for heterogeneity; random- and fixed-effects
1597 models will give identical results when there is no heterogeneity.

1598 Following this and considerations in the [Cochrane handbook](#), an inverse variance weighted
1599 random-effects model using estimates and standard errors was used. Where standard errors
1600 and effect estimates were not available for a study (e.g., confidence intervals and odds
1601 ratios were available), these were back calculated manually. For binary outcomes, the
1602 relevant summary method was used for odds ratios, risk ratios, and hazard ratios etc. For
1603 continuous outcomes, the mean difference was used for the underlying summary method.
1604 For completeness, and given this is a hypothesis-generating process, a multiple testing
1605 threshold was not used. For both binary and continuous outcomes, the Hartung and Knapp
1606 method to adjust confidence intervals to reflect uncertainty in the estimation of between-study
1607 heterogeneity^{422,423}, which is recommended for random effects models^{424,425}, was used
1608 where ≥ 5 studies were included in the meta-analysis⁴²⁴. Between study variance was
1609 estimated for all meta-analyses using the Paule-Mandel estimator⁴²⁶, for which simulation
1610 studies have shown good performance compared to other estimators⁴²⁷.

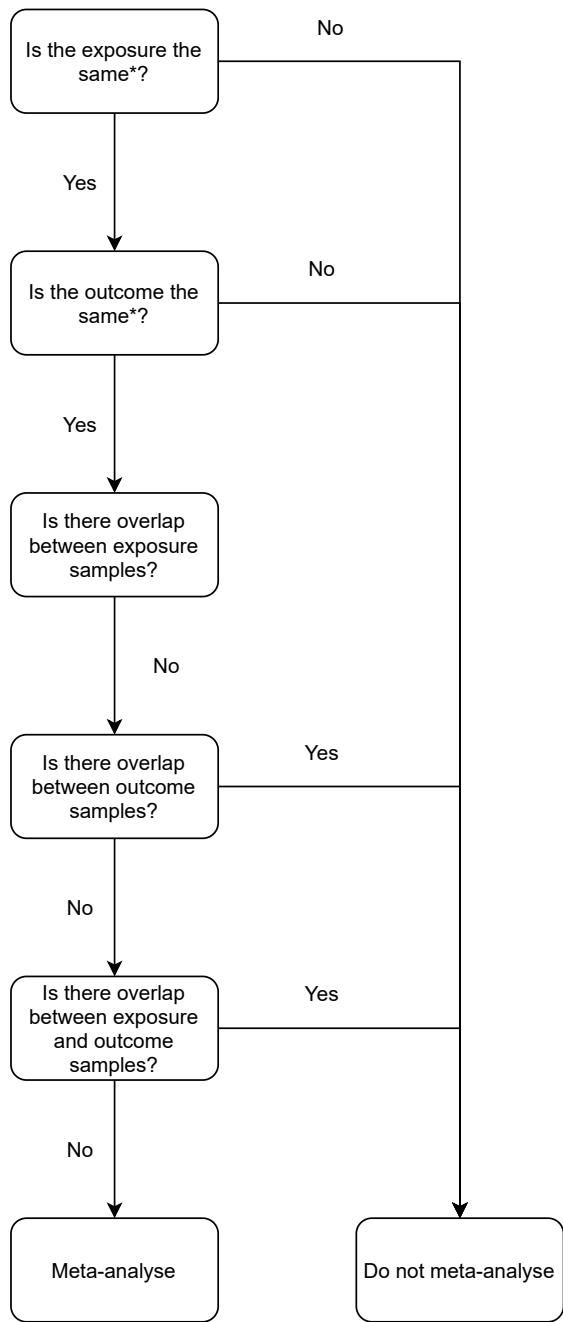


Figure 2.1: Inclusion criteria for meta-analysis: flowchart. Mendelian randomization (MR) analyses were included in meta-analyses if they met the conditions set out in the flowchart and in Section 2.2.5. * = MR analyses had to use the same exposure and the same outcome to be compatible, e.g. for the exposure, body mass index could not be meta-analysed with any other exposure that was not body mass index. This also applies to outcomes, e.g., the outcome oestrogen negative breast cancer could not be meta-analysed with breast cancer, it could only be meta-analysed with oestrogen negative breast cancer.

1611 **2.2.6 Narrative synthesis**

1612 In order to gain a global picture of reported causal effects, a narrative synthesis was
1613 performed. All articles that were not included in the meta-analyses were included in the
1614 narrative synthesis where individual studies were summarised. Due to this latter point, the
1615 narrative synthesis therefore will have included non-independent estimates but is there to
1616 give a global synthesis of the causal relevance of adiposity on health outcomes. The outcome
1617 categories were used to guide the synthesis. MR analyses that were included in the meta-
1618 analyses were not included in the narrative synthesis. The direction of effect estimates
1619 across outcome categories were summarised across the exposures used for these analyses.
1620 Given that studies may not report p-values these were not the focus here. The synthesis is
1621 presented in alphabetical order of the outcome categories.

1622 **2.3 Results**

1623 **2.3.1 Literature search**

1624 In total, 8,377 articles were returned from the combined search of EMBASE (N = 3,772),
1625 MEDLINE (N = 3,638), and bioRxiv (N = 966). After combining the articles from EMBASE and
1626 MEDLINE, de-duplication resulted in the removal of 1,500 articles (N = 5,910). De-duplication
1627 of bioRxiv search results removed an 293 articles (N = 673). Published bioRxiv articles were
1628 dealt with at the data extraction stage. The 5,910 articles from EMBASE and MEDLINE were
1629 combined with the 673 articles from bioRxiv and titles and abstracts were screened. A total of
1630 277 articles were retained after title and abstract screening (data available on [GitHub](#)).

1631 Of the 277 articles included in the full text screening, a total of 104 articles were removed
1632 for the following reasons (with number of articles removed indicated in brackets): a conference
1633 abstract with no response within 6 months from the author or there was no data available
1634 from the authors (25), conference abstracts with the full paper included in the search (23),
1635 duplicates not excluded by the de-duplication process (23), not using a measure of adiposity
1636 as the exposure (15), a commentary (8), erratum (3, the corrected papers were identified in
1637 the search), not MR (i.e., regression of SNP on trait; 4), a conference proceeding (1), not
1638 available in English (1), and preprint paper in which the published paper did not include an
1639 MR of an adiposity measure (1). After full text screening, 173 articles were included in the
1640 analysis (data available on [GitHub](#)).

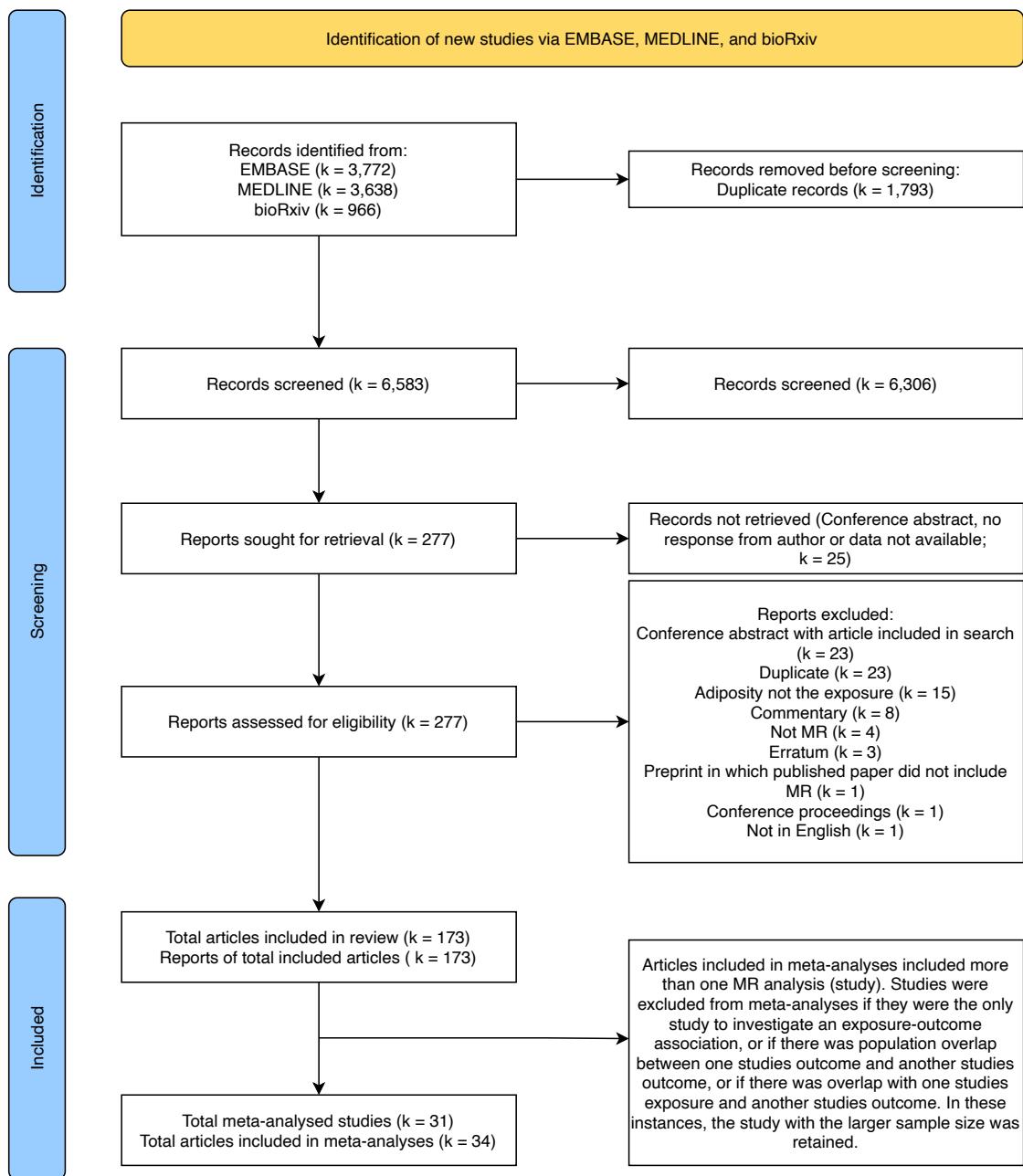


Figure 2.2: PRISMA flowchart. *k* gives the number of articles at each stage

1641 **2.3.2 Data extraction**

1642 Articles from bioRxiv included in data extraction were replaced with their published version
1643 if available. Of the 23 included bioRxiv articles, 18 were published once data extraction
1644 began and were included instead of the bioRxiv article. One bioRxiv article was excluded
1645 as the published version did not include the MR analysis. The remaining 4 bioRxiv articles
1646 were included. The majority of articles were published in the past 5 years (Figure 2.3) and
1647 one-sample MR was the predominant analysis performed across the 173 studies (Figure 2.4).

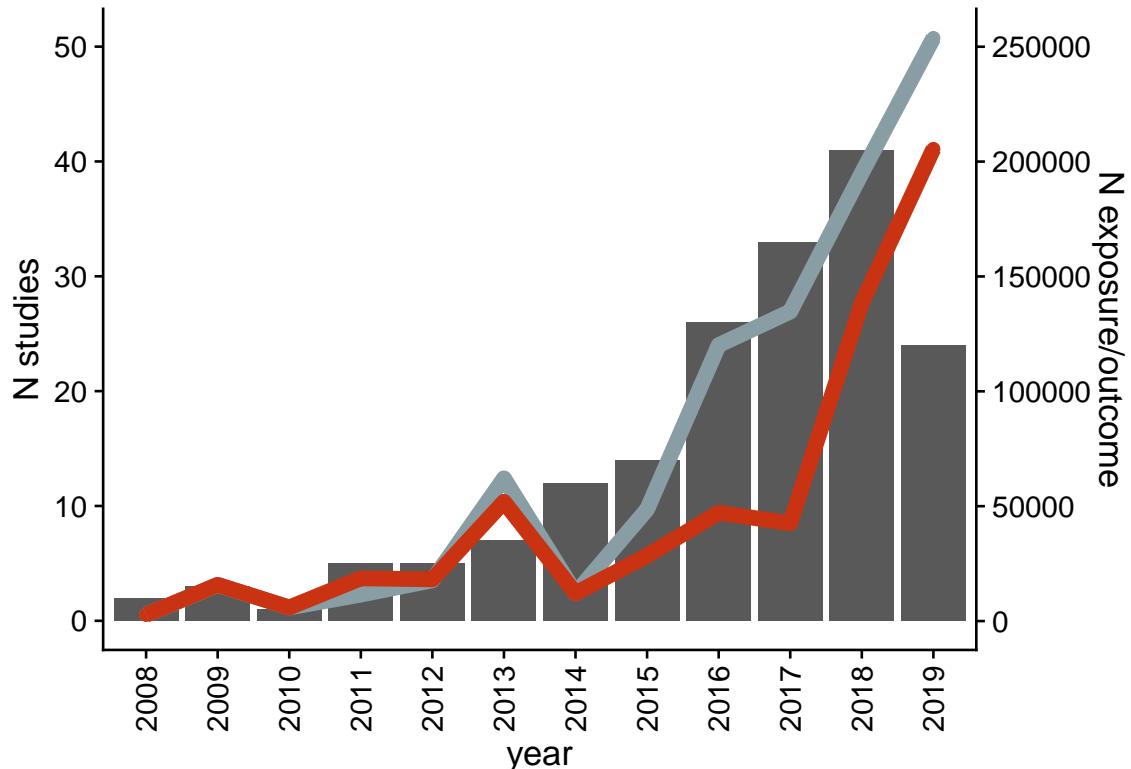


Figure 2.3: **Distribution of publication year per article and average exposure and outcome sample sizes across included studies up to February 2019.** The number of articles included per year is given on the left Y axis; the right Y axis gives the average sample size for exposure (grey) and outcome (red) for each year. Outcome cases and controls were summed within analyses.

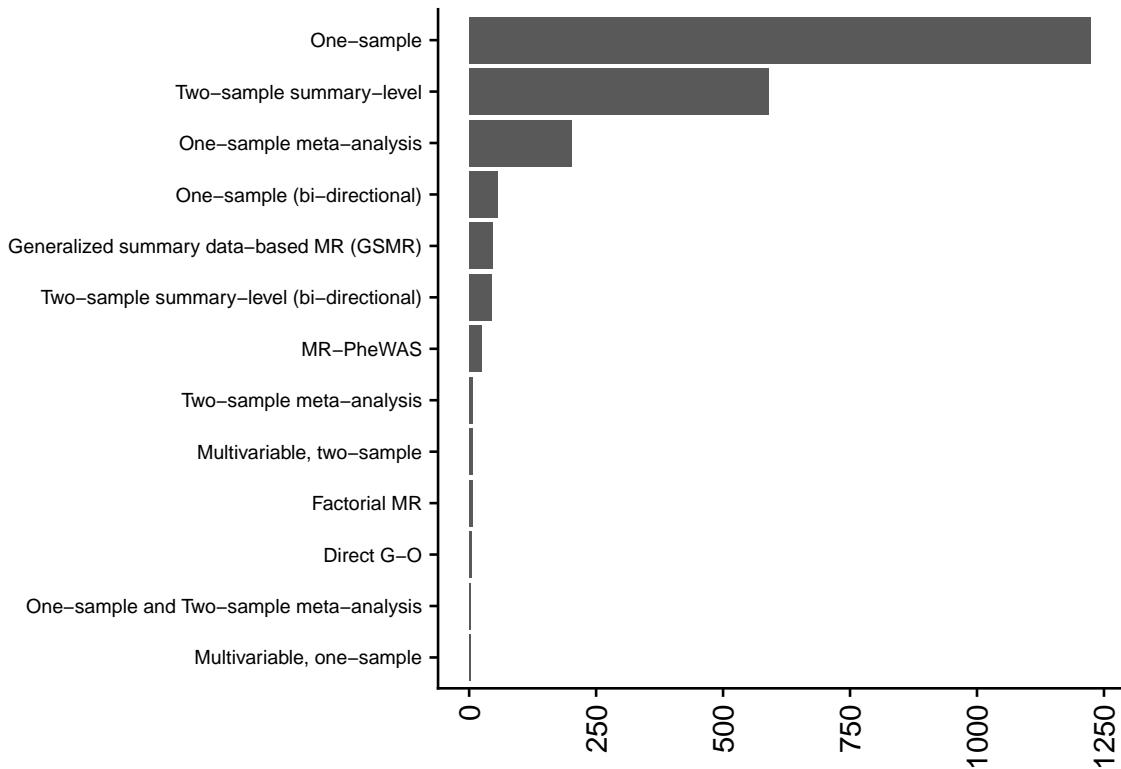


Figure 2.4: Distribution of study design across 173 included articles. The majority of the 173 included articles reported more than one Mendelian randomization (MR) analysis. Where a study performed a bi-directional MR analysis and adiposity was the secondary analysis (i.e., to check for reverse causation), this was recorded as a bi-directional MR analysis. One-sample and two-sample meta-analysis indicates the meta-analysis included MR analyses that were both one- and two-sample designs.

1648 A total of 2214 MR analyses were performed across the 173 studies (i.e., many studies
 1649 conducted multiple MR analyses). This included 31 exposures and 659 outcomes. The
 1650 majority of the 2214 MR analyses used BMI as the exposure (Table 2.1). After formatting
 1651 the outcome data into three columns of (column 1) general outcome (e.g., breast cancer),
 1652 (column 2), analysis specific outcome (eg., ER-), and outcome category (e.g., cancer), a
 1653 total of 311 general outcomes were available and grouped into 16 outcome categories. Of
 1654 the 311 outcomes, smoking was used in the most MR analyses followed by asthma and
 1655 DNA methylation (Table 2.2). The largest proportion of outcomes were grouped into the
 1656 metabolic and cancer categories (Table 2.3). The other category included 118 methylation
 1657 outcomes, 68 mortality outcomes, and a handful of the following outcomes: age related
 1658 macular degeneration, cataract, disease count, hernia, sleep, and physical activity. Categories,
 1659 as discussed in the methods, were assigned based on prior biological knowledge.

Table 2.1: Number and frequency of exposures used across all 2214 MR analyses

Exposure	N	%
BMI	1509	68.16
WHRadjBMI	156	7.05
WHR	112	5.06
birth weight	102	4.61
WC	50	2.26
body fat percentage	45	2.03
fat mass	37	1.67
BMI increasing and WHR decreasing	20	0.90
BMI increasing and WHR increasing	20	0.90
obesity	15	0.68
WCadjBMI	14	0.63
fat percentage	10	0.45
HC	10	0.45
hepatic fat	10	0.45
non-fat mass	10	0.45
sum of skinfolds	10	0.45
total body fat	10	0.45
fat mass index	9	0.41
fat-free mass	9	0.41
HCadjBMI	9	0.41
favourable adiposity	7	0.32
overweight	7	0.32
fat free mass	6	0.27
lean mass	6	0.27
body fat mass	5	0.23
central obesity	4	0.18
adiponectin	3	0.14
Obesity class 1	3	0.14
weight	3	0.14
body non-fat mass	2	0.09
body fat	1	0.05

BMI = body mass index; WHR = waist hip ratio;
 WHRadjBMI = WHR adjusted for BMI; WC = waist circumference; WCadjBMI = WC adjusted for BMI; HC = hip circumference; HCadjBMI = HC adjusted for BMI.

Table 2.2: Number and frequency of 10 most used outcomes across all 2214 MR analyses

Outcome	N	%
smoking	175	7.90
asthma	122	5.51
methylation (cpg)	118	5.33
coronary artery disease	87	3.93
breast cancer	80	3.61
mortality	68	3.07
depression	58	2.62
lung cancer	52	2.35
stroke	49	2.21
osteoarthritis	47	2.12

Table 2.3: Number and frequency of outcomes within each outcome category across all 2214 MR analyses

Group	N	%
metabolic	404	18.25
cancer	352	15.90
respiratory	318	14.36
cardiovascular	285	12.87
other	235	10.61
mental health	127	5.74
skeletal	95	4.29
anthropometric	85	3.84
brain	73	3.30
hepatic	71	3.21
social	71	3.21
renal	34	1.54
reproductive	19	0.86
gastrointestinal	17	0.77
skin	16	0.72
immune	12	0.54

1660 **2.3.3 Quality assessment**

1661 Studies that contributed to the meta-analyses were assessed for quality using a modified
1662 version of the assessment criteria devised by Mamluk et al. (2020)⁴⁰⁶. Studies were assessed
1663 on a 3-point scale across 12 questions, with values ranging from 12-36. Analyses with lower
1664 scores (12-19) were considered to be of higher quality, with high scoring (28-36) studies
1665 considered lower quality. Scores in between were of medium quality. The average assessment
1666 was 24 (Figure 2.5). Individual studies were assessed as opposed to the article, as most
1667 articles conducted multiple studies. Only the study of the effect of body mass index (BMI)
1668 on hemorrhagic stroke by Dale et al (2017)⁴²⁸ was ranked as high quality. The majority of
1669 studies (24) were assigned a medium quality score. All of the six low scoring studies showed
1670 consistent directions of effect with the other studies they were meta-analysed with. Quality
1671 scores are presented alongside the meta-analysis results (Figures 2.6 and 2.7).

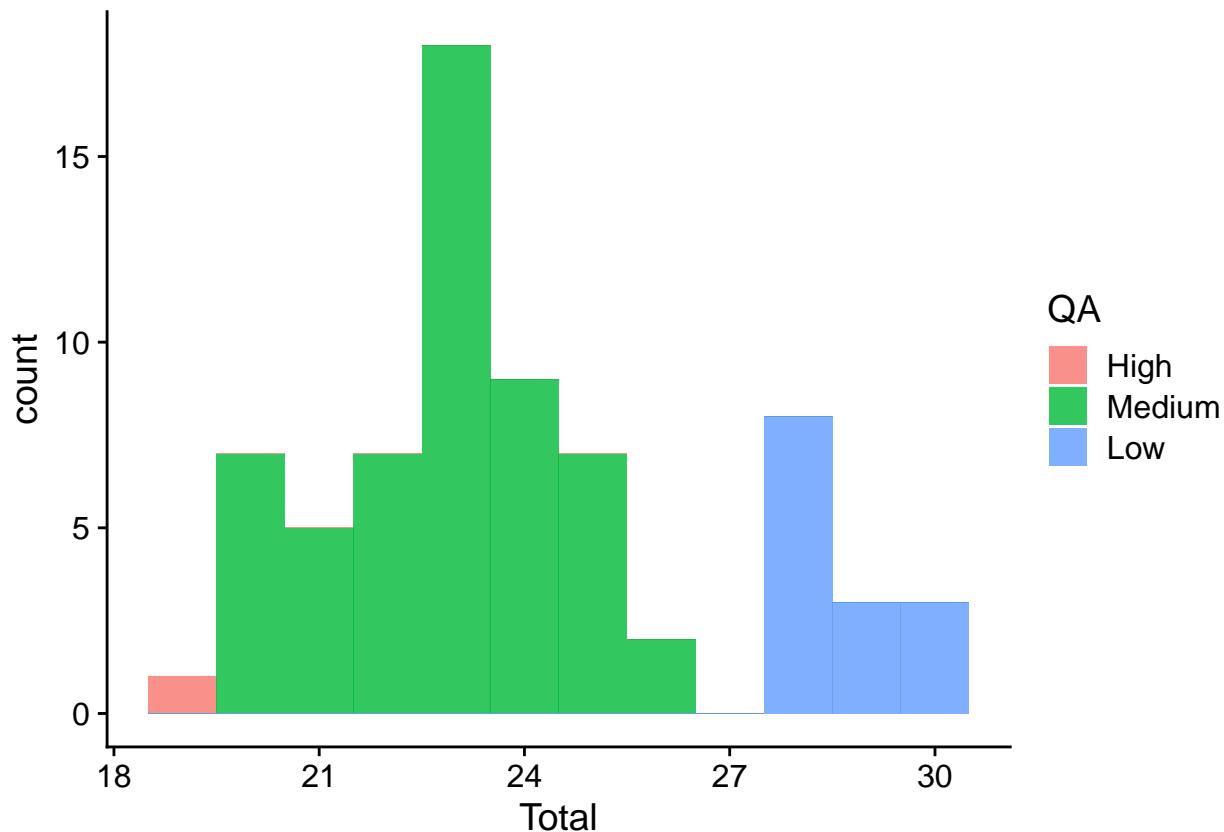


Figure 2.5: **Quality assessment: distribution of quality assessment scores for studies included in the meta-analyses.** High indicates a study scored highly; low indicates a study scored poorly. QA = quality assessment score.

1672 **2.3.4 Meta-analysis**

1673 In total, 31 meta-analyses were conducted using data from 34 studies. A majority of MR
1674 analyses were excluded due to a lack of meta-analysable data (i.e. only one MR analysis
1675 looked at an exposure-outcome pair). Additional reasons for exclusion were: population
1676 overlap, incompatible units, and incompatible MR model. A majority of the 34 studies
1677 contributed to just one meta-analysis. A number of studies contributed multiple MR analyses
1678 which were included in meta-analyses: four studies contributed to two meta-analyses, three
1679 studies to three meta-analyses, two studies to four meta-analyses, two studies to seven
1680 meta-analyses, and one study to eight meta-analyses (Table 2.4).

Table 2.4: Number of times a study was used in meta-analyses

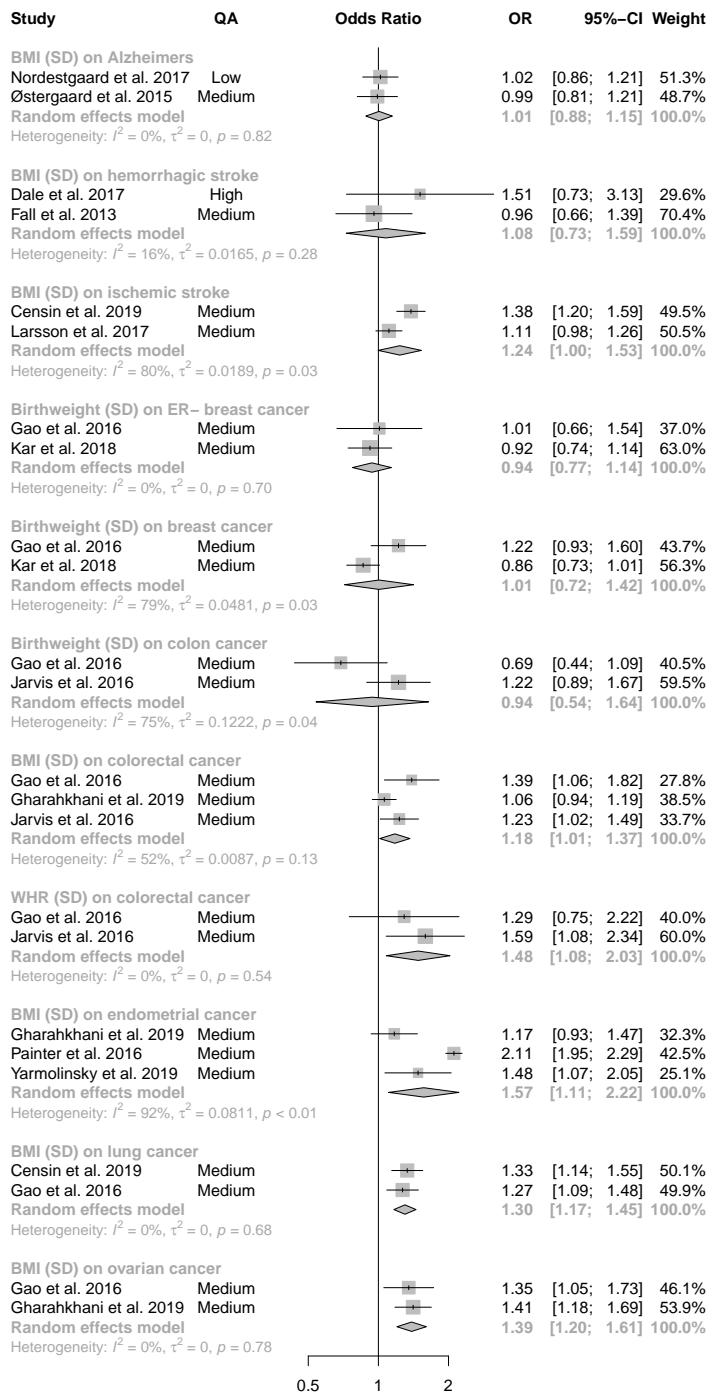
Study	N
Gao et al. 2016 ⁴²⁹	8
Censin et al. 2019 ⁴³⁰	7
Fall et al. 2013 ³⁸⁹	7
Gharahkhani et al. 2019 ⁴³¹	4
Holmes et al. 2014 ³⁹⁰	4
Jarvis et al. 2016 ⁴³²	3
Wang et al. 2018 ⁴³³	3
Xu et al. 2017 ³⁵⁷	3
Dale et al. 2017 ⁴²⁸	2
Kar et al. 2018 ⁴³⁴	2
Shu et al. 2018 ⁴³⁵	2
Wurtz et al. 2014 ¹⁵⁷	2
Ostegaard et al. 2015 ⁴³⁶	1
Brower et al. 2018 ⁴³⁷	1
Day et al. 2018 ⁴³⁸	1
Kivimaki et al. 2008 ⁴³⁹	1
Klarin et al. 2017 ⁴⁴⁰	1
Larsson et al. 2017 ⁴⁴¹	1
Larsson et al. 2018 ⁴⁴²	1
Lindstrom et al. 2017 ⁴⁴³	1
Lv et al. 2018 ⁴⁴⁴	1
Lyall et al. 2016 ⁴⁴⁵	1
Nordestgaard et al. 2017 ⁴⁴⁶	1
Painter et al. 2016 ⁴⁴⁷	1
Palmer et al. 2011 ⁴⁴⁸	1
Richardson et al. 2019 ⁴⁴⁹	1
Shapland et al. 2019 ⁴⁵⁰	1
Skaaby et al. 2018 ⁴⁵¹	1
Speed et al. 2019 ⁴⁵²	1

Study	N
Tyrrell et al. 2019 ⁴⁵³	1
van den Broek et al. 2018 ⁴⁵⁴	1
Wade et al. 2018 ⁴⁵⁵	1
Wang et al. 2018 ⁴⁵⁶	1
Yarmolinsky et al. 2019 ⁴⁵⁷	1

1681 For all binary outcomes, results are given per standard deviation (SD) unit increase, such
 1682 that an odds ratio (OR) is the change in the outcome per SD unit increase in the exposure.
 1683 Studies which used risk ratios and hazard ratios were excluded from the meta-analysis
 1684 following the rules set out in Figure 2.1, e.g., sample overlap. For continuous outcomes,
 1685 results are given as the mean difference (MD) and reflect a change in the outcome unit per
 1686 SD unit increase in the exposure - effect estimate is used throughout.

1687 In the meta-analyses, there were 22 binary outcomes and 9 continuous outcomes. For
 1688 the 22 binary outcomes, 2 tests (birthweight on ER- breast cancer and CAD) had negative
 1689 effect estimates. Both tests had confidence intervals which spanned the null. The remaining
 1690 20 tests had positive effect estimates, 13 of which had confidence intervals that did not span
 1691 the null. All of these reached the nominal p-value threshold (Figure 2.6). The majority of MR
 1692 analyses included in these meta-analyses had a medium quality assessment score. The MR
 1693 analysis by Dale et al., (2017)⁴²⁸ of BMI on hemorrhagic stroke was the only analysis to score
 1694 highly. The four studies which had a low quality assessment score did not have weights that
 1695 were drastically different compared to the other MR analyses in those meta-analyses; all but
 1696 one had tight confidence intervals which did not cross the null.

1697 For the 9 continuous outcomes, 3 tests (BMI on high density lipoprotein (HDL; SD and
 1698 mmol/L) and low density lipoprotein (LDL; mmol/L)), had negative effect estimates with
 1699 confidence intervals which spanned the null. The remaining 6 tests had positive effect
 1700 estimates, 2 (BMI on systolic blood pressure (SBP; mmHg) and fasting glucose (mmol/L)) of
 1701 which had confidence intervals that did not span the null. Both reached the nominal p-value
 1702 threshold (Figure 2.7). Two included MR analyses ranked low for quality assessment; the study
 1703 by Shapland et al., (2018)⁴⁵⁰ had comparable weight to the other two studies investigating
 1704 BMI on SBP. The study by Wang et al., (2018)⁴⁵⁶ investigating HOMA IR (Homeostatic Model
 1705 Assessment for Insulin Resistance) had a much larger weight than the study by Kivimaki et
 1706 al., (2008)⁴³⁹, a result of a larger population. The remaining studies had a medium quality
 1707 assessment score, none of which showed estimates that deviated strongly from the effects of
 1708 the other MR analyses in those meta-analyses.



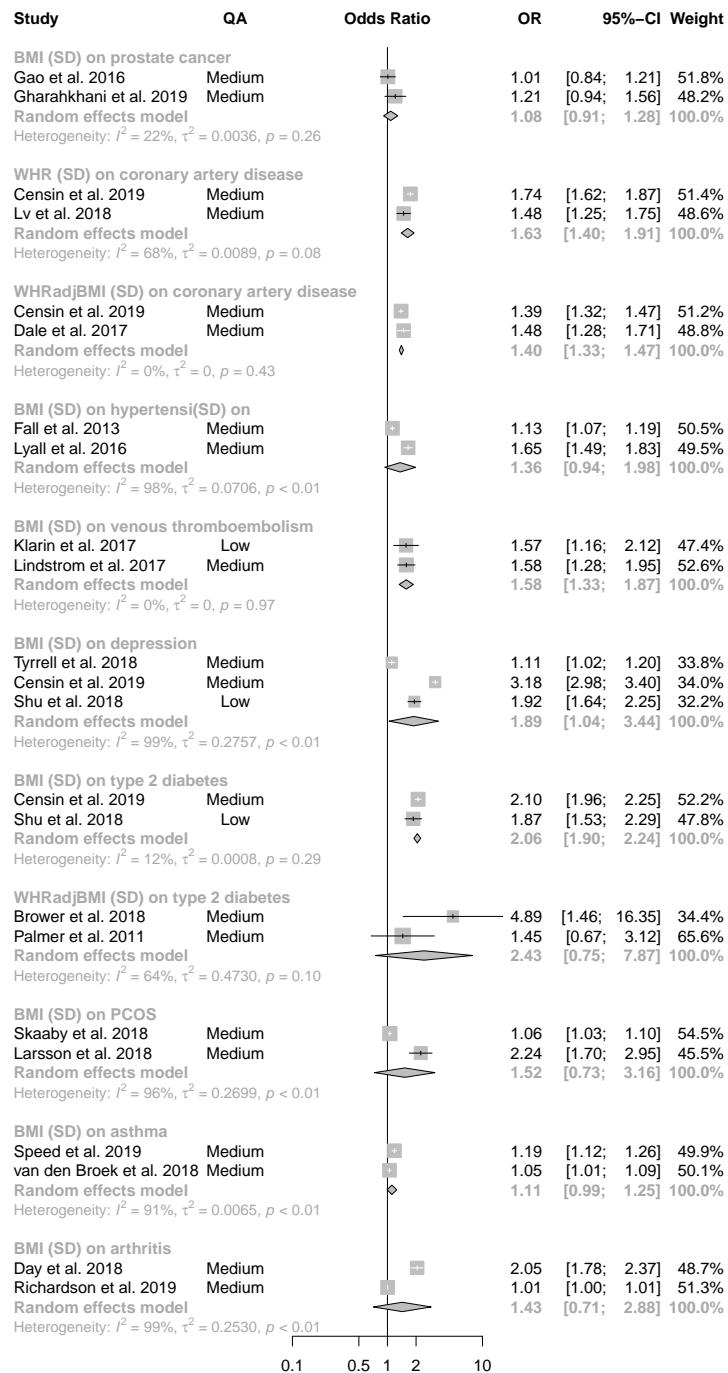


Figure 2.6: Meta-analysis: effect estimates and 95% confidence intervals for binary outcomes. Forestplot shows effect estimates and 95% confidence intervals from a meta-analysis of 22 different exposure-outcome pairs. MR analyses included based on criteria in Figure 2.1. QA = quality assessment score; OR = odds ratio; CI = confidence interval.

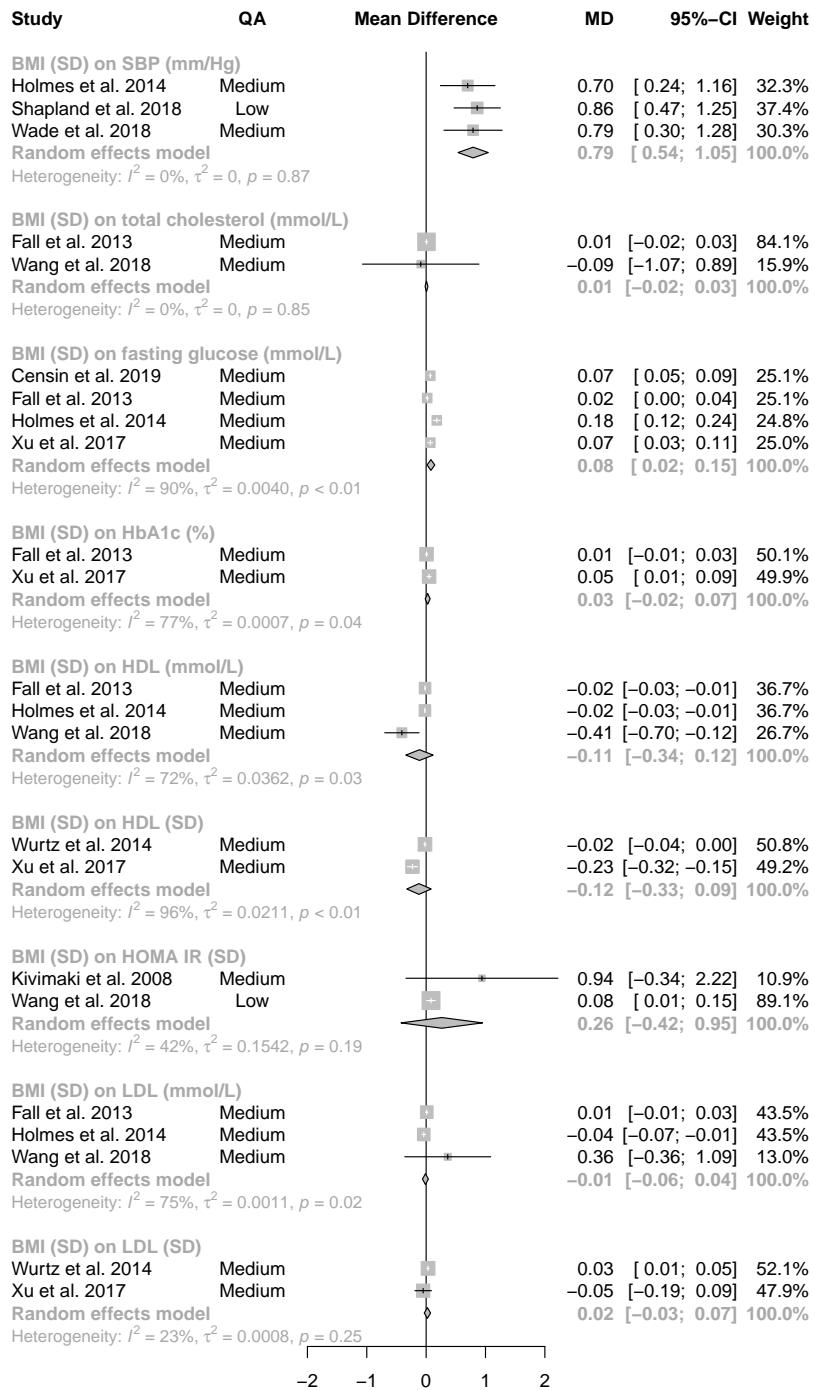


Figure 2.7: Meta-analysis: effect estimates and 95% confidence intervals for continuous outcomes. Forestplot shows effect estimates and 95% confidence intervals from a meta-analysis of 9 different exposure-outcome pairs. MR analyses included based on criteria in Figure 2.1. QA = quality assessment score; MD = mean difference; CI = confidence interval.

1710 Three outcomes were investigated using more than one exposure, coronary artery disease
1711 with waist hip ratio (WHR) and waist hip ratio adjusted for body mass index (WHRadjBMI),
1712 colorectal cancer with BMI and WHR, and type 2 diabetes with BMI and WHRadjBMI. There
1713 was strong evidence for an effect of WHR (OR in the outcome per SD unit increase in the
1714 exposure = 1.63; 95% confidence interval (CI) = 1.4 – 1.91) and WHRadjBMI (OR = 1.4; 95%
1715 CI = 1.33 – 1.47) on coronary artery disease. When looking at colorectal cancer, WHR (OR
1716 = 1.48; 95% CI = 1.08 – 2.03) and BMI (OR = 1.18; 95% CI = 1.01 – 1.37) showed similar
1717 effects with overlapping confidence intervals. For type 2 diabetes, BMI (OR = 2.48; 95% CI =
1718 1.52 – 4.07) and WHRadjBMI (OR = 2.06; 95% CI = 1.9 – 2.24) both show strong estimates
1719 with overlapping confidence intervals.

1720 All remaining tests were conducted with BMI as the exposure. Except for the negative
1721 effect on breast cancer, BMI was found to be associated with an increases with all cancers
1722 tested (colorectal, endometrial, lung, ovarian, and prostate), confidence intervals crossed the
1723 null only for prostate cancer (OR = 1.08; 95% CI = 0.91 – 1.28). The strongest evidence for
1724 an effect was found for venous thromboembolism (OR = 1.58; 95% CI = 1.33 – 1.87) and type
1725 2 diabetes (OR = 2.48; 95% CI = 1.52 – 4.07) - asthma also showed strong evidence but with
1726 a small effect size (OR = 1.06; 95% CI = 1.03 – 1.1). There was weak evidence for an effect
1727 of BMI on ischemic and hemorrhagic stroke, hypertension, arthritis, and Alzheimer's disease,
1728 with effect estimates close to the null and confidence intervals spanning the null.

1729 The weights for each study included in the individual meta-analyses were broadly even
1730 (e.g., in a meta-analysis of three studies each study had a weighting of roughly 33%). The
1731 exceptions, where one study had a much larger or much smaller weight than the other(s) was
1732 for: BMI and asthma, BMI and polycystic ovary syndrome (PCOS), BMI and hemorrhagic
1733 stroke, BMI and total cholesterol, BMI and HOMA IR, and BMI and LDL. There is evidence
1734 of heterogeneity within the included studies, 8 of 22 binary outcomes and 5 of 9 continuous
1735 outcomes had heterogeneity statistics with p-values < 0.05, e.g., BMI on endometrial cancer
1736 $\chi^2 = 92\%$ (p-value < 0.01; Table A.4). However, given no meta-analysis met the requirements
1737 for heterogeneity statistics (> 5 studies; See [Cochrane Handbook](#))⁴⁵⁸, these results should
1738 be used with caution. Heterogeneity statistics and weights are presented in Figures 2.6 and
1739 2.7, as well as Appendix Table A.4.

1740 2.3.5 Narrative synthesis

1741 A total of 2144 studies where not included in the meta-analyses. Though many of these
1742 could not be included because they did not meet certain requirements for inclusion (e.g.,
1743 overlapping populations), they still provide information on the potential effects of adiposity.
1744 Using the 16 categories used to group outcomes, these effects are summarised here, in
1745 alphabetical order with the other category at the end. Where there are a large number of
1746 studies a summary of the directions of effect are given. Given that studies use a variety of
1747 transformations, units, and models, comparison of the magnitude of effect is not appropriate.

1748 Instead, the focus here is on directions of effect estimates and whether evidence is consistent
1749 across studies.

1750 **Anthropometric**

1751 A total of 85 studies were reported across 9 articles for anthropometric outcomes. This
1752 included analyses of birthweight, BMI, hip circumference adjusted for BMI (HCadjBMI), waist
1753 circumference adjusted for BMI (WCadjBMI), and WHRadjBMI on similar anthropometric traits
1754 such as adipose tissue volume, birth length, body fat, head circumference, leg fat, and trunk fat.
1755 The majority of effect estimates were positive (N = 60; negative = 24). A single MR analysis
1756 (BMI on offspring BMI) had an effect estimate of 0. A number of other analyses focussed on
1757 offspring traits as the outcome with both positive and negative effect estimates. The study by
1758 Winkler et al. (2018)⁴⁵⁹ used a unique instrumentation method, using a composite measure
1759 of BMI, WHR, and WHRadjBMI, for example BMI increasing and WHR increasing SNPs were
1760 used as a genetic instrument. Generally, the effect of adiposity on anthropometric traits is to
1761 increase them, but this is likely a reciprocal relationship, i.e., increased BMI leads to increased
1762 WHR and increased WHR leads to increased BMI.

1763 **Neurological/behavioural**

1764 A total of 67 studies were reported across 21 articles for brain related outcomes. This
1765 included analyses of Alzheimer's disease, amyotrophic lateral sclerosis (ALS), dementia,
1766 multiple sclerosis (MS), Parkinson's, and stroke. Bipolar disorder, schizophrenia, cognitive
1767 ability, grey matter volume, and migraine were also present. Exposures included birth weight,
1768 BMI, WHR, and WHRadjBMI - BMI was used in the majority of analyses. For WHR on grey
1769 matter volume and BMI on cognitive ability, effect estimates were negative. The majority
1770 of effect estimates were positive (N = 45; negative = 17). Two analyses (BMI on stroke
1771 (ischemic small vessel) and dementia) had an OR of 1. Effect estimates appeared larger on
1772 the whole when in the positive direction, however in many cases across both the positive and
1773 negative estimates, confidence intervals spanned the null. On balance, results suggest an
1774 increased adiposity increases the risk of all types of stroke. However for all other outcomes
1775 there appears conflicting or weak evidence for an effect.

1776 **Cancer**

1777 A total of 332 studies were reported across 39 articles for cancer related outcomes. This
1778 included analyses of all cancers, cancer mortality, cancer types such as breast and prostate,
1779 and subtypes such as ER- and ER+ breast cancer. A majority of effect estimates were positive

1780 (N = 189; negative = 137), while six studies of breast, kidney, lung, and prostate cancer
1781 showed effect estimates approximately equal to 1. A majority of analyses with positive effect
1782 estimates had confidence intervals which spanned the null. The same was true for negative
1783 effect estimates.

1784 Of the 31 cancer outcomes, three showed negative effect estimates – cervical (with BMI
1785 and WHRadjBMI), clear cell (BMI), gastric (BMI) – while fourteen showed only positive effect
1786 estimates. This included overall cancer mortality (with BMI) and cancer risk (with BMI), as
1787 well as an “other cancer” category with two measures: not including lung, breast, colorectal,
1788 skin, prostate, and cervical; and not including any non-skin cancer, non-melanoma skin
1789 cancer, lung cancer, other smoking related cancers, colon, kidney, breast, and prostate. The
1790 remaining cancer types with positive effect estimates included Barrett’s esophagus (with
1791 BMI), colon (with BMI), esophageal (with BMI), lymphoid (with BMI), meningioma (with BMI,
1792 WC, and body fat percentage (BF)), rectal (with BMI), renal (with BMI, WHR, and BF), skin
1793 (including melanoma; with BMI), and stomach and esophageal (with BMI). Low malignant
1794 potential tumors also showed a positive effect estimate with BMI. The remaining 13 cancer
1795 types had positive and negative effect estimates. This included any cancer, breast, colorectal,
1796 endometrial, glioma, kidney, lung, multiple myeloma, ovarian, pancreatic, prostate, testicular,
1797 and upper aerodigestive.

1798 Results suggest increased adiposity increases the risk of overall cancer risk and mortality.
1799 However, this risk is modulated by cancer type and subtype. In the case of cancers with only
1800 negative effect estimates, these cancers were only analysed once, whereas cancers like
1801 breast and lung, which were measured multiple times, showed both positive and negative
1802 effect estimates.

1803 **Cardiovascular**

1804 A total of 274 studies were reported across 50 articles for cardiovascular related out-
1805 comes. This included analyses of 11 continuous traits and 19 binary outcomes. Exposures
1806 included birth weight, BMI, body fat mass measures, HC, WC, WCadjBMI, weight, WHR, and
1807 WHRadjBMI.

1808 In total, 83 studies investigating the effect of adiposity with continuous traits were reported
1809 across 22 articles. The majority of these studies reported positive effect estimates (N = 69;
1810 negative = 14). Of the 11 traits, four had a single reported MR result (left ventricular mass,
1811 mean arterial pressure, pulse pressure, and pulse wave velocity) – all had positive effect
1812 estimates except pulse wave velocity. Of the seven remaining traits, five had both positive
1813 and negative effect estimates. Effects on heart rate were negative with wide confidence
1814 intervals, while effects on carotid intermedia thickness (IMT) were positive with some studies
1815 reporting effects estimates with confidence intervals that did not overlap the null for BMI and
1816 WHRadjBMI.

1817 Of the five traits with positive and negative effect estimates, diastolic blood pressure
1818 (DBP) showed a negative effect estimate solely in relation to the effect of birthweight, and
1819 SBP showed weak evidence for a decreasing effect of BMI and birthweight. There was
1820 much stronger evidence for an increasing effect on SBP and DBP across BMI, WHR, and
1821 WHRadjBMI. The positive and negative effect estimates associated with heart beat are
1822 associated with confidence intervals which span the null for both BMI and WHRadjBMI. For
1823 carotid IMT, evidence appears stronger, narrower confidence intervals which do not span
1824 the null, for an increasing effect of BMI. Evidence for an effect of BMI and WHRadjBMI
1825 on left ventricular hypertrophy was weak and dependent upon the method used to assess
1826 hypertrophy.

1827 In total, 191 studies investigating the effect of adiposity with binary outcomes were
1828 reported across 35 articles. Nine of these studies were from a single article which only
1829 reported p-values for the effect of BMI on coronary artery disease (CAD). Of the remaining
1830 182 studies, the majority reported positive effect estimates (N = 139; negative = 43). There
1831 was strong evidence across multiple studies for the effect of BMI on CAD and CVD and results
1832 also support an effect of WHRadjBMI, WCadjBMI, and WHR. Though there was evidence for
1833 an effect of BMI on heart failure, there was weak evidence for a similar effect of WHRadjBMI on
1834 the same outcome. There was strong evidence for an effect of fat mass and fat free mass on
1835 increased risk of arrhythmia, but only weak evidence for a similar effect from WHRadjBMI and
1836 birthweight. There was conflicting evidence for an effect of increased BMI on MI. When using
1837 BMI increasing and WHR decreasing instruments myocardial infarction (MI) risk decreased
1838 while when using BMI and WHR increasing instruments MI risk increased. There was also
1839 weak evidence of increased birthweight reducing MI risk. There was strong evidence across
1840 many different adiposity measures for an increased risk of deep vein thrombosis (DVT).

1841 Though there were some conflicting results, BMI and MI for example, and some analyses
1842 reported both positive and negative effect estimates for the same exposure-outcome pairs,
1843 BMI and SBP for example, on balance reported results support an increasing effect of
1844 adiposity on cardiovascular traits. Evidence was strongest for the effect of adiposity on CAD,
1845 CVD, and DVT.

1846 **Gastrointestinal**

1847 A total of 17 studies were reported across 5 articles for gastrointestinal related outcomes.
1848 This included analyses of inflammatory bowel disorders, *Helicobacter pylori* infection mea-
1849 sures, gallstone disease, and peptic ulcers. There was evidence for an effect of birthweight on
1850 inflammatory bowel disease and some evidence for an effect of BMI on peptic ulcers, however
1851 weak evidence for an effect of WHRadjBMI on peptic ulcers. All other analyses showed weak
1852 evidence of effect. As no exposure-outcome pairs were analysed by more than one study it is
1853 hard to draw conclusions from the available evidence, however effect estimates were mostly
1854 positive across studies.

1855 **Hepatic**

1856 A total of 71 studies were reported across 6 articles for hepatic related outcomes. In total,
1857 11 outcomes were reported, of which the majority of analyses (N = 40) were for three liver
1858 markers: alanine transaminase (ALT), aspartate transaminase (AST), and gamma-glutamyl
1859 Transferase (GGT). AST was reported once with strong evidence of a reducing effect of
1860 BMI. Analyses of ALT and GGT used multiple measures across multiple studies, for example
1861 adjusting for alcohol consumption. Results support an increasing effect of increased BMI
1862 on ALT and GGT, which persisted after adjustment for alcohol consumption. The remaining
1863 8 outcomes were investigated by a single article. Evidence was found for BMI and
1864 WHR on chronic liver disease and BMI, WHR, and WHRadjBMI on NAFLD. There was strong
1865 evidence for an effect of adiposity on all hepatic traits.

1866 **Immune**

1867 A total of 12 studies were reported across 5 articles for immune related outcomes. In
1868 total, 8 outcomes were reported. There was weak evidence for an effect of adiposity on all
1869 outcomes, except for birthweight on celiac disease, BMI on dermatophytosis (though weak
1870 evidence for an effect of WHRadjBMI), and BMI on psoriasis.

1871 **Mental health**

1872 A total of 124 studies were reported across 22 articles for mental health related outcomes.
1873 In total, 27 outcomes were reported, though 16 of these were reported only once - all showed
1874 weak evidence of an effect (e.g., attention deficit hyperactivity disorder, anorexia nervosa, be-
1875 ing a worrier/nervous person, body dissatisfaction (evidence from weight and shape concern
1876 analyses showed a negative effect of BMI), and happiness). Of the remaining 11 outcomes,
1877 the majority of analyses focused on depression. Across the 11 articles which looked at
1878 depression, there was strong evidence for an effect of adiposity increasing depression. When
1879 excluding non-neuronal SNPs (which will influence adiposity at a cellular as opposed to
1880 behavioural level), the effect of BMI was reduced and confidence intervals crossed the null⁴⁵³.
1881 This would suggest that the association with depression is not a result of behavioural changes
1882 associated with adiposity. Rather, the association is likely due to the physicality of adiposity
1883 and probably the stigmatization associated with that. There was weak evidence for an effect
1884 of BMI on anxiety. There was weak evidence for an effect of increased BMI on increased
1885 loneliness. Similarly, there was weak evidence for a decreasing effect of BMI, WHR, WC,
1886 and BF on subjective wellbeing. There was some evidence for a decreasing effect of BMI
1887 and WHR on stress/nervous feelings, however weak evidence was found for replications and
1888 for all other psychological distress traits. Binge eating and overeating increased as a result

1889 of increased BMI. On balance, there appears to be an association between adiposity and
1890 mental health traits, particularly body image related traits. However, this association is likely
1891 not a direct result of adipose tissue, but is perhaps a result of sociological factors.

1892 **Metabolic**

1893 A total of 380 studies were reported across 51 articles for metabolic related outcomes. In
1894 total, 27 outcomes were reported. A majority of these were metabolites, many of which were
1895 reported once. Although the majority of metabolite effect estimates were positive, confidence
1896 intervals for many spanned the null. There was, for example, weak evidence for an increasing
1897 effect of BMI on cholesterol, however strong evidence for an effect of WHRadjBMI. C-reactive
1898 protein (CRP) was investigated with BMI across 9 studies with all but two reporting strong
1899 evidence for an increasing effect of BMI on CRP levels. BMI was found to decrease levels
1900 of apolipoprotein A-I and increase apolipoprotein B levels; there was weak evidence for an
1901 increasing effect of BMI and WHRadjBMI on apolipoprotein A-IV. There was strong evidence
1902 for a decreasing effect of BMI, WHRadjBMI, and birthweight on HDL levels; there was weaker
1903 evidence for an overall effect of adiposity on LDL – WHRadjBMI was strongly associated with
1904 a increase in LDL, while birthweight showed a decreasing effect on LDL. BMI showed weak
1905 evidence of both increasing and decreasing effects on LDL. There was also strong evidence
1906 for an increasing effect of BMI, WHR, and WHRadjBMI on triglycerides.

1907 There was strong evidence for an effect of increased BMI, WHR, WHRadjBMI on fasting
1908 glucose; there was weaker evidence for an effect of childhood BMI and birthweight on fasting
1909 glucose. There was weak evidence for an effect of BMI (adult and childhood) on two hour
1910 glucose test (there was evidence for a decreasing effect of birth weight), and weak evidence
1911 for an increasing effect of BMI on non-fasting glucose. There was strong evidence for an
1912 effect of BMI on hyperuricaemia as well as uric acid. Weaker evidence was reported for an
1913 effect of BMI (adult and childhood) and WHRadjBMI on glomerular filtration rate, creatine, and
1914 creatinine. There was strong evidence for an increasing effect of BMI, WHR, and WHRadjBMI
1915 on fasting insulin. There was however weak evidence for an increasing effect of BMI on insulin
1916 secretion. Binary outcomes reported broadly increasing effects of adiposity. For example,
1917 there was strong evidence for an effect of increased BMI on increased diabetes (type1, type2,
1918 all). Similarly strong evidence was reported when using birthweight, childhood BMI, WHR,
1919 WHRadjBMI, and WC. Strong evidence for an effect of BMI and WHRadjBMI on increased
1920 dyslipidemia and metabolic syndrome was reported, but there was weak evidence for an
1921 effect of BMI and WHRadjBMI on hyper- and hypo-thyroidism and iron deficiency. The effect
1922 of adiposity appears far reaching in regards to metabolic traits. This effect is generally to
1923 increase levels of traits that are themselves associated with poor health outcomes.

1924 **Renal**

1925 A total of 34 studies were reported across 4 articles for renal related outcomes. A
1926 majority of analyses looked at renal failure (N = 20) which showed strong evidence for an
1927 increasing effect of BMI, WHR, and WHRadjBMI. These analyses were however from a single
1928 study⁴³⁰. A similar picture is present for BMI and renal disease which was investigated by a
1929 single study⁴⁶⁰, as well as macroalbuminuria and BMI⁴⁶⁰. There was weak evidence for an
1930 increasing effect of childhood BMI and birth weight on chronic kidney disease. As few articles
1931 looked at renal related outcomes it is difficult to draw conclusions given a lack of replication.
1932 However, the general trend is for an increasing effect of adiposity on the risk of renal related
1933 traits and renal diseases.

1934 **Reproductive**

1935 A total of 17 studies were reported across 5 articles for primarily menarche (age at and
1936 early onset). One study reported evidence of an increasing effect of BMI on PCOS⁴³⁸, while
1937 another reported weak evidence for an increasing effect of WHRadjBMI on uterine fibroids⁴⁶¹.
1938 The two studies reporting on age at menarche and BMI and childhood BMI found strong
1939 evidence that adiposity decreased age at menarche,^{462,463} which is associated with poor
1940 health outcomes in later life. The remaining 15 analyses on early menarche were reported by
1941 one study⁴⁶⁴ and showed evidence that BMI, total body fat, fat free mass, sum of skinfolds, HC,
1942 and WHR all lead to an earlier menarche. There is clear evidence that adiposity increases the
1943 likelihood of early menarche. There is also compelling evidence that adiposity is detrimental
1944 in regards to all reproductive traits.

1945 **Respiratory**

1946 A total of 316 MR analyses were reported across 13 studies for respiratory related
1947 outcomes. A majority of these analyses were for smoking outcomes (N = 175) such as age
1948 at initiation, status, number of cigarettes per day, as well as comparisons between smoking
1949 status (e.g., ever vs never). There was strong evidence for an effect of BMI, WHR, and
1950 WHRadjBMI on current smoking status. There was also evidence for a positive effect of BMI
1951 on lifetime smoking. There was weak evidence for an effect of BMI on former vs current
1952 and experimental vs never smoking. There was some evidence for an effect of BMI on ever
1953 vs never smoking, increasing the odds of being an ever smoker. There was similarly an
1954 increasing effect on ever being a smoker for BMI, WC, and BF – this effect modulated when
1955 including/excluding neuronal/deprivation related SNPs. The majority of the remaining MR
1956 analyses were for asthma and asthma subtypes. There was broadly weak evidence for an
1957 increasing effect of BMI on asthma. Strong evidence for an increasing effect of BMI on chronic

1958 obstructive pulmonary disorder. There was also evidence for an effect of BMI on wheezing,
1959 and lung volume measures (forced vital capacity and forced expiratory volume).

1960 **Skeletal**

1961 A total of 93 studies were reported across 13 articles for skeletal related outcomes. A
1962 majority of these were arthritic outcomes (arthritis and osteoarthritis), though evidence was
1963 conflicting. There was some evidence for an increasing effect of BMI on rheumatoid arthritis
1964 and gout, however weak evidence for an effect of WHRadjBMI. Strong evidence was reported
1965 for an increasing effect of BMI, WC, and HC on osteoarthritis (self report, hospital diagnosed:
1966 hip, knee), however weak evidence for an effect of WHR and birth weight. One study reported
1967 an effect of BMI on osteoporosis, there was evidence of an increasing effect. There was strong
1968 evidence for an increasing effect of BMI and fat mass on bone mineral density (including site
1969 specific bone mineral density), and some evidence for an increasing effect of trunk fat mass
1970 on bone mineral content. On balance, the effect of adiposity is detrimental to skeletal traits.
1971 This is especially true for arthritic traits, where body composition as opposed to deposition
1972 appears to be more important.

1973 **Skin**

1974 A total of 16 studies were reported across 1 article⁴⁶⁵. Budu-Aggrey et al., (2019)
1975 investigated the effect of BMI on psoriasis using one-sample and two-sample MR analyses.
1976 To strengthen evidence for a causal effect, they meta-analysed one- and two-sample MR
1977 results and performed the reverse MR investigating the effect of psoriasis on BMI. There was
1978 strong evidence for an increasing effect of BMI on psoriasis and weak evidence for an effect
1979 of psoriasis on BMI.

1980 **Social**

1981 A total of 71 studies were reported across 12 articles for social related traits such as
1982 income, education, and employment. Overall there was evidence across 12 studies for a
1983 decreasing effect of increased BMI on income. There was weak evidence for an effect of
1984 BMI on cohabitation and for an increasing effect on socioeconomic status. Evidence was
1985 conflicting for an effect of BMI on education traits such as years in education and degree status.
1986 There was weak evidence for a decreasing effect of BMI on employment traits such as years
1987 employed, employment status, and job class. Data on physical activity was not well reported.
1988 There was weak evidence for a decreasing effect of BMI on risk taking behavior, satisfaction
1989 with family, friends, finances and work. However there was evidence for a decreasing effect

1990 on health satisfaction. On balance, the effect of adiposity was detrimental for social related
1991 traits. Similar to the evidence for mental health traits, these results are unlikely to be a
1992 consequence of adipose tissue and are instead likely consequences of sociological factors
1993 such as stigmatization.

1994 **Other**

1995 Where outcomes could not easily be grouped into one of the previous categories there
1996 were grouped into the other category. A total of 235 studies were reported across 82 articles
1997 for outcomes that could not easily be grouped into one of the previous categories. A majority
1998 ($N = 118$) of these MR analyses were a hypothesis free investigation of the effect of BMI on
1999 DNA methylation. Few of these analyses reported an effect estimate ($N = 34$), those that did
2000 showed there was weak evidence for an effect of BMI. Of the remaining 117 MR analyses,
2001 68 looked at the effect of BMI on mortality and cause specific mortality. There was weak
2002 evidence for an effect of BMI on cause specific mortality across the board, including for all
2003 cancer, all cardiovascular, cancer specific, respiratory, and stroke. There was also weak
2004 evidence for an effect on all cause mortality.

2005 Of the remaining 49 MR analyses, there was weak evidence for an effect of increased
2006 BMI on multiple sleep traits (over-/under-sleeper, hours slept, chronotype etc.). There was
2007 however evidence for an increasing effect of BMI on daytime sleepiness. There was some
2008 evidence for a decreasing effect of BMI on physical activity, including moderate to vigorous
2009 physical activity. Fat mass index showed similar effects, however childhood BMI did not
2010 appear to show a similar effect on physical activity. There was weak evidence for an effect of
2011 BMI on cataract and macular degeneration.

2012 **2.4 Discussion**

2013 **2.4.1 Summary of the evidence**

2014 Observational studies have highlighted numerous risk factors and diseases associated
2015 with adiposity. However, observational studies are limited, for example by confounding and
2016 reverse causation, and can lead to biased results^{145,301,302,333–335}. Many MR analyses have
2017 been conducted which add to the body of evidence in a way that is analogous to RCTs.
2018 Here, 173 articles and over 2,000 MR analyses were reviewed. Meta-analyses and narrative
2019 synthesis of the MR analyses provide an overview of the causal landscape of adiposity,
2020 revealing adiposity to increase the risk of many cancers as well as cardiovascular traits, type
2021 2 diabetes, and depression. Results are broadly consistent with the observational literature
2022 (See Chapter 1).

2023 The meta-analyses included 70 studies from 31 articles investigating the effect of adiposity
2024 on 34 outcomes. A majority of the 31 articles contributed to just one meta-analysis of a
2025 particular outcome. Where articles contributed to more than one meta-analysis, these were
2026 related meta-analyses i.e., an article contributing to a meta-analysis of colorectal cancer
2027 also contributes to a meta-analysis of endometrial cancer. There was strong evidence that
2028 adiposity increased many cancer types (e.g., colorectal, endometrial, lung, and ovarian
2029 cancers), cardiovascular disease (e.g., ischemic stroke, CAD, and hypertension), metabolic
2030 factors (e.g., type 2 diabetes) and neurological disorders (e.g., depression). There was
2031 weaker evidence for an effect of adiposity on breast and prostate cancer, haemorrhagic stroke,
2032 arthritis, asthma, HDL, and LDL.

2033 There was general consistency between results from the meta-analyses and the narrative
2034 synthesis. However, given there were many more studies included in the narrative synthesis
2035 their was variability within outcomes. For example, there was strong evidence for an increasing
2036 effect of adiposity on endometrial and colorectal cancer, but within the narrative synthesis
2037 there were studies which reported evidence of an increasing, protective, and null effect
2038 of adiposity on both cancers. For colon cancer there was weak evidence for an effect of
2039 adiposity, while the narrative synthesis suggested there was evidence for an increasing effect
2040 of adiposity. In the narrative synthesis, effect estimates crossed the null in many analyses of
2041 the effect of adiposity on cancers, whereas in the meta-analyses this occurred less frequently.
2042 Broadly, the narrative synthesis highlighted that associations varied depending on the type
2043 and subtype of the cancer. This is reflected in the observational literature, for example
2044 increased BMI is associated with a reduced risk of prostate cancer¹⁰⁸, but also with an
2045 increased risk of advanced prostate cancer¹¹⁰.

2046 Differences in the effect of adiposity on cancers across meta-analyses and the narra-
2047 tive synthesis are observed for many other traits. For example, the narrative synthesis
2048 suggested evidence for an increasing effect of adiposity on haemorrhagic and ischemic
2049 stroke, while meta-analyses suggest this association is present for ischemic stroke only.
2050 For the majority of cardiovascular traits however there was broad consistency across the
2051 meta-analyses and narrative synthesis for a broad effect of adiposity, including effects on
2052 SBP, CAD, and atherosclerosis. These findings are consistent with those from observa-
2053 tional studies, with evidence suggesting adiposity increases risk of CVD^{122,123}, as well
2054 as thrombosis¹²⁶, atherosclerosis¹³⁶, hypertension¹²⁹, and ischemic stroke^{142,143}. There
2055 are some inconsistencies with the observational literature however, notably for the effect of
2056 adiposity on haemorrhagic stroke, where evidence for an effect of adiposity was weak in
2057 meta-analysis but is strong in observational analyses^{142,143}.

2058 Broadly speaking, there was greater similarity between results from the meta-analyses
2059 and evidence from the narrative synthesis with the observational literature than there were
2060 differences. For instance, there is a large body of evidence for an increasing effect of adiposity
2061 on type 2 diabetes and fasting glucose¹⁵³ in the observational literature which was evident
2062 from the narrative synthesis and meta-analyses. There was also evidence in the narrative
2063 synthesis for an effect of adiposity on a broad number of metabolites which is also found in

2064 the observational literature^{157,158}. However, evidence for an effect of BMI on HDL (decrease)
2065 and LDL (increase), which is repeatedly found in observational studies^{157,158}, was weak in
2066 the meta-analysis. These inconsistencies may be a result of the unbiased estimates that MR
2067 analyses are able to obtain in comparison to observational studies.

2068 Of particular note are the effects of adiposity on depression. In meta-analysis, there
2069 was evidence for an increasing effect of BMI on depression. In the narrative synthesis there
2070 was also strong evidence for an increasing effect of BMI. In observational studies there
2071 is also strong evidence for an increasing effect of BMI on depression²²⁷. However, when
2072 excluding SNPs that are associated with the physicality of BMI (i.e., SNPs associated with
2073 adipose tissue and not behavioural change), the effect of BMI on depression was attenuated,
2074 and confidence intervals spanned the null⁴⁵³. This would suggest the effect of BMI on
2075 depression is not a result of behavioural characteristics associated with BMI, and is instead
2076 a consequence of physical changes and thus sociological factors. This example highlights
2077 the strength, and importance, of using multiple methods to obtain evidence for an effect. Of
2078 particular importance with these analyses is the prior knowledge of the genetics of BMI and
2079 the understanding that SNPs identified in GWAS do not associate with the phenotype through
2080 the same pathways, rather, these associations can be both biological and sociological.

2081 2.4.2 Quality

2082 Although not included in the meta-analyses due to sample overlap, the study by Tyrrell et
2083 al., (2019)⁴⁵³ would likely have received a high quality score due to the detailed investigation
2084 of the three MR assumptions and potential biases associated with their analysis of the effect
2085 of BMI on depression. The majority of studies included in the meta-analyses did not rank
2086 highly; one analysis was ranked as high quality and the majority were ranked as medium. The
2087 study by Dale et al., (2017)⁴²⁸ of the effect of BMI on hemorrhagic stroke only just received a
2088 high score, scoring 19 (12-19 = high quality); seven studies scored 20. The areas in which
2089 studies could have improved in their scoring was in relation to the selection of exposure
2090 and outcome samples, data availability, and statistical paramaters. Specifically, few studies
2091 provided full and accurate information on the data they used, this included not-providing a list
2092 of exposure instruments. There were also studies which did not fully or accurately report on
2093 the statistical methods and paramaters used, for example, studies employing two-sample MR
2094 rarely reported whether they allowed for the use of proxy SNPs. In addition, some studies
2095 did not report how they had identified SNPs as being independent of one another. These
2096 examples are unlikely to affect the results of an analysis, but they do call into question the
2097 reliability of results and whether they can be replicated.

2098 The analyses that ranked low did not appear to have undue influence on the resulting
2099 meta-analyses. Quality assessment focussed solely on the MR analyses and the information
2100 reported by the studies. As such, missing or incomplete information resulted in analyses
2101 scoring poorly. Complete data extraction was not possible for any MR analysis. Given

2102 that data extraction was based upon the STROBE-MR guidelines, this suggests important
2103 information was missing from all analyses. Most commonly data was not extracted because
2104 the authors did not report it. As the STROBE-MR guidelines have now been published⁴¹⁷ it is
2105 expected that the reporting quality of studies will improve, especially if journals and reviewers
2106 require a STROBE-MR checklist be reported.

2107 Many of the MR analyses included in the meta-analyses used summary statistics from
2108 publicly available GWAS. Those that did not, performed their own GWAS. The quality as-
2109 sessment did not however include evaluation of these GWAS. Given, MR analyses rely upon
2110 GWAS to identify instruments the quality of the identifying GWAS is of importance. It is clear
2111 from the reporting of the MR studies included here, especially those performing two-sample
2112 MR, that there is some confusion around GWAS papers; GWAS papers are not necessarily
2113 written with epidemiologists in mind. A number of the MR analyses incorrectly reported
2114 information. This was primarily found in the reporting of instruments. For example stating
2115 the use of BMI in European ancestries from Locke et al (2015)³⁶ but reporting data on BMI
2116 in All ancestries from Locke et al (2015). As no study provided the code, and not all studies
2117 provided detailed information on instruments (i.e. a table of SNPs), it was not possible to check
2118 whether studies incorrectly reporting instrument details also reported different instruments to
2119 those they used in their analyses. For example, some studies stated the use of instruments
2120 from Locke et al (2015) but included information from other BMI GWAS which identify different
2121 numbers of SNPs.

2122 2.4.3 Limitations

2123 This systematic review and meta-analysis is the first to investigate the causal effect of
2124 adiposity across all outcomes. In total, 173 articles performed 2,214 MR analyses, of which
2125 70 studies form 34 articles were included in meta-analyses of 31 outcomes. Although a
2126 majority of the 31 meta-analyses included just two studies, this work is the largest assessment
2127 of adipositys effects to date.

2128 Meta-analysis was only possible for 31 exposure-outcome pairs, the majority of which
2129 included just two MR analyses. This was primarily a result of overlapping outcome samples
2130 across studies which would ultimately bias results towards the observational confounded
2131 estimate. This is reflective of replication but primarily the use of meta-GWAS which incorporate
2132 prior GWAS and meta-analysis results into ever larger GWAS, for example the GWAS of
2133 BMI by Yengo et al., (2018)⁴¹ is a combination of the previous BMI GWAS by Locke et al.,
2134 (2015)³⁶ and data from UK Biobank. The limited number of analyses included in each meta-
2135 analysis prevents meaningful interpretation of heterogeneity statistics, where > 5 studies are
2136 recommended to achieve reliable estimates. As such, when fewer than 5 studies are included,
2137 the power to detect effects that are greater than the number of studies that contribute to the
2138 meta-analysis is not sufficient^{458,466}.

2139 On balance, across the meta-analyses and narrative synthesis, adiposity appeared to
2140 have an increasing effect on the majority of outcomes. There were a number of studies which
2141 showed conflicting evidence, however as few studies were available for meta-analyses it was
2142 not possible to assess publication bias through funnel plots. Additionally, a majority of studies
2143 did not perform power calculations prior to or after their analyses. As such, it is difficult to say
2144 whether studies were underpowered. In general, two-sample MR studies are well powered
2145 given the use of large publicly available summary statistics, however there are instances, for
2146 example cancer subtypes, where outcome samples (where the power in an MR analysis is
2147 derived) are low.

2148 There were some inconsistencies between evidence from the meta-analyses and narrative
2149 synthesis. This is expected to some degree due to the fact that in meta-analyses the sample
2150 size is taken into account and studies are weighted by this. Whereas, in the narrative synthesis
2151 only the direction of effect was used to summarise the effect of adiposity. Additionally, studies
2152 included in the meta-analyses were non-overlapping, whereas the narrative synthesis will have
2153 included numerous studies of the same exposure-outcome pair with overlapping samples. As
2154 a result, effects from the same population are likely repeated in the narrative synthesis and
2155 so, if a negative effect of adiposity was found for the same sample over two studies, this will
2156 have biased the summation of the overall effect of adiposity.

2157 The systematic search used a broad array of adiposity related terms in order to capture
2158 all possible adiposity exposures. As there was no restriction on the type of adiposity exposure
2159 used, a large body of work was identified. Although 31 adiposity exposures were identified,
2160 the majority (68%) of MR analyses used BMI as the exposure. Although BMI shows similar
2161 relationships to other anthropometric measures with many diseases⁶³, it does not accurately
2162 reflect body composition, and observational studies have highlighted the potential role for
2163 fat deposition, as well as overall fat mass, in the development of many diseases. Obtaining
2164 evidence from multiple measures of adiposity, which capture variation in body composition
2165 and fat deposition in different ways, can prove informative in assessing the underlying
2166 mechanisms of associations. For instance, in the meta-analyses of type 2 diabetes, colorectal
2167 cancer, and coronary artery disease BMI showed similar results to WHR and WHRadjBMI.
2168 Where evidence is consistent across adiposity measures, in particular consistency of body
2169 composition measures and fat depositions measures as in this example, this is likely to
2170 suggest that fat deposition is not as important as overall body composition in the association
2171 with disease. If however the effect of WHR was found to be stronger than BMI in the
2172 association with colorectal cancer this would suggest that fat deposition plays a potentially
2173 more important role in disease development. In the case of type 2 diabetes, colorectal cancer,
2174 and coronary artery disease there was little difference in the size of effect estimates, though
2175 the effect of BMI resulted in tighter confidence intervals.

2176 Although MR studies are robust to confounding and other biases (See 1.8), they are
2177 subject to a number of limitations pertinent to this systematic review. Results of MR studies
2178 may represent different underlying processes to that of observational studies, as exposures in
2179 MR studies reflect a lifetime exposure. In observational studies, the exposures are determined

2180 by genetic and non-genetic factors at that point in time. Additionally, genetic instruments
2181 used in MR analyses must be robust and appropriate. Given the incomplete, and often poor
2182 reporting of MR analyses, results must be interpreted cautiously. This is especially true in
2183 regards to instruments with many hundreds of SNPs, which although are likely to be robust
2184 and appropriate, will undoubtedly contain many pleiotropic SNPs. This is because, few studies
2185 actively investigated pleiotropy outside of sensitivity models such as MR-Egger. Studies were
2186 excluded from meta-analysis if there was overlap between the outcomes of studies or between
2187 the exposures and outcomes of studies. However, it is likely this was not completely accurate
2188 given not all studies reported the cohorts used in their analyses. Additional limitations of MR
2189 analyses, discussed in Chapter 1 (Section 1.8), including homogeneity and monotonicity may
2190 be especially important in these analyses given effects among different populations may not
2191 be homogeneous (i.e., the effect of the IV or exposure is not the same for all populations) or
2192 monotonic (i.e., the effect of the IV is differential among populations). The main challenge in
2193 appraising MR assumptions is the quality (including written quality) of the studies, but with
2194 the implementation of the STROBE-MR guidelines it is hoped this will improve.

2195 2.4.4 Conclusion

2196 In this systematic review and meta-analysis, adiposity is shown to exert its effect on
2197 numerous outcomes including many cancers, cardiovascular outcomes, and many metabolic
2198 traits. Meta-analyses of 31 exposure-outcome pairs highlighted predominantly increasing
2199 effects of BMI. Results are broadly consistent with the observational literature and provide
2200 corroborative evidence for association with a number of traits including endometrial cancer.
2201 Evidence, from meta-analyses and the narrative synthesis, which was corroborated by
2202 observational studies, was particularly strong for the effect of BMI on endometrial and
2203 colorectal cancer, as well as CAD. There is also evidence that these diseases are associated
2204 with metabolic changes (Chapter 1 Section 1.9). The recent availability of a large GWAS for
2205 endometrial cancer⁴⁶⁷ will enable the investigation of the potential intermediary effects of
2206 metabolites in the adiposity endometrial cancer relationship in Chapter 6.

2207 Even though there were conflicting results in the narrative synthesis for a number of
2208 outcomes, these results should be taken with caution as these summations focussed primarily
2209 on the direction of effect, unlike the meta-analyses which included weights based on sample
2210 size. The lack of high quality studies, and an abundance of missing and incorrect data
2211 reported by studies, limits the inferences that can be made from results. In particular, the
2212 limited number of studies included in the meta-analyses prohibits meaningful interpretation
2213 of heterogeneity statistics. Inclusion of non-independent samples in the narrative synthesis
2214 means results must be interpreted cautiously. Taken together, meta-analysis results are
2215 useful but only if the studies have been conducted appropriately as to mitigate any bias. As
2216 the majority of studies were not of high quality, focus should be given evidence from the
2217 meta-analyses and narrative synthesis fits with the wider literature, including observational
2218 analyses. Given many MR analyses are replications which use the same data-sets, future

2219 meta-analyses will become increasingly difficult without the ability to separate out cohort
2220 specific estimates. There is thus a need for future studies to (i) replicate their work in
2221 independent sources (ii) or use datasets that are independent of previously published results.

2222 Chapter 3

2223 EpiViz: a tool to visualise large 2224 association analyses

2225 Chapter summary

2226 The exploration, interpretation, display, and communication of large association analyses
2227 is challenging. Previous studies have used Circos plots, however the process of producing
2228 Circos plots is cumbersome and, through personal experience⁴⁶⁸, has been time consuming
2229 and inefficient. This chapter details EpiViz, a web application and R package that efficiently
2230 produces Circos plots. EpiViz is used in the following chapters to gain a global overview
2231 and identify patterns of association in observational and Mendelian randomization analyses.
2232 EpiViz considerably improves the speed and efficiency of producing Circos plots and has
2233 been used in multiple studies^{1,468,469} since its development. EpiViz is available on [GitHub](#).

2234 **3.1 Introduction**

2235 Large epidemiological analyses, involving potentially hundreds of exposure-outcome
2236 associations, pose a significant challenge for the digestion and interpretation of results. Data
2237 visualization is key to improving the exploration, interpretation, communication, and display
2238 of data⁴⁷⁰. Traditionally, forest plots have been used to effectively communicate association
2239 analyses and, have been used successfully for analyses involving metabolites^{25,382,468,471-473}.
2240 These studies have however been limited to a handful of exposure-outcome associations.
2241 When dealing with hundreds of variables forest plots can become cumbersome, requiring
2242 many separate plots to be created in order to present all of the analyses. This makes
2243 comparison and the ability to gain global overview of large analyses difficult.

2244 An alternative approach is to compress information into a circular rather than vertical
2245 or horizontal form. Circos plots⁴⁷⁴, as implemented in many genetics studies to condense
2246 large genomic information into informative visuals⁴⁷⁵, provide an efficient visualisation tool for
2247 incorporating hundreds to thousands of data points. Circos plots have been used effectively to
2248 communicate results from large metabolomic analyses^{1,317,318,468,469,476}. However, the Circos
2249 software is designed for genomic data and written in programming languages unfamiliar to
2250 many epidemiologists. The R package *Circlize*⁴⁷⁷ provides most of the functionality of the
2251 Circos software and is publicly available. However, it is not designed for association analyses
2252 as performed in this thesis. As such, it can be time consuming and inefficient to produce
2253 Circos plots.

2254 In order to efficiently and reproducibly create Circos plots that enable the exploration,
2255 interpretation, communication, and display of large epidemiological analyses as conducted in
2256 this thesis, *EpiViz* was developed. *EpiViz* is a web application and R package, that builds
2257 on the *Circlize*⁴⁷⁷ and *ComplexHeatmap*⁴⁷⁸ R packages. It enables data to be presented in
2258 a compact form that allows for the efficient identification of profile changes in metabolomic
2259 data specifically for this thesis, as well as global patterns of associations in high-dimensional
2260 epidemiological data.

2261 **3.2 Methods**

2262 **3.2.1 Circos plots for association analyses**

2263 *EpiViz* composes Circos plots using six elements: template, plotting space, tracks,
2264 sections, data, and legend (Figure 3.1). The template element is a square of defined
2265 proportions within which information is plotted. Each additional element is layered onto the
2266 template one after the other. The plotting space element is an empty circle which is layered
2267 and centred on top of the template. Data is plotted on to the plotting space. An optional extra

2268 of the Circos plot, the legend element, takes the dimensions of the template and creates a
2269 separate plotting space that can be layered on to the bottom of the template element.

2270 The plotting space is separated into tracks and sections. Tracks are laid down as rings
2271 within the plotting space. Each track represents a single element of information such as an
2272 exposure or method. Tracks are numbered from the outside to the centre of the circle and
2273 can be coloured separately. Sections divide the plotting space into distinct areas, much like a
2274 pie chart. Sections are defined by a grouping variable such as a metabolite profile (e.g., class
2275 or sub class). A section track is placed at the outside of the tracks to give a header for each
2276 section. The header is referenced in the legend element.

2277 Once the template, plotting space, tracks and sections are laid down, coordinates for each
2278 section and track location can be called to plot the data element. Each track and section
2279 coordinate, e.g., track 2 section 3, is treated as an individual plotting space. As such, data
2280 can be plotted based on the following coordinates: track, section, X , Y . The X axis of each
2281 track is defined by the number of rows in the data frame, i.e., a data frame with 100 rows will
2282 have an X axis of length 100, with each row given an X axis coordinate from 1–100. The Y
2283 axis is defined by the minimum and maximum of the data for that track. As such, each track
2284 and section coordinate, e.g., track 2 section 3, can be considered an individual plot with a Y
2285 axis that is shared by all of the sections in that track. For each position on the X axis, the
2286 label element of each row is plotted outside of the section header.

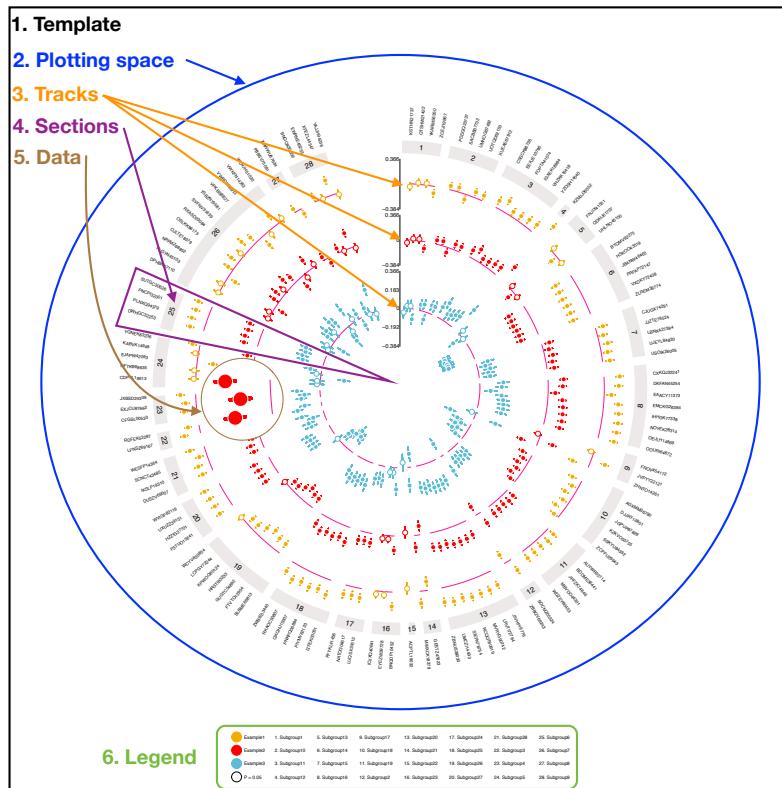


Figure 3.1: **Circos plot highlighting the six elements used to plot data.** The Circos plot shows three tracks and 28 sections of data simulated to give an example of the effect of three exposures on over 100 outcomes.

2287 3.2.2 Implementation

2288 EpiViz is a Shiny web application and R package. R⁴⁷⁹ version 3.6.2 and Shiny⁴⁸⁰
 2289 version 1.4.0 were used to develop the web application; R version 3.6.2 was used to develop
 2290 the R package. Shiny is an R package that enables the development and deployment of web
 2291 applications written in the R programming language. Development of EpiViz was progressive
 2292 and feedback from colleagues was vital in this process.

2293 3.2.3 Operation

2294 The **web application** is publicly accessible and held under an **MIT license**. The web
 2295 application has been tested on computers running macOS (version 10.14) and Windows
 2296 (version 10) using: Internet Explorer (version 11; Windows), Google Chrome (version 79;

2297 macOS and Windows), Safari (version 13; macOS).

2298 The [R package](#) is publicly accessible through GitHub and held under an [MIT license](#). The
2299 R package is accessible on all computers with R version 3.3.0 or higher and has been tested
2300 on macOS (version 10.14) and Windows (version 10) running R version 3.3.0 or higher.

2301 A legend function is available for both the web application and R package and is imple-
2302 mented using functions from the [ComplexHeatmap](#)⁴⁷⁸ R package. By default, the colours used
2303 for the Circos plot in both the web-application and R package are accessible colours identified
2304 using [i want hue](#). Data (example data), of the effect of BMI on metabolites from Kettunen
2305 et al., (2016)³¹⁸ using male, female, and sex combined instruments, is provided on the web
2306 application *Home* tab and with the R package using the function `EpiViz::EpiViz_data*`()
2307 where * is 1-3. Example data can be produced for use with the web-application and R package
2308 using code on [GitHub](#).

2309 **Web application**

2310 In order to use the [web application](#), a web-browser and an internet connection of at
2311 least 1Mbps is required. No other system requirements are needed. Upon opening the web
2312 application, users are shown example Circos plots created using simulated and the example
2313 data described above, and are directed towards the *Home* tab. The *Home* tab provides users
2314 with a short summary of the application, a link to the R package, and example data for use
2315 with the app.

2316 The *How to* tab provides instructions for using the application. Step 1 deals with the
2317 preparation of data to be used with the application, Step 2 deals with how to use the
2318 application, and Step 3 provides information on potential next steps. Users are instructed to
2319 upload one data frame per track of the Circos plot. Each data frame should be a tab delimited
2320 text file and R code is provided for users to achieve this. The user is guided through an
2321 example utilizing the example data.

2322 Once userse have uploaded their data, descriptive information, including a volcano plot,
2323 will be produced automatically. Users then select the *Plot* tab, from which they choose the
2324 names of the columns for the: label, group, estimate, p-value, lower confidence interval, and
2325 upper confidence interval. A p-value adjustment (X) can be provided to indicate a p-value
2326 threshold in the format $\frac{0.05}{X}$. On the Circos plot, a solid point is indicated as reaching the
2327 p-value threshold. An optional legend function is provided and users can choose the labels
2328 for the legend points. The legend is auto-populated using the levels in the group column
2329 of the uploaded data frame. Finally, an option to use colours accessible to individuals with
2330 colour impaired sight is provided.

2331 **R package**

2332 The EpiViz R package is accessed using R version 3.3.0 or above. Documentation for
2333 using the package is available as a README on [GitHub](#). The README includes use cases and
2334 troubleshooting. The R package can be installed and loaded into the current R session directly
2335 from GitHub into with the following code:

```
# Install devtools
install.packages("devtools")
library(devtools)

# Install directly from GitHub
devtools::install_github("mattlee821/EpiViz/R_package")
library(EpiViz)
```

2336 Once installed, users can use the example data provided with the package to produce
2337 Figure 3.2, which illustrates the three different types of track available (point, line, bar), using
2338 the following R code:

```
circos_plot(track_number = 3,
            track1_data = EpiViz::EpiViz_data1,
            track2_data = EpiViz::EpiViz_data2,
            track3_data = EpiViz::EpiViz_data3,
            track1_type = "points",
            track2_type = "lines",
            track3_type = "bar",
            label_column = 1,
            section_column = 9,
            estimate_column = 2,
            pvalue_column = 3,
            pvalue_adjustment = 1,
            lower_ci = 4,
            upper_ci = 5,
            lines_column = 2,
            lines_type = "o",
            bar_column = 2)
```

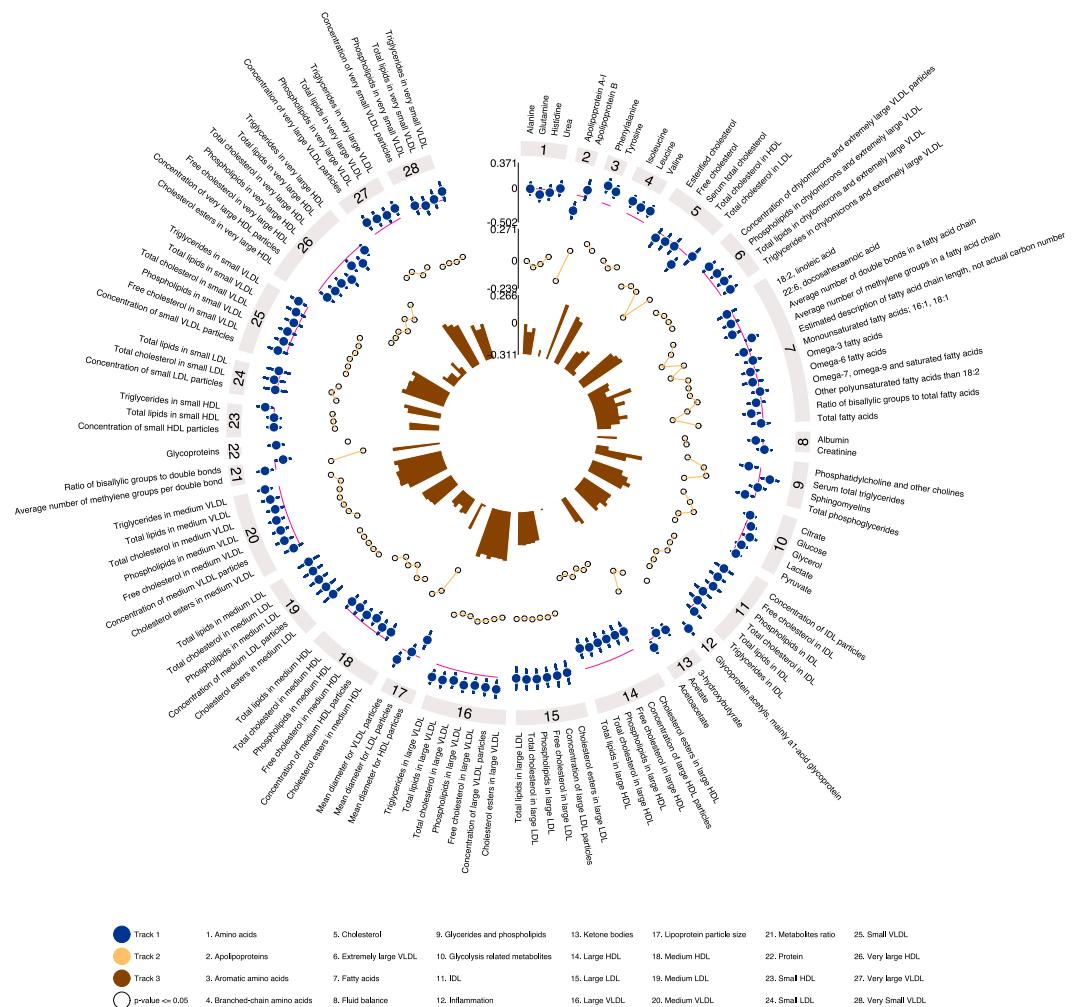


Figure 3.2: **Example Circos plot using EpiViz R package and example data.** Data presented are results from Mendelian randomization analyses of the effect of BMI on metabolites from Kettunen et al., (2016)³¹⁸ using male (blue), female (yellow), and sex combined (brown) instruments.

2339 In producing plots using the R package, users are advised to save as PDF using the below
 2340 code. This is because the Circos plot is designed to be larger than normal plots, as such
 2341 viewers like the R Studio Viewer pane display the plot in a compressed form, squashing the
 2342 plot. Viewers like this should be used only as a guide when making the plot. PDF files can be
 2343 converted to other image formats without compression. Users are advised to iterate the sizing
 2344 of the plot, adjusting the width and height arguments to get the desired plot size and then

2345 adjust the point size argument to increase and decrease the size of the information in the plot
2346 (labels, points, lines etc.). The values provided in the below code work with most plots and
2347 were used for all examples in this Chapter. In addition, users can adjust the size of the Circos
2348 plot directly using the argument `circle_size`. The default for `circle_size` is 25, smaller
2349 numbers increase the size of the circle and larger numbers decrease the size of the circle:

```
pdf("my_circos_plot.pdf",
  width = 30, height = 30, pointsize = 35)
circos_plot(
  circle_size = 25)
dev.off()
```

2350 Finally, users can adjust the height of tracks individually using `track*_height` where *
2351 refers to the track number. The default for each track is 20 percent of the total size of the
2352 circle (the section track is fixed at 5 percent):

```
circos_plot(
  track1_height = 0.2,
  track2_height = 0.2,
  track3_height = 0.2)
```

2353 In order to minimise the time required to maintain the R package, further customisation
2354 is achieved by the users themselves. The function is written to aid this customisation with
2355 *Default parameters* and *Customisable parameters* located at the top of the `circos_plot()`
2356 function. Guidance and code is provided on [GitHub](#) to aid user customisation, for example,
2357 changing track colours requires users to navigate to the *Customisable parameters* section
2358 and then to the *Colours* section and replace the following lines with their desired colours:

```
discrete_palette <- c("#00378f", # track 1 colour
  "#ffc067", # track 2 colour
  "#894300") # track 3 colour
```

2359 3.3 Discussion

2360 Circos plots have been used to visualise and provide overview of large metabolomic
2361 association analyses^{1,317,318,468,476}. The creation of Circos plots for Taylor et al. (2019)⁴⁶⁸
2362 was cumbersome and inefficient with each Circos plot requiring bespoke code and revisions
2363 were time consuming. EpiViz simplifies and streamlines the process of making Circos plots. In
2364 Bos et al. (2021)¹, EpiViz was used to summarise and visually compare observational and

2365 Mendelian randomization analyses of multiple sleep traits on over 100 metabolites. EpiViz
2366 made producing and refining Circos plots for this analysis faster and more efficient.

2367 The web application is intended for researchers with limited to no experience of R and
2368 is in a stable release. Additions to the web applications functionality are therefore not
2369 envisioned. The R package, also in a stable release, will be the focus of further development
2370 as maintenance costs are lower as there is no requirement for the additional coding of the user
2371 interface as with the web application. The focus of future changes should be on converting
2372 the current code style to the [tidy style](#) to improve readability and consistency. This will also
2373 improve the ease with which contributions can be implemented. Additional features should be
2374 directed by the needs of the users, this is likely to include additional plotting mechanics such
2375 as the chord diagram, which shows relatedness between nodes. In the case of metabolites, a
2376 chord diagram would provide greater understanding about the relationship between different
2377 sections. In addition, the ability to filter and choose specific sections to display would simplify
2378 the presentation of key results – currently this is achieved with supplementary forestplots.

2379 In order to achieve the desired goal of providing global overview of large association
2380 analyses, Circos plots are larger than traditional plots. This poses a challenge for their use
2381 in print as there are size restrictions. Circos plots are therefore ideally suited to online use,
2382 however, if large enough, and with proper sizing of text, they can be used as overview figures
2383 in printed media. When used online there is the potential for creating interactive Circos plots,
2384 allowing users to expand and filter the Circos plot as desired. Interactive plots are can be
2385 used in [online publications](#), and can enhance the reader experience, allowing readers to gain
2386 a better understanding of the presented research.

2387 EpiViz is simple and efficient to use. The web app provides a platform for quick and
2388 simple plots to be generated while the R applicaton provides customisation. EpiViz has been
2389 successfully used to interrogate large metabolomic association analyses¹ and will be used
2390 throughout this thesis to achieve the same goals.

2391 **Chapter 4**

2392 **Associations between multiple** 2393 **measures of adiposity and** 2394 **metabolites: observational analysis**

2395 **Chapter Summary**

2396 In Chapter 1 and 2, the link between adiposity and disease was presented both in obser-
2397 vational and causal analysis frameworks. This work highlighted that adiposity is associated,
2398 likely causally, with many outcomes. A key takeaway from Chapter 1 was that the underlying
2399 mechanisms of disease development are not well understood, but that some adiposity-disease
2400 relationships may be explained by changes in metabolic, immune, and protein pathways.
2401 In this chapter, observational epidemiological methods are used to explore associations
2402 between one of the potential pathways linking adiposity to disease, metabolites. The aim of
2403 this chapter is to provide an observational grounding for subsequent causal analysis work in
2404 Chapter 5 and 6.

2405 **4.1 Introduction**

2406 Adiposity is associated with an increased risk of numerous diseases, as well as
2407 mortality^{71,75–77,89–94} (see Chapter 1 and 2). There is a need to understand the mechanisms
2408 underlying these associations so that intervention strategies, which are challenging to
2409 implement effectively, can be targeted and more efficacious.

2410 Adiposity has been linked with many downstream measures that could serve as interme-
2411 diates of disease and thus be useful targets for reducing the burden of adiposity. Notably,
2412 previous work has highlighted metabolites^{88,274,275} as a possible pathway linking adiposity
2413 with diseases such as cancer¹⁰⁸, coronary heart disease (CHD)^{122,123}, and type 2 diabetes¹⁵³.
2414 Metabolites are intermediate or end products of cellular processes with multiple functions
2415 including as energy, signalling, transportation, and structural components. Metabolic effects
2416 can be far reaching^{311,312} and during homeostasis are tightly controlled. The many func-
2417 tions metabolites play mean that imbalances can be detrimental^{308,311,312}. Measurement of
2418 metabolites has become increasingly common among cohort studies with the advancement,
2419 including reduction in costs, of mass spectrometry (MS) and nuclear magnetic resonance
2420 (NMR) platforms able to perform targeted, semi-targeted, and untargeted assays.

2421 Adiposity has been associated with many metabolites^{157,158,328,363–366,368–372} and, in the
2422 largest study to date by Wurtz et al., (2014)¹⁵⁷, influences whole classes of metabolites
2423 including amino acids, fatty acids, hormones, inflammatory markers, and lipids. Although
2424 there is evidence to show these associations vary between sexes and over time¹⁵⁷, the focus
2425 of many analyses has been on body mass index (BMI) as a measure for overall adiposity, has
2426 included small sample sizes, and has looked at a single time-point.

2427 While BMI correlates well with, and is a predictor of, many health outcomes^{61,62}, it is a
2428 crude measure of adiposity with several issues, not least the inability to differentiate lean and
2429 fat mass⁶⁴, and evidence highlights the importance and utility of combining complimentary
2430 assessments of adiposity using multiple measures^{83,88}. As many studies have focussed on
2431 the effect of adiposity on metabolites in adulthood and given the effects of adiposity are shown
2432 to present early in life^{481–485}, investigation of the effect of adiposity in younger ages may
2433 provide additional information on the association with metabolites. The Avon Longitudinal
2434 Study of Parents and Children (ALSPAC), a longitudinal birth cohort study, with repeated
2435 metabolomic and anthropometric measures, provides an opportunity to expand on the current
2436 literature. Here, observational analysis of measures of adiposity and metabolites provide a
2437 basis from which to investigate causal associations. In addition, comparison with previous
2438 work by Wurtz et al.,¹⁵⁷ is possible for a number of metabolites.

2439 **4.2 Methods**

2440 Data were available for exposures (measures of adiposity), outcomes (metabolites), and
2441 potential confounders from ALSPAC. Exposures included body mass index (BMI), waist hip
2442 ratio (WHR) and body fat percentage (BF). Metabolomics data were available for up to 234
2443 metabolites. These metabolites include directly measured metabolites such as the amino
2444 acids tyrosine and phenylalanine, as well as derived (not-directly measured) metabolite
2445 measures such as the ratio of saturated fatty acids to total fatty acid. Covariate data included
2446 age, sex, mothers or own education, smoking history, alcohol history, diet, and physical activity.
2447 All analyses and data manipulation were performed using R (version 3.6.2)⁴⁷⁹. Specific R
2448 packages are described where appropriate. All code for this work is available on [GitHub](#).

2449 **4.2.1 Data overview**

2450 The Avon Longitudinal Study of Parents and Children^{486–488} is a prospective cohort study
2451 that invited women resident in Avon, UK with expected dates of delivery between 1st April
2452 1991 and 31st December 1992 to participate. The initial number of pregnancies enrolled
2453 was 14,541 (for these at least one questionnaire has been returned or a “Children in Focus”
2454 clinic has been attended by 19/07/99). Of these initial pregnancies, a total of 14,676 foetuses,
2455 resulted in 14,062 live births and 13,988 children alive at one year of age. The mothers
2456 and fathers associated with each pregnancy are referred to as generation 0 (G0) while the
2457 children of each eligible pregnancy (including individuals from subsequent recruitment drives)
2458 are referred to as generation 1 (G1).

2459 When the oldest G1 individuals were approximately seven years of age, an attempt
2460 was made to bolster the initial sample with eligible cases who had failed to join the study
2461 originally. As a result, when considering variables collected from the age of seven onwards
2462 (and potentially abstracted from obstetric notes) there are data available for more than the
2463 14,541 pregnancies mentioned above. The number of new pregnancies not in the initial
2464 sample (known as Phase I enrolment) that are currently represented on the built files and
2465 reflecting enrolment status at the age of 24 is 913 (456, 262, and 195 recruited during
2466 Phases II, III, and IV respectively), resulting in an additional 913 G1 individuals being enrolled.
2467 The phases of enrolment are described in more detail in the cohort profile paper and its
2468 update^{486–488}. The total sample size for analyses using any data collected after the age of
2469 seven is therefore 15,454 pregnancies, resulting in 15,589 foetuses, of which 14,901 were
2470 alive at one year of age.

2471 The study [website](#) contains details of all the data that is available through a fully search-
2472 able data dictionary and [variable search tool](#). Ethical approval for the study was obtained
2473 from the ALSPAC Ethics and Law Committee and the [Local Research Ethics Committees](#).
2474 Informed consent for the use of data collected via questionnaire and clinics was obtained

2475 from participants following recommendations of the ALSPAC Ethics and Law Committee at
2476 the time. Full details of the ALSPAC consent procedures are available on the [study website](#).

2477 Data in ALSPAC are split by clinic visits. For this work, metabolomics data were available
2478 for G1 individuals from the following clinics: Focus at 7 (~8 years old), Focus at 8 (~9 years
2479 old), Before Breakfast Study (~8 years old), Teen Focus 3 (~18 years old), Teen Focus 4
2480 (~17 years old), and Focus at 24 (~24 years old). Metabolomics data for G0 individuals were
2481 available from: Focus on Mothers 1 (~48 years old), Focus on Mothers 2 (~51 years old), and
2482 Focus on Fathers 1 (~53 years old). All data on exposures and covariates were obtained from
2483 the same clinic from which metabolomics data were collected. Where data on exposures
2484 and covariates were not available at the metabolomics clinic visit they were obtained from
2485 the most recent clinic with available data. The Before Breakfast Study, unlike the other
2486 clinics, only collected metabolomics data, as such data on exposures and covariates were
2487 extracted for these individuals from the Focus at 8 clinic. Metabolomics data for each time
2488 point were extracted first and subsequent data on exposures and covariates were extracted
2489 for individuals with metabolomics data. The metabolomics data were provided with standard
2490 exclusions for identifiable individuals and those with withdrawn consent already excluded.

2491 4.2.2 Exposures: adiposity

2492 Measures of adiposity (BMI, WHR, and BF) were obtained for all individuals with available
2493 metabolomics data; identifiable individuals and those with withdrawn consent were removed
2494 when obtaining the metabolomics data. Data were obtained from the same clinics as data for
2495 the metabolomics data, e.g., metabolomics data was obtained from Focus at 7 and so was
2496 anthropometric data for the same individuals. No anthropometric data were available for the
2497 Before Breakfast Study, so the Focus at 8 clinic, as the most age appropriate clinic, was used
2498 instead. Data on WHR was not available for the Teen Focus 3 and 4 clinics.

2499 BMI was calculated as $\frac{weight(kg)}{height(m^2)}$ and WHR as $\frac{waist\ circumference\ (cm)}{hip\ circumference\ (cm)}$. Height
2500 was measured to the last complete mm using the Harpenden Stadiometer. Individuals were
2501 positioned with their feet flat and heels together, standing straight so that their heels, calves,
2502 buttocks and shoulders came into contact with the vertical backboard of the stadiometer. The
2503 headboard was lowered until it touched the individuals head and a 1Kg weight was placed on
2504 the headboard to ensure head contact and to minimise hair thickness. Weight was measured
2505 using the Tanita Body Fat Analyser (Models TBF 305 and 401A) or electronic bathroom scales,
2506 if the individual had a pacemaker. Individuals were encouraged to pass urine and undress to
2507 their underclothes. For all G1 individuals 'Female Standard' was entered into the machine.
2508 For all individuals and their height was entered to the nearest cm. Individuals stepped onto
2509 the measuring platform which had been wiped with disinfecting alcohol and positioned so
2510 that both feet were located in parallel with the toe and heel in contact with their respective
2511 electrodes. Measurement was completed when the weight and fat ratio readings were fixed

2512 and the buzzer beeped. Weight was measured to the nearest 50g for G1 individuals and to
2513 the nearest 0.1kg for G0 individuals. Circumferences were measured using the Seca 200 or
2514 201 body tension tape and were repeated twice for accuracy.

2515 BF was measured using dual-energy x-ray absorptiometry in all individuals except for indi-
2516 viduals from Foucs at 7. Briefly, measurement required individuals to be prone and stationary
2517 while a Lunar prodigy narrow fan beam densitometer performed a whole body DXA scan. Data
2518 were processed using Lunar Prodigy software. Individuals did not have measurements taken
2519 if they: were pregnant, had a radiological investigation using contrast media within the week
2520 before the DXA scan, had a recent nuclear medicine investigation with persistent radioactivity,
2521 weighed greater than 159kg. BF was calculated as
$$\frac{\text{fat mass}}{\text{fat mass} + \text{fat free mass}} \times 100.$$

2522 For Focus at 7, BF was not measured, instead bioelectrical impedance data, which were
2523 converted to estimates of BF, were available. Briefly, individuals were encouraged to pass
2524 urine and undress to their underclothes. A Tanita Body Fat Analyser (Model TBF 305) was
2525 used to measure weight and impedance. Height was entered to the nearest cm and 'female
2526 standard' was used as the sex variable for all individuals. The Tanita Body Fat Analyser TBF
2527 305 is a single frequency (50kHz) leg-to-leg device. In single frequency devices, impedance
2528 is a representation of resistance which is related to the volume of water (which one assumes
2529 makes up the majority of fat free mass (FFM)), as such, the higher the resistance/impedance
2530 the greater the amount of FFM. Calculation of BF from the impedance measure is only
2531 possible at the time of measurement, however these derived BF measures were not stored
2532 and the equation to calculate them was not available from the manufacturer. Previous
2533 work⁸⁶ has shown that comparison of BF derived from the manufacturer's equation and an
2534 alternative⁸⁵ showed little difference in resulting BF estimates. The equation was derived in a
2535 study involving 205 (101 women) healthy adults with a mean age of 43.8 (SD = 16) for men
2536 and 40.4 (SD = 13.6) for women. The equation, where Z is the impedance measure from the
2537 device in ohms, height is in meters, weight is in kilograms, age is in years, and female-specific
2538 components are given as $19.6 + \ln(\text{height})$, is given as:

$$\begin{aligned} BF = & -156.1 - 89.1 \ln(\text{height}) \\ & + 45.6 \ln(\text{weight}) \\ & + 0.120 \text{ age} \\ & + 0.0494 Z \\ & + (19.6 \ln(\text{height})) \end{aligned} \tag{4.1}$$

2539 Given that the equation was derived from adult data, its application to child data was explored.
2540 A raw impedance measure, from a similar model (Tanita Body Fat Analyser (Model TBF
2541 401A)), were obtained for individuals from Teen Focus 3 and 4 (i.e., where both DXA and
2542 impedance measures were available) and the equation was used to compare BF derived
2543 from the impedance device and BF measured with DXA in adolescents. Exploration involved

2544 visual inspection of distribution and Spearman's correlation with BMI, height, weight and other
2545 BF measures from Teen Focus 3 and 4. The same observations were carried out for raw
2546 impedance. The calculated BF estimates correlated well with height and weight, however they
2547 resulted in negative estimates of BF. In a linear model the estimate is based on the per-unit
2548 increase, the absolute value of the exposure therefore does not need to be positive. As such,
2549 BF calculated using equation (4.1) was used in subsequent analyses as a measure of BF in
2550 children.

2551 4.2.3 Outcomes: metabolites

2552 Metabolomics data were measured using the same NMR platform for all individuals. Briefly,
2553 high-throughput proton (^1H) NMR assays were performed on ethylenediaminetetraacetic acid
2554 (EDTA) plasma/serum samples. Samples were fasted. Measurements were taken at three
2555 molecular windows (lipoprotein lipids, low molecular-weight metabolites, and lipid extracts)
2556 enabling broad quantification of metabolomic measures. These measures also included
2557 derived measures, lipoprotein particle sizes, and fatty acid ratios, inclusion of which has
2558 shown to increase overall power in statistical analyses^{489–491}. Full details on the NMR
2559 methodology has previously been described^{476,492–494} and is available from the [ALSPAC](#)
2560 [data dictionary](#) (data dictionary identifiers: children = D5704, mothers = D5705, fathers =
2561 D5700). The Before Breakfast study does not have a documentation file and is described
2562 elsewhere⁴⁹⁵; fasted metabolomics data, not the post glucose challenge metabolomics data,
2563 were taken from The Before Breakfast Study. Descriptions of metabolites are available in the
2564 Appendix ??.

2565 The spectral NMR data were processed by Nightingale Health and provided as a pro-
2566 cessed file with identifiable individuals (triplets/quadruplets) and individuals who had with-
2567 drawn consent removed. Some mothers and fathers were duplicated in this data due to the
2568 way in which mothers were originally enrolled into the study and assigned IDs. If a mother
2569 enrolled with two different pregnancies (both having an expected delivery date within the
2570 recruitment period (April 1991–December 1992)), she will have two separate IDs. A father
2571 associated with both of these pregnancies will also be duplicated. Duplicate measurements
2572 for mothers and fathers were removed.

2573 In order to maximize the sample size at each clinic, data were combined where clinics
2574 were within a similar age range. For these combined data sets, duplicate individuals (i.e.,
2575 those attending both clinics) were identified, and the measurement from the most recent clinic
2576 was dropped. No metabolites were excluded at this stage, however the number of metabolites
2577 available for each clinic visit differed as metabolites are added and removed over time by
2578 Nightingale as validations change.

2579 Pre-analysis processing of metabolomics data is important as there can be quality issues
2580 with sample and metabolite data⁴⁹⁶. However, there is no standardised method for performing

2581 pre-analysis processing and for deciding what thresholds to use to exclude metabolites
2582 and individuals from downstream analyses. The R package `metaboprep`⁴⁹⁶ can be used to
2583 process data before analyses using a transparent and reproducible workflow. Here, after
2584 combining clinic visit data where appropriate, `metaboprep` (version 0.0.1) was used to identify
2585 and exclude individuals and metabolites that did not meet certain requirements. This process
2586 was performed twice, firstly including and secondly excluding the derived metabolomics
2587 measures from missingness and clustering analyses to see if these derived measures were
2588 unduly influencing exclusions. Briefly, individuals and then metabolites, with high missingness
2589 ($\geq 80\%$) were removed. Missingness was then re-calculated for individuals and metabolites,
2590 with removal based on 20% missingness. Individuals were then removed based on total
2591 sum abundance, considering outliers as ≥ 5 standard deviations (SD) away from the mean.
2592 Using this metabolite dataset, a dendrogram based on a Spearman's rho distance matrix
2593 was produced, and a set of clusters identified based on a Spearman's rho of 0.5. For each
2594 cluster, the metabolite with the least missingness was tagged as the representative feature.
2595 Finally, principal component (PC) analysis was conducted using the representative features
2596 to evaluate structure among individuals. Outliers were identified as being ≥ 5 standard
2597 deviations away from the mean of PC 1 and 2, and were excluded. Additionally, in order to
2598 gain an overview of the variability of metabolite concentrations, the mean, SD of the mean,
2599 median, and range of metabolite concentrations were visually compared across age groups.

2600 4.2.4 Covariates

2601 Evidence has shown that age^{497,498}, sex⁴⁹⁸, education⁴⁹⁹, smoking⁵⁰⁰, alcohol⁵⁰⁰, diet⁵⁰¹,
2602 and physical activity⁵⁰² all influence the metabolomic profile and adiposity. Data for each of
2603 these covariates were obtained for all individuals with metabolomics data. Age was taken from
2604 the metabolomics clinic visit as months since birth. Sex was taken from the initial assessment
2605 questionnaire for G1 individuals completed by their parents. For G0 individuals sex was
2606 self-reported at the metabolomics clinic visit.

2607 Maternal education was used for children, adolescents and young adults. Own education
2608 was used for mothers and mother reported partner education was used for fathers. Specifically,
2609 mothers were asked, during their pregnancy, 'What educational qualifications do you, your
2610 partner, your mother, and your father have?' with possible answers: CSE or GCSE (D, E, F or
2611 G); O-level or GCSE (A, B, or C); A-level; qualifications in shorthand/ typing/or other skills
2612 e.g. hairdressing; apprenticeship; state enrolled nurse; state registered nurse; City and Guilds
2613 intermediate technical; City & Guilds final technical; City & Guilds full technical; teaching
2614 qualification; university degree; no qualification; qualifications not known; not applicable; other
2615 (please describe). This data was available as a recoded variable of 5 categories of lowest (1)
2616 to highest (5) level of education.

2617 Smoking was binary; adolescents (at the metabolomics clinic), young adults (at the
2618 metabolomics clinic), and adults (mothers were asked during pregnancy; fathers were asked

2619 in at a clinic prior to the metabolomics clinic) were asked whether they had ever smoked a
2620 cigarette before.

2621 Alcohol was assessed differently for Teen Focus 3 and Teen Focus 4, Focus at 24, and
2622 G0 individuals. Individuals from Teen Focus 3 were asked what their alcohol drinking pattern
2623 was with possible answers: only ever tried drinking once/twice, used to drink sometimes
2624 [but] never drink now, sometimes drink but less than once a week, usually drink on 1/2 days
2625 a week, usually drink on >2 days a week but not every day, and usually drink every day.
2626 Individuals from Teen Focus 4, Focus at 24, and G0 individuals (mothers were asked at a
2627 prior clinic visit to the metabolomics clinic and fathers at the metabolomics clinic) were asked
2628 the frequency they had drinks containing alcohol with possible answers: never, monthly or
2629 less, two to four times a month, two to three times a week, four or more times a week.

2630 Diet data derived by Anderson et al., (2013)⁵⁰³ were available for G1 individuals aged 7
2631 and 13. Data from age 7 was matched with metabolomics data for Focus at 7 while data from
2632 age 13 was matched with metabolomics data for Teen Focus 3 and 4. Diet data were not
2633 available for young adults or adults. Data is given as predicted kilo-calories consumed per
2634 day.

2635 Data on physical activity were collected differently for different clinic visits. For G1
2636 individuals from Teen Focus 3 and Focus at 24, accelerometry data were collected at the same
2637 clinic for which metabolomics data were collected. Briefly, individuals wore an accelerometry
2638 device for the 7 days following their clinic visit whilst keeping a diary of the times they wore and
2639 took off the device. Individuals were advised to wear the accelerometer device if the following
2640 days were part of a 'normal week' with regards to their activity. For Focus at 24 individuals,
2641 physical activity data are the average number of minutes per day spent doing moderate to
2642 vigorous physical activity. For Teen Focus 3 individuals, physical activity data are the mean
2643 counts per minute spent doing moderate to vigorous physical activity for the whole week. For
2644 G0, individuals were asked 'Do you take part in physical activity (e.g., running, swimming,
2645 dancing, golf, tennis, squash, jogging, and bowls)?' with possible answers: no, occasionally
2646 (less than monthly), and frequently (once a month or more). Data for mothers were available
2647 in 2010, fathers data were available at the metabolomics clinic. Physical activity data were
2648 not available for Focus at 7.

2649 4.2.5 Statistical analysis

2650 Data

2651 Metabolomics data were obtained first. This data were obtained with identifiable individuals
2652 (triplets/quadruplets) and individuals who had withdrawn consent removed. Metabolomics
2653 data were available for: Focus at 7 (n = 5518; metabolites = 230), Before Breakfast Study (n

2654 = 640; metabolites = 228), Teen Focus 3 (n = 3371; metabolites = 230), Teen Focus 4 (n =
2655 3175; metabolites = 230), Focus at 24 (n = 3269; metabolites = 224), Focus on Mothers 1 (n
2656 = 4362; metabolites = 230), Focus on Mothers 2 (n = 2708; metabolites = 230), Focus on
2657 Fathers (n = 1833; metabolites = 230). Metabolites includes derived ratios, for example ratio
2658 of saturated fatty acids to total fatty acid.

2659 In order to maximise sample size, data for similarly aged clinics were combined. After
2660 combining the G1 clinics Focus at 7 and Before Breakfast Study (children), and Teen Focus 3
2661 and Teen Focus 4 (adolescents), metabolomics data were available for 5656 children (mean
2662 age (SD) = 7.56 (0.36); metabolites = 234), 4489 adolescents (mean age (SD) = 16.06
2663 (1.11); metabolites = 230), and 3269 young adults from the Focus at 24 clinic (mean age
2664 (SD) = 24.03 (0.85); metabolites = 224); Table 4.1). After combining the G0 clinics Focus on
2665 Mothers 1, Focus on Mothers 2, and Focus on Fathers 1, metabolomics data were available
2666 for 6406 adults (mean age (SD) = 49.53 (5.32); metabolites = 232). G0 and G1 individuals
2667 are independent, however there is familial overlap between them. Importantly, G1 individuals
2668 are not independent, that is, the same individuals will have attended multiple clinics. As such,
2669 any effects specific to an individual will be propagated through all G1 clinics that individual
2670 attends.

2671 All subsequent data on exposures and covariates were obtained for the individuals
2672 included in each of the combined groups. All data on exposures and covariates were obtained
2673 from the clinic closest in date to that of the metabolomics clinic.

2674 **Statistical analysis**

2675 To estimate the association between measures of adiposity and metabolites, all exposures
2676 were *Z*-scored and linear regression was performed. Metabolites were not transformed as
2677 the majority did not appear to have skewed distributions after pre-analysis processing using
2678 metaboprep. Variables known to influence the metabolomic profile and adiposity (age, sex,
2679 education, smoking, alcohol, diet, and physical activity), were included as covariates. Three
2680 linear models were used to investigate potential effects of these covariates. Model 1 included
2681 age and sex. Model 2 included variables in model 1 and mothers/own level of education,
2682 whether respondent had ever smoked, frequency respondent had a drink containing alcohol,
2683 and predicted kilo-calories consumed per day. Model 3 comprised all variables included
2684 in model 2 and physical activity. To maximise sample size for analyses, model 1 and 2
2685 comprised individuals with data on all covariates except physical activity. Model 3 comprised
2686 all individuals with data on all covariates. Model 2, as the most adjusted and given the
2687 reduced sample size in model 3, is presented as the main analysis for this work.

2688 For all analyses, units represent the change in each metabolite per standard deviation
2689 change in the exposure. Metabolite abbreviations are used in figures and tables for space. For
2690 complete labels, class, subclass, and units for each metabolite see [GitHub](#). 95% confidence

2691 intervals (CI) were calculated and a multiple testing threshold specific for each age group was
2692 applied. Multiple testing thresholds were calculated as the number of independent metabolites
2693 within the raw metabolomics data given a Spearman's rho of approximately 0.75 among the
2694 metabolites with data for at least 20% of samples – this was calculated during metabolite
2695 pre-analysis processing using `metaborpep`. The number of independent metabolites in each
2696 age group were calculated as: children = 42, adolescents = 42, young adults = 40, adults =
2697 44.

2698 Metabolites were grouped into subclasses (grouping data provided by the metabolomics
2699 platform) based on biological pathway. Consistency in the direction of effect estimates
2700 across models within each exposure and age group was investigated, as was the number
2701 of tests reaching a multiple testing threshold. In addition, directional consistency for results
2702 for exposures across age groups (using results from model 2) was also investigated. A
2703 Spearman's rho correlation analysis was used to investigate the correlation between effect
2704 estimates across exposures and age groups. Circos plots (via the `EpiViz` R package) were
2705 used to visualize and compare global metabolic profiles within age groups across exposures.
2706 Forestplots, created using the `ggforestplot` R package, were used to examine specific
2707 subclasses where variation among metabolites within the subclass and strong effects were
2708 identified. Results for derived measures and lipoprotein particle size and fatty acid ratios are
2709 presented in the Appendix. Directions of effect estimates were compared to that of previous
2710 work by Wurtz et al. (2014)¹⁵⁷ where appropriate.

2711 4.3 Results

2712 4.3.1 Data overview

2713 In pre-analysis processing of metabolomics data, derived metabolite measures drastically
2714 increased the number of representative metabolites in the dataset. As such, metabolomics
2715 data that underwent pre-analysis processing with the inclusion of derived features were used
2716 throughout. The total number of individuals with metabolomics data, as well as the number of
2717 metabolites available prior to pre-processing is given in Table 4.1.

Table 4.1: Metabolomics data available prior to pre-analysis processing

Age group	N	Metabolites	Age/clinic subgroup	N	Unique N	Metabolites	Unique metabolites
Children	5656	234	7	5518	5016	230	5
			8	640	138	228	7
Adolescents	4489	230	15	3371	1314	230	0
			17	3175	1118	230	0
Young adults	3269	224	24	3269	–	224	–
Mothers	4573	230	Focus on mothers 1	4362	1865	230	0
			Focus on mothers 2	2708	211	230	0
Fathers	1833		Focus on fathers	1833	–	230	–
Adults	6406	232	Mothers	4573	–	230	2
			Fathers	1833	–	230	2

Data were available from multiple time points. Generation 1 individuals were measured at 5 time points and combined into three groups (Children, Adolescents, Young adults). Generation 0 individuals were measured at three time points and combined (Adults). N = number of individuals in each group; Metabolites = the number of metabolites measured for each group; Subgroup = age or clinic identifier; Unique N = the number of individuals in the subgroup who do not appear in the other subgroup; Unique metabolites = Number of metabolites unique to that subgroup.

2718 Pre-analysis processing of metabolomics data resulted in the following exclusions: 6
2719 individuals and 4 metabolites were removed from the children's data (4 samples excluded for
2720 $\geq 80\%$ missingness, 1 sample excluded for total sum abundance ≥ 5 SD from the mean, 1
2721 sample excluded as a result of being ≥ 5 SD from the mean of PC1 and 2, and 4 metabolites
2722 removed for $\geq 20\%$ missingness); 5 individuals and 0 metabolites were removed from the
2723 adolescents data (1 sample excluded for total sum abundance ≥ 5 SD from the mean and 4
2724 samples excluded as a result of being ≥ 5 SD from the mean of PC1 and 2); 4 individuals
2725 and 0 metabolites were removed from the young adults data (1 sample excluded for total sum
2726 abundance ≥ 5 SD from the mean and 3 sample excluded as a result of being ≥ 5 SD from
2727 the mean of PC1 and 2); 7 individuals and 4 metabolites were removed from the adults data
2728 (1 sample excluded for total sum abundance ≥ 5 SD from the mean, 6 samples excluded as
2729 a result of being ≥ 5 SD from the mean of PC1 and 2, and 4 metabolites removed for $\geq 20\%$
2730 missingness). Processed metabolomics data available for analysis is given in Table 4.2.

Table 4.2: Metabolomics data available after pre-analysis processing

Group	N	Metabolites
Children	5650	230
Adolescents	4484	230
Young adults	3265	224
Adults	6399	228

Number of individuals (N) and metabolites available after pre-analysis processing of metabolomics data was performed using the 'metaboprep' 'R' package and including derived measures.

2731 Of individuals with metabolomics data, a majority also had data on measures of adiposity
2732 (Table 4.3), except for adolescents where data on WHR was not available. Adiposity data
2733 were normally distributed (Appendix Figure ??). Data on covariates were also available in
2734 the majority of individuals, the exception being physical activity where fewer individuals had
2735 available data and no data was available for children (Table 4.3).

Table 4.3: Measures of adiposity available for individuals with metabolomics data

Age group	BMI				WHR				BF		
	N	N	mean	SD	N	mean	SD	N	mean	SD	
Children	5650	5622	16.20	2.00	5589	0.86	0.04	5381	-	-	
Adolescents	4484	4404	21.71	3.68	-	-	-	4210	25.65	11.7	
Young adults	3265	3230	24.73	4.88	3223	0.8	0.07	3153	31.75	9.22	
Adults	6399	6352	26.83	4.98	6360	0.85	0.09	6138	34	9.08	

Measures were obtained for all individuals with processed metabolomics data. Wait hip ratio (WHR) was not available for adolescents. Body fat percentage (BF) was not available for children but a raw impedance measure was available. Age group = the combined age group clinic visits were combined into; BMI = body mass index; N = the number of individuals with available data; mean = the mean of the anthropometric measure; SD = standard deviation of the mean; - = data not available.

Table 4.4: Covariates available for individuals with metabolomics data

		Children	Adolescents	Young adults	Adults
Metabolomics	N	5650	4484	3265	6399
	N	5634	4474	3264	6381
Age	Mean	7.56	16.06	24.03	49.53
	SD	0.36	1.11	0.85	5.32
Sex	N	5634	4474	3265	6390
	Female N	2727	2340	1966	4557
Education	1 (low)	483	344	201	412
	2	434	308	202	392
	3	1798	1370	978	1785
	4	1369	1168	856	1717
Smoking	5 (high)	866	745	612	1366
	N	-	2499	3219	5854
	1 (low)	-	394	88	364
Alcohol frequency	2	-	336	668	587
	3	-	2037	1223	1042
	4	-	759	998	1731
	5 (high)	-	153	172	1209

	N	5577	4338	-	-
Diet	Mean	1672.32	2252.32	-	-
	SD	134.94	190.62	-	-
Physical activity	N/1	-	1768	672	1563
	Mean/2	-	483.67	50.23	659
	SD/3	-	179.39	30.24	2330

Education is highest level of education (1-5); for children, adolescents, and young adults education is maternal education; for adults education is own education. Smoking is binary. Alcohol is frequency respondent consumes an alcoholic drink, with 1 being low and 5 high. Diet is predicted kilo-calories consumed per day. Physical activity in adolescents is mean counts per minute of activity across 7 days. For young adults physical activity is the average number of valid minutes per day spent doing moderate to vigorous activity. For adults, physical activity is: no physical activity (1), occasionally (2), frequently (3). Group = the groups clinic visits were combined into; N = the number of individuals with available data; mean = the mean of the measure; SD = standard deviation of the mean. - = data not available.

2736 When looking at the variability in metabolite concentrations across age groups, the mean
 2737 and SD of the mean metabolite concentration, as well as the median metabolite concentration
 2738 and range between the minimum and maximum metabolite concentration for each metabolite
 2739 were generally similar across all age groups. Variability in metabolite concentrations tended to
 2740 increase (larger SD of mean metabolite concentration and range of metabolite concentrations
 2741 for each metabolite) with age however, with the largest variability predominantly seen in adults.
 2742 Variability appeared much larger when looking exclusively at derived metabolite measures as
 2743 opposed to directly measured metabolites. All plots are available on [GitHub](#).

2744 **4.3.2 Statistical analysis**

2745 In total, metabolomics data were available for 5,650 children, 4,484 adolescents, 3,265
 2746 young adults, and 6,399 adults. After obtaining data on covariates data were available
 2747 for between 575–4450 individuals (Table 4.5). Model 1 and model 2 were restricted to all
 2748 individuals with complete data for adiposity and covariate data for model 2. Model 3 was
 2749 restricted to individuals with complete data for all adiposity measures and all covariates.

Table 4.5: Total number of individuals with metabolomics data and total number of individuals included in each linear model

Age group	N	Model 1 & 2	Model 3
-----------	---	-------------	---------

Children	5650	4450	–
Adolescents	4484	1762	588
Young adults	3265	2589	575
Adults	6399	3505	2981

The total number of individuals with available metabolomics data (N) and the sample size for each linear model (Model 1 & 2 and Model 3).

2750 Across all models and exposures, between 33.91–86.4 % (median = 61.3 %) of metabo-
 2751 lites reached a multiple testing threshold (multiple testing threshold for children = 0.0012,
 2752 adolescents = 0.0012, young adults = 0.0013, adults = 0.0011). In summary, between 83-193,
 2753 139-193 and 75-197 metabolites reached a multiple testing threshold across all age groups
 2754 for BMI, WHR, and BF respectively (Table 4.6).

Table 4.6: Number of tests reaching a multiple testing threshold

	N	BMI			WHR			BF		
		1	2	3	1	2	3	1	2	3
Children	230	141	137	–	148	148	–	141	132	–
Adolescents	230	138	156	83	–	–	–	150	159	75
Young adults	224	173	172	139	173	172	139	183	180	135
Adults	228	193	191	183	193	191	183	197	191	186

Number of metabolites (N) reaching a multiple testing threshold for each model (1-3) within each age group for each exposure. Multiple testing thresholds: children = 0.0012, adolescents = 0.0012, young adults = 0.0013, adults = 0.0011. BMI = body mass index; WHR = waist hip ratio; BF = body fat percentage; 1 = model 1 adjustment for age and sex; 2 = model 2 adjustment for model 1 plus maternal/own education, smoking status, alcohol frequency, diet (where available); 3 = model 3, adjustment for model 2 plus physical activity (where available).

2755 **Directional consistency between effect estimates across three linear models within**
 2756 **exposures and age groups**

2757 Across models, within each exposure and age group, the majority of tests resulted in
 2758 directionally consistent effect estimates (Figure 4.1). Of these directionally consistent effects,
 2759 the majority of effect estimates were positive. That is, across a majority of metabolites,
 2760 adiposity was associated with an increase in metabolites. The strength of these effect
 2761 estimates were broadly consistent across models; confidence intervals overlapped across

2762 the majority of metabolites for all models within each group and exposure (Appendix A.3.2).
2763 Of the 220 metabolites measured in all age groups, directional consistency was supported by
2764 strong evidence of correlation across models within exposures and age groups (Appendix
2765 A.3.2).

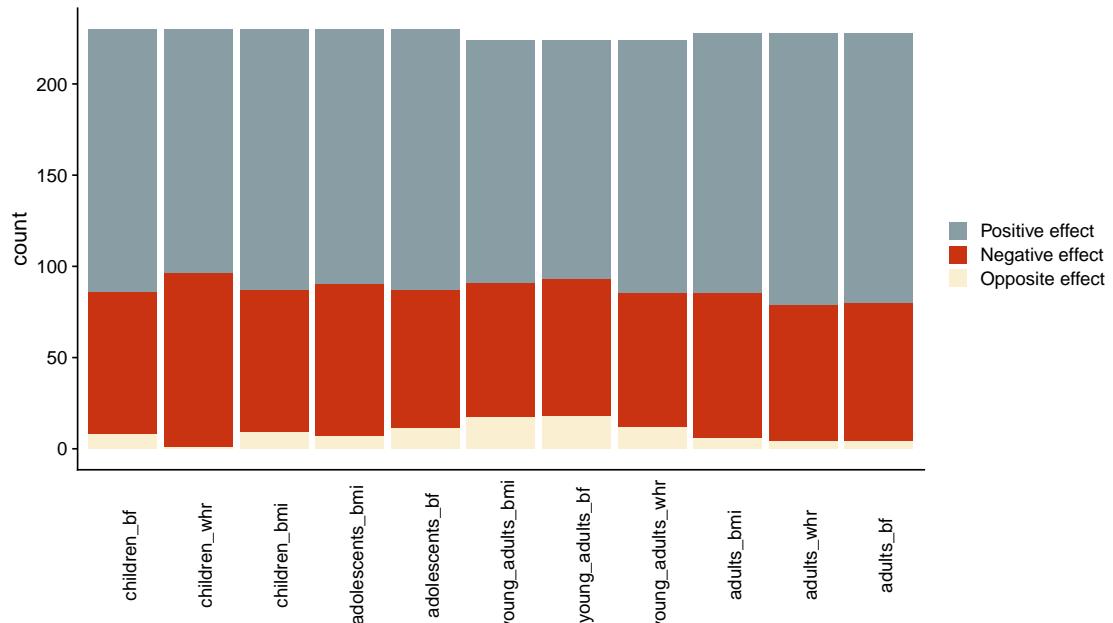


Figure 4.1: Directional consistency between effect estimates across three linear models within exposures and age groups. A positive effect reflects all model betas being in the positive direction; a negative effect reflects all model betas being in a negative direction; opposite effect reflects different directions for the model betas. BMI = body mass index; WHR = waist hip ratio; BF = body fat percentage.

2766 **Directional consistency between effect estimates for linear model 2 across exposures**
2767 **within age groups**

2768 Given the broad agreement between models, including the high correlation observed for
2769 effects between model 2 and model 3 (Spearmans rho = 0.81–0.99; Appendix A.3.2), and
2770 the fact that model 3 was not run for children as data on physical activity were not available,
2771 results here on are presented for model 2 (all results are available on [GitHub](#)).

2772 Across the three exposures (BMI, WHR, and BF) within each age group, effects showed
2773 mostly consistent directions of effect. The majority of effects were to increase (positive)
2774 metabolites (Figure 4.2). The most inconsistent directions across exposures were found for
2775 children, where 21% of metabolites showed inconsistent directions of effect. For adolescents,

2776 young adults, and adults 5, 9, and 5% of metabolites showed inconsistent directions of effect
2777 across measures of adiposity respectively. A large proportion of metabolites with evidence for
2778 a consistent direction of effect across all three exposures also reached the multiple testing
2779 threshold: children = 110, adolescents = 148, young adults = 164, and adults = 180.

2780 Of the 220 metabolites measured across all age groups, directional consistency was
2781 supported by strong evidence of correlation across exposures within age groups (Spearmans
2782 rho = 0.61–0.97) and within exposures across age groups (Spearmans rho = 0.48–0.82;
2783 Appendix A.3.2).

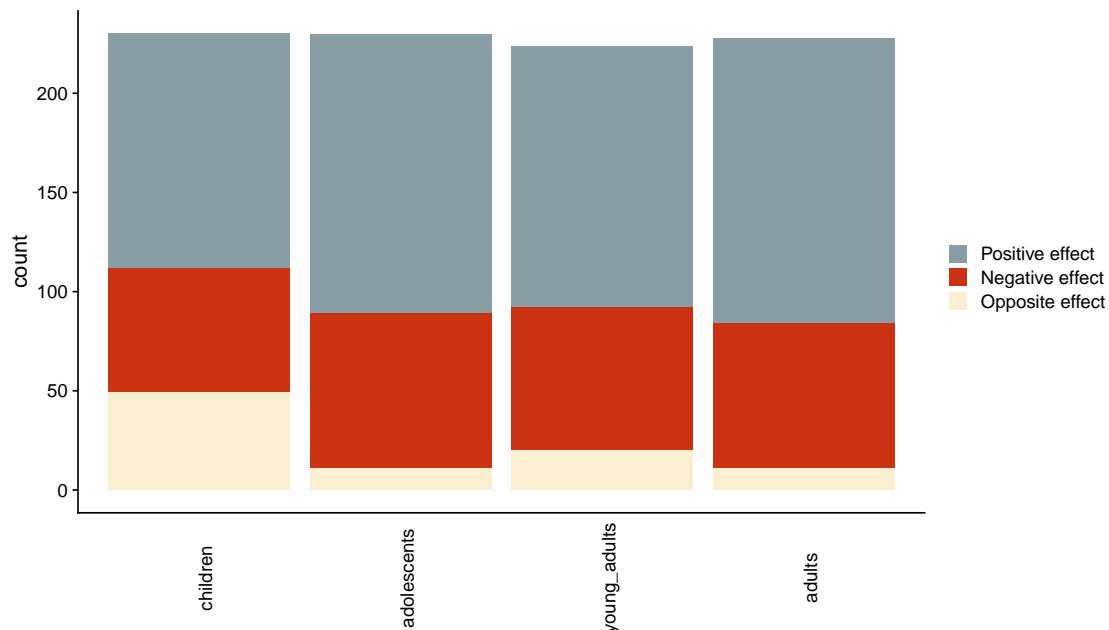


Figure 4.2: **Directional consistency between effect estimates for linear model 2 across exposures within age groups.** A positive effect reflects all model betas being in the positive direction; a negative effect reflects all model betas being in a negative direction; opposite effect reflects different directions for the model betas. BMI = body mass index; WHR = waist hip ratio; BF = body fat percentage.

2784 Global metabolic profile

2785 To aid visual comparison between metabolite subclasses, directly measured metabolites
2786 were visualised separately to derived measures which are presented in the Appendix A.3.
2787 Derived metabolites are metabolites that are not directly measured by the metabolomics array
2788 and are instead derived during the processing of the raw NMR spectra. Derived measures
2789 include ratios such as Cholesterol esters in very large VLDL to total lipids in very large VLDL

2790 ratio and Total cholesterol in very large VLDL to total lipids in very large VLDL ratio.

2791 Overall, the global pattern of association was very similar for all measures of adiposity for
2792 children (Figure 4.3), adolescents (Figure 4.4), young adults (Figure 4.5), and adults (Figure
2793 4.6) across all directly measured metabolites. Across all age groups, the largest effects were
2794 found for the fatty acids subclass; the metabolite total fatty acids showed the largest effect
2795 across all exposures for each age group. Metabolites in subclasses small VLDL, medium
2796 VLDL, large VLDL, and very large VLDL were the only ones to reach the specified multiple
2797 testing thresholds across all exposures and age groups. Across age groups, the direction of
2798 effect was consistent for the majority of metabolites within exposures (Appendix A.3.2). On
2799 the whole, effect sizes were lowest in children and increased with age. The largest effect in
2800 adults, total fatty acids (change in metabolite per standard deviation change in the exposure
2801 (beta) = 0.51), was twice that observed in children (beta = 0.21).

2802 Effect estimates for derived measures were much higher for all exposures and age groups
2803 compared to directly measured metabolites. However, the global pattern of association was
2804 very similar within age groups across exposures and across age groups within exposures
2805 (Appendix A.3.2). There was also considerable variation within subclasses for derived
2806 measures. The largest effect across all age groups and exposures among derived measures
2807 was observed for cholesterol esters in very large VLDL to total lipids in very large VLDL ratio
2808 and total cholesterol in very large VLDL to total lipids in very large VLDL ratio.



Figure 4.3: Effect estimates from linear regression of adiposity measures on metabolites in children. The Circos plot shows each track as one of the measures of adiposity; the outer track is body mass index (BMI), the middle track is waist hip ratio (WHR), the inner track is body fat percentage (BF). Units are the change in metabolite per standard deviation change in the exposure; 95% confidence intervals are shown and may be hidden by the point estimate if very tight. Solid points indicate a multiple testing threshold has been reached (0.05/42).

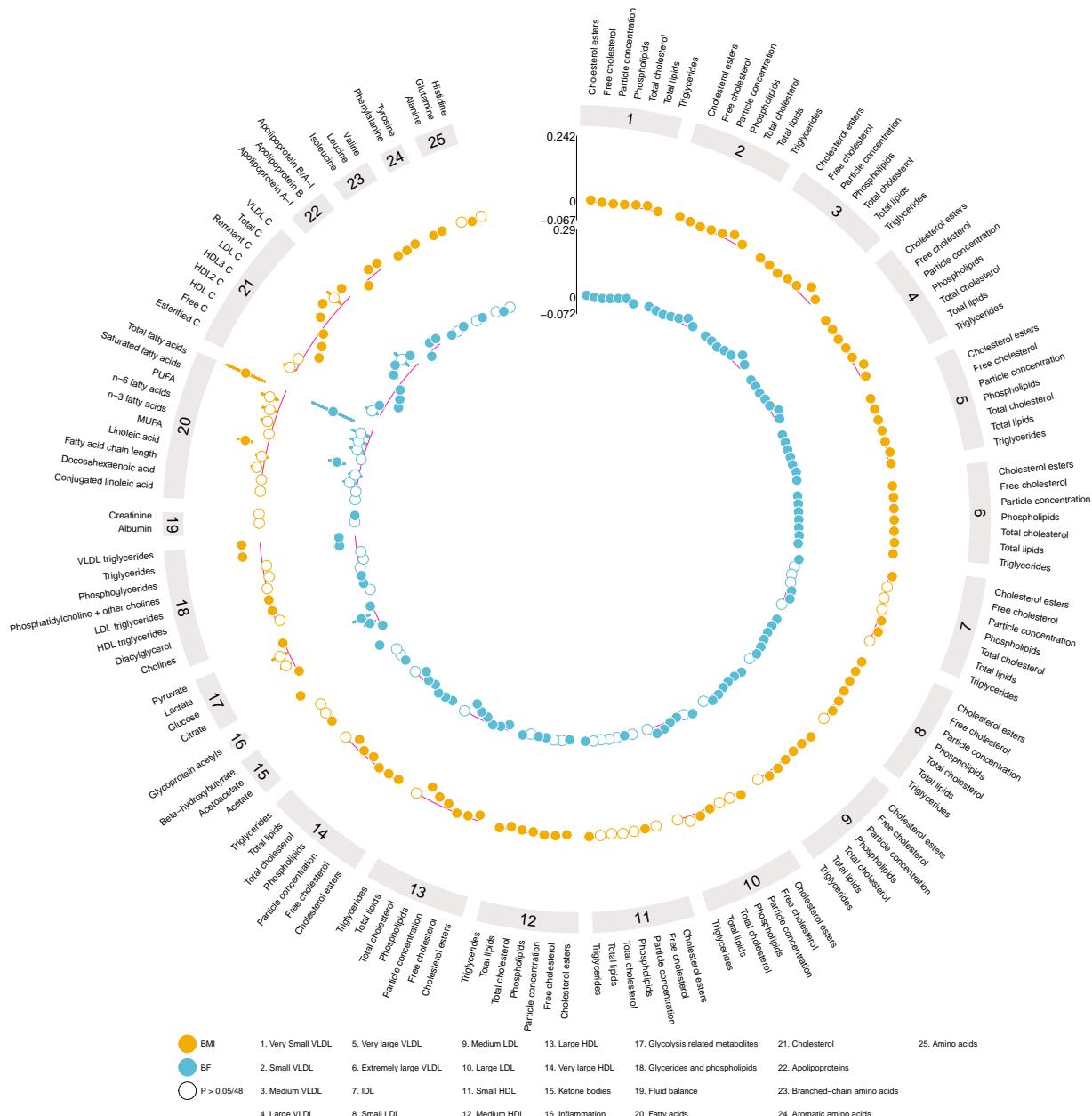


Figure 4.4: Effect estimates from linear regression of adiposity measures on metabolites in adolescents. The Circos plot shows each track as one of the measures of adiposity; the outer track is body mass index (BMI), the middle track is waist hip ratio (WHR), the inner track is body fat percentage (BF). Units are the change in metabolite per standard deviation change in the exposure; 95% confidence intervals are shown and may be hidden by the point estimate if very tight. Solid points indicate a multiple testing threshold has been reached (0.05/42).



Figure 4.5: Effect estimates from linear regression of adiposity measures on metabolites in young adults. The Circos plot shows each track as one of the measures of adiposity; the outer track is body mass index (BMI), the middle track is waist hip ratio (WHR), the inner track is body fat percentage (BF). Units are the change in metabolite per standard deviation change in the exposure; 95% confidence intervals are shown and may be hidden by the point estimate if very tight. Solid points indicate a multiple testing threshold has been reached (0.05/40).

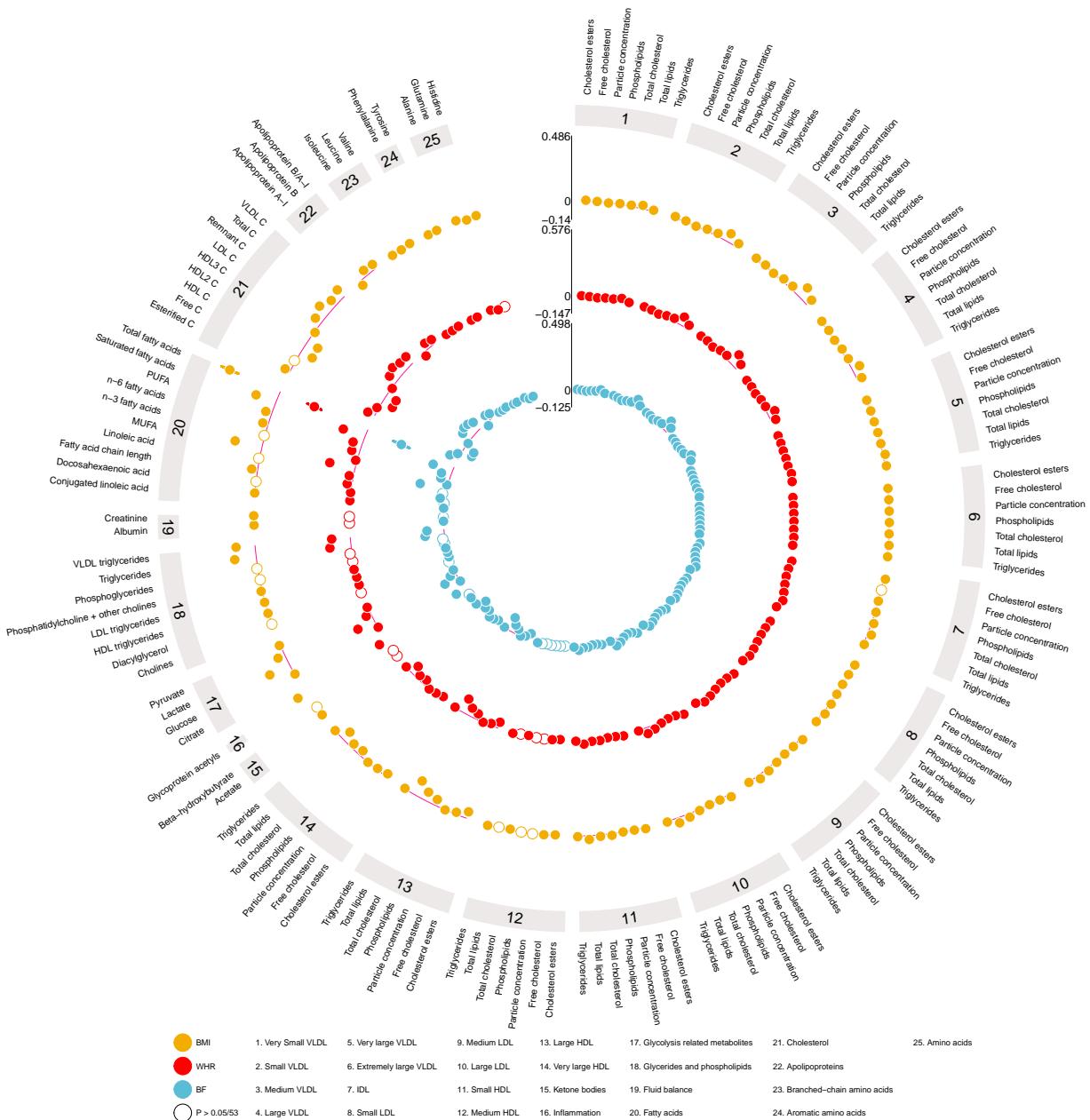


Figure 4.6: Effect estimates from linear regression of adiposity measures on metabolites in adults. The Circos plot shows each track as one of the measures of adiposity; the outer track is body mass index (BMI), the middle track is waist hip ratio (WHR), the inner track is body fat percentage (BF). Units are the change in metabolite per standard deviation change in the exposure; 95% confidence intervals are shown and may be hidden by the point estimate if very tight. Solid points indicate a multiple testing threshold has been reached (0.05/44).

2809 **Subclass results**

2810 The Nightingale NMR platform is able to separate out metabolites into subclasses. When
2811 looking at directly measured metabolite subclasses, there were associations between mea-
2812 sures of adiposity and all subclasses except for large LDL in children where confidence
2813 intervals for all metabolites crossed the nul. For the derived measure subclasses, associa-
2814 tions were observed across measures of adiposity for all subclasses except extremely large
2815 VLDL ratios in young adults and adults.

2816 Across all age groups and exposures, associations with every metabolite in a particular
2817 subclass was observed for small VLDL, medium VLDL, large VLDL, very large VLDL, and
2818 extremely large VLDL. As age increased, the number of associations within subclasses
2819 tended to increase across all measures of adiposity. For example, across small LDL, medium
2820 LDL, and large LDL few associations were observed across measures of adiposity in children,
2821 however in adolescents and young adults, a majority of metabolites showed evidence of
2822 association; while in adults, all metabolites showed evidence of association. Effect size
2823 tended to increase with age as well, for example, the largest effects seen for small, medium,
2824 and large LDL metabolites were in adults and young adults (Figure 4.7).

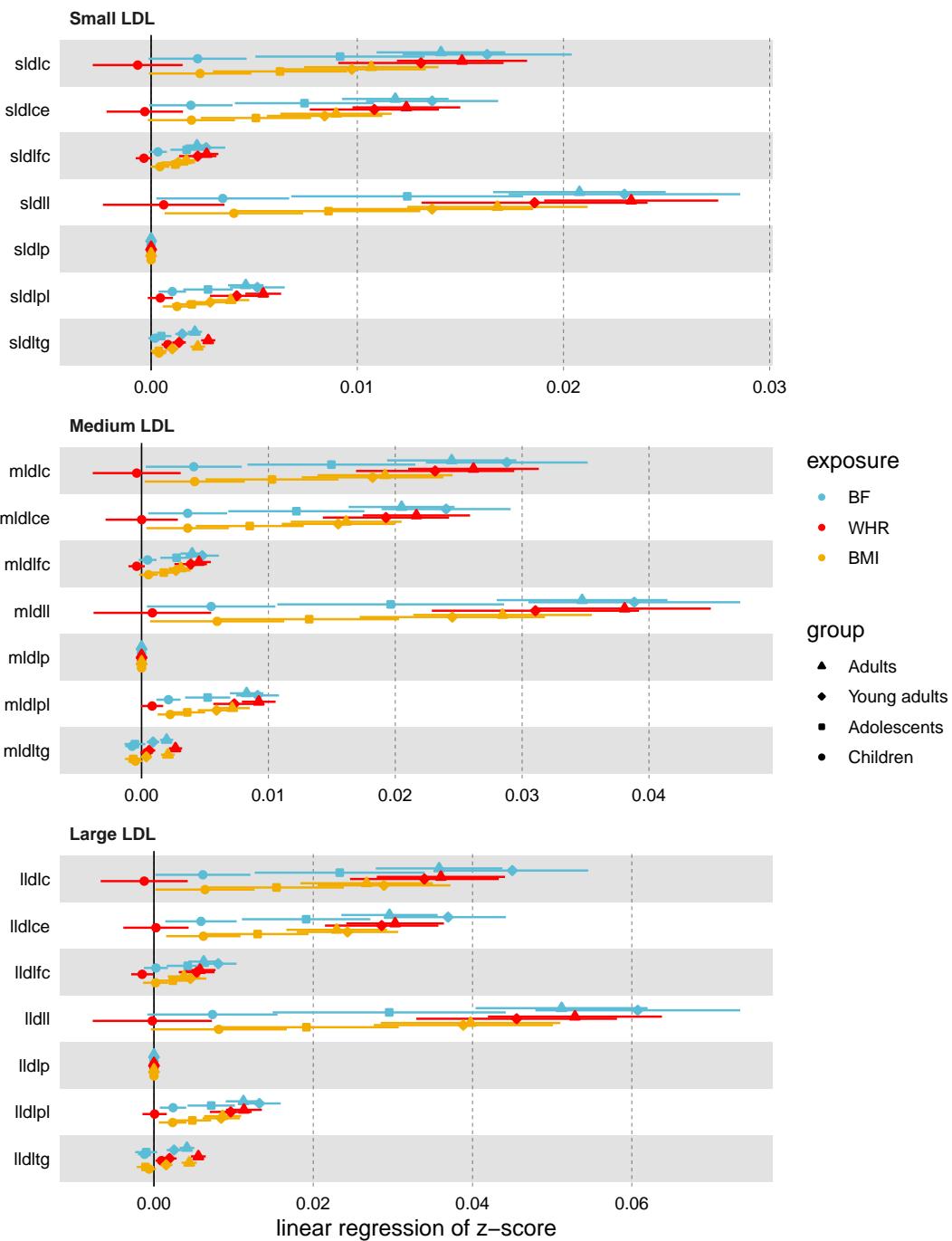


Figure 4.7: **Effect estimates from linear regression of adiposity measures on metabolites in all age groups for small, medium, and large LDL subclasses.** The forestplot shows the change in metabolite per standard deviation change in the exposure. 95% confidence intervals are given. BMI = body mass index; WHR = waist hip ratio; BF = body fat percentage.

2825 **4.3.3 Comparison with previous work**

2826 A total of 82 metabolites were measured by Wurtz et al. (2014)¹⁵⁷. Of these, 42 were
2827 also measured across children, adolescents, young adults, and adults. Across all analyses
2828 conducted here for model 2, a majority of metabolites showed a consistent direction of effect
2829 with the study by Wurtz et al.¹⁵⁷ (Figure 4.8). However, there were some metabolites where
2830 the effects from Wurtz et al., were opposite to the majority of effects found here, such as fatty
2831 acid chain length (falen) and albumin (alb).

2832 The majority of effect estimates were in the positive direction, i.e. an increase in adiposity
2833 increased metabolites; 30 metabolites had primarily positive effects as a result of adiposity,
2834 and 12 metabolites had primarily negative effects as a result of adiposity. A total of 18
2835 metabolites were associated with a positive effect across analyses by Wurtz et al., and all
2836 age groups and exposures here, this included phenylalanine, tyrosine, and apolipoprotein
2837 B. Seven metabolites were associated with a negative effect across all analyses, including
2838 citrate and apolipoprotein A1. Looking at the overall effects for each subclass, there were
2839 primarily positive effects of adiposity for subclasses aromatic and branched chain amino
2840 acids, cholesterol, fatty acids, glycerides and phospholipids, and glycolysis related metabolites.
2841 Primarily negative effects were found for lipoprotein particle size and fatty acids ratios. The
2842 remaining subclasses were closely split (amino acids and fluid balance) or composed of one
2843 metabolites (IDL, inflammation, and ketone bodies).

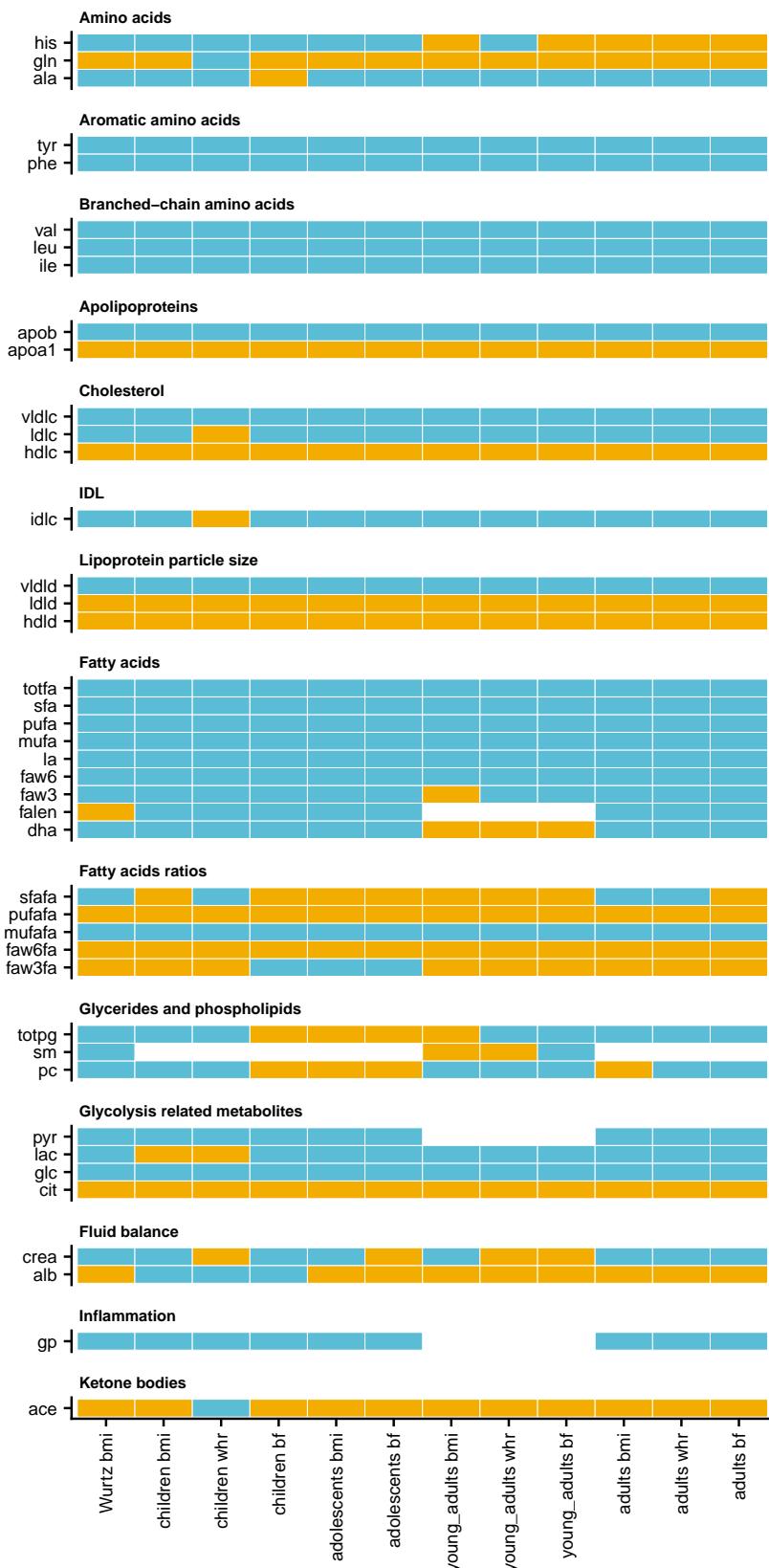


Figure 4.8: Comparison of the directions of effect estimates from linear regression of adiposity measures on metabolites across all age groups and analysis by Wurtz et al. (2014)¹⁵⁷.

2844 The tile plot shows the direction of effect estimate for 42 metabolites as positive (blue) or
2845 negative (yellow). If all analyses show the same direction of effect then that row will be the
2846 same colour, e.g., all analyses show a decreasing effect on apoa1 and an increasing effect
2847 on apob. BMI = body mass index; WHR = waist hip ratio; BF = body fat percentage; IDL
2848 = intermediate density lipoproteins. White space indicates no analysis. For full names of
2849 metabolites see appendix.

2850 4.4 Discussion

2851 In this chapter, the influence of adiposity on the metabolic profile is demonstrated in
2852 an observational framework. These effects persist, not only when measured at different
2853 ages, but also when adjusting for covariates such as smoking, physical activity, and diet.
2854 Effects are similar across multiple measures of adiposity and to those previously reported¹⁵⁷.
2855 The association between adiposity and metabolites is global, with effects seen across all
2856 subclasses of metabolites. These effects are broadly consistent between metabolites within
2857 each subclass.

2858 When looking at each exposure across age groups (e.g., the effects of BMI in children,
2859 adolescents, young adults, and adults), there were similar effects seen in individuals measured
2860 at multiple time points. As age increased, the number of associations within subclasses
2861 tended to increase across all measures of adiposity. Fewer associations were observed in
2862 children than in adolescents and young adults. A similar metabolic profile is apparent in adults,
2863 though effect sizes appear slightly larger. These results may reflect (i) an effect of prolonged
2864 adiposity exposure or (ii) increased variation in metabolites with age. Given that longitudinal
2865 work has shown that BMI tracks over time^{504,505} there may be a dose response relationship
2866 here whereby adiposity has a compounding effect on metabolites over time. Alternatively,
2867 metabolite concentrations have been shown to increase in older populations over time³³¹ and
2868 there was evidence that metabolite variation tended to increase as age increased.

2869 Within each age group, there was directional consistency across the three exposures.
2870 Effect sizes across exposures were similar within age groups, with overlapping confidence
2871 intervals. BMI showed a broadly stronger association on metabolites across age groups.
2872 Effects for BF appeared to be closer to, and crossed, the null more often than BMI and WHR.
2873 This may suggest that overall body composition in addition to detrimental deposition, may
2874 be driving these effects. That is, the compounding effect of increased overall body mass
2875 (primarily through increased fat mass; as estimated by BMI) and the predilection to store this
2876 additional mass viscerally (as estimated by WHR), may be more powerful than either of these
2877 effects alone on the effect of metabolites.

2878 Results were consistent across models. Effect estimates showed consistent directions for
2879 the majority of tests. Effect sizes were broadly similar in the case of models 1 and 2. Effect

2880 sizes in model 3 were smaller compared with models 1 and 2. However, confidence intervals
2881 overlapped across the majority of tests even if they crossed the null in some instances for
2882 model 3. Consistency across models suggests weak evidence of confounding. Whether
2883 these results represent a true causal effect or are subject to reverse causation or residual
2884 confounding requires further investigation.

2885 The key strengths of this work however are the use of (i) ALSPAC and (ii) multiple
2886 measures of adiposity to examine adipositys effects (iii) across the lifecourse. ALSPAC is a
2887 prospective cohort study with a general population base and here, analyses were performed
2888 on up to 4,450 individuals. The breadth of data measured across multiple generations
2889 enabled linear models to be adjusted for covariates that have been known to affect adiposity
2890 and metabolite concentrations^{497,498,498–502}. These analyses were performed across two
2891 generations of individuals, G0 and G1. Data on G1 individuals were available from multiple
2892 time points and, in combination with data from G0 individuals, enabled the investigation of
2893 adipositys effects across the lifecourse from childhood to adulthood. The use of multiple
2894 measures of adiposity, as has been recommended previously^{83,88} but which has not been
2895 explored in relation to metabolic effects, may indicate that overall body composition and
2896 deposition is more important in metabolic changes than either component alone.

2897 Previous work by Wurtz et al. (2014)¹⁵⁷ identified numerous associations between BMI and
2898 metabolites in a large population. Results here show a broadly similar pattern of association
2899 with those by Wurtz et al., with the majority of the 42 comparable metabolites showing
2900 consistent directions of effect. This includes phenylalanine and tyrosine, both of which are
2901 known to be increased by adiposity and age^{386,387,506–511}, as well as consistent negative
2902 effects on seven metabolites, including apolipoprotein A1 which is the major component of
2903 HDL particles and enables uptake of lipids by HDL and the subsequent recycle and excretion
2904 of lipids^{512,513}. Differences in effect estimate direction was minimal and split relatively evenly
2905 across age groups. Only one metabolite from the Wurtz et al study, fatty acid chain length, had
2906 a direction of effect (negative) that was inconsistent across all age groups here. In follow-up
2907 Mendelian randomization analysis by Wurtz et al., fatty acid chain length was found to have a
2908 positive direction of effect.

2909 4.4.1 Limitations

2910 Observational analyses are limited due to issues of confounding and reverse causation.
2911 Prospective studies, such as ALSPAC, are able to provide some separation between the
2912 measurement of an exposure and outcome, and thus mitigate the potential effects of reverse
2913 causation. However, as data on adiposity measures and metabolites were collected at the
2914 same time reverse causation remains a limitation. It was however possible to account for
2915 the potential effects of confounding by adjusting linear models for potential confounders.
2916 Although many potential confounders were included across the models, it is likely these
2917 analyses will not have fully accounted for confounding, either due to measurement error or

2918 unmeasured confounding. For example, adults were asked 'do you take part in physical
2919 activity (e.g. running, swimming, dancing, golf, tennis, squash, jogging, bowls)?', with possible
2920 answers of 'no', 'occasionally (less than monthly)' and 'frequently (once a month or more)'.
2921 Broad categories such as these are unlikely to capture the full impact of physical activity
2922 given that 'frequently' will encompass individuals who exercise once a month as well as every
2923 day. In addition, although it was possible to investigate the effect of adiposity across the life
2924 course, it is likely that G1 individuals with measurement data from their teens will have been
2925 undergoing puberty. Puberty is likely to have an affect on adiposity and covariates such as
2926 physical activity, as well as the metabolome.

2927 There are two key limitations in regards to metabolomics data. The first is that these
2928 data are limited in their breadth. Although a relatively large number of metabolites were
2929 investigated, they were predominantly lipid based. This leaves a broad array of metabolites
2930 that have not been investigated. As such, the metabolites investigated here are broadly re-
2931 reflective of the lipidome and not a global metabolic profile. The Nightingale NMR platform does
2932 however provides many derived measures, such as ratios and number of bonds. However,
2933 there was considerable variation in the concentration of these derived measures compared to
2934 directly measured metabolites. Given adiposity's relationship with lipids it is unsurprising there
2935 were a large number of associations identified. There is thus a clear need to expand beyond
2936 lipid based platforms in investigating the effects of adiposity, as these effects are likely not
2937 reflective of a global metabolic profile. MS platforms, such as those used by Metabolon, are
2938 a potential next step as studies investigating adiposity's effects on MS derived metabolites
2939 (which are few and far between) have used case control analyses⁵¹⁴, looked at small number
2940 of individuals¹⁵⁸, or focussed on children⁵¹⁵.

2941 There is no standardised approach, nor a gold standard, for performing metabolomics
2942 quality control. Here, quality control, including outlier detection and removal, was performed
2943 using the metaboprep R package. The default settings for exclusions based on metabolite
2944 missingness (20%), sample missingness (20%), total sum abundance (5 SD), and principal
2945 components (5 SD) were used. Most samples were removed for having missingness > 20%
2946 compared to total peak area and PC. 20% sample missingness is arbitrarily defined and used
2947 among other studies³²¹, however a more stringent threshold (e.g. 5%) may have resulted in
2948 additional sample exclusions. Metadata, such as batch and run day, was not available and
2949 could not be included in quality control, though this is likely more an issue with MS data.
2950 metaboprep calculated the number of independent metabolites using a clustering dendrogram
2951 and a tree cut height based on a Spearman's Rho of 0.5. Although most metabolites
2952 were shared across age groups, differences in the number of independent metabolites
2953 were found. There were also differences in the number of clusters and truly independent
2954 metabolites. Similarly, inclusion of derived metabolites resulted in differing numbers of
2955 independent metabolites. Metabolites did not undergo transformation to standardise their
2956 concentrations in order to preserve clinical utility and improve interpretation of results.

2957 Metabolomics measures were taken at specific time points in ALSPAC children and at
2958 ~50 years of age in adults. Though measures of adiposity were available at the same time

2959 points, data on confounders were not. Smoking status for example was available for adult
2960 males at the metabolomics clinic but the closest available measure for adult females was a
2961 number of years earlier, in which time (though unlikely) they may have taken up smoking.
2962 Although data were obtained where available from the closest time point, the mismatch in
2963 timings may lead to confounded estimates, increasing the number of false positives.

2964 In all age groups, the availability of data was limited. Absence of physical activity and
2965 diet data meant model 3 was not performed for children and model 2 and 3 for young adults
2966 and adults were not adjusted for diet. However, adjusting for diet, where available (children
2967 and adolescents), had little impact on results (Table 4.6) and is therefore likely not to have
2968 changed findings in young adults and adults.

2969 In children, BF was not available. A raw impedance measure was available and derivation
2970 of BF using Equation (4.1) showed positive correlation with BMI and weight in children and
2971 BF measures in adolescents. However, this derived measure included numerous negative
2972 estimates of BF. The equation performed well in adolescents, correlating highly with DXA
2973 derived BF (Figure ??). The negative estimates of BF are likely a result of Equation (4.1)
2974 being derived in an adult population. Brief investigation showed negative estimates of BF
2975 remained when using adolescent age with child height and weight (data not shown). Given
2976 that in single frequency devices, impedance is based on the volume of an individual, it is
2977 probable that the values used in the equation do not accurately reflect the proportions of
2978 children pre-puberty. Child-specific equations were not available and the manufacturer was
2979 unwilling to share the equation used by their impedance devices. However, given that in a
2980 linear model, the estimate is based on the per-unit increase, the absolute value negative or
2981 otherwise, will likely not alter the strength of association.

2982 The distribution of BMI at each age group was very similar across sexes. However, the
2983 distribution of WHR and BF differed among sexes in adolescents, young adults and adults;
2984 males had on average a higher WHR, while females had a higher BF. Though Z-scores were
2985 used and sex was included as a covariate, the differences in WHR and BF distributions may
2986 highlight an underlying difference which, if associated with metabolites, may have confounded
2987 the relationship between adiposity and metabolites. For example, hormonal contraceptive
2988 use has shown to influence the metabolome⁵¹⁶ and may be used less often by individuals
2989 with a high BMI⁵¹⁷.

2990 There was little difference in results from the three models in regards to direction of
2991 effect estimates. However, when including only individuals with physical activity (model 3),
2992 the size of effects and the number of associations reaching a multiple testing threshold
2993 decreased. This is likely an effect of reduced power given the smaller sample size for model 3.
2994 Given the consistency in the direction of effects across models and the highly similar effects
2995 across model 1 and 2, there is likely little effect of confounding. However, the possibility of
2996 unmeasured confounding can not be ruled out.

2997 **4.4.2 Conclusion**

2998 The large number of associations identified using multiple exposures adds weight to the
2999 evidence, shown previously for BMI, of association between adiposity and metabolites. As with
3000 previous work, there were associations between adiposity and many metabolites, including
3001 increased phenylalanine and tyrosine and decreases with HDL components. This work
3002 suggests the broad and consistent effects of adiposity on a broad array of lipid metabolites
3003 likely persist over time. Though covariates were included, the potential for unmeasured
3004 confounding is likely, and follow-up analysis will require this to be taken into account. Of
3005 particular note is the increasing effect size and number of associations as a result of adiposity
3006 across the metabolic profile with age. Given that many adiposity associated diseases occur
3007 later in life, exposure to an altered metabolic profile over time may be important in disease
3008 development. This is especially true as weight loss in overweight and obese individuals is
3009 associated with a normalizing of metabolite changes³⁷².

3010

3011

3012 **Chapter 5**

3013 **Associations between multiple** 3014 **measures of adiposity and** 3015 **metabolites: Mendelian** 3016 **randomization analysis**

3017 **Chapter summary**

3018 Chapter 4 explored the association between multiple measures of adiposity (body mass
3019 index (BMI), waist hip ratio (WHR), body fat percent (BF)) and up to 230 metabolites in the
3020 Avon Longitudinal Study of Parents and Children (ALSPAC). Across four time points (children
3021 (~7 years), adolescents (~15 years), young adults (~24 years), adults (~50 years)) three linear
3022 models, adjusting for age and sex (models 1, 2, 3), education, smoking, alcohol, and diet
3023 (models 2, 3) and physical activity (model 3), found a large number of associations across
3024 metabolite subclasses. Poor measurement, residual confounding, reverse causation, and
3025 missing data are all limitations of observational analyses which could bias estimates. Perhaps
3026 most pertinent to analysis in Chapter 4 is residual confounding and reverse causation. As
3027 discussed in 1.6 and 1.8, methods which try to account for these issues, such as Mendelian
3028 randomization (MR), provide an opportunity to further interrogate relationships. MR analyses
3029 hold a different set of assumptions and limitations; triangulation of evidence across multiple
3030 study designs with different sources of bias can strengthen confidence in results⁵¹⁸. In
3031 this chapter, MR is employed to investigate the association between multiple measures of
3032 adiposity (BMI, WHR, BF) and metabolites. Consistent results between observational and
3033 Mendelian randomization analyses strengthens evidence of association between adiposity
3034 and metabolites and may help to better characterize the changing metabolic profile as a result
3035 of adiposity which may later inform disease analysis.

3036 **5.1 Introduction**

3037 Evidence from observational analyses, including Chapter 4, highlights the broad effect
3038 of adiposity on metabolites^{157,158,328,363-372}. These effects are consistent across overall
3039 body composition and site-specific adiposity, and persist over time, and after adjustment for
3040 covariates. However, observational analyses are subject to a number of limitations including
3041 measurement error and unmeasured confounding (Section 1.8). These limitations could have
3042 resulted in biased estimates of the effect of adiposity on metabolites XXX.

3043 Mendelian randomization (MR), as discussed in Section 1.8 is able to reduce the is-
3044 sues traditionally observed in observational analyses (namely confounding and reverse
3045 causation)³³⁶. MR has previously been used to investigate the causal effect between adipos-
3046 ity and metabolic profile^{157,392,519}. These studies used body mass index (BMI)¹⁵⁷ alongside
3047 waist hip ratio (WHR)^{392,519} with 82¹⁵⁷, 123³⁹², and 249⁵¹⁹ metabolic measures in up to
3048 12,664¹⁵⁷, 24,925³⁹², and 109,532⁵¹⁹ individuals. They show similar evidence of association
3049 to observational analyses with systemic changes across the metabolome. However, these
3050 studies either used sole measures of adiposity or did not include a measure that accurately
3051 capture the total body fat content of the body, such as body fat percentage (BF). The
3052 studies are therefore unable to comment on the similarities or differences between adiposity
3053 measures.

3054 Here, a parallel MR analysis using multiple measures of adiposity (BMI, WHR, BF) with
3055 123 metabolites from Kettunen et al. (2016)³¹⁸ (N = 24,925) and 230 metabolites from the
3056 INTERVAL trial⁵²⁰ (Efficiency and safety of varying the frequency of whole blood donation;
3057 unpublished; N = 40,849) is performed. For all metabolites that were shared between
3058 these two studies, follow-up meta-analyses were conducted in up to 65,774 individuals of
3059 European ancestries. Triangulation of these results with observational analyses in adults
3060 (Chapter 4), were used to identify persistent effects across studies and methods to take
3061 into analyses presented later in this thesis. Metabolites investigated here include directly
3062 measured metabolites such as the amino acids tyrosine and phenylalanine, as well as derived
3063 (not-directly measured) metabolite measures such as the ratio of saturated fatty acids to total
3064 fatty acid.

3065 **5.2 Methods**

3066 This chapter details hypothesis-free, two-sample summary-level MR analyses and subse-
3067 quent meta-analyses (Figure 5.1). Genetic variants, single nucleotide polymorphisms (SNPs),
3068 associated with variation in three measures of adiposity (BMI⁴¹, WHR⁴², and BF³⁹) measured
3069 in individuals of European ancestries were used as genetic instrumental variables for the
3070 three adiposity exposures. Each of the exposures (BMI, WHR, and BF) capture different

3071 components of adiposity, and the associated genetic variants explain different amounts of
3072 variation in the respective trait.

3073 A parallel MR analysis was performed on the following outcomes: (1) 123 nuclear magnetic
3074 resonance (NMR) derived metabolites measured in 24,925 individuals of European ancestries
3075 from Kettunen et al. (2016)³¹⁸; (2) 230 NMR derived metabolites from 40,849 individuals of
3076 European ancestries from INTERVAL (unpublished). The main MR analysis consisted of an
3077 inverse variance weighted multiplicative random effects (IVW-MRE) model and sensitivity
3078 analyses using additional models (MR-Egger, weighted median, and weighted mode). Meta-
3079 analyses of all metabolites shared between the Kettunen and INTERVAL studies was also
3080 performed. Meta-analyses included up to 65,774 individuals of European ancestries. These
3081 datasets, including the adiposity data detailed later on, were inverse rank normally transformed
3082 prior to genome-wide analysis. Traditionally, and throughout this Chapter, units from these
3083 GWAS are given as standard deviations (SD). However, due to this transformation the data
3084 are in essence dimensionless and have no units⁵²¹. The benefit of this transformation is that
3085 estimates from these GWAS can be easily compared across other GWAS using the same
3086 transformation.

3087 All data manipulation and analyses were performed using R⁴⁷⁹ (version 3.5.3). MR
3088 analysis was performed using the TwoSampleMR⁵²² (version 0.4.22) R package. Data for
3089 exposures were obtained from published summary statistics (Table ??). Summary statistics
3090 for the Kettunen et al. (2016)³¹⁸ metabolites were obtained from MR Base⁵²² (accessed
3091 26/07/2019), while those for INTERVAL were unpublished and obtained from collaborators
3092 (Adam Butterworth, University of Cambridge). A list of all metabolites is available in the
3093 Appendix (Table A.5) and on [GitHub](#). A list of all genetic variants used as instrumental
3094 variables are available on [GitHub](#). Meta analyses were conducted using the meta (version
3095 4.18-0) R package. Results were visualised using the EpiViz (version 0.0.0.9; detailed in
3096 Chapter 3) and ggforestlot (version 0.1.0) R packages.

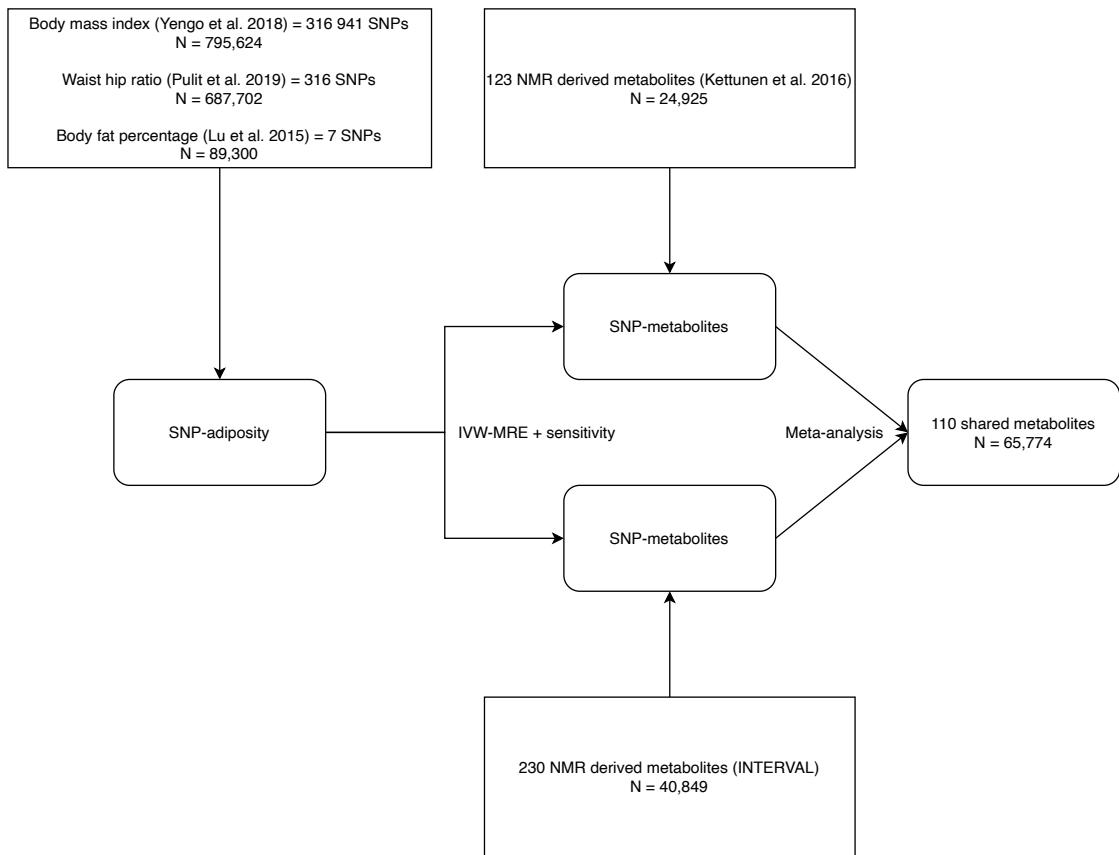


Figure 5.1: Analysis overview: Effect of adiposity on metabolites. The main analysis consisted of two parallel two-sample Mendelian randomization (MR) analyses using three exposures (BMI (body mass index), WHR (waist hip ratio), and BF (body fat percentage)). Two outcome metabolomics data sets were used: (1) 123 nuclear magnetic resonance (NMR) derived metabolites measured in 24,925 individuals of European ancestries from Kettunen et al. (2016)³¹⁸; (2) 230 NMR derived metabolites from 40,849 individuals of European ancestries from INTERVAL³²¹. Four models were used: inverse variance weighted, multiplicative random effects (IVW-MRE) model was the main analysis; MR-Egger, weighted median, and weighted mode were used as sensitivity analyses. The number of SNPs used for each exposure along with the N for the exposure and outcome data sets are given. A meta-analysis of the IVW-MRE results of 110 metabolites that were shared across the two studies was conducted.

3097 **5.2.1 Instrumentation**

3098 Instrumentation of exposures is primarily achieved using either a single genetic instru-
 3099 mental variable or multiple genetic instrumental variables. A limitation of using a single
 3100 genetic variant to instrument an exposure is that the proportion of variance explained by

3101 that variant on the exposure will likely be low (although some cases exist where cis-variants
3102 near or in the protein coding region of the gene lead to changes in protein levels, such as
3103 CRP). This will result in a weak instrument which can lead to biased estimates towards the
3104 confounded observational effect³³⁸. Using multiple genetic instruments which, collectively,
3105 explain a greater proportion of the trait variance than any one individual variant can mitigate
3106 weak instrument bias.

3107 In Chapter 2, 173 studies which used a measure of adiposity as an exposure in 2,214
3108 MR analyses were reviewed. The majority of these 2,214 MR analyses used BMI (N =
3109 1,509) as the exposure; 112 analyses used WHR and 45 used BF, the remaining analyses
3110 used differing measures of fat mass and depositions such as hepatic fat. The majority of
3111 analyses used multiple instruments. In many instances, it was not possible to identify the
3112 instruments used due to absence of information or miss-reporting. On the whole however,
3113 the main instrumentation strategy was to obtain instruments from the most recent GWAS with
3114 the largest sample size. This resulted in most analyses using a genome-wide significance
3115 threshold of 5×10^{-8} . Few studies discussed the independence of their genetic variants, and
3116 those that did mostly reported that the original GWAS declared the SNPs to be independent
3117 from one another - very few analyses performed clumping of their obtained instruments. In
3118 order to allow comparison between results here and those previously reported, the instru-
3119 mentation approach used most commonly in previous analyses was used here. For the main
3120 analysis here, instruments were obtained directly from the most recently published GWAS
3121 with the largest sample size. Additional analyses used instruments obtained from GWAS
3122 which did not use UK Biobank, due to potential concerns over population structure^{346,523–525},
3123 and performed clumping of identified instruments to ensure independence of instruments in
3124 the lists.

3125 The majority of MR analyses identified in Chapter ?? investigated the effects of adiposity
3126 on health outcomes using sex combined exposure and outcome data. Where the outcome was
3127 sex specific, the majority of analyses used sex specific exposure data. In the analyses detailed
3128 here in, exposure and outcome data were sex combined. However, this does not necessarily
3129 mean that results can not be used to inform subsequent analyses which may involve sex
3130 specific outcomes. Many studies identified in Chapter ?? used sex combined exposure
3131 data to investigate sex specific outcomes such as ovarian,⁵²⁶ breast⁴³⁵, and endometrial
3132 cancer^{457,527}, as well as cardiovascular disease⁵²⁸.

3133 5.2.2 Data

3134 The following details a number of GWAS and meta-analyses of previously published
3135 studies which were used in this chapter to perform MR analyses and are required reporting
3136 as per STROBE-MR guidelines⁴¹⁷. I did not perform these analyses. The total N and sex
3137 specific N does not always tally, this is because of variation in the N for each SNP. Where this
3138 is the case, 'N up to' is used.

3139 **Exposures: measures of adiposity**

3140 **Body mass index** Robustly associated genetic variants associated with BMI were ex-
3141 tracted from Yengo et al. (2018)⁴¹, who analysed data from 515,509–795,624 individuals of
3142 European ancestries from two studies, the Genetic Investigation of Anthropometric Traits
3143 (GIANT) consortium³⁶ and UK Biobank. In both studies BMI was calculated as $\frac{\text{weight (kg)}}{\text{height (m}^2)}$.
3144 Yengo et al. (2018)⁴¹ performed a fixed effect inverse variance weighted meta-analysis of
3145 BMI using GWAS results generated from UK Biobank (N = 456,426) and results from The
3146 Genetic Investigation of ANthropometric Traits (GIANT) consortium, Locke et al. (2015)³⁶
3147 (N = 322,154). These genetic variants were used as genetic instrumental variables in MR
3148 analyses.

3149 In UK Biobank, BMI was adjusted for age, sex, recruitment centre, genotyping batch, and
3150 the first 10 principal components (PC) calculated from 132,102 (out of 147,604) genotyped
3151 SNPs pre-selected by the UK Biobank quality control team⁵²⁹. In GIANT, BMI was adjusted
3152 for age, age squared, and study specific covariables (e.g. genotype-derived principal com-
3153 ponents). The residuals from both UK Biobank and GIANT were inverse normally transformed
3154 and therefore represent normalised standard deviation (SD) units. In total, 681,275 individ-
3155 uals of European ancestry and ~2.4 million HapMap-2 imputed SNPs were included in the
3156 meta-analysis. The intercept for linkage disequilibrium score regression (LDSC 1.03; SE =
3157 0.02) suggested population stratification however, LDSC can rise above 1 as sample sizes
3158 and heterogeneity increase⁵³⁰.

3159 In the combined meta-analysis, a total of 656 primary associations reaching a genome
3160 wide significance threshold of 5×10^{-8} were identified. Approximate conditional and joint
3161 multiple-SNP (COJO) analysis identified a further 285 independent SNPs reaching an ad-
3162 justed genome-wide significance threshold of 1×10^{-8} . Together, these 941 associations
3163 explain 6% (SE = 0.8%) and 22.4% (SE = 3.7%) of the variance and heritability of BMI
3164 respectively. COJO-specific summary statistics for all 941 SNPs were used in the main
3165 analysis.

3166 **Waist hip ratio** Robustly associated genetic variants associated with WHR were extracted
3167 from Pulit et al. (2019)⁴², who analysed data from 485,486—697,702 individuals of European
3168 ancestries from two studies: UK Biobank and the GIANT consortium³⁷. In both studies, WHR
3169 was calculated as $\frac{\text{waist circumference (cm)}}{\text{hip circumference (cm)}}$. Pulit et al. (2019)⁴² performed a fixed effects
3170 inverse variance weighted meta-analysis of WHR using GWAS results generated from UK
3171 Biobank (N up to = 485,486 (men up to = 263,148; women up to= 222,338)) and results from
3172 the GIANT consortium (N = up to212,248 (women up to= 118,004; men up to= 94,434))³⁷
3173 using METAL⁵³¹. SNPs with a frequency difference > 15% between the two studies were
3174 removed prior to meta-analysis. In both studies, WHR was adjusted for sex (UK Biobank

3175 only), age at assessment, age at assessment squared, and assessment centre (UK Biobank
3176 only); study specific covariables were included in the GIANT GWAS where appropriate. The
3177 residuals from both UK Biobank and GIANT were inverse normally transformed and therefore
3178 represent normalised SD units. In GIANT residuals were calculated for men and women
3179 separately. In total, 316 associations reaching a genome-wide significance threshold of 5×10^{-9}
3180 were identified. The p-value was adjusted to account for denser imputation data⁵³².
3181 These associations explain 3% of the phenotypic variance as calculated in an independent
3182 dataset (N = 7,721). SNP based heritability across all SNPs was estimated to be 22.7%.
3183 These genetic variants were used as genetic instrumental variables in MR analyses.

3184 For the UK Biobank GWAS, the second release (June 2017) of UK Biobank data, which
3185 did not have corrected imputation at non-HRC sites, was used. Pulit et al performed a linear
3186 mixed model (LMM) using BOLT-LMM⁵³³ on SNPs with imputation quality score > 0.3, MAF >
3187 0.01%, and SNPs present in the HRC imputation reference panel. The LMM was adjusted
3188 for the SNP array that was used to genotype each sample, i.e. genome-wide or metabochip
3189 array. No other covariables were included in the mode. They used an infinitesimal model in
3190 BOLT-LMM. In GIANT, individual studies recruited participants and undertook sample and
3191 SNP quality control. In studies using genome-wide arrays, imputation was performed with
3192 CEU (Utah residents of northern and western European ancestry) haplotypes from HapMap.
3193 Assuming an additive genetic model, each study ran a linear regression GWAS. Sex-specific
3194 summary statistics were corrected for population structure using the genomic control inflation
3195 factor. Prior to meta-analysis of studies using genome-wide array or Metabochip, SNPs
3196 were removed if they had a minor allele count ≤ 3 , were not in Hardy-Weinberg equilibrium
3197 (p-value $< 10^{-6}$), had a call rate $< 95\%$ or an imputation quality score < 0.3 for MACH, < 0.4 for
3198 IMPUTE, and < 0.8 for PLINK. In step 1, a meta-analysis of each array was performed using
3199 a fixed effects inverse variance weighted model and corrected for genomic control to account
3200 for structure between cohorts. In step 2, the meta-analysed genome wide array studies and
3201 meta-analysed Metabochip studies were meta-analysed using a fixed effects inverse variance
3202 weighted model using METAL. In this second step genomic correction was not performed.

3203 **Body fat percentage** Robustly associated genetic variants associated with BF were ex-
3204 tracted from Lu et al. (2016)³⁹. In this GWAS individuals with dual-energy X-ray absorptiometry
3205 (DXA) or bioelectrical impedance measured BF were included. The number of individuals
3206 with DXA and impedance measured BF was not reported, instead a total N was reported (N
3207 up to = 89,297). Lu et al. (2016)³⁹ performed a fixed effects inverse variance weighted meta-
3208 analysis of BF using two meta-analyses generated from genome-wide array (N = 65,831)
3209 and Metabochip array GWAS (N = 23,468) using METAL⁵³¹. In total, up to 89,297 (men up
3210 to= 44,429; women up to= 45,525) individuals of European ancestries were included in the
3211 inverse variance weighted fixed effects meta-analysis. In total 7 SNPs reached a genome
3212 wide significance threshold (p-value $< 5 \times 10^{-8}$) and were considered independent ($\pm 500\text{kb}$ of
3213 the most significant SNP). Estimation of variance explained was not available in the European
3214 ancestries meta-analysis. In a meta-analysis of individuals of all ancestries, which included up
3215 to 11,419 additional individuals of non-European ancestries, these 7 SNPs explained 0.416%

3216 of the variance in BF. The additional 5 SNPs identified in this all ancestries meta-analysis
3217 explained 0.58% of the variance in BF. These 7 genetic variants, identified in the European
3218 ancestries meta-analysis, were used as genetic instrumental variables in MR analyses.

3219 In each cohort, BF was measured via a bioelectrical impedance device or DXA as
3220 described previously³⁴. For each study, BF was adjusted for age, age squared, and study-
3221 specific covariables (e.g. genotype-based principle components and study centre) if necessary.
3222 These residuals, for studies of unrelated individuals, were calculated separately in men and
3223 women, and separately in cases and controls. For studies of family-based design, the
3224 residuals were calculated in men and women together, and sex was additionally included in
3225 the model. The residuals were then inverse rank normally transformed prior to association
3226 testing, and therefore represent normalised SD units. For studies of family-based design, the
3227 family relatedness was additionally included in the association testing.

3228 Lu et al performed their meta-analysis in two stages. First, two parallel meta-analyses
3229 were conducted; one meta-analysis combined summary statistics from genome-wide array
3230 GWAS, totalling up to 65,831 individuals of European ancestries and the other meta-analysis
3231 combined summary statistics from Metabochip array GWAS, totalling up to 23,469 individuals
3232 of European ancestries. Each study performed study specific quality control. Imputation
3233 was performed in each study using the European ancestries HapMap Phase II (Release 22)
3234 reference panel. Associations between individual SNPs and inverse normally transformed
3235 BF residuals were identified using linear regression assuming an additive genetic model. All
3236 SNPs with low imputation scores (MACH r2-hat < 0.3, IMPUTE proper_info < 0.4, or PLINK
3237 info < 0.8) and a MAC \leq 3 were removed. In the second stage, meta-analysis of these two
3238 meta-analyses was performed using an inverse variance-weighted fixed-effect model using
3239 METAL.

3240 Each unique locus was defined as $\pm 500\text{kb}$ on either side of the most significant SNP
3241 that reached the genome wide significance threshold ($p\text{-value} < 5 \times 10^{-8}$) in the meta-
3242 analysis. Genotype data for genome-wide significant SNPs were of high quality with a median
3243 imputation score of ≥ 0.95 .

3244 Outcomes: metabolites

3245 **Kettunen et al. (2016)³¹⁸** Genome-wide summary level data for 123 NMR derived metabo-
3246 lites, which includes derived metabolite measures (Appendix Table A.5), were available from
3247 Kettunen et al. (2016)³¹⁸. Kettunen et al (2016)³¹⁸ performed a fixed effects meta-analysis of
3248 123 serum and EDTA (Ethylenediaminetetraacetic acid) serum NMR quantified metabolites
3249 from 14 GWAS. In total, up to 24,925 individuals of European ancestries were included in the
3250 meta-analysis.

3251 In each of the 14 cohorts 123 metabolites were quantified from human blood. Four cohorts

3252 used EDTA-plasma, the other 10 used serum. The majority of blood samples were fasted.
3253 Where a study did not have over night fasted samples, correction for fasting time effect was
3254 performed using the *gam* R package and fitting a smoothed spline to adjust for fasting. The
3255 NMR analysis was performed with the comprehensive quantitative serum/plasma platform
3256 described by Soininen et al. (2009)⁴⁹². All metabolites were adjusted for age, sex, time from
3257 last meal if applicable, and the first ten PCs. The residuals for each adjusted metabolite were
3258 inverse rank normally transformed and therefore represent normalised SD units. Each of the
3259 14 cohorts performed a univariate GWAS assuming an additive genetic model. SNPs were
3260 imputed up to 39 million markers using the 1000 Genomes Project, March 2012 version.

3261 For the meta-analysis, SNPs with accurate imputation (*info* > 0.4) and minor allele
3262 count > 0.3 were combined using double genomic control correction, that is, both individual
3263 cohort results and meta-analysis results were corrected for the genomic inflation factor as
3264 implemented in *GWAMA*. Up to 12,133,295 SNPs were included in the meta-analysis. Variants
3265 present in more than seven studies after filtering and meta-analysis were considered for the
3266 final results. All traits gave genomic inflation factors in the meta-analysis < 1.034 showing
3267 little evidence of systematic bias in the test statistic. A genome-wide significance threshold
3268 (*p*-value < 2.27 x 10⁻⁹) after correcting for 22 independent tests (22 being the number of
3269 principal components explaining over 95% of variation in the metabolite data) was used.

3270 **INTERVAL** Genome-wide summary level data for 230 NMR derived metabolites, which
3271 includes derived metabolites measures (Appendix Table A.5), were available from the IN-
3272 TERVAL study. INTERVAL is a randomised trial of ~50,000 individuals recruited from June
3273 2012–June 2014 in the United Kingdom to investigate the safety of different blood donation
3274 intervals⁵²⁰. In INTERVAL, a LMM GWAS of 230 serum NMR quantified metabolites was
3275 performed. In total, up to 40,849 individuals of European ancestries were included. The
3276 INTERVAL GWAS data were unpublished and provided by collaborators (Adam Butterworth,
3277 University of Cambridge).

3278 SNPs were imputed using a combined 1000 Genomes + UK10K panel, which captured
3279 87,696,910 variants. A LMM was fit using BOLT-LMM. Variants that were poorly imputed (*info*
3280 score < 0.3 or R^2 < 0.3), variants with unrealistic results (e.g. standard error < 0, standard error
3281 > 10, beta = infinity), and variants with minor allele count < 5 were excluded. A genome-wide
3282 significance threshold was not set. A total of 28 PC explained over 95% of variation in the
3283 metabolite data.

3284 The NMR analysis was performed with the comprehensive quantitative serum/plasma
3285 platform described by Soininen et al. (2009)⁴⁹². Samples or analytes with potentially unreliable
3286 data were flagged. Flagged metabolites included acetate, pyruvate, glucose, and lactate.
3287 Individuals with flagged values for any of these metabolites had those values set to missing. In
3288 addition, metabolites with values of 0 or values > 10 SD from the mean after log transformation
3289 were set to missing. Individuals were excluded if they had > 30% analyte missingness, and

3290 one NMR PC outlier. All metabolites were log transformed and then adjusted for age, sex,
3291 recruitment centre, time between blood draw and sample processing, and the first 10 PCs of
3292 genetic ancestry. Residuals were then inverse normal rank transformed prior to GWAS and
3293 therefore represent log normalised SD units.

3294 **5.2.3 Two-sample Mendelian randomization: effect of adiposity measures on**
3295 **metabolites**

3296 For all exposures, the following data were obtained from the original GWAS publications:
3297 rsID, effect allele, other/non-effect allele, effect allele frequency, effect estimate, standard error
3298 of the effect estimate, p-value, N, and units. The same data for these genetic instrumental
3299 variables were obtained from each metabolite for each exposure separately. Clumping of
3300 SNPs was not performed as the studies from which they were obtained stated they were
3301 independent or near-independent and had already been clumped. For BMI, near-independent
3302 SNPs were identified by Yengo et al. as those with the smallest p-value within a 1 mega-base
3303 (Mb) window, as well as those additional SNPs within that same window that reached the
3304 p-value threshold (1×10^{-8}) after COJO analysis⁴¹. For WHR, lead SNPs (p-value $< 5 \times 10^{-9}$),
3305 and all SNPs in LD ($r^2 > 0.05$) with the lead SNPs, within a 10Mb window were identified
3306 by Pulit et al. using the PLINK clumping algorithm. The genomic span of each LD-based
3307 clump was identified and a 1kb buffer was added up- and down-stream. Where windows
3308 overlapped, they were merged into a single locus⁴². For BF, Lu et al.³⁹ defined each locus by
3309 the SNP with the smallest p-value that reached the GWS threshold (p-value $< 5 \times 10^{-8}$) and a
3310 1Mb window centered around that SNP. F-statistics for each SNP and an average for each
3311 exposure were calculated.

3312 Genetic instrumental variables were extracted from the metabolite GWAS and where
3313 these were not present, proxy SNPs were included if linkage disequilibrium was ≥ 0.8 . For
3314 proxy SNPs, the inclusion of SNPs where the reference strand was ambiguous (strand flips)
3315 was allowed and the direction was inferred using a minor allele frequency threshold. That is,
3316 the direction was inferred using a minor allele frequency, so long as that frequency was not \geq
3317 0.3 or ≤ 0.7 , in which case it was excluded. Exposure and outcome SNPs were harmonized
3318 in reference to the exposure effect allele being on the increasing scale. For included alleles
3319 where the reference strand was ambiguous, the positive strand was inferred using effect allele
3320 frequency.

3321 An inverse variance weighted (IVW), multiplicative random effects (IVW-MRE) model was
3322 used to investigate the effect of each exposure on each metabolite. The model assumes that
3323 the strength of the association of the genetic instruments with the exposure is not correlated
3324 with the magnitude of the pleiotropic effects and that the pleiotropic effects have an average
3325 value of zero⁵³⁴. A multiple testing threshold of p-value < 0.0023 was applied to the Kettunen
3326 et al. (2016) metabolomics data. This is based on the number of principal components (22) in
3327 the Kettunen et al. (2016) meta-analysis that explained 95% of the variation in metabolite data.

3328 A multiple testing threshold of p-value < 0.0018 was applied to the INTERVAL metabolomics
3329 data. This is based on the number of principal components (28) in the INTERVAL GWAS that
3330 explained 95% of the variation in metabolite data.

3331 The metabolic profile of adiposity was visualised using Circos plots. Directions of effect
3332 were compared across exposures for analysis using Kettunen data and analysis using
3333 INTERVAL data. Tests which reached a multiple testing threshold within each analysis are
3334 presented and the effect of adiposity on subclasses were explored in regard to the multiple
3335 testing threshold.

3336 **Sensitivity analysis**

3337 Where possible, the assumptions of no pleiotropy among genetic instruments and out-
3338 comes were explored using: MR-Egger⁵³⁵, weighted median⁵³⁶, and weighted mode⁵³⁷ based
3339 estimators. Sensitivity analysis was performed for all exposure-outcome pairs, but focus was
3340 given to those pairs which met a multiple testing threshold in the main analysis. For these
3341 sensitivity models, no threshold requirements were set, instead, consistency between the
3342 IVW-MRE model and these methods was investigated.

3343 MR-Egger provides an estimate of unbalanced/directional horizontal pleiotropy via the
3344 intercept of a linear regression of the SNP-exposure and SNP-outcome association. In
3345 the presence of pleiotropy the intercept will bias away from the origin. MR-Egger gives
3346 consistent estimates when 100% of genetic instruments are invalid⁵³⁵. The weighted median
3347 is complimentary to MR-Egger but does not rely on the “instrument strength independent of
3348 direct effect” (InSIDE) assumption. It calculates the median of an empirical distribution of
3349 the causal effect estimates weighted for precision. It provides consistent estimates when at
3350 least 50% of the weight comes from valid genetic instruments and as long as no one genetic
3351 instrument contributes > 50% of the weight⁵³⁶. The weighted mode assumes the true causal
3352 effect is the most common effect, it is robust when the majority of effect estimates are derived
3353 from valid instruments⁵³⁷.

3354 In these analyses, it is assumed that the instruments influence the exposure first. Their
3355 effect on the outcome through the exposure is secondary. Given the large number of
3356 genetic instruments used, as well as the feed-back and feed-forward loops present in the
3357 metabolome, the directionality of association may be difficult to evaluate. For example, a SNP
3358 which is strongly associated with adiposity may also have an effect on the metabolite under
3359 investigation via that SNPs association with another metabolite which is up- or down-stream
3360 of that metabolite. The Steiger test⁵³⁸, applied here using the TwoSampleMR package, can
3361 be used to estimate whether the test under investigation is the “true” causal direction. The
3362 Steiger test calculates the variance explained in the exposure and the variance explained in
3363 the outcome by the exposure related instruments. If the variance explained in the outcome
3364 is less than that explained in the exposure, then the direction of effect can be considered

3365 to be from the exposure to the outcome. In order to estimate the variance explained the N
3366 is required. For the Kettunen dataset, the N for each individual SNP was available. For the
3367 INTERVAL dataset, the overall sample size (N = 40,849) was used as the individual SNP
3368 sample sizes were not available.

3369 **Additional sensitivity analysis**

3370 The effects of heterogeneity in the exposure instruments was investigated using Cochran's
3371 Q statistic for IVW-MRE and MR-Egger models. A single SNP MR, whereby the association
3372 of each SNP individually is estimated using the IVW and MR-Egger models, and Wald ratio
3373 where appropriate, was also used to investigate heterogeneity. A "leave-one-out" MR analysis
3374 was performed whereby each SNP was sequentially left out of the MR analysis and the causal
3375 effect estimated absent of that SNP. If the estimated effect is substantially altered after the
3376 removal of a single SNP, this may imply that SNP is driving the association between the
3377 exposure and outcome. Finally, each of these additional sensitivity analyses were visualised
3378 using a funnel plot and using forest plots of the single SNP and leave-one-out MR analysis to
3379 identify potential pleiotropic effects. These additional sensitivity analysis are presented as a
3380 summary for each analysis.

3381 **Additional analyses**

3382 There were a number of potential limitations with the genetic instrumental variables used
3383 in the main analysis. In order to test whether these limitations produced biased results the
3384 following post-hoc additional analyses were performed. Firstly, both BMI and WHR genetic
3385 instrumental variables were obtained from studies using UK Biobank, which has shown
3386 evidence of latent population structure^{346,347}. BF instruments were obtained from a study
3387 which used different measures of BF; two of the included SNPs have also been associated
3388 previously with 'favourable adiposity' and may therefore not be reflections of total body fat^{35,539}.
3389 Additional genetic instrumental variables for BMI, WHR and BF were obtained from additional
3390 published summary statistics (Appendix Table ??). Briefly, SNPs for BMI were obtained
3391 from the initial non-COJO analysis by Yengo et al 2018⁴¹, and a separate set of SNPs were
3392 obtained from Locke et al. (2015)³⁶ - which did not use UK Biobank; for WHR, SNPs were
3393 obtained from Shungin et al. (2015)³⁷ - which did not use UK Biobank; for BF, the SNPs
3394 associated with 'favourable adiposity' were removed and an additional SNP set identified in a
3395 GWAS using a single measure of BF was obtained from Hubel et al. (2019)⁴³. Outcome data
3396 extraction and harmonization was performed as for the main analysis. F-statistics (Figure
3397 A.12) and detailed information on each additional exposure is available in the Appendix,
3398 section ?? . The main analysis (IVW-MRE, MR-Egger, weighted median, weighted mode and
3399 additional sensitivity analysis) was re-run using these additional exposures.

3400 In addition, studies used a variety of methods and definitions of independence for SNPs.
 3401 In order to ensure SNPs were independent to the same degree across all studies, genetic
 3402 instrumental variables for all exposures (those used in the main analysis and in additional anal-
 3403 yses described here) were clumped and the main analysis (IVW-MRE, MR-Egger, weighted
 3404 median, weighted mode and additional sensitivity analysis) was re-run using these clumped
 3405 instruments. Clumping was performed using the R package TwoSampleMR⁵²² setting an LD
 3406 R² threshold of 0.001 for SNPs within a 10,000 base window of each other. Spearman's
 3407 correlation between MR results from the non-clumped and clumped SNP lists was performed.

3408 **5.2.4 Meta-analysis of two-sample Mendelian randomization results**

3409 Metabolomics data from Kettunen et al. (2016)³¹⁸ were inverse rank normally transformed
 3410 prior to GWAS. For INTERVAL, metabolomics data were log transformed and then inverse
 3411 rank transformed prior to GWAS. As these transformations use different scales, meta-analysis
 3412 using MR effect estimates was not appropriate. Instead, a meta-analysis of each exposure-
 3413 outcome pair was performed with p-values using the metap (version 1.4) R package using
 3414 Fisher's method for combining p-values. Meta-analysis of the IVW-MRE results from the main
 3415 analysis was performed.

3416 Fisher's method (5.1) combines p-values from each test into one test statistic (X^2), where
 3417 p_i is the p-value for the i^{th} test, k is the number of tests, and $2k$ is the degrees of freedom.
 3418 When all the null hypotheses are true, and the p_i are independent, X^2 has a chi-squared
 3419 distribution with $2k$ degrees of freedom.

$$X_{2k}^2 \sim -2 \sum_{i=1}^k \ln(p_i), \quad (5.1)$$

3420 The chi-squared statistic will follow a chi-squared distribution with $2k$ degrees of freedom if
 3421 there is no effect in every study, where k is the number of tests. A large chi-squared statistic
 3422 compared to the degrees of freedom (with a corresponding low p-value) provides evidence of
 3423 an effect in at least one study. The null hypothesis is that each tests null hypothesis is true;
 3424 the alternative is that at least one of the tests null hypotheses is true. Acceptance of the null
 3425 hypothesis is not interpreted as evidence of no effect in all studies. Given only two studies
 3426 are included in any one exposure-outcome pair in the current work, the potential effects of
 3427 heterogeneity are difficult to interpret and are therefore not focused on here. Meta-analysis
 3428 results are presented as directions of effect and p-values. A Bonferroni corrected p-value
 3429 threshold was used to identify evidence of association.

3430 **5.2.5 Comparison of two-sample Mendelian randomization and observational**
3431 **results from Chapter 2**

3432 Metabolites included in the meta-analysis, which showed consistent directions of ef-
3433 fect across the Kettunen and INTERVAL data and which reached a Bonferroni corrected
3434 meta-analysis p-value threshold, were compared with observational results from Chapter
3435 4. Direction of effect in the meta-analysis was compared with direction of effect in the ob-
3436 servational analysis to triangulate evidence, where consistency across methods improved
3437 confidence in causal effects.

3438 **5.3 Results**

3439 The effects of BMI, WHR and BF on a total of 123 metabolites from Kettunen et
3440 al. (2016)³¹⁸ and 230 metabolites from INTERVAL were investigated using an IVW-MRE
3441 model. F-statistics for all SNPs used as genetic instruments for each exposure were above
3442 10 (mean F-statistics: BMI = 73, WHR = 66, BF = 44; Figure A.12; [GitHub](#)).

3443 For MR analyses using the Kettunen et al. (2016) metabolites, effect estimates are given
3444 in SD units per SD higher BMI/WHR/BF. For MR analyses using the INTERVAL metabolites,
3445 effect estimates are given in log SD units per SD higher BMI/WHR/BF. Metabolites were
3446 grouped into subclasses by the metabolomics analytic platform. The Kettunen data included
3447 two subclasses (Metabolites ratio and Protein) which were not present in the INTERVAL data
3448 In the INTERVAL data, 16 subclasses, which were all derived metabolic measures (i.e. ratios
3449 of one measured metabolite to another), were not available in the Kettunen data

3450 **5.3.1 Two-sample Mendelian randomization: effect of adiposity measures on**
3451 **metabolites**

3452 **Metabolic profile**

3453 Effect estimates from the IVW-MRE model revealed evidence for a broad effect of adiposity
3454 on the metabolic profile (Figure 5.2 and 5.3). The pattern of association was visually similar
3455 when using the Kettunen data for BMI and WHR, and when using the INTERVAL data for
3456 BMI and WHR. When using the Kettunen data, the effect of BF on metabolites was visually
3457 distinct to that of BMI and WHR effects, with many effect estimates appearing to be inverse to
3458 those for BMI and WHR. However, when using the INTERVAL data, the effect of BF looked
3459 more similar in regards to the direction of effect to that of BMI and WHR compared to effects
3460 in the Kettunen data. However, these effects appear to cross the null much more often.

3461 Effects for WHR were generally larger with wider confidence intervals, whereas those for
3462 BMI were smaller with tighter confidence intervals. Effects for BF were much larger across
3463 both metabolite datasets, with wide confidence intervals which spanned the null. The wide
3464 confidence intervals observed for BF and the tighter confidence intervals for BMI are perhaps
3465 expected given the variance explained by instruments for each measure of adiposity varies
3466 from 0.4% for BF, 3% for WHR, and 6% for BMI.

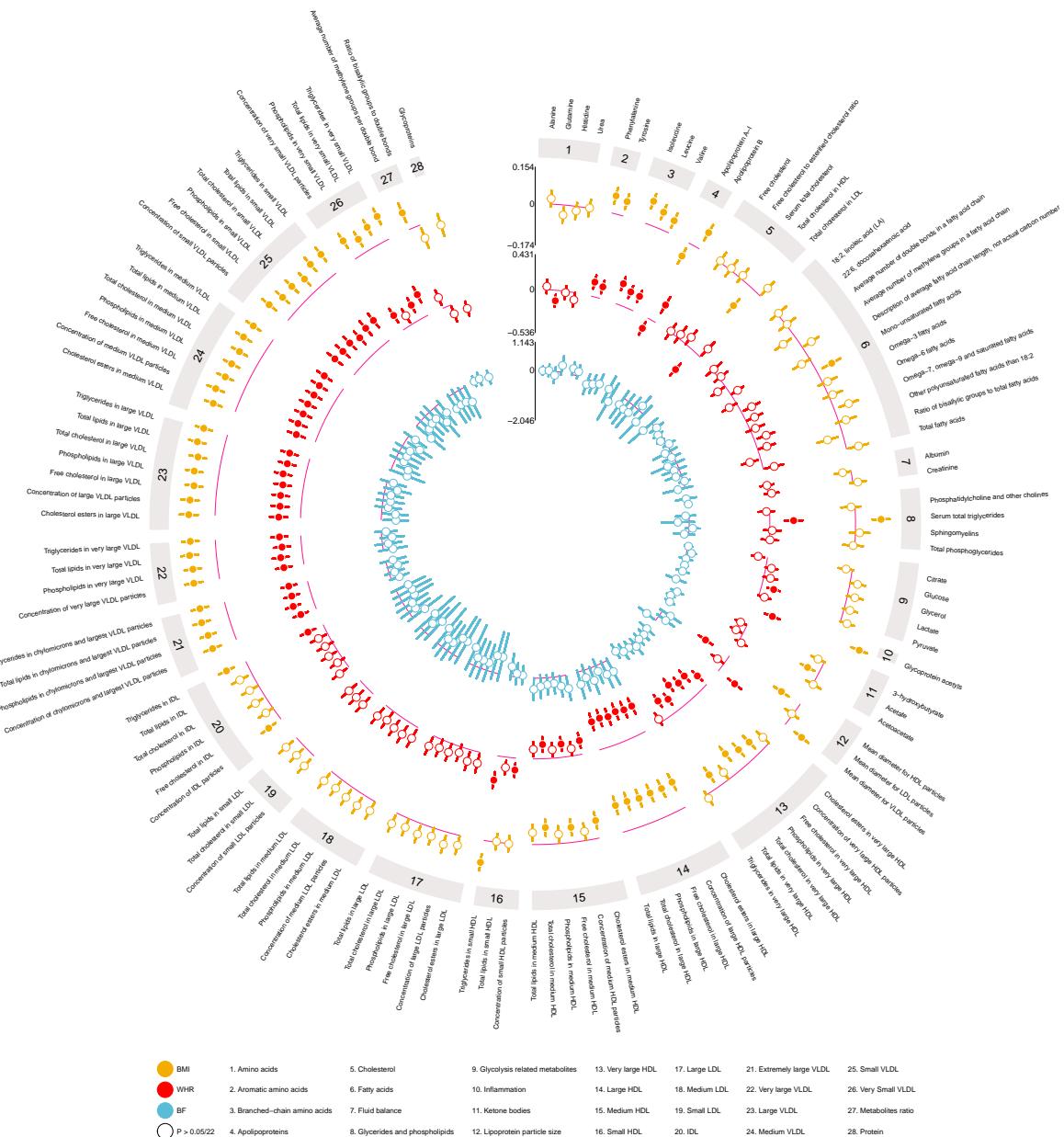


Figure 5.2: **Mendelian randomization analysis: Effect of adiposity measures on 123 NMR derived metabolites using Kettunen data.** Circos plot shows each track as one measures of adiposity; the outer track is body mass index (BMI), the middle track is waist hip ratio (WHR), the inner track is body fat percentage (BF). Solid points indicate a multiple testing threshold (0.0023) has been reached. Effect estimates are given in SD units per SD higher BMI/WHR/BF. 95% confidence intervals shown. Metabolites are grouped by subclass and arranged alphabetically within each subclass.

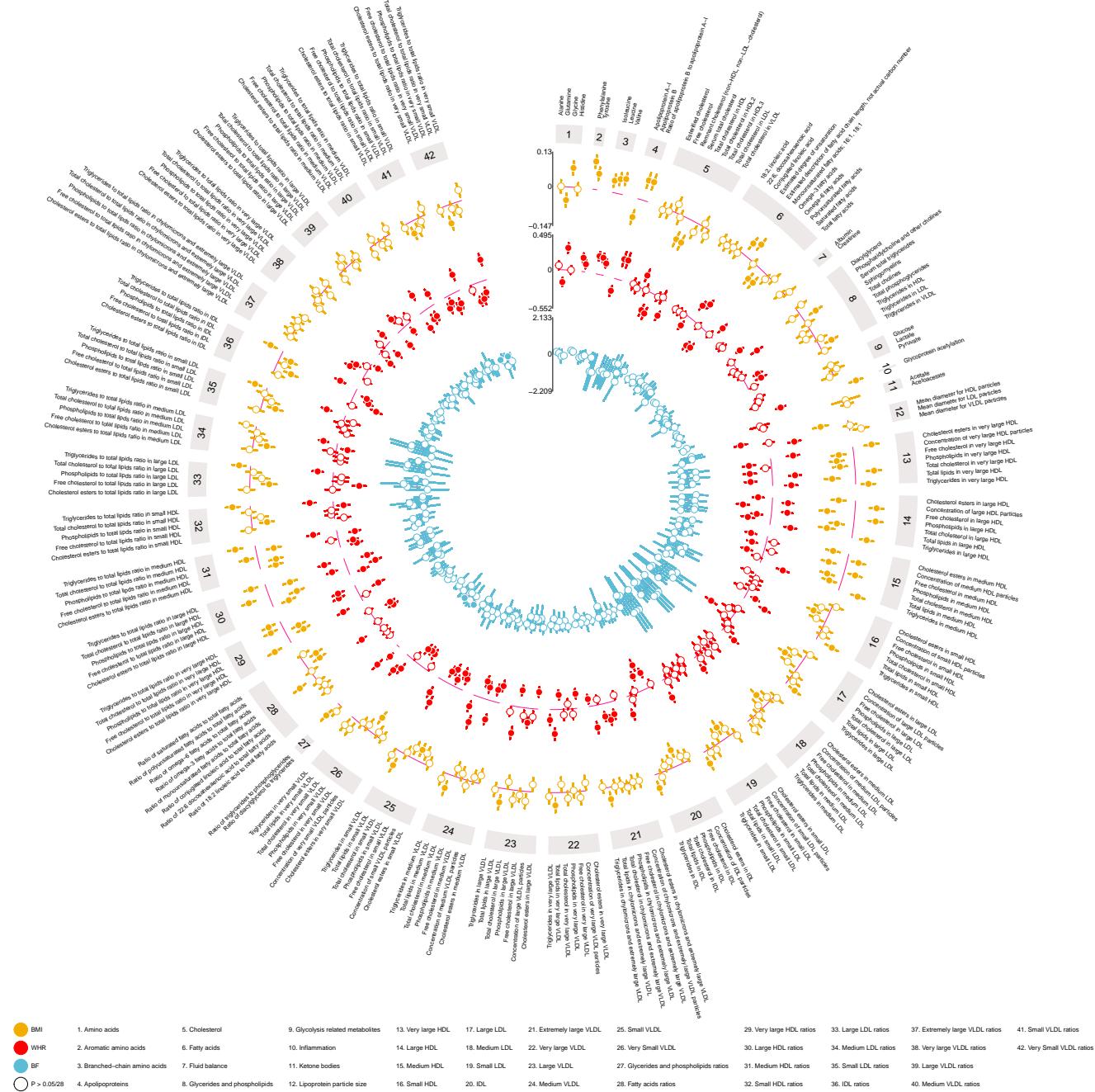


Figure 5.3: Mendelian randomization analysis: Effect of adiposity measures on 230 NMR derived metabolites using INTERVAL data. Circos plot shows each track as one measures of adiposity; the outer track is body mass index (BMI), the middle track is waist hip ratio (WHR), the inner track is body fat percentage (BF). Solid points indicate a multiple testing threshold (0.0018) has been reached. effect estimates are given in log SD units per SD higher BMI/WHR/BF. 95% confidence intervals shown. Metabolites are grouped by subclass and arranged alphabetically within each subclass.

3467 **Consistency of effect direction**

3468 Given adiposity measures used here are highly correlated, and evidence from Chapter
3469 ?? highlighted broad consistency in the direction of effect estimates across adiposity mea-
3470 sures, consistency of effect direction across exposures within datasets was investigated.
3471 Directional concordance or discordance may help to understand the underlying mechanisms
3472 of the adiposity-metabolite relationship, for example, if a metabolite is decreased by BMI
3473 but increased by WHR the effect of deposition over composition will be more important in
3474 downstream analyses. Here, a positive effect is identified if there is a positive effect for all
3475 exposures being assessed, i.e. if BMI, WHR, and BF all had a positive effect on metabolite A
3476 this would be recorded as a positive effect. If however, the effect of BF on metabolite A was
3477 negative this would be recorded as an 'opposite effect'.

3478 Across all exposures, directional consistency of effect estimates from the IVW-MRE
3479 model was low for both metabolite sources from the Kettunen (N opposite effect = 105) and
3480 INTERVAL (N = 178) data (Figure 5.4). This was the same for BF and WHR and for BMI and
3481 BF. Directional consistency was much higher for BMI and WHR for both analyses using the
3482 Kettunen (N = 5) and INTERVAL (34) data

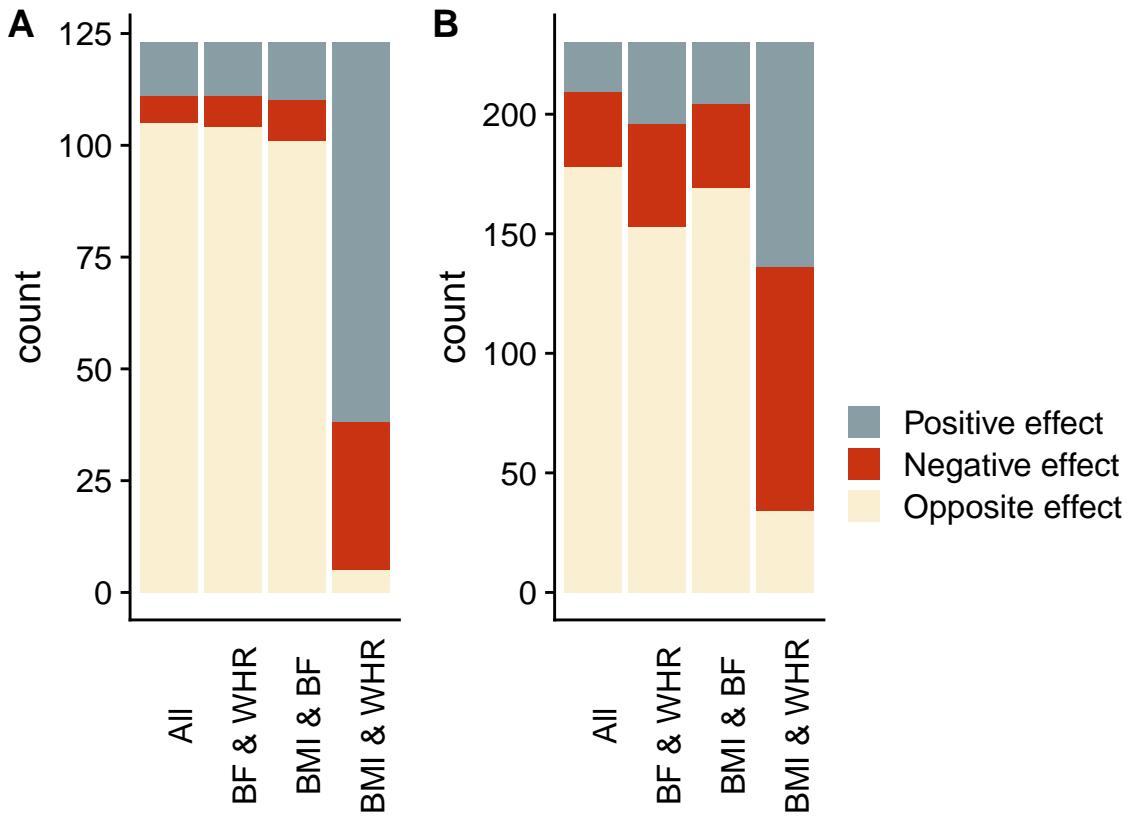


Figure 5.4: **Directional consistency of two-sample MR effect estimates.** A positive effect reflects the effect estimates being either all positive or both positive, depending on whether the comparison is with all three exposure (All) or between just two exposures; a negative effect reflects effect estimates being in a negative direction; opposite effect reflects different directions for the effect estimates across the comparisons. **A:** Two-sample MR IVW-MRE for 123 metabolites from Kettunen et al. (2016)³¹⁸; **B:** Two-sample MR IVW-MRE for 230 metabolites from INTERVAL. BMI = body mass index; WHR = waist hip ratio; BF = body fat percentage.

3483 **Multiple testing threshold**

3484 Multiple testing thresholds of 0.0023 and 0.0018 were used for the analyses using the
 3485 Kettunen and INTERVAL data, respectively. For the 123 metabolites in the Kettunen data,
 3486 a total of 63, 63, and 0 tests reached a multiple testing threshold for BMI, WHR, and BF
 3487 respectively. Of these tests, 58 reached a multiple testing threshold across both BMI and
 3488 WHR. A total of 5 and 5 tests reached a multiple testing threshold for BMI only and WHR only
 3489 respectively. For the 230 INTERVAL metabolites, a larger proportion of metabolites reached
 3490 the multiple testing threshold. A total of 88, 138, and 4 tests reached a multiple testing
 3491 threshold for BMI, WHR, and BF respectively. Of these tests, 76 reached a multiple testing

3492 threshold across both BMI and WHR. A total of 12 and 62 tests reached a multiple testing
3493 threshold for BMI only and WHR only, respectively. None of the 4 tests for BF overlapped with
3494 BMI or WHR tests.

3495 For the 58 tests reaching a multiple testing threshold for BMI and WHR in analysis using
3496 the Kettunen data, the strongest positive and negative effects for BMI were for Total lipids
3497 in chylomicrons and extremely large VLDL ($\beta = 0.12$) and Cholesterol esters in large
3498 HDL ($\beta = -0.12$) respectively. For WHR the strongest positive and negative effects were
3499 found for Triglycerides in small VLDL ($\beta = 0.31$) and Free cholesterol in large HDL ($\beta = -0.38$)
3500 respectively. For the 76 tests reaching a multiple testing threshold for BMI and
3501 WHR in analysis using the INTERVAL data, the strongest positive and negative effects for
3502 BMI were found for Phenylalanine ($\beta = 0.10$) and Mean diameter for HDL particles ($\beta = -0.10$)
3503 respectively. For WHR the strongest positive and negative effects were found for
3504 Ratio of triglycerides to phosphoglycerides ratio ($\beta = 0.40$) and Free cholesterol in large
3505 HDL to total lipids in large HDL ratio ($\beta = -0.41$) respectively. For BF, all 4 tests reaching
3506 the multiple testing threshold were in the subclass Very small VLDL. Three of these (Total
3507 cholesterol in very small VLDL, Cholesterol esters in very small VLDL, and the ratio of these
3508 two) had the same effect sizes and p-value ($\beta = 0.2241230$; p-value = 5.27×10^{-7}). The
3509 fourth metabolite, Total cholesterol in very small VLDL to total lipids in very small VLDL ratio
3510 had a similar effect size ($\beta = 0.27$).

3511 **Subclass results**

3512 When using the Kettunen data, tests reaching the multiple testing threshold (p-value
3513 < 0.0022727) were observed for at least one exposure in 23 of 28 subclasses across the
3514 three exposures. No tests reached the multiple testing threshold for subclasses: small LDL,
3515 medium LDL, large LDL, protein, and fluid balance. However, a number of metabolites in
3516 these subclasses had confidence intervals which did not span the null. For subclasses IDL,
3517 metabolites ratio, ketone bodies, glycolysis related metabolites, glycerides and phospholipids,
3518 cholesterol, and amino acids, only a small number of metabolites within each subclass
3519 reached the multiple testing threshold. Whereas, for subclasses small VLDL, medium VLDL,
3520 large VLDL, very large VLDL, extremely large VLDL, large HDL, very large HDL, lipoprotein
3521 particle size, branched-chain amino acids, and aromatic amino acids a majority of metabolites
3522 reached the multiple testing threshold.

3523 The INTERVAL data grouped metabolites into 42 subclasses. The additional subclasses
3524 to that in the Kettunen data were comprised of ratios (e.g., small HDL ratios). A protein
3525 subclass, available in the Kettunen data, was not available in the INTERVAL data. Of the
3526 42 subclasses, tests reached a multiple testing threshold ($0.05/28$) in 39 subclasses. No
3527 tests reached the multiple testing threshold for subclasses fluid balance, glycolysis related
3528 metabolites, and ketone bodies. A majority of tests did not reach the multiple testing threshold
3529 in a majority of subclasses. Only 12 subclasses (amino acids, apolipoproteins, aromatic

3530 amino acids, branched chain amino acids, glycerides and phospholipids ratios, inflammation,
3531 large HDL, large HDL ratios, medium HDL, medium HDL ratios, very large HDL, and very
3532 large HDL ratios) had a majority of tests reaching the multiple testing threshold.

3533 **Sensitivity analysis**

3534 Assumptions of no pleiotropy were explored using MR-Egger⁵³⁵, weighted median⁵³⁶
3535 and weighted mode⁵³⁷ based estimators. Globally, sensitivity analysis was similar to that
3536 of the main analysis for each exposure, though with wider confidence intervals (Appendix
3537 ??). Confidence intervals for sensitivity analyses tended to cross the null and were widest for
3538 MR-Egger, which is unsurprising given the lower power afforded with this model. Sensitivity
3539 results for WHR appeared to show most consistency with the main analysis for analyses
3540 using both Kettunen and INTERVAL data; confidence intervals for weighted median and
3541 mode models did not cross the null in a majority of results for subclasses. When looking at
3542 concordance in effect direction across all models (sensitivity and main analysis), consistency
3543 in the direction of effect was highest for WHR across both datasets (Figure 5.5. For both
3544 datasets, effects for BF were the least consistent.

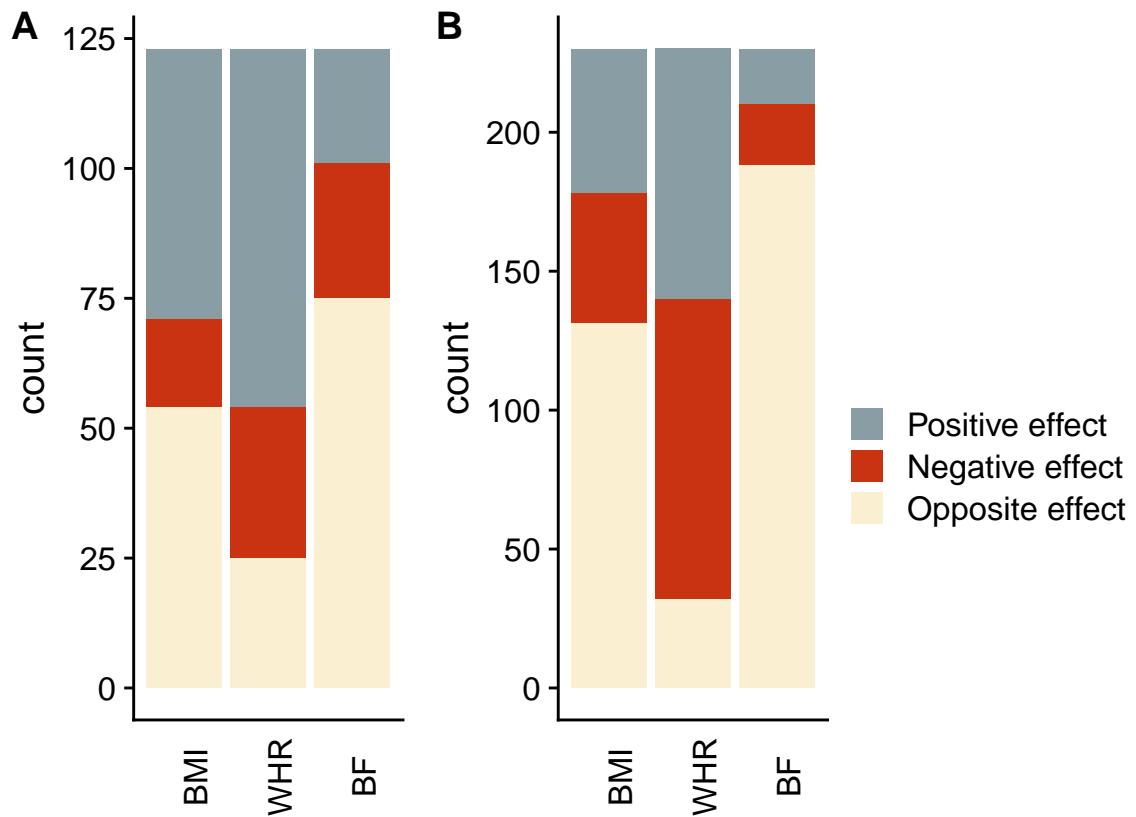


Figure 5.5: **Directional consistency of main and sensitivity analyses from two-sample Mendelian randomization results using Kettunen and INTERVAL data.** Plot shows the directional consistency across all 4 models (IVW-MRE, MR-Egger, Weighted Mode, and Weighted median) within each exposure. A positive effect reflects the effect estimate from all four models being in the positive direction; a negative effect reflects betas being in a negative direction; opposite effect reflects different directions for the effect estimates. **A:** Two-sample MR IVW-MRE for 123 metabolites from Kettunen et al. (2016)³¹⁸; **B:** Two-sample MR IVW-MRE for 230 metabolites from INTERVAL. BMI = body mass index; WHR = waist hip ratio; BF = body fat percentage.

3545 Of these directionally consistent tests, a total of 29 tests were directionally consistent
 3546 across methods for all three exposures when using the Kettunen data (Figure 5.6). The
 3547 direction of effect for BF was on the whole opposite to BMI and WHR for all methods. Of
 3548 these 29 tests, only valine was also found to reach the multiple testing threshold (0.05/22) for
 3549 both BMI and WHR in the main analysis. Sensitivity analysis showed a consistent positive
 3550 direction of effect with the main analysis for the effect of BMI, WHR, and BF on valine. The
 3551 effect of WHR appeared to show the strongest evidence for an association with valine, with
 3552 confidence intervals for all models away from the null. When using the INTERVAL data, of the
 3553 directionally consistent results, a total of 9 tests were directionally consistent across methods
 3554 for all three exposures (Figure 5.7). Of these 9 tests, valine and Tyrosine were also found

3555 to reach the multiple testing threshold (0.05/28) for both BMI and WHR in the main analysis
3556 where models were consistent. Sensitivity analysis showed a consistent positive direction of
3557 effect with the main analysis for the effect of BMI, WHR, and BF on valine and tyrosine. The
3558 effect of WHR appeared to show the strongest evidence for an association with valine, with
3559 confidence intervals for all models away from the null.

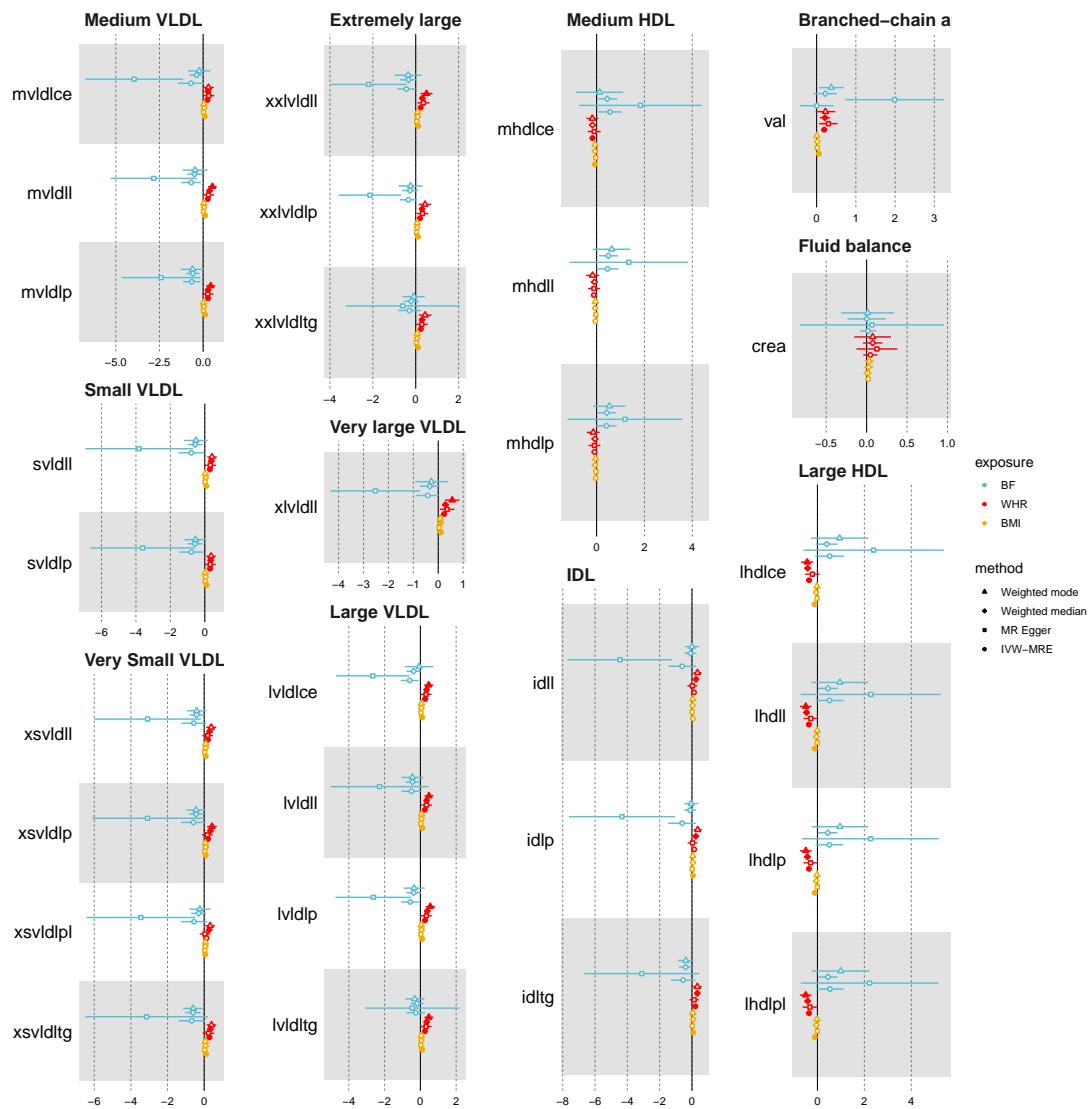


Figure 5.6: Directionally consistent effects across all two-sample Mendelian randomization models using the Kettunen data. Effect estimates and 95% confidence intervals for metabolites which showed directionally consistent results across the main (IVW-MRE = inverse variance weighted multiplicative random effects) and sensitivity analysis (MR-Egger, weighted median, and weighted mode) within each exposure. Solid points indicate a multiple testing threshold (0.0023) has been reached. BMI = body mass index; WHR = waist hip ratio; BF = body fat percentage. Names for metabolites are available in Appendix Table A.5.

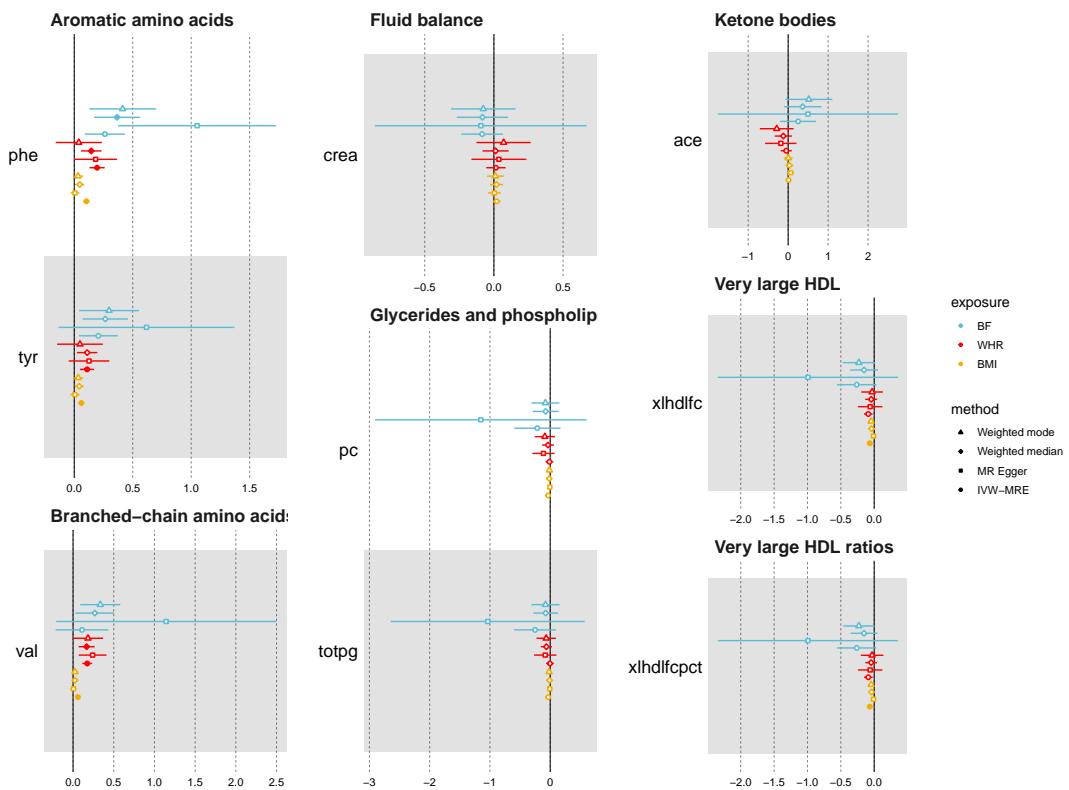


Figure 5.7: Directionally consistent effects across all two-sample Mendelian randomization models using the INTERVAL data. Effect estimates and 95% confidence intervals for metabolites which showed directionally consistent results across the main (IVW-MRE = inverse variance weighted multiplicative random effects) and sensitivity analysis (MR-Egger, weighted median, and weighted mode) within each exposure. Solid points indicate a multiple testing threshold (0.0018) has been reached. BMI = body mass index; WHR = waist hip ratio; BF = body fat percentage. Names for metabolites are available in Appendix Table A.5

3560 The causal direction between the exposure and outcomes were assessed using the
 3561 Steiger test. In the Kettunen data, a total of 123 tests were performed for each exposure,
 3562 the causal direction of effect from the exposure to the outcome was “true” for 0 tests for BMI
 3563 (i.e., the Steiger test results reflect that a change in the outcome is a consequence of the
 3564 exposure), and 5 tests for WHR. In contrast, the causal direction of effect from the exposure to
 3565 the outcome was “true” for 141 tests for BF. For INTERVAL data, a majority of test directions
 3566 were found to be “true.” A total of 230 tests were performed for each exposure, the causal
 3567 direction of effect from the exposure to the outcome was “true” for 200 tests for BMI, 110 tests
 3568 for WHR, and 141 tests for BF.

3569 **Additional sensitivity analysis**

3570 The effects of heterogeneity in each exposures instruments were investigated using
3571 Cochran's Q statistic for IVW-MRE and MR-Egger models. Heterogeneity of the genetic
3572 instruments for each exposure was greater than the degrees of freedom for a majority of
3573 metabolites for all three exposures across both the Kettunen and INTERVAL datasets (Table
3574 5.1). When using the Kettunen data, for BMI and WHR all IVW and MR-Egger tests with
3575 Q greater than the degrees of freedom overlapped. That is, for each exposure-metabolite
3576 pair, Q was greater than the degrees of freedom for the same exposure-metabolite pair in the
3577 IVW and MR-Egger tests. For BF, there were 5 tests which did not overlap. When using the
3578 INTERVAL data, for BMI 1 test did not overlap, for BF 7 tests did not overlap, and for WHR
3579 all tests overlapped. The remaining sensitivity analyses are presented individually for each
3580 outcome dataset.

Table 5.1: Number of tests in which Cochran's Q exceeds genetic instrumental variable degrees of freedom

		IVW	MR-Egger
BMI		120	120
WHR	Kettunen	121	121
BF		111	112
BMI		229	230
WHR	INTERVAL	229	229
BF		213	220

Table gives the number of tests for each exposure in which heterogeneity, measured by Cochran's Q, was greater than the degrees of freedom (number of SNPs - 1) for each exposure. That is, if BMI has 941 SNPs then the degrees of freedom is 940, and if Cochran's Q for the effect of BMI on metabolite 1 is 1000 then there is evidence of heterogeneity in the genetic instruments in relation to that test. Total number of exposure-outcome tests for Kettunen and INTERVAL data was 123 and 230 respectively. IVW = Inverse variance weighted method; BMI = body mass index; WHR = waist hip ratio; BF = body fat percentage

3581 **Analysis using Kettunen data:** In single SNP MR using Kettunen data, visual inspection
3582 of forest plots showed *S* shaped distributions of effect estimates for all tests (Schematic
3583 representative Figure 5.8). Effect estimates for some SNPs in the single SNP MR analysis
3584 appeared to be outliers. For example, for BMI on glycoproteins, rs4673553 showed a
3585 disproportionately larger effect estimate of 22 SD units increase per normalised SD higher
3586 BMI (Figure 5.9; standard error = 0.85; p-value = 5.66×10^{-148}) when compared to other
3587 SNPs.

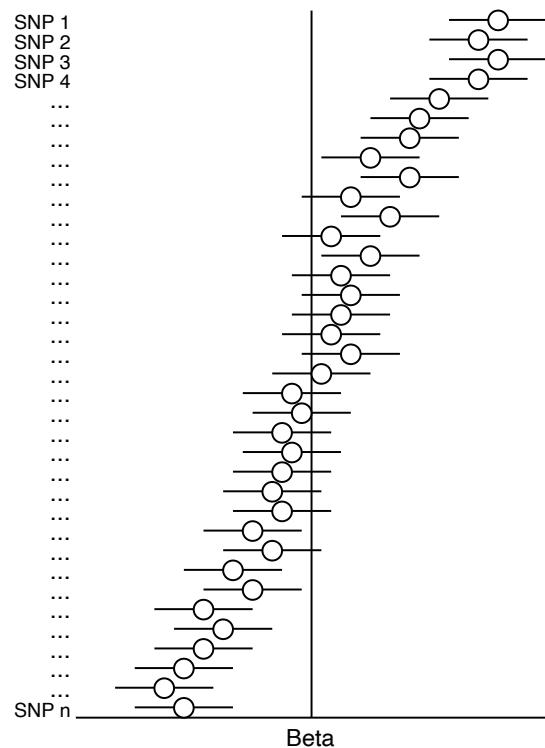


Figure 5.8: **Schematic representative figure: Single-SNP MR analysis of the effect of adiposity on a metabolite.** In single-SNP MR analysis, the effect of each SNP on the outcome is investigated. In these analyses, SNP effects are organised from largest to smallest. Here, the figure shows an *S* shaped distribution of effects which is representative of the single-SNP MR analyses performed for Kettunen and INTERVAL data.

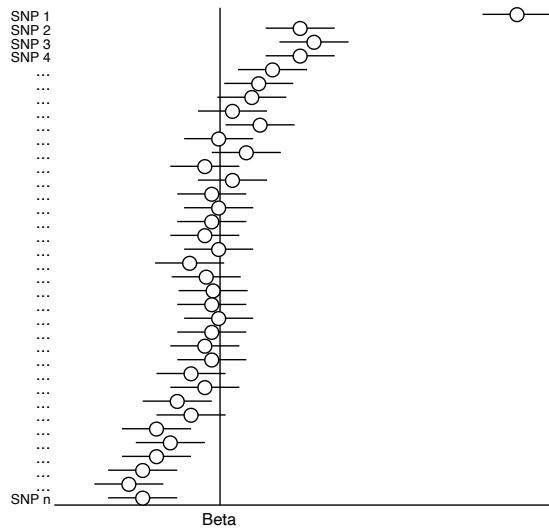


Figure 5.9: Schematic representative figure: Single-SNP MR analysis of the effect of adiposity on a metabolite. In single-SNP MR analysis, the effect of each SNP on the outcome is investigated. In these analyses, SNP effects are organised from largest to smallest. Here, the figure shows an *S* shaped distribution of effects which is representative of the single-SNP MR analyses performed for Kettunen and INTERVAL data. In addition, a single SNP is shown to have a much larger effect than the others.

3588 When looking at the median effect size across all metabolites for each SNP, a number of
 3589 SNPs showed larger median effect sizes across a majority of metabolites. The total number
 3590 of SNPs with the largest positive (effect sizes in the top 95%) and largest negative (effect
 3591 sizes in the bottom 5%) effect sizes were: BMI = 46 (5%) and 46 (95%); WHR = 46 (5%)
 3592 and 46 (95%); BF = 1 (5%) and 1 (95%). Funnel plots did not however highlight outlying
 3593 SNPs. Instead, funnel plots were reflective of some SNPs having larger effect estimates more
 3594 broadly (Representative Figure 5.10). The low number of SNPs used for BF did not result in
 3595 meaningfully interpretable funnel plots (Representative Figure 5.11).

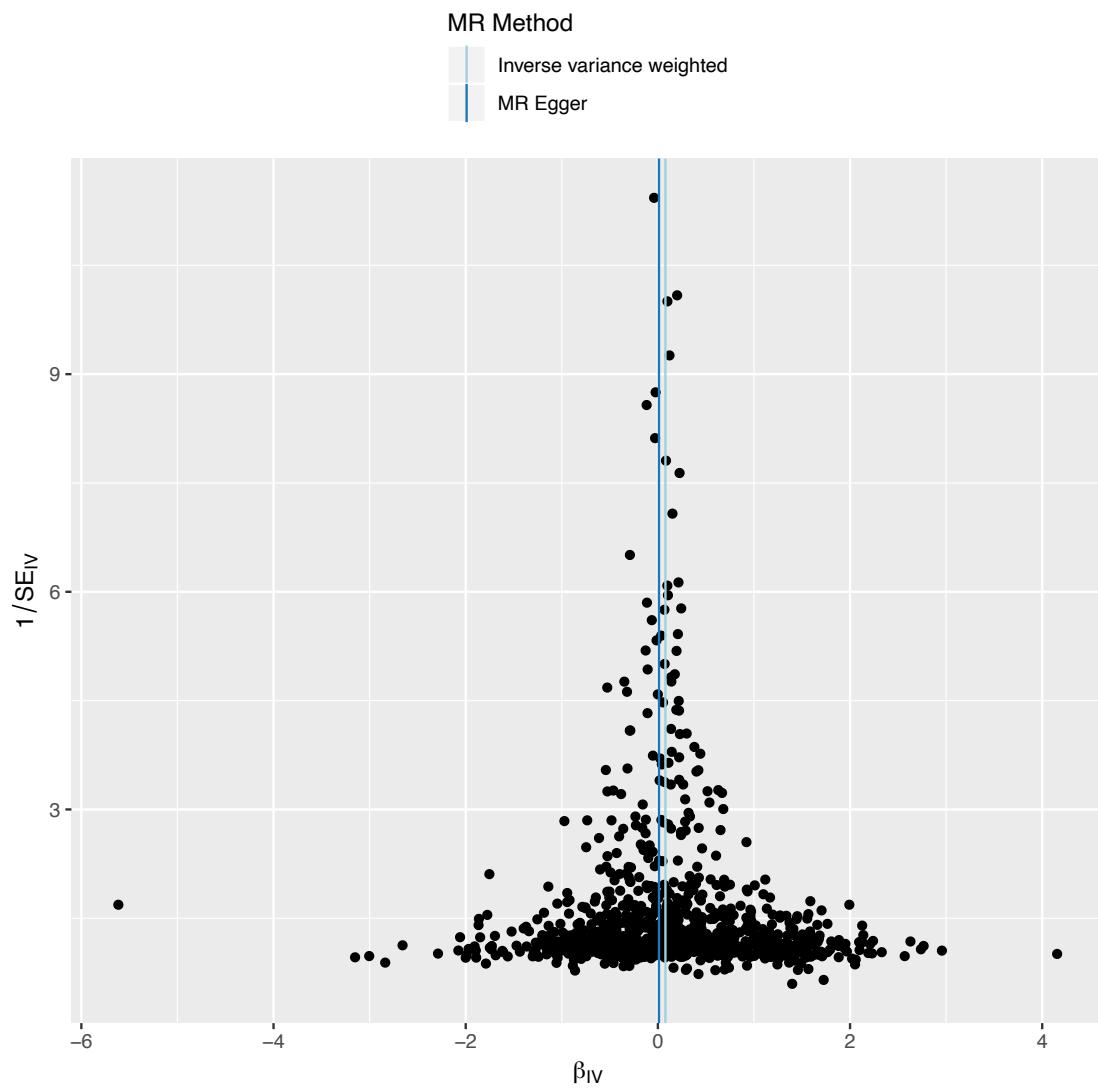


Figure 5.10: Representative figure: Funnel plot of BMI on one metabolite using Kettunen data. Funnel plot shows the results of a single SNP MR with effect estimate and standard error. Asymmetry in the funnel indicates unreliable effect estimates. The representative plot illustrates a SNP (bottom left) with a larger effect estimate.

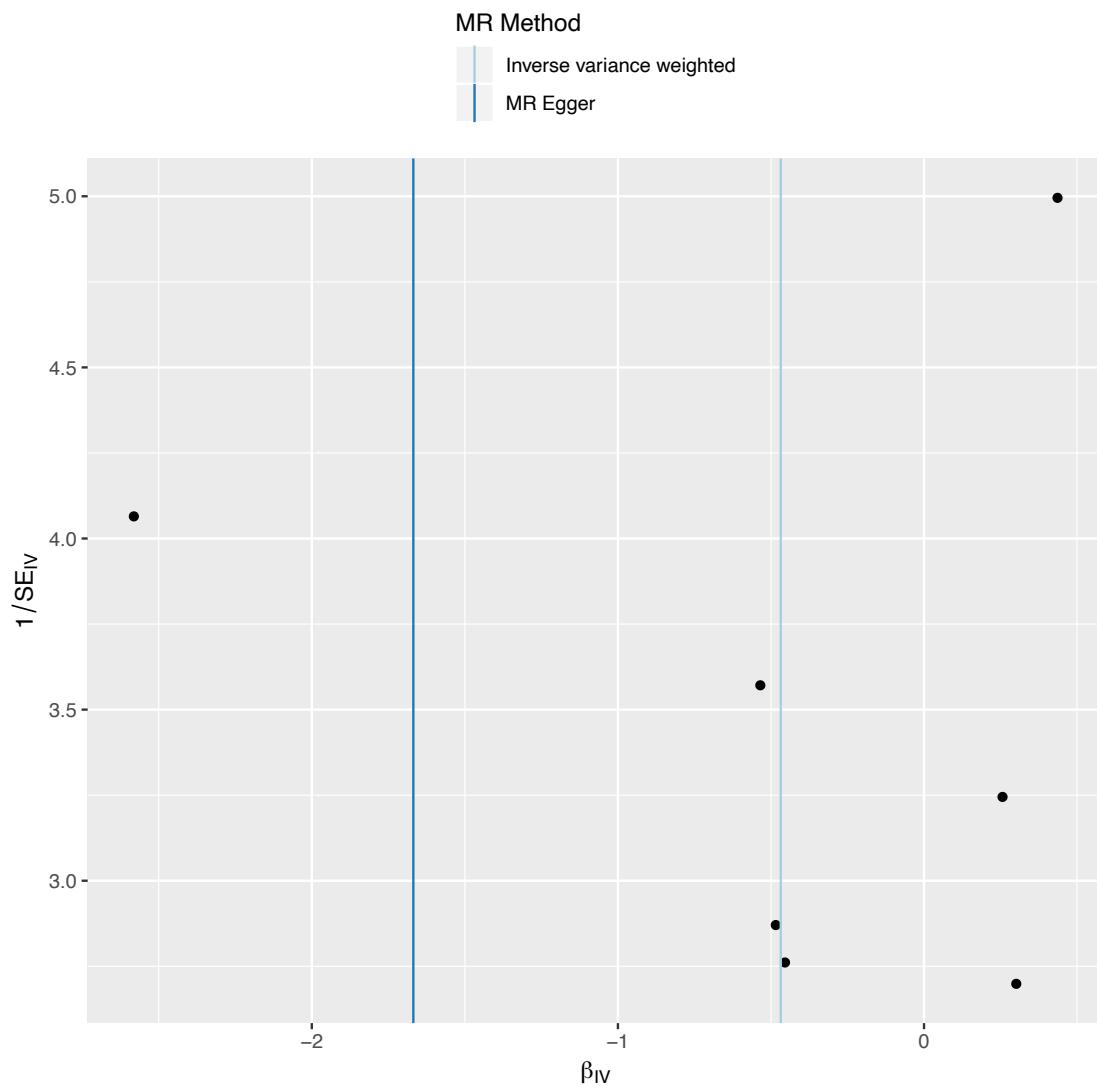


Figure 5.11: **Representative figure: Funnel plot of BF on one metabolite using Kettunen data.** Funnel plot shows the results of a single SNP MR with effect estimate and standard error. Asymmetry in the funnel indicates unreliable effect estimates.

3596 Although a number of SNPs showed disproportionately larger effect sizes, in leave-one-
 3597 out analysis, visual inspection of forest plots showed that no single SNP altered the direction
 3598 of effect for any metabolite across exposures. For BF, confidence intervals for one or more
 3599 SNPs crossed the null for every metabolite tested (Representative Figure 5.12). This was not
 3600 the case for BMI and WHR, where for many metabolites confidence intervals did not cross
 3601 the null for any SNPs (Representative Figure A.22).

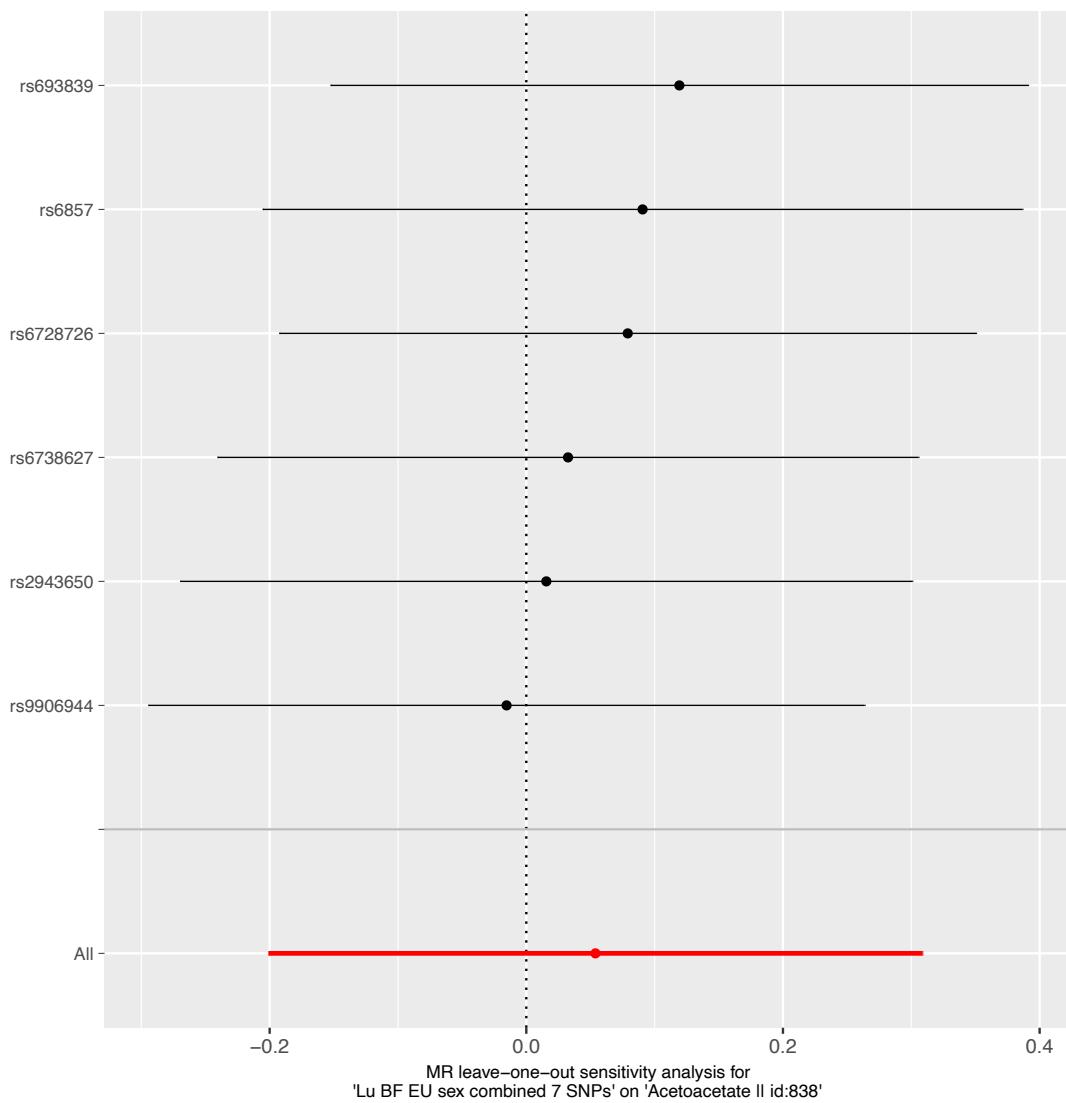


Figure 5.12: Representative figure: Leave-one-out MR analysis of BF on acetoacetate using Kettunen data. A leave-one-out analysis performs an MR of exposure on outcome for all SNPs excluding a different SNP each time. Forest plot shows the effect estimate and 95% confidence interval for each SNP exclusion on acetoacetate. Effect estimates represent the change in the inverse rank position of each metabolite per change in the inverse rank position of the exposure.

3602 **Analysis using INTERVAL data:** Broadly speaking, additional sensitivity analyses using
 3603 INTERVAL data were similar to that of the additional sensitivity analyses using the Kettunen
 3604 data. In single SNP MR visual inspection of forest plots showed *S* shaped distributions of
 3605 effect estimates for all tests (Representative Figure A.23, figure also shows outlier SNP with

3606 effect estimate close to -6). As with the Kettunen data, effect estimates for some SNPs in the
3607 single SNP MR analysis were much greater than others (See Representative Figure A.23).

3608 When looking at the median effect size across all metabolites for each SNP, a number of
3609 SNPs showed larger median effect sizes across a majority of metabolites. The total number
3610 of SNPs with the largest positive (effect sizes in the top 95%) and largest negative (effect
3611 sizes in the bottom 5%) effect sizes were: BMI = 46 (5%) and 46 (95%); WHR = 46 (5%)
3612 and 46 (95%); BF = 1 (5%) and 1 (95%). Many of the SNPs identified as having larger
3613 effects are shared across both the Kettunen and INTERVAL datasets: BMI = 23 (5%) and 9
3614 (95%); WHR = 23 (5%) and 9 (95%); BF = 0 (5%) and 1 (95%). As an example, for BF, more
3615 often than not, rs6857 showed an effect estimate greater than the 6 other SNPs and did not
3616 overlap confidence intervals with them or the null. rs6857 was found in both the Kettunen and
3617 INTERVAL data to have a disproportionately larger effect estimate than other SNPs. For BMI
3618 and WHR, SNPs with disproportionately larger effect sizes tended to have confidence intervals
3619 which overlapped other SNPs. The degree of overlap was minimal however and mostly at the
3620 tail-end of the confidence interval.

3621 Funnel plots did not highlight outlying SNPs, but did reflect some SNPs having larger
3622 effect estimates across the board (Representative Figure A.24). The low number of SNPs
3623 used for BF did not result in meaningfully interpretable funnel plots (Representative Figure ??).
3624 In leave-one-out analysis, visual inspection of forest plots showed that no single SNP altered
3625 the direction of effect for any metabolite across exposures. For BF, confidence intervals for
3626 one or more SNPs crossed the null for a majority of metabolite tested (Representative Figure
3627 A.25). This was not the case for BMI and WHR, where for many metabolites confidence
3628 intervals did not cross the null for any SNPs (Representative Figure A.26).

3629 Additional analyses

3630 **Additional exposures** A number of additional SNP lists were used to instrument BMI,
3631 WHR, and BF to explore the validity of the instruments used in the main analyses. Additional
3632 SNP lists for BMI and WHR were selected on the basis that they did not contain UK Biobank
3633 individuals, which were included in SNP lists for the main analysis. For BF, an additional SNP
3634 list which did contain UK Biobank individuals was chosen. For BMI and WHR, additional SNP
3635 lists contained fewer SNPs than the SNP list used in the main analysis. For BF, the additional
3636 SNP list contained more SNPs than the main analysis. Each SNP list explained a different
3637 amount of variation in its respective trait in comparison the the SNP list used in the main
3638 analysis. All details on the additional SNP lists are presented in the Appendix (A.4.1).

3639 All analyses, including sensitivity analyses, were repeated for these additional SNP lists.
3640 Focus here is on the Kettunen data as results were similar for the INTERVAL data. A table of
3641 all SNP lists can be found on [GitHub](#).

3642 Results from additional instruments for BMI showed broadly larger effect estimates but
3643 consistent directions of effect across metabolites (Appendix Figure ??). For the BMI SNPs
3644 obtained from a non-UK Biobank GWAS, effect estimates had much wider confidence intervals.
3645 Spearman's Rho correlation of MR results was highest between the two SNP lists (N SNP
3646 = 941 and 656) from Yengo et al. (2018) (0.98). Correlation between the Locke et al (2014)
3647 SNP list (N SNP = 77) and the COJO SNP list from Yengo et al (N SNP = 941) (0.9) and
3648 the non-COJO SNP list from Yengo et al. (N SNP = 656) (0.93) were also high. For WHR
3649 the pattern of association was similar between both the main analysis (N SNP = 316) and
3650 analysis using the additional SNP list from Shungin et al. (2014; N SNP = 26) (Figure ??) with
3651 high correlation between MR results (0.9). Effect estimates were larger, however confidence
3652 intervals were wider and crossed the null more often when using the additional SNP list from
3653 Shungin et al compare to the main analysis.

3654 For BF, there was considerable similarity between the main analysis and the additional
3655 analysis when using SNPs from Lu et al (2016) which did not include two SNPs previously
3656 identified as being associated with 'favourable adiposity' (Figure ??). More tests reached
3657 the multiple testing threshold when using the 5 SNPs from Lu et al., as opposed to the
3658 full 7 SNPs, this included associations with Apolipoprotein A1, Phenylalanine, Tyrosine,
3659 Glucose, and Cholesterol esters in very large HDL. For the additional analysis, which used
3660 76 SNPs from Hubel et al (2016), MR results were considerably smaller and appeared to
3661 show conflicting directions of effect with that of the Lu et al. (2016) SNPs (both using 7 and
3662 5 SNPs). Confidence intervals were much tighter and two metabolites (phenylalanine and
3663 glycoprotein acetyls) reached the multiple testing threshold. Correlation between the two Lu
3664 et al (2016) SNP lists was high (0.93), however both the 5 (-0.64) and 7 (-0.52) SNP lists from
3665 Lu et al. (2016) showed weaker inverse correlations with the SNP list from Hubel et al (2016).

3666 Steiger directionality tests were variable across analyses (Kettunen vs. INTERVAL) and
3667 instrument lists. Variability between instrument lists is perhaps unsurprising given that a
3668 "true" causal direction from the exposure to the outcome is identified when the SNP variance
3669 explained is greater in the exposure than in the outcome. When using large instrument
3670 lists, as the number of SNPs increases the proportion of variance explained by each SNP
3671 will decrease. A result of this will be that many SNPs will explain a very low amount of the
3672 variance in the exposure, this will increase the likelihood of that SNP explaining a larger
3673 proportion of variance in the outcome phenotype thus resulting in the "true" causal direction
3674 being from the outcome to the exposure.

3675 In the main analysis using BMI with 941 SNPs and the Kettunen data, 0 tests were
3676 shown to reflect the "true" causal direction of exposure to outcome. This did not change
3677 after clumping. However, when using the non-COJO SNPs (N SNP = 656) from the same
3678 study, 4 tests were found to reflect the "true" causal direction. This increased to 5 tests after
3679 clumping. Additionally, when using 77 SNPs from Locke et al. (2014)³⁶ to instrument BMI a
3680 total of 76 tests were shown to reflect the "true" causal direction. This increased to 79 tests
3681 after clumping. There was no difference in the non-clumped (N SNP = 316) and clumped
3682 SNP list for WHR used in the main analysis (N test that reflect the "true" causal direction = 4).

3683 However, when using the 26 SNP instrument from Shungin et al. (2015)³⁷, 102 and 112 tests
3684 were found to reflect the “true” causal direction for the non-clumped and clumped SNP lists.
3685 For BF, the 7 SNP instrument resulted in 80 true tests while the 5 SNP instrument resulted in
3686 76 true tests. The BF GWAS from Hubel et al. (2019)⁴³ did not have an N available for the
3687 summary statistics, instead the total N for the GWAS (155,961) was used. All 123 tests were
3688 found to reflect the “true” causal direction when using both the non-clumped (N SNP = 76)
3689 and clumped instruments from Hubel et al. (2016) for BF.

3690 When using the INTERVAL data, a different picture was found, with a majority of tests
3691 shown to reflect the “true” causal direction across all instrument lists. The exception was when
3692 instrumenting WHR with SNPs from Pilit et al. (2019)⁴². After clumping, the number of tests
3693 that reflected the “true” causal direction for WHR reduced from 91 to 75. There was also little
3694 difference in the number tests that reflected the “true” causal direction when instrumenting
3695 BMI using 77 SNPs from Locke et al. compared to the 941 SNPs used in the main analysis.
3696 As with the Kettunen data, all tests reflected a “true” causal direction when instrumenting BF
3697 using the 76 SNPs from Hubel et al.

3698 **Clumped exposures** All SNP lists used as instruments, including those described above,
3699 underwent clumping and all analyses were repeated. Clumping resulted in the removal of the
3700 following number of SNPs due to LD ($R^2 \geq 0.001$) with other variants or absence from the LD
3701 reference panel: BMI Locke et al. (2014) = 14, BMI Yengo et al. (2018) using COJO SNPs =
3702 583, BMI Yengo et al. (2018) using non-COJO SNPs = 336, WHR Pilit et al. (2018) = 234,
3703 WHR Shungin et al. (2014) = 17, BF Hubel et al. (2018) = 4. No SNPs were removed due to
3704 clumping for BF from Lu et al. (2016). All SNPs, including whether they were removed due to
3705 clumping, are presented in the Appendix (Table ??).

3706 When using the Kettunen data, correlation for BMI results between the Yengo COJO
3707 (0.97), non-COJO (0.97), and Locke (0.98) non-clumped and clumped MR results was high.
3708 Similarly, for WHR MR results from non-clumped and clumped analyses correlation was
3709 high for the main exposure (Pilit et al. (2018) = 0.97) and for the additional exposure (0.98).
3710 For BF, clumping was not possible for the main exposure, however correlation between the
3711 non-clumped and clumped SNP list from Hubel et al. (2018) was high (0.98)

3712 When using the INTERVAL data, correlation for BMI results between the Yengo COJO
3713 (0.96), non-COJO (0.91), and Locke (0.93) non-clumped and clumped MR results was high.
3714 Similarly, for WHR MR results from non-clumped and clumped analyses correlation was
3715 high for the main exposure (Pilit et al. (2018) = 0.99) and for the additional exposure (0.99).
3716 For BF, clumping was not possible for the main exposure, however correlation between the
3717 non-clumped and clumped SNP list from Hubel et al. (2018) was high (0.996).

3718 5.3.2 Meta-analysis of two-sample Mendelian randomization results

3719 In total, 110 metabolites were shared across the Kettunen and INTERVAL metabolite
 3720 data. Meta-analysis of p-value was performed using Fisher's method (Equation (5.1)). Across
 3721 all 3 exposures (330 tests), a total of 120 tests had a positive direction of effect when using
 3722 the Kettunen and INTERVAL data, 91 tests had a negative direction of effect. For BMI, 48
 3723 metabolites had positive directions of effect and 20 metabolites had negative directions of
 3724 effect. Similar results were found for WHR (positive = 50; negative = 18). For BF, a larger
 3725 number of metabolites had consistent negative directions of effect (N = 53) than positive (N =
 3726 22).

3727 Across the 330 tests, a total of 141 tests reached a Bonferroni (0.05/110) multiple testing
 3728 threshold. Of these, 63 tests had positive directions of effect and 31 tests had negative
 3729 directions of effect when using the Kettunen and INTERVAL data. For BMI, 30 metabolites
 3730 had consistent positive directions of effect and met the multiple testing threshold, while 16
 3731 metabolites had consistent negative directions of effect and met the multiple testing threshold.
 3732 Similar results were found for WHR (positive = 33; negative = 15). For BF, no metabolites
 3733 reached the multiple testing threshold.

3734 Across both BMI and WHR, a total of 33 metabolites had a consistent positive direction of
 3735 effect and met the multiple testing threshold, while 16 metabolites had a consistent negative
 3736 direction of effect and reached the multiple testing threshold. As such, a total of 49 metabolites
 3737 were associated with BMI or WHR in meta-analysis. Table 5.2 gives a list of all 49 metabolites
 3738 associated with BMI and WHR along with their directions of effect, metabolites are ordered
 3739 alphabetically by class, subclass, and then metabolite label, and are presented with their
 3740 originally measured units.

Table 5.2: Metabolites with a consistent direction of effect between BMI and WHR which reached a multiple testing threshold (p-value < 0.00045) in meta-analysis

Metabolite	Class	Subclass	
Glutamine (mmol/l)	Amino acids	Amino acids	-
Phenylalanine (mmol/l)	Amino acids	Aromatic amino acids	+
Tyrosine (mmol/l)	Amino acids	Aromatic amino acids	+
Isoleucine (mmol/l)	Amino acids	Branched-chain amino acids	+
Leucine (mmol/l)	Amino acids	Branched-chain amino acids	+
Valine (mmol/l)	Amino acids	Branched-chain amino acids	+
Apolipoprotein A-I (g/l)	Apolipoproteins	Apolipoproteins	-
Apolipoprotein B (g/l)	Apolipoproteins	Apolipoproteins	+
Total cholesterol in HDL (mmol/l)	Cholesterol	Cholesterol	-
Monounsaturated fatty acids; 16:1, 18:1 (mmol/l)	Fatty acids	Fatty acids	+
Total fatty acids (mmol/l)	Fatty acids	Fatty acids	+
Serum total triglycerides (mmol/l)	Glycerides and phospholipids	Glycerides and phospholipids	+
Lactate (mmol/l)	Glycolysis related metabolites	Glycolysis related metabolites	+
Mean diameter for HDL particles (nm)	Lipoprotein particle size	Lipoprotein particle size	-
Mean diameter for VLDL particles (nm)	Lipoprotein particle size	Lipoprotein particle size	+

Table 5.2: Metabolites with a consistent direction of effect between BMI and WHR which reached a multiple testing threshold (p -value < 0.00045) in meta-analysis (continued)

Metabolite	Class	Subclass	
Concentration of chylomicrons and extremely large VLDL particles (mol/l)	Lipoprotein subclasses	Extremely large VLDL	+
Phospholipids in chylomicrons and extremely large VLDL (mmol/l)	Lipoprotein subclasses	Extremely large VLDL	+
Total lipids in chylomicrons and extremely large VLDL (mmol/l)	Lipoprotein subclasses	Extremely large VLDL	+
Triglycerides in IDL (mmol/l)	Lipoprotein subclasses	IDL	+
Cholesterol esters in large HDL (mmol/l)	Lipoprotein subclasses	Large HDL	-
Free cholesterol in large HDL (mmol/l)	Lipoprotein subclasses	Large HDL	-
Total cholesterol in large HDL (mmol/l)	Lipoprotein subclasses	Large HDL	-
Total lipids in large HDL (mmol/l)	Lipoprotein subclasses	Large HDL	-
Concentration of large VLDL particles (mol/l)	Lipoprotein subclasses	Large VLDL	+
Free cholesterol in large VLDL (mmol/l)	Lipoprotein subclasses	Large VLDL	+
Phospholipids in large VLDL (mmol/l)	Lipoprotein subclasses	Large VLDL	+
Total lipids in large VLDL (mmol/l)	Lipoprotein subclasses	Large VLDL	+
Triglycerides in large VLDL (mmol/l)	Lipoprotein subclasses	Large VLDL	+
Cholesterol esters in medium HDL (mmol/l)	Lipoprotein subclasses	Medium HDL	-
Free cholesterol in medium HDL (mmol/l)	Lipoprotein subclasses	Medium HDL	-
Total cholesterol in medium HDL (mmol/l)	Lipoprotein subclasses	Medium HDL	-
Total lipids in medium HDL (mmol/l)	Lipoprotein subclasses	Medium HDL	-
Free cholesterol in medium VLDL (mmol/l)	Lipoprotein subclasses	Medium VLDL	+
Total lipids in medium VLDL (mmol/l)	Lipoprotein subclasses	Medium VLDL	+
Triglycerides in medium VLDL (mmol/l)	Lipoprotein subclasses	Medium VLDL	+
Total lipids in small HDL (mmol/l)	Lipoprotein subclasses	Small HDL	+
Triglycerides in small HDL (mmol/l)	Lipoprotein subclasses	Small HDL	+
Total lipids in small VLDL (mmol/l)	Lipoprotein subclasses	Small VLDL	+
Triglycerides in small VLDL (mmol/l)	Lipoprotein subclasses	Small VLDL	+
Concentration of very large HDL particles (mol/l)	Lipoprotein subclasses	Very large HDL	-
Free cholesterol in very large HDL (mmol/l)	Lipoprotein subclasses	Very large HDL	-
Phospholipids in very large HDL (mmol/l)	Lipoprotein subclasses	Very large HDL	-
Total lipids in very large HDL (mmol/l)	Lipoprotein subclasses	Very large HDL	-
Concentration of very large VLDL particles (mol/l)	Lipoprotein subclasses	Very large VLDL	+
Phospholipids in very large VLDL (mmol/l)	Lipoprotein subclasses	Very large VLDL	+
Total lipids in very large VLDL (mmol/l)	Lipoprotein subclasses	Very large VLDL	+
Triglycerides in very large VLDL (mmol/l)	Lipoprotein subclasses	Very large VLDL	+
Total lipids in very small VLDL (mmol/l)	Lipoprotein subclasses	Very Small VLDL	+
Triglycerides in very small VLDL (mmol/l)	Lipoprotein subclasses	Very Small VLDL	+

3741 **5.3.3 Comparison of two-sample Mendelian randomization and observational**
 3742 **results from Chapter 2**

3743 All 104 metabolites identified in the meta-analysis as associated with adiposity were
 3744 analysed in the observational analysis in Chapter 4. In the observational analysis, a multiple
 3745 testing threshold of $0.05/53 (9 \times 10^{-4})$ was used. For the 46 metabolites associated with BMI
 3746 in the meta-analysis, all showed consistent directions of effect in the observational analysis.
 3747 Only one metabolite (Total lipids in medium HDL, p -value = 0.1) did not reach the multiple

3748 testing threshold in the observational analysis for BMI. All 48 metabolites associated with
3749 WHR in the meta-analysis had consistent directions of effect in the observational analysis
3750 and met the multiple testing threshold in the observational analyses (Figure 5.13).

3751 As no associations reached the multiple testing threshold for BF in the meta-analysis, all
3752 metabolites with consistent directions of effect in the meta-analysis for BF were compared
3753 for consistent directions of effect with the observational analysis. A total of 74 metabolites
3754 directionally consistent in the MR analyses for BF were available in the observational analysis.
3755 Of these, 4 had consistent negative directions across MR and observational analyses and 7
3756 had consistent positive directions. All 4 metabolites with a negative direction of effect reached
3757 the multiple testing threshold, while 5 of the 7 metabolites with positive directions of effect
3758 reached the multiple testing threshold.

3759 Across the 45 BMI, 48 WHR, and 9 BF associated metabolites, a total of 55 metabolites
3760 were associated with at least one measure of adiposity. For BMI and WHR, associations
3761 were identified through directional consistency and a multiple testing threshold in the MR
3762 meta-analysis followed by whether these metabolites had a consistent direction of effect
3763 and reached a multiple testing threshold in the observational analysis. For BF, associated
3764 metabolites were identified by consistent directions of effect across the MR meta-analysis and
3765 observational analysis, and which reached a multiple testing threshold in the observational
3766 analysis. Figure 5.13 shows all 55 adiposity associated metabolites and their directions
3767 of effect in each analysis. Associations were identified as: having consistent directions of
3768 effect within exposures across MR and observational analyses and which reached a multiple
3769 testing threshold (BMI and WHR) in the meta-analysis (p -value $\leq 5 \times 10^{-4}$) and observational
3770 analysis (p -value $\leq 9 \times 10^{-4}$; BMI and WHR) and just in the observational analysis (BF).
3771 Yellow tiles indicate a negative direction of effect; blue tiles indicate a positive direction of
3772 effect.

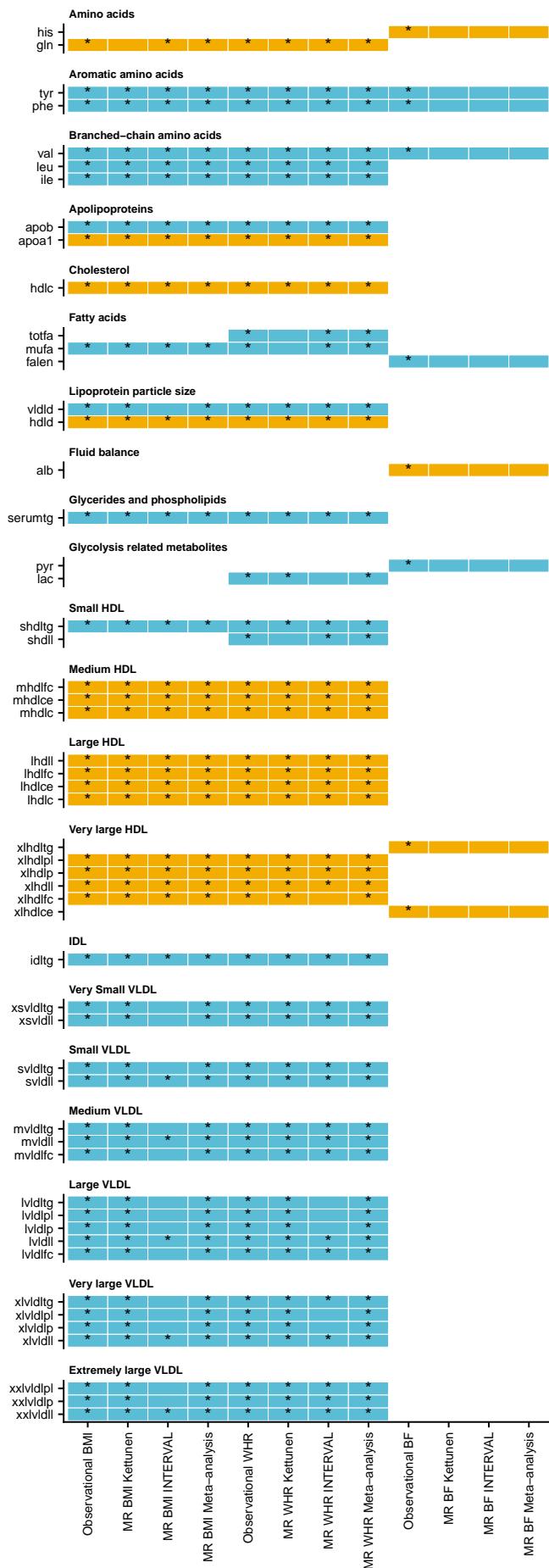


Figure 5.13: Metabolites associated with measures of adiposity in observational, MR, and MR meta-analysis.

3773 Tile plot shows all metabolites with consistent directions of effect within exposures across MR
3774 and observational analyses and which reached a multiple testing threshold (BMI and WHR)
3775 in the meta-analysis and observational analysis (BMI and WHR) and just in the observational
3776 analysis (BF). Yellow tiles indicate a negative direction of effect; blue tiles indicate a positive
3777 direction of effect. * = multiple testing threshold reached: observational = 9×10^{-4} ; Kettunen
3778 = 0.0023; INTERVAL = 0.0018; meta-analysis = 5×10^{-4} .

3779 5.4 Discussion

3780 In this chapter, the influence of adiposity on the metabolic profile is demonstrated in an
3781 MR framework. The use of MR allowed the interrogation of causality of various measures of
3782 adiposity on the metabolic profile, while accounting for limitations in observational analyses
3783 (discussed in Chapter 1 and 4). Data on adiposity measures were available for: BMI from
3784 up to 795,624 individuals of European ancestries from GIANT[Yengo2018], WHR from up to
3785 697,702 individuals of European ancestries from GIANT⁴², BF from up to 89,297 individuals
3786 European ancestries from Lu et al. (2016)³⁹. Two parallel MR analyses of 123 NMR derived
3787 metabolites measured in up-to 24,925 individuals of European ancestries from Kettunen
3788 et al. (2016)³¹⁸ and 230 NMR derived metabolites measured in up-to 40,905 individuals of
3789 European ancestries from INTERVAL (unpublished) were conducted. Meta-analysis of 110
3790 metabolites measured in both the Kettunen and INTERVAL datasets and comparison with
3791 observational analyses from Chapter 4 identified 56 associations between adiposity and
3792 metabolite measures that were consistent in direction of effect across MR and observational
3793 analyses and passed multiple testing thresholds for both analyses. Positive associations were
3794 found for metabolites in VLDL (small, very small, medium, large, and very large), as well as
3795 aromatic and branched chain amino acid subclasses. While negative associations were found
3796 for HDL (medium, large and very large) subclasses.

3797 In MR analyses, there was evidence for a broad effect of BMI and WHR on the metabolic
3798 profile. Both adiposity measures showed positive and negative effects on whole subclasses
3799 of metabolites, such as small and very small VLDL, branched chain amino acids (positive),
3800 and large and very large HDL (negative). However, there were many effect estimates across
3801 the Kettunen and INTERVAL datasets that did not show consistent directions of effect. For
3802 example, when using the Kettunen data, positive directions of effect were found for BMI and
3803 WHR with very large VLDL. But when using the INTERVAL data, a mix of positive, negative,
3804 and null effects were found for BMI and WHR with very large VLDL.

3805 On the whole, MR results revealed WHR to have a larger effect across metabolites
3806 compared to BMI and BF. However, confidence intervals for WHR and BMI mostly overlapped
3807 with one another. Evidence for an association between BF and the metabolic profile was
3808 weak. Generally, directions of effect for BF conflicted with those of BMI and WHR. Analyses
3809 using BF resulted in larger effect estimates and wider confidence intervals which spanned

3810 the null for the majority of metabolites. For example, when using Kettunen data, BF was
3811 negatively associated with metabolites in the medium VLDL subclass, while BMI and WHR
3812 were positively associated. When using INTERVAL data, BF was negatively associated with
3813 creatinine while BMI and WHR were positively associated.

3814 Consistent directions of effect across all adiposity measures were found for the amino
3815 acids tyrosine, phenylalanine, and valine only. There were no other metabolites across
3816 all three measures of adiposity with a consistent direction of effect. All three metabolites
3817 showed an increase as a result of adiposity. The only other amino acids associated with
3818 adiposity were decreased as a result of BMI and WHR (glutamine) and by BF (histidine).
3819 Results for tyrosine, phenylalanine, and valine are consistent with previous findings, includ-
3820 ing those by Wurtz et al. (2014)¹⁵⁷. Although confidence intervals were much wider than
3821 results here, Wurtz et al. (2014) found glutamine was increased in MR and decreased in
3822 observational analyses. Weak evidence for an increase in histidine was found by Wurtz et
3823 al. (2016). Of these associated amino acids, all but glutamine and tyrosine are essential.
3824 Increased tyrosine, phenylalanine, valine, leucine, and isoleucine have been associated with
3825 increased risk of numerous diseases such as colorectal cancer^{540–543}, pancreatic cancer⁵⁴⁰,
3826 preeclampsia^{544,545}, irritable bowel syndrome^{546,547}, and crohns and ulcerative colitis⁵⁴⁸. Re-
3827 ductions in glutamine and Histidine meanwhile have been associated with colorectal^{541,542,549}
3828 and pancreatic cancer^{549,550}. There is some conflicting evidence however, with higher his-
3829 tidine levels also found to be associated with colorectal cancer⁵⁵¹. A recent prospective
3830 analysis found associations between 12 metabolites and increased risk of endometrial cancer,
3831 including, tyrosine, phenylalanine, leucine, and isoleucine. However, after adjusting for BMI
3832 only the association between glycine and serine with endometrial cancer remained⁵⁵². Given
3833 many of these metabolites were associated with increased adiposity it is possible they may
3834 be intermediates in the adiposity relationship with endometrial cancer.

3835 For BMI and WHR, directions of effect were the same across all metabolites. BMI and
3836 WHR were both associated with reductions of metabolites in medium HDL, large HDL, and
3837 very large HDL subclasses, as well as reductions in apolipoprotein A1, mean diameter
3838 for HDL particles, and total cholesterol in HDL. In addition, BMI and WHR were both as-
3839 sociated with increases in numerous LDL subclasses (e.g., IDL and very small VLDL) as
3840 well as apolipoprotein B. These results are supported by the underlying relationship be-
3841 tween HDL and LDL metabolites as well as similar results in numerous MR^{157,392,519} and
3842 observational^{157,158,328,363–372} analyses. Apolipoprotein A1 is the major component of HDL
3843 particles and enables uptake of lipids by HDL from cells. HDL particles primarily transport
3844 lipids away from the cells, in what is known as reverse cholesterol transport. These lipids end
3845 up at the liver where they are recycled and excreted^{512,513}. Apolipoprotein B on the other
3846 hand is the major component of VLDL, IDL, and LDL particles, enabling lipid uptake, and thus
3847 increased transport of lipids around the body^{512,553}.

3848 Observational studies have highlighted increased HDL to be associated with a protective
3849 effect on cardiovascular disease (CVD)⁵⁵⁴ while increased LDL is shown to increase CVD
3850 risk⁵⁵⁵. MR studies support observational results for LDL⁵⁵⁵, but are conflicting for HDLs

3851 protective effect⁵⁵⁶. Randomised controlled trials have also not found strong evidence for
3852 an effect of HDL lowering drugs on CVD risk⁵⁵⁷. Some focus has been given to a measure
3853 of HDLs contribution to reverse cholesterol transport, HDL cholesterol efflux capacity (HDL-
3854 CEC), which is shown to be protective for CVD^{558,559}. Estimation of HDL-CEC however was
3855 not available in the current analyses. There are also associations between reduced HDL
3856 and many cancers, however there is some evidence to suggest this may be bi-directional⁵⁶⁰.
3857 Recent work has suggested that increased HDL does not confer a protective benefit on
3858 mortality⁵⁶¹, and instead there is evidence that apolipoprotein B may underlie the association
3859 between many lipids and diseases such as CHD⁵⁶². Apolipoprotein B and its associated
3860 lipids, such as VLDLs, may therefore be potential intermediates in disease associations.

3861 Sensitivity analyses were broadly consistent with the main analysis when investigating
3862 the impact of each exposure on metabolites from both the Kettunen and INTERVAL datasets.
3863 There was considerable heterogeneity in the effect estimates derived from the different genetic
3864 instruments used in the main analysis. In single SNP MR analysis, a number of SNPs showed
3865 disproportionately larger effect estimates across many metabolites. Leave-one-out analysis
3866 did not highlight any individual SNP driving the effect for any one metabolite.

3867 Steiger directionality tests can provide an estimate as to whether the “true” causal direction
3868 of an MR analyses has been tested. That is, if we perform the analysis of exposure *A* on
3869 outcome *B* and our hypothesis is that *A* causes *B*, the Steiger test will estimate whether this
3870 direction (*A* to *B*) is “true”. The test relies on using effect size to estimate variance explained
3871 in the exposure and outcome. Results from the Steiger test varied across analyses using
3872 Kettunen and INTERVAL data. For analyses using the Kettunen data, a majority of tests did
3873 not reflect the “true” causal direction. This was primarily a result of almost all tests using
3874 BMI and WHR not reflecting a causal direction. When using the INTERVAL data, a majority
3875 of tests reflected the “true” causal direction. As the majority of analyses using additional
3876 instrument lists for BMI and WHR, for which all instruments were included in the larger main
3877 analysis instruments, reflected a “true” causal direction these results should be interpreted
3878 with caution. As both BMI and WHR used large SNP lists, it is likely that many of those
3879 SNPs will have larger effect sizes for metabolites given their effects on BMI and WHR are
3880 very small - effect sizes are small for BMI and WHR for numerous SNPs because they have
3881 been identified in large well powered GWAS which enable identification of SNPs with small
3882 effect size. For BF, results of the Steiger test are likely more complex, given that many tests
3883 found to be “true” when using the Kettunen data were also found to be “true” when using the
3884 INTERVAL data even though directions of effect were different. It is unclear why there is a
3885 difference here, but it may be due to the instrumentation of BF, which, unlike BMI and WHR,
3886 has been linked with ‘favourable adiposity’.

3887 **5.4.1 Limitations**

3888 MR relies upon the use of an instrument to model the effect of an exposure on an outcome.
3889 This relationship is dependent upon the genetic architecture of the instruments and the traits
3890 under investigation and power (predominantly sample size of the instrument-outcome effect),
3891 both of which influence the utility of the various MR methods. The first MR assumption,
3892 that the instrument is robustly associated with the exposure, is generally measured via an
3893 *F* statistic, for which an arbitrary threshold of > 10 denotes a *strong instrument*. All of the
3894 instruments used here exceeded this threshold. In addition, the variance explained in an
3895 exposure by a GWAS can indicate the power afforded in the MR analysis; the variance
3896 explained for exposures used in the main analysis varied from 6% for BMI, to 3% for WHR,
3897 and $\sim 0.4\%$ for BF. The considerably lower variance explained for BF may have impacted on
3898 results; when using additional instrument lists for BMI and WHR which explained a lesser
3899 percentage of variance confidence intervals became wider. Winners curse is unlikely given
3900 summary statistics used here are well powered and underwent replication by the authors of
3901 the original GWAS publications.

3902 The consistent directions of effect observed across BMI and WHR measures adds weight
3903 to their associations with metabolites. For BF, conflicting directions of effect were observed
3904 when comparing to BMI and WHR. However, confidence intervals were wider and, although
3905 they spanned the null, included the estimates and confidence intervals for BMI and WHR
3906 in a majority of cases. There was considerable difference in the sample sizes used in the
3907 GWAS for BMI, WHR, and BF. In addition, whereas BMI and WHR were measured in only
3908 one way for their respective GWAS, the BF GWAS included measures of BF from DXA and
3909 impedance devices. Though, as shown in Chapter ??, DXA and impedance measures of BF
3910 are highly correlated, the additional analysis for BF which used only impedance measures
3911 in the GWAS showed much greater directional consistency with BMI and WHR and also
3912 included a number of metabolites which reached the multiple testing threshold. Given the
3913 highly correlated nature of the exposures, and the consistency in observational estimates
3914 with metabolites, results for BF here appear counter-intuitive. Additional analyses failed to
3915 appropriately account for the differences in results and further investigation of instrumentation
3916 is warranted.

3917 The instrumentation approach used in the main analysis followed that used by the majority
3918 of analyses reviewed in Chapter 2. That is, instruments were obtained directly from the most
3919 recently published GWAS with the largest sample size. As GWAS sample sizes increase
3920 so does the likelihood of sample overlap which can bias estimates towards the confounded
3921 observational effect. Additionally, there is the concern of population structure within all
3922 GWAS, both large^{346,523–525} and small³⁵¹, which if not accounted for appropriately can lead
3923 to violation of the independence assumption and bias estimates towards the observational
3924 confounded estimate. Additional analyses, which used a number of different instrumentation
3925 practices, including using GWAS which did not include UK Biobank participants and using
3926 a standardised clumping strategy, aimed to reduce the impact of population structure that

3927 has been demonstrated in UK Biobank and that independence of SNPs was the same within
3928 and between exposure instrument lists. Additional BMI and WHR exposures, which did not
3929 include UK Biobank participants, showed consistent results with the main analysis. For BF
3930 there was little difference between the main analysis and the additional analysis which used 5
3931 SNPs from Lu et al. (2015)¹⁹² having removed two SNPs that likely have different biological
3932 function. However, there was considerable difference in directions of effect when using the
3933 additional SNP list from Hubel et al. (2019)⁴³.

3934 There is clearly complexity in the choice of instrument, especially with complex traits such
3935 as adiposity. This complexity is perhaps well demonstrated through the BF analysis. For
3936 BF, there was considerable difference in effect estimates compared to the BMI and WHR
3937 results, with a large proportion of estimates in the opposite direction to those of BMI and WHR.
3938 This is counter-intuitive given the strong correlation between BMI, WHR, and BF, and the
3939 consistent results obtained in observational analyses in Chapter 4 between BMI, WHR, and
3940 BF. Removal of two SNPs previously associated with 'favourable adiposity'^{35,539} resulted in a
3941 global tightening of confidence intervals and a number of effect estimates changing direction
3942 to be more consistent with BMI and WHR. A number of these effects subsequently reached
3943 the multiple testing threshold. However, the majority of MR results remained opposite to that
3944 of BMI and WHR. To further investigate the difference in effects observed for BMI and WHR,
3945 and BF post-hoc interrogation of results from a single, well characterised, adiposity SNP
3946 (rs1558902) was undertaken. Measures of adiposity have been associated with numerous
3947 mutations in the *FTO* locus, with studies highlighting major roles in neural signalling of appetite
3948 suppression, alongside roles in adipocyte browning⁵⁵. Single SNP MR analysis using the
3949 *FTO* locus (rs1558902; BF beta = 0.051; BMI beta = 0.082) and Kettunen metabolites showed
3950 highly consistent results for BF instrumented using the Lu et al. (2016) estimate and BMI
3951 instrumented using the Locke et al. (2015) estimate. These results differed only in their effect
3952 estimate and standard error which was in line with the difference in the SNP beta for the traits
3953 - BMI effect estimates were 62% larger than BF effect estimates. Leave-one-out analysis for
3954 BF instrumented using the 7 genetic variants identified in Lu et al. (2016) did not indicate a
3955 single SNP that could be driving a pleiotropic association. However, median effect estimates
3956 for rs6857 (rs6857 is associated with *NECTIN2* and has previously been associated with a
3957 number of diseases including Alzheimers⁵⁶³) were much larger than those for other SNPs,
3958 both with and without exclusion of the two 'favourable adiposity' SNPs. In many cases, rs6857
3959 did not span the null. It is possible that tests for pleiotropy were underpowered however. The
3960 unexpected results for BF instrumented using the Lu et al. (2016) genetic variants may be
3961 due to a variety of reasons, not least measurement error, sample size differences between BF
3962 (up-to 89,297), BMI (up-to 795,624), and WHR (up-to 697,702), and the variance explained
3963 by the respective instruments: BF = 0.416%, WHR = 3%, BMI = 6%. The complexity of
3964 instrumentation is discussed further in the discussion Chapter 3.3 as this is also relevant to
3965 Chapter 6.

3966 As discussed in the limitations section of Chapter 4, there is no standardised approach,
3967 nor a gold standard, for performing metabolomics quality control. In Chapter 4, quality control,
3968 including outlier detection and removal, was performed using the metaboprep R package.

3969 Metabolite data here however were from summary statistics, the data that went into these
3970 GWAS will have undergone different quality control procedures for the metabolomics data.
3971 Additionally, the metabolomics data were inverse rank normally transformed in the Kettunen
3972 data and inverse normally log transformed in the INTERVAL data, they were also adjusted for
3973 different covariables, it was therefore not appropriate to meta-analyse using effect estimates.
3974 Future work should look to use a standardised method for pre-analysis quality control of
3975 metabolomics data, such as metaboprep while ensuring study specific adjustments are made
3976 where appropriate. On the latter point, a centralised approach to analysis of metabolomics
3977 data, such as that being employed by the Consortium of Metabolomics Studies, will allow for
3978 more efficient comparisons and meta-analyses of studies.

3979 MR analyses are subject to a number of assumption the main three being: (i) the instru-
3980 mental variable (Z) is robustly associated with the exposure (X), (ii) there is no independent
3981 association of the instrumental variable with the outcome (Y) other than through the exposure,
3982 (iii) the instrumental variable is independent of common causes of the instrument-outcome
3983 association (U). In this work, instruments were obtained from large well-powered GWAS,
3984 and the F statistic for each instrument was above a nominal threshold of 10 meaning the
3985 first assumption was likely met. In regards the other two assumptions, formal testing is not
3986 possible. However, sensitivity analyses can provide an indication of pleiotropy and, though not
3987 comprehensive, sensitivity analyses conducted here were concordant with the main analysis,
3988 suggesting that there is no large impact of pleiotropy in these results. However, as discussed,
3989 there may be additional cautions to be taken with the BF analysis, where instrumentation is
3990 currently likely to be inadequate.

3991 An additional consideration with regards to instrumentation is the possibility that SNPs
3992 associated with adiposity traits may also be associated directly with metabolites or metabolic
3993 pathways. The Steiger directionality test can be used to test whether the “true” causal
3994 direction is the one under investigation, i.e. the effect of adiposity measures on metabolites.
3995 There are a number of limitations associated with the Steiger test⁵³⁸ one of which is that it
3996 assumes that there are no pleiotropic effects. When using large SNP lists, the likelihood of an
3997 association between any one of the exposure associated SNPs with the outcome is likely to
3998 increase. The Steiger test uses effect size to estimate the variance explained by each SNP in
3999 the exposure and outcome respectively. In a well powered study, as more SNPs are included
4000 in an instrument there is a greater chance of including SNPs with small effect sizes for the
4001 exposure, as such the likelihood of the test returning a “true” causal direction will decrease.
4002 Although the additional analyses for BMI and WHR, which used smaller SNP lists compared
4003 to the main analysis, were found to reflect the “true” causal direction more often than the
4004 main analysis, the high correlation between the two analyses both in terms of effect size
4005 and direction of effect would suggest that any associations between the SNPs used in the
4006 main analysis with the outcomes did not drastically affect results. This does not fully address
4007 the potential for direct associations between exposure SNPs and the outcome however, and
4008 future work should look at exploring the genetic correlation between the traits.

4009 **5.4.2 Conclusion**

4010 Adiposity exerts a systemic impact on the metabolome, which is consistent across multiple
4011 measures of adiposity (namely, BMI and WHR). Results here highlight a number of metabolites
4012 associated with adiposity which have consistent directions of effect and effect sizes with
4013 previous MR analyses^{157,392}. Some of these metabolites have previously been shown to be
4014 associated with adiposity-related diseases such as colorectal³⁹² and endometrial cancer⁵⁶⁴,
4015 meaning that they could play an intermediary role in these relationships. However, this
4016 possible intermediary role of adiposity-associated metabolites on various health outcomes
4017 are yet to be explored systematically, especially in an MR context. Although adiposity
4018 instrumentation has been well established, and shown here to give consistent results across
4019 many different instrument lists and exposures, results for the main BF analysis presented here
4020 remain unexpected and inconsistent with BMI and WHR. Therefore, proper consideration of
4021 the appropriate use of genetic variants when compiling instruments for use in MR analyses
4022 is required, especially when instrumenting complex traits such as those presented here.
4023 By extension, instrumentation is of particular importance when investigating associations
4024 between adiposity-associated metabolites identified here and diseases identified in Chapter
4025 2. In Chapter 6, investigation of the intermediary role of these adiposity-related metabolites
4026 and an exemplar outcome perturbed by adiposity as identified in meta-analysis in Chapter 2,
4027 endometrial cancer, will be explored.

4028 **Chapter 6**

4029 **Associations between adiposity** 4030 **associated metabolites and** 4031 **endometrial cancer: Mendelian** 4032 **randomization analysis**

4033 **Chapter summary**

4034 This Chapter pulls together information from previous Chapters to investigate the effects
4035 of adiposity associated metabolites on endometrial cancer. In Chapter 2, a systematic review
4036 and meta-analysis showed that adiposity was likely causally implicated in 13 diseases as well
4037 as other traits including systolic blood pressure and fasting glucose. Of these, endometrial
4038 cancer, with which body mass index (BMI) was most strongly related (OR = 1.57; 95% CI
4039 = 1.11 – 2.22; p-value = 0.01), was selected for further investigation in this Chapter. Here,
4040 two-sample Mendelian randomization (MR) was used to estimate the association between
4041 adiposity (BMI, waist-hip ratio (WHR), and body fat percentage (BF)) and endometrial cancer.
4042 Next, two-sample MR was performed to estimate the effect of adiposity-related metabolites,
4043 identified in Chapter 5, and endometrial cancer. To understand whether metabolites, with
4044 evidence of association on endometrial cancer, play an intermediary role in the relationship
4045 between adiposity and endometrial cancer, multivariable MR was performed.

4046 **6.1 Introduction**

4047 In Chapter 1 and 2, I showed that the number of adiposity-associated diseases is high,
4048 consistent with adiposity being a global health concern^{2,3,5-7} and estimated to be responsible
4049 for 8% of global deaths⁴. Many observational studies have highlighted associations with com-
4050 mon diseases, such as cardiovascular disease^{93,96,97,105,122-140,142-144}, cancers^{93,95,108-121},
4051 and common risk factors such as hypertension^{129,157}. Many of these associations have been
4052 replicated in Mendelian randomization (MR) studies as discussed in Chapter 2.

4053 Previous work^{157,158,328,363-372,392,519}, and work conducted in Chapter 4 and 5, has high-
4054 lighted the numerous metabolites associated, in observational and MR analyses, with mea-
4055 sures of adiposity. Many of these adiposity-associated metabolites have been linked with
4056 adiposity-associated diseases including type 2 diabetes^{391,394}, fasting glucose^{391,394}, col-
4057 orectal cancer³⁹², and coronary heart disease³⁹⁵, implying a potential intermediate role for
4058 metabolites. There is also evidence that metabolites can be used to distinguish cancers¹⁸⁸.

4059 To date, few studies have investigated the causal relationship between adiposity-
4060 associated metabolites and adiposity-associated diseases. As an exemplar analysis to test
4061 the potential intermediary role of adiposity-associated metabolites, identified in Chapter 4
4062 and 5, and adiposity-associated diseases, identified in Chapter 2, endometrial cancer was
4063 chosen. This was because endometrial cancer met four key requirements: there was strong
4064 evidence in Chapter 2 for an effect of one or more adiposity measures, there is consistent
4065 evidence across observational⁵⁶⁵ and MR^{431,447,457} analyses that adiposity is associated
4066 with an increased risk of endometrial cancer, there is a large and publicly available GWAS
4067 on endometrial cancer, and the extent to which circulating metabolites may play a role in
4068 the relationship between adiposity and endometrial cancer has not been assessed in the
4069 literature previously.

4070 The World Cancer Report (2020) has identified endometrial cancer (cancer of the inner
4071 lining of the uterus) as one cancer that would benefit greatly from reductions in adiposity, as
4072 approximately 30-40% of cases are a result of obesity⁵⁶⁵. The endometrium is the inner lining
4073 of the uterus, consisting of an epithelial cell layer and mucous membrane. The epithelial cell
4074 layer consists of two parts, a basal layer and a functional layer which thickens and sheds
4075 during menstruation in response to oestrogen and progesterone⁵⁶⁶. Broadly, endometrial
4076 cancer is separated into two types: endometrioid (type 1) and non-endometrioid (type 2); the
4077 latter being the more aggressive form of the disease.

4078 Endometrioid cancer is the more common type of endometrial cancer and is consid-
4079 ered hormone dependent. That is, most endometrioid cancers develop from endometrial
4080 hyperplasia which is associated with pro-longed and unopposed oestrogen exposure⁵⁶⁷.
4081 Non-endometrioid cancer is much less common (~10%) and has typically been considered
4082 hormone independent⁵⁶⁷, however, recent studies suggest that non-endometrioid cancers
4083 are very similar to endometrioid cancers in terms of their hormone dependence⁵⁶⁵. Though

4084 the extent to which they are hormonally driven is likely lesser than endometrioid cancers⁵⁶⁵.
4085 Molecular differences are apparent between the two types, with *PTEN* mutations (associated
4086 with oestrogen exposure) found in many endometrioid cancers and *TP53* and *HER2* (*TP53*
4087 is a tumour suppressor gene and *HER2* is a proto-oncogene that promotes cell growth)
4088 mutations found in non-endometrioid cancers⁵⁶⁷.

4089 A leading cause of endometrioid cancer in pre- and post-menopausal women is adiposity:
4090 in pre-menopausal women, adiposity leads to insulin resistance, increased androgen,
4091 anovulation, and decreased progesterone; in post-menopausal women, adiposity leads to
4092 increased oestrogen via excess androgen conversion. In both instances, these changes
4093 result in an increase in oestrogen and subsequent endometrial hyperplasia⁵⁶⁷. Additionally,
4094 there is evidence of oestrogen independent activation of the oestrogen-receptor, primarily
4095 through insulin-like growth factor 1⁵⁶⁸, which is also up-regulated as a result of adiposity.
4096 Although there is evidence that non-endometrioid cancers are hormonally driven, and would
4097 therefore likely share risk factors with endometrioid cancer, there is weaker evidence of an
4098 association with adiposity compared with endometrioid cancer⁵⁶⁹.

4099 In Chapter 2, meta-analysis of three MR studies^{431,447,457} found a 1.57 increased risk of
4100 endometrial cancer with each SD increase in BMI (95% CI = 1.11 – 2.22). Previous work has
4101 shown different metabolite profiles between overall endometrial cancer cases and controls
4102 independent of obesity and other risk factors such as diabetes^{564,570}. More recently, adiposity-
4103 associated metabolites, identified in Chapter 4 and 5, have been linked with endometrial
4104 cancer risk in a prospective cohort⁵⁵². Whether there is a causal effect of metabolites on
4105 endometrial cancer development however is unclear. Here, I combine work undertaken
4106 in the previous Chapters, which identified metabolites causally related to adiposity, and
4107 utilise two-sample MR and two-sample multivariable MR (MVMR) to understand the possible
4108 intermediary role played by adiposity-related metabolites on endometrial, endometrioid, and
4109 non-endometrioid cancers using data from the largest publicly available metabolomics and
4110 endometrial cancer GWAS to date.

4111 6.2 Methods

4112 This Chapter details hypothesis-driven MR analyses investigating associations between
4113 adiposity-related metabolites (here in referred to as metabolites) and endometrial cancer
4114 (Figure 6.1). Firstly, a two-sample MR analysis of the effect of multiple measures of adiposity
4115 on endometrial cancer was performed to independently validate findings from the meta-
4116 analysis performed in Chapter 2. Secondly, a two-sample MR analysis of the effect of multiple
4117 measures of adiposity on metabolites was performed to obtain the best estimates of the
4118 effect of adiposity on metabolites for MR analyses with endometrial cancer. This analysis was
4119 performed using recently available data which included the largest currently available GWAS
4120 of circulating metabolites conducted in UK Biobank. Thirdly, metabolites with consistent

4121 directions of effect across results performed in Chapter 5 and this Chapter, were used in a
4122 two-sample MR analysis to investigate their effect on endometrial cancer. Finally, metabolites
4123 for which there was evidence of an effect on endometrial cancer were taken forward and
4124 used in a MVMR analysis with adiposity measures to estimate their intermediate effects on
4125 endometrial cancer.

4126 All data manipulation and analyses were performed using R⁴⁷⁹ (version 3.5.3) and bash.
4127 All code used is available on [GitHub](#). Two-sample MR analyses were performed using the
4128 TwoSampleMR⁵²² (version 0.4.22) R package. MVMR analyses were performed using the
4129 MVMR (version 0.3) R package. Summary statistics for adiposity measures and metabolites
4130 were obtained from the original GWAS sources (described in the next section). Summary
4131 statistics for metabolites were from UK Biobank and were obtained from collaborators prior
4132 to publication of the GWAS (unpublished; Carolina Borges, University of Bristol). Summary
4133 statistics for endometrial cancer were available from MR-Base⁵²² (accessed 17/07/2021). A
4134 list of metabolites used in this Chapter is available in the Appendix (Table A.6) and on [GitHub](#).
4135 Results were visualised using the ggforestplot (version 0.1.0) R package.

Chapter 5

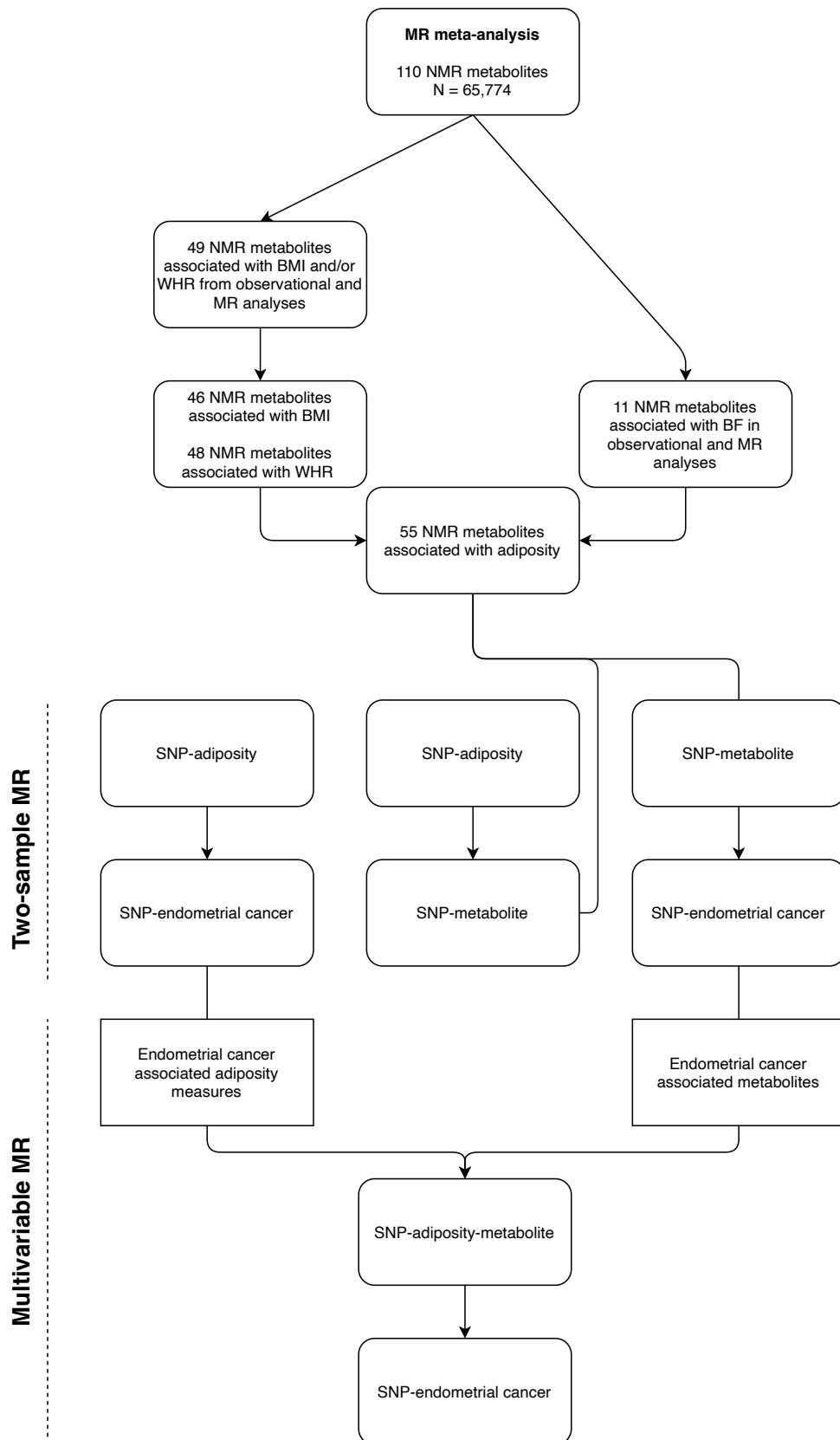


Figure 6.1: Analysis overview. In Chapter 5, 55 metabolites were identified as being associated with adiposity. For body mass index (BMI) and waist hip ratio (WHR) these metabolites were identified through directional consistency and multiple testing thresholds across Mendelian randomization and observational analyses. For body fat percentage (BF) metabolites were identified through directional consistency and a multiple testing threshold in the observational analysis only. These 55 metabolites were taken forward and used in two-sample MR and multivariable MR analyses to investigate intermediate effects on endometrial cancer.

4136 **6.2.1 Data**

4137 The following details a number of GWAS and meta-analyses, as per STROBE-MR⁴¹⁷,
4138 which I did not conduct but that were available to me either via publicly available resources
4139 or collaborations. The total N and sex specific N does not always tally which is because of
4140 variation in the N for each SNP. Where this is the case, 'N up to' is used.

4141 **Exposures: Adiposity**

4142 **Body mass index** Detailed information is presented in Chapter 5 Section 5.2.2. Briefly,
4143 summary statistics for BMI were obtained from Locke et al. (2015)³⁶, in which 322,154
4144 individuals of European ancestries were included in an inverse variance weighted fixed effects
4145 meta-analysis. The reason that this data was used instead of the largest GWAS of BMI
4146 conducted by Yengo et al. (comprising both GIANT and UK Biobank) was to avoid sample
4147 overlap with the largest available endometrial cancer GWAS conducted in UK Biobank used
4148 in this Chapter (see below). A total of 82 GWAS and 43 studies using the Metabochip
4149 array were included in the meta-analysis. Individual GWAS were adjusted for age, age²,
4150 and study specific covariates with residuals inverse normally transformed. Imputation was
4151 performed using HapMap phase II CEU reference panel. Each study used a linear regression
4152 model assuming an additive genetic model with quality control following procedures outlined
4153 previously⁵⁷¹. A fixed effect inverse variance weighted meta-analysis was performed using
4154 METAL for the 82 GWAS and 43 studies using the Metabochip array separately. The final
4155 meta-analysis combined the SNPs found in both the meta-analyses of GWAS and Metabochip
4156 studies that were corrected for genomic control. A total of 77 loci reaching genome-wide
4157 significance (p-value $\leq 5 \times 10^{-8}$) and separated by at least 500 kilobases were identified.

4158 **Waist hip ratio** Detailed information is presented in Chapter 5 Section 5.2.2. Briefly,
4159 summary statistics for WHR were obtained from Shungin et al. (2016)³⁷, in which 210,088
4160 individuals of European ancestries were included in an inverse variance weighted fixed
4161 effects meta-analysis. The reason that this data was used instead of the largest GWAS
4162 of WHR conducted by Pulit et al. (comprising both GIANT and UK Biobank) was to avoid
4163 sample overlap between the largest available endometrial cancer GWAS conducted in UK
4164 Biobank used in this Chapter (see below). A total of 57 GWAS and 44 studies using the
4165 Metabochip array were included in the meta-analysis. WHR was adjusted for age, age²,
4166 study-specific covariates if necessary with residuals inverse normally transformed. Imputation
4167 was performed using HapMap phase II CEU reference panel. Each study used a linear
4168 regression model assuming an additive genetic model. The final meta-analysis combined the
4169 SNPs found in both the meta-analyses of GWAS and Metabochip studies that were corrected
4170 for genomic control. A total of 26 loci reaching genome-wide significance (p-value $\leq 5 \times 10^{-8}$)
4171 and separated by at least 500 kilobases were identified.

4172 **Body fat percentage** Detailed information is presented in Chapter 5 Section 5.2.2. Briefly,
4173 Lu et al. (2016) identified 7 independent SNPs in 89,300 individuals of European ancestries
4174 (in a meta-analysis of studies not including UK Biobank). Subsequent work has identified two
4175 of these SNPs (rs6738627 and rs2943650) to be associated with *favourable adiposity*^{35,539}.
4176 *Favourable adiposity* loci have been associated with increased BF, BMI, fat mass, and reduced
4177 fat free mass while simultaneously being associated with reduced risk of type-2 diabetes,
4178 hypertension, and heart disease, as well as more favourable blood pressure⁵³⁹. In additional
4179 analyses conducted in Chapter 5, comparison of the effect of BF on metabolites, instrumenting
4180 BF with the 7 SNPs identified by Lu et al. and a 5 SNP instrument that did not include the 2
4181 *favourable adiposity* SNPs, resulted in tighter confidence intervals for the 5 SNP instrument.
4182 Additionally, a number of effect estimates reversed in direction when instrumenting BF using
4183 the 5 SNPs. As these *favourable adiposity* SNPs are likely adding noise to the instrumentation
4184 of BF, the 5 SNP instrument was used here-in.

4185 **Intermediates: Metabolites**

4186 A total of 56 metabolites were identified as being associated with adiposity in Chapter
4187 5; one metabolite was not available in the UK Biobank GWAS. Female-specific summary
4188 statistics for the 55 available metabolites were obtained from UK Biobank (Borges 2021,
4189 unpublished), in which, 118,466 women of European ancestries were included in a linear
4190 mixed model GWAS. UK Biobank is a prospective cohort study of ~500,000 individuals from
4191 the United Kingdom aged 37–70 with a host of genetic and phenotypic data^{529,572}. Un-fasted
4192 individuals were selected at random to undergo high-throughput NMR metabolomics⁵⁷³ using
4193 the Nightingale Health Ltd biomarker quantification^{492,574} (version 2020). All metabolites
4194 were inverse normally transformed prior to genome wide analysis. Genome-wide association
4195 analysis was performed using the MRC IEU, UK Biobank GWAS pipeline⁵⁷⁵. A linear mixed
4196 model using BOLT-LMM, adjusting for genotype array and fasting time was fit for 118,466
4197 individuals. Population structure was controlled for using 143,006 directly genotyped SNPs
4198 (minor allele frequency > 0.01; genotyping rate > 0.015; Hardy-Weinberg equilibrium p-value
4199 < 0.0001 and linkage disequilibrium (LD) pruning to an R² threshold of 0.1 using PLINKv2.00).
4200 Borges et al. had not identified lead SNPs at the time of writing. Lead SNPs were thus
4201 identified as those reaching a genome wide significance threshold (p-value $\leq 5 \times 10^{-8}$) and
4202 which were retained after clumping using a LD R² threshold of 0.001 for SNPs within a 10,000
4203 base window of each other and using the 1000 Genomes V3 reference panel.

4204 **Outcomes: Endometrial cancer**

4205 Female-specific summary statistics for endometrial cancer were available from O'Mara et
4206 al. (2018)⁴⁶⁷. This data included summary statistics on overall endometrial cancer as well
4207 as endometrioid and non-endometrioid cancer. Briefly, 17 studies on endometrial cancer

4208 including 12,906 cases and 108,979 country-matched controls of European ancestries were
4209 included in an inverse variance weighted fixed effects meta-analysis. Genotypes were imputed
4210 using the 1000 Genomes Project v3 reference panel or a combined 1000 Genomes (V3)
4211 UK10K reference panel. In each study, univariate GWAS were conducted adjusting for
4212 principal components. The analysis was repeated for endometrioid only (cases = 8758;
4213 controls = 108,979) and non-endometrioid only (case = 1230; controls = 108,979) cancers.
4214 Overall endometrial cancer included the cases from the endometrioid and non-endometrioid
4215 cancers alongside a number of other unclassified cases. Of the 12,906 cases and 108,979
4216 controls for overall endometrial cancer, 636 cases (5%) and 62,853 controls (58%) were from
4217 UK Biobank. For endometrioid and non-endometrioid cancer, no cases and 62,853 controls
4218 (58%) were from UK Biobank.

4219 **Overlap**

4220 In a two-sample MR analyses, overlap in the exposure and outcome datasets will bias
4221 estimates towards the observational confounded effect and increase the false positive rate.
4222 This bias will be exacerbated by weak instruments as weak instrument bias will be in the
4223 direction of the confounded observational estimate with greater overlap in the two samples
4224 (given a strong enough instrument bias as a result of overlap will be close to the un-biased
4225 estimate)⁴¹⁸. For continuous outcomes, bias away from the null is a linear function of the
4226 sample overlap (e.g., sample overlap of 50% leads to a bias of 5%). For binary outcomes,
4227 when the SNP-outcome association is estimated in all participants, bias is similar to that for
4228 a continuous outcome. Where the SNP-outcome association is estimated in controls only,
4229 unbiased estimates can be obtained⁴¹⁸. In the analyses discussed in this Chapter, there is
4230 potential overlap between metabolite data (exposure) and overall endometrial cancer data
4231 (outcome) of a maximum 5%. This 5% maximum overlap would equate to ~0.5% increased
4232 false positive rate.

4233 **6.2.2 Two-sample Mendelian randomization**

4234 All data manipulation and analyses were performed using R⁴⁷⁹ (version 3.5.3) and the
4235 TwoSampleMR⁵²² (version 0.4.22) R package. Summary statistics for BMI³⁶, WHR³⁷, and
4236 BF³⁹ were obtained from published GWAS, as described above, all genetic instruments are
4237 available on [GitHub](#). Summary statistics from UK Biobank for metabolites were unpublished,
4238 all genetic instruments are available on [GitHub](#), while a list of metabolites used are available
4239 in the Appendix (Table A.6). Results were visualised using the ggforestplot (version 0.1.0)
4240 R package.

4241 For all exposures, the following data were obtained from the original GWAS publications:
4242 rsID, effect allele, other/non-effect allele, effect allele frequency, effect estimate, standard error

4243 of the effect estimate, p-value, N, and units. The same data for genetic instrumental variables
4244 were obtained for the outcome for each exposure separately. Where genetic instrumental
4245 variables were not present in the outcome GWAS, proxy SNPs were included if LD was ≥ 0.8 .
4246 For proxy SNPs, the inclusion of SNPs where the reference strand was ambiguous (strand
4247 flips) was allowed and the direction was inferred using a minor allele frequency threshold
4248 of 0.3. That is, the direction was inferred using a minor allele frequency, so long as that
4249 frequency was not ≥ 0.3 or ≤ 0.7 , in which case it was excluded.

4250 SNPs for all adiposity measures were not clumped as the studies from which they were
4251 obtained stated they were independent or near-independent and analysis in Chapter 5
4252 indicated clumping did not considerably alter associations with metabolites. SNPs for all
4253 metabolites were identified using a genome-wide significance threshold of p-value $\leq 5 \times 10^{-8}$.
4254 Associated SNPs were then clumped using the TwoSampleMR⁵²² R package setting a LD R²
4255 threshold of 0.001 for SNPs within a 10,000 base window of each other and using the 1000
4256 genomes reference panel. Exposure and outcome SNPs were harmonized in reference to the
4257 exposure effect allele being on the increasing scale. For included alleles where the reference
4258 strand was ambiguous, the positive strand was inferred using effect allele frequency.

4259 An inverse variance weighted (IVW), multiplicative random effects (IVW-MRE) model was
4260 used to estimate the effect of each exposure on the outcome. The model assumes that the
4261 strength of the association of the genetic instruments with the exposure is not correlated with
4262 the magnitude of the pleiotropic effects and that the pleiotropic effects have an average value
4263 of zero⁵³⁴. Where the number of instruments was not sufficient for an IVW-MRE model, the
4264 Wald ratio was used. As analysis of the effect of adiposity on endometrial cancer aimed to
4265 validate results presented in Chapter ??, and analysis of the effect of adiposity on metabolites
4266 aimed to obtain updated estimates for metabolites with evidence of association presented in
4267 Chapter 5, and analysis of the effect of metabolites on endometrial cancer aimed to identify
4268 which of those adiposity-associated metabolites were also associated with endometrial cancer,
4269 no multiple testing threshold was set for any analysis. For adiposity measures and metabolites,
4270 residuals were inverse rank normally transformed prior to genome wide analysis, as such the
4271 units represent a standard deviation (SD). For MR analyses using these measures, effect
4272 estimates are given in SD units. Results testing for the causal effect of both adiposity and
4273 metabolites on endometrial cancer are given as odds ratios (OR) and represent the odds of
4274 developing endometrial cancer per SD unit increase in the exposure.

4275 **Sensitivity analysis**

4276 Where possible (i.e. 3 or more instruments), the assumptions of no pleiotropy among
4277 genetic instruments and outcomes were explored using: MR-Egger⁵³⁵, weighted median⁵³⁶,
4278 and weighted mode⁵³⁷ based estimators. These methods are sensitive to the effects of
4279 potential pleiotropy. No p-value threshold requirements were set for these methods, instead
4280 consistency between the IVW-MRE model and these methods was investigated. Briefly, MR-

4281 Egger provides an estimate of unbalanced/directional horizontal pleiotropy via the intercept of
4282 a linear regression of the SNP-exposure and SNP-outcome association. In the presence of
4283 pleiotropy, the intercept will bias away from the origin. MR-Egger gives consistent estimates
4284 when 100% of genetic instruments are invalid⁵³⁵. The weighted median is complimentary to
4285 MR-Egger but does not rely on the “instrument strength independent of direct effect” (InSIDE)
4286 assumption. It calculates the median of an empirical distribution of the causal effect estimates
4287 weighted for precision. It provides consistent estimates when at least 50% of the weight
4288 comes from valid genetic instruments and as long as no one genetic instrument contributes
4289 > 50% of the weight⁵³⁶. The weighted mode assumes the true causal effect is the most
4290 common effect and it is robust when the majority of effect estimates are derived from valid
4291 instruments⁵³⁷.

4292 6.2.3 Two-sample multivariable Mendelian randomization

4293 All adiposity measures were included in the MVMR analysis. Of the metabolites included
4294 in the two sample MR analysis estimating the causal effect of adiposity on metabolites and of
4295 metabolites on endometrial cancer, only those which showed a consistent direction of effect
4296 with both adiposity and endometrial cancer were included in the MVMR analyses. That is, if
4297 adiposity increased a metabolite, that metabolite increased endometrial cancer, and adiposity
4298 increased endometrial cancer, then the metabolite was considered as a potential intermediate.
4299 Otherwise, the metabolite was excluded from MVMR analyses.

4300 In MVMR, SNPs for the exposure (adiposity) and proposed intermediate (metabolites) are
4301 combined to create a combined instrument of the exposure and intermediate. This combined
4302 SNP list is used to extract instruments from the exposure GWAS from the intermediate GWAS.
4303 The resulting instrument is then clumped to remove duplicate SNPs and SNPs in LD with one
4304 another to avoid double counting effects. The following process was followed for obtaining
4305 instruments:

4306 The resulting instrument contains SNPs for both the exposure and intermediate; extracting
4307 data from the exposure GWAS gives an instrument for the exposure adjusted for the intermediate
4308 and extracting data from the intermediate GWAS gives an instrument for the intermediate
4309 adjusted for the exposure (Figure 6.2). As such, instruments for BMI and instruments for
4310 metabolites will use the same SNPs but different estimates. For example, a BMI instrument
4311 adjusted for metabolite 1 will include the SNPs associated with BMI and the SNPs associated
4312 with metabolite 1, with all data extracted from the BMI GWAS. An instrument for metabolite 1
4313 adjusted for BMI will include the SNPs associated with BMI and the SNPs associated with
4314 metabolite 1, with all data extracted from the metabolite 1 GWAS.

- 4315 1. SNPs associated with each adiposity measure and each metabolite were identified in
4316 the same way as for the two sample MR analysis described previously (6.2.2)

- 4317 2. Identified SNPs for each adiposity and metabolite pair were combined to create a SNP
 4318 list for each adiposity and metabolite pair
 4319 3. These SNP lists were used to extract summary level data for instruments (rsID, effect
 4320 allele, other/non-effect allele, effect allele frequency, effect estimate, standard error of
 4321 the effect estimate, p-value, N, and units) from the GWASs of each adiposity measure
 4322 and each metabolite individually
 4323 4. Each instrument was clumped using the TwoSampleMR⁵²² R package setting an LD R²
 4324 threshold of 0.001 for SNPs within a 10000 base window of each other and using the
 4325 1000 genomes reference panel
 4326 5. Each adiposity and metabolite instrument pair were harmonised in reference to the
 4327 adiposity instrument effect allele being on the increasing scale
 4328 6. Instruments for each adiposity and metabolite pair were extracted from the endometrial
 4329 cancer GWAS using MR-Base. Where SNPs were not present in the outcome GWAS,
 4330 proxy SNPs were included if LD was ≥ 0.8 . For proxy SNPs, the inclusion of SNPs
 4331 where the reference strand was ambiguous (strand flips) was allowed and the direction
 4332 was inferred using a minor allele frequency threshold of 0.3
 4333 7. The outcome instruments were harmonised in reference to the adiposity instrument
 4334 effect allele being on the increasing scale

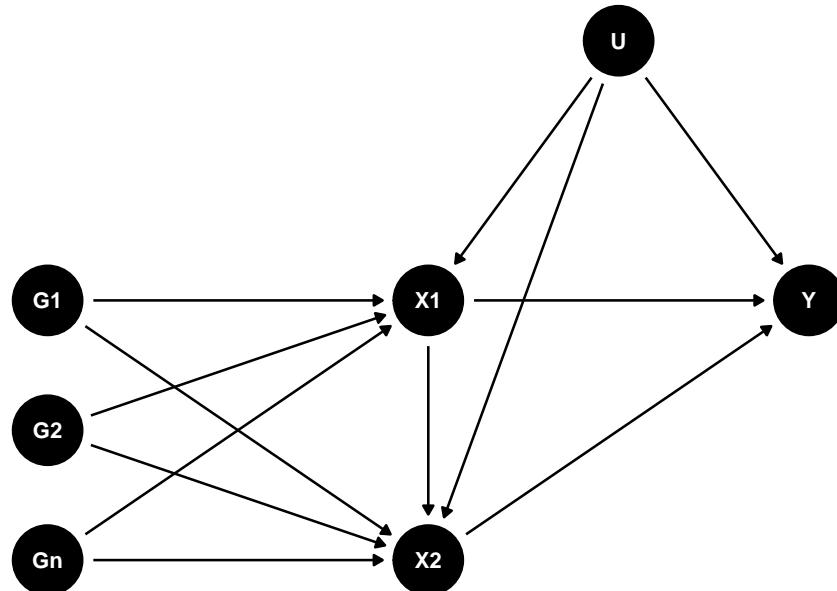


Figure 6.2: **Directed acyclic graph of multivariable MR principle.** The DAG illustrates the principle of multivariable MR using an exposure (X1) and a mediator (X2) on a single outcome. Genetic instruments (G) associated with X1 and X2 are combined into an instrument, in this example the instrument includes G1, G2, and any number of other SNPs (Gn). The instrument is extracted from summary statistics for each of X1 and X2. G = single nucleotide polymorphism; X1 = exposure, X2 = mediator; Y = outcome; U = unmeasured confounding

4335 For MVMR analysis, the SNP, effect estimate, and standard error for each exposure
4336 (adiposity and metabolite) and outcome are required. Instrument strength for each exposure
4337 was estimated using a generalized version of Cochran's Q^{576,577} using the `strength_mvmr()`
4338 function in the MVMR R package assuming a pairwise covariance of 0. The assumption
4339 being that the associations between the SNP and exposure are estimated in independent
4340 samples. Horizontal pleiotropy was evaluated using a modified form of Cochran's Q using the
4341 `pleiotropy_mvmr()` function in the MVMR R package and assuming a pairwise covariance of
4342 0.

4343 An IVW MVMR model was used to obtain the direct causal effect of each adiposity
4344 measure adjusted for each metabolite and each metabolite adjusted for each adiposity
4345 measure on endometrial cancer risk. Results from these analyses will inform as to whether
4346 the effect of adiposity on endometrial cancer is partly explained by the effect of adiposity on
4347 metabolites and those metabolites effects on endometrial cancer. Additionally, it is possible
4348 that the effect of metabolites on endometrial cancer may be independent of the effect of
4349 adiposity on metabolites and the metabolites are therefore not solely intermediates but are
4350 also direct causes of endometrial cancer. Results are presented as OR and represent the
4351 odds of developing endometrial cancer per SD unit increase in the exposure.

4352 6.3 Results

4353 For this section, results are presented first for the two-sample MR analyses of the effect
4354 of adiposity on endometrial cancer, the effect of adiposity on metabolites in UK Biobank,
4355 and the effect of those metabolites for which there was evidence for a causal effect of
4356 adiposity (consistent across analyses conducted in Chapter 5 and UK Biobank here) and
4357 endometrial cancer. Finally, results from the two-sample MVMR analyses unpicking the
4358 potential intermediary role of these metabolites in the relationship between adiposity and
4359 endometrial cancer are presented.

4360 6.3.1 Two-Sample Mendelian randomization: effect of adiposity measures on 4361 endometrial cancer

4362 Here, using the largest endometrial cancer GWAS to date, the MR analyses performed as
4363 part of this thesis provided strong evidence for a causal effect of BMI on endometrial cancer
4364 (OR per SD unit increase in BMI = 1.91; 95% CI = 1.75–2.07; p-value = 1.74 x 10⁻¹⁵; (Table
4365 6.1). A similar effect (OR = 2.02; 95% CI = 1.83–2.21; p-value = 1.59 x 10⁻¹³) was found for
4366 endometrioid cancer, while a smaller effect was observed for non-endometrioid cancer (OR =
4367 1.63; 95% CI = 1.25–2.01; p-value = 0.01).

4368 The effect of WHR on the three outcomes was less clear. Though effect estimates were
 4369 all in line with those for BMIF, only the effect of WHR on non-endometrioid cancer (OR = 2.31;
 4370 95% CI = 1.68–2.94; p-value = 0.01) had confidence intervals that did not cross the null. The
 4371 effect of BF on endometrial cancer (OR = 2.54; 95% CI = 2.32–2.76; p-value = 1.02×10^{-16}),
 4372 endometrioid (OR = 2.73; 95% CI = 2.33–3.14; p-value = 9.93×10^{-7}), and non-endometrioid
 4373 (OR = 2.01; 95% CI = 0.66–3.35; p-value = 0.31) were more similar to that of BMI though
 4374 with larger effect sizes and wider confidence intervals.

4375 Sensitivity analyses, which included weighted median, weighted mode, and MR-Egger
 4376 models, were consistent with the IVW-MRE model for each exposure (Appendix Figure A.33).
 4377 All figures associated with sensitivity analyses (e.g. leave one out analyses) are available on
 4378 [GitHub](#) with representative figures presented in the appendix (Appendix A.5). All results are
 4379 available on [GitHub](#).

Table 6.1: Two-sample Mendelian randomisation results: the effect of adiposity measures on endometrial cancer

		OR	lowe ci	upper ci	p-value
BMI	Endometrial cancer	1.91	1.75	2.07	0.00
BMI	Endometrioid	2.02	1.83	2.21	0.00
BMI	Non-endometrioid	1.63	1.25	2.01	0.01
BF	Endometrial cancer	2.54	2.32	2.76	0.00
BF	Endometrioid	2.73	2.33	3.14	0.00
BF	Non-endometrioid	2.01	0.66	3.35	0.31
WHR	Endometrial cancer	1.26	0.98	1.53	0.10
WHR	Endometrioid	1.25	0.93	1.58	0.17
WHR	Non-endometrioid	2.31	1.68	2.94	0.01

Odds ratios (OR) and associated 95% confidence intervals (CI) per standard deviation unit increase in body mass index (BMI), waist hip ratio (WHR), or body fat percentage (BF)

4380 **6.3.2 Two-sample Mendelian randomization: effect of adiposity measures on**
 4381 **metabolites**

4382 In Chapter 5, 46, 49, and 9 metabolites were associated with BMI, WHR, and BF re-
 4383 spectively. Of these metabolites, two (serum total triglycerides and estimated description of
 4384 fatty acid chain length not actual carbon number) were not measured in UK Biobank and so
 4385 were not available for analyses. As such, a total of 45, 48, and 8 metabolites, which were
 4386 associated with BMI, WHR, and BF, respectively, in Chapter 5, were available for analysis in
 4387 UK Biobank.

4388 For BMI, of the 45 metabolites, 4 had directions of effect that were not consistent with the
4389 previous MR analysis (Chapter 5), meta-analysis (Chapter 5), and observational (Chapter 4)
4390 analyses: apolipoprotein B, free cholesterol in medium VLDL, total lipids in medium VLDL,
4391 total lipids in very small VLDL. For WHR, of the 48 metabolites, 2 had directions of effect
4392 that were not consistent with the previous MR analysis (Chapter 5), meta-analysis (Chapter
4393 5), and observational (Chapter 4) analyses: phenylalanine and tyrosine. For BF, all 8 of
4394 the analysed metabolites had consistent directions of effect with the previous MR analysis
4395 (Chapter 5), meta-analysis (Chapter 5), and observational (Chapter 4) analyses. Sensitivity
4396 analyses did not highlight metabolite analyses which violated the MR assumptions. As such,
4397 a total of 41, 46, and 8 metabolites were taken forward for MR analysis to estimate the effect
4398 of these metabolites on endometrial cancer outcomes.

4399 All results for the investigation of the effect of adiposity measures on metabolites measured
4400 in UK Biobank are presented in Appendix Figure A.38 and are available on [GitHub](#). All
4401 sensitivity plots are available on [GitHub](#) and representative figures are presented in the
4402 appendix (Appendix A.5).

4403 **6.3.3 Two-sample Mendelian randomization: effect of adiposity-associated 4404 metabolites on endometrial cancer**

4405 Two-sample MR analysis was used to investigate the causal effect of 53 metabolites
4406 (i.e., 41, 46, and 8 metabolites driven by BMI, WHR, and BF, respectively) on endometrial
4407 cancer. A total of 7 and 5 metabolites were only found to be associated with WHR and
4408 BF respectively. Two metabolites were found to be associated with BMI and BF and 38
4409 metabolites were associated with BMI and WHR. One metabolite, valine, was associated with
4410 all three measures. Instruments for metabolites were obtained from an unpublished GWAS of
4411 249 circulating metabolites performed in UK Biobank (Borges, unpublished). Using a genome
4412 wide p-value threshold of 5×10^{-8} and an LD R² clumping threshold for SNPs within a 10,000
4413 base window, to ensure independence of SNPs, identified 2595 total associations with 53
4414 metabolites (unique number of SNPs = 934; minimum number of SNPs associated with a
4415 metabolite = 6, maximum number of SNPs associated with a metabolite = 84; Figure 6.3).
4416 The greater the number of SNPs associated with a trait, the larger the variance explained by
4417 those SNPs and thus the power afforded by that instrument in an MR analysis. Of the 934
4418 unique SNPs identified, 54% of SNPs were associated with just one metabolite each (Figure
4419 6.4). The remaining 46% of SNPs were associated with 2 or more metabolites, with one SNP
4420 (rs12154627) associated with 33 metabolites.

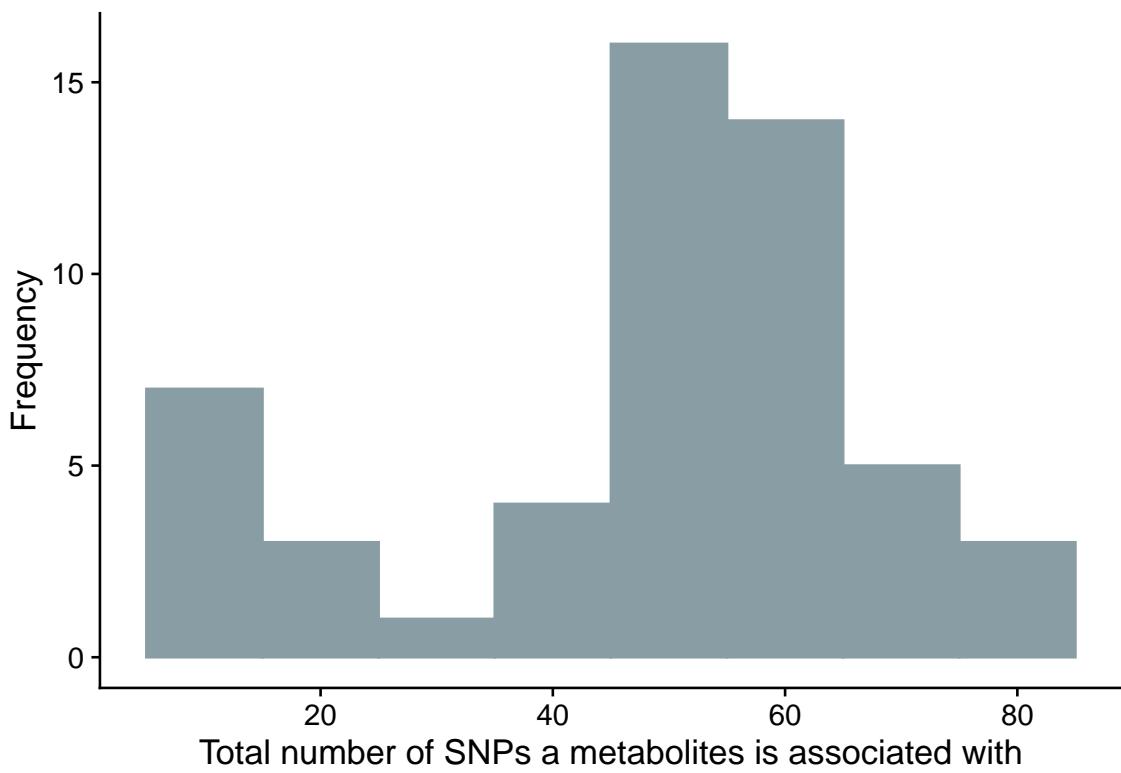


Figure 6.3: **Distribution of the number of SNPs individual metabolites were associated with.** A total of 934 unique SNPs were identified across 53 metabolites. The number of SNPs associated with each metabolite varied between 6 and 84 with a mean of 49. Bin size = 10.

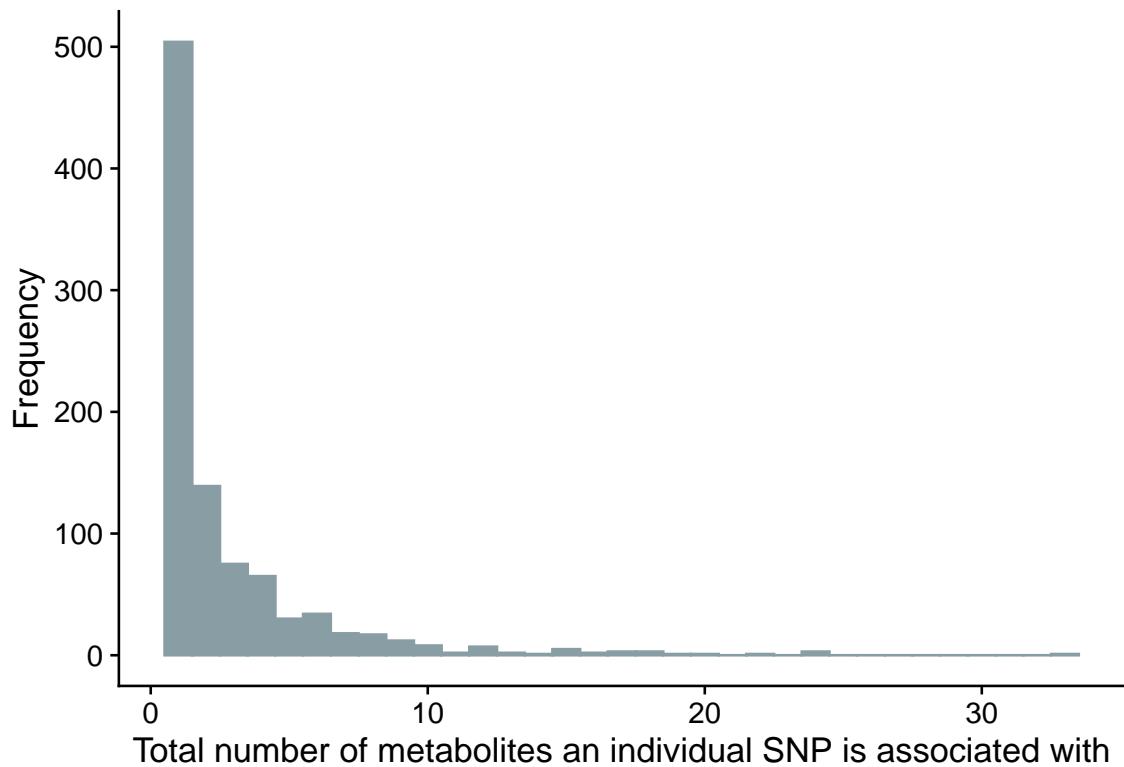


Figure 6.4: **Distribution of the number of metabolites individual SNPs were associated with.** The majority of the 934 identified SNPs were associated with 1 metabolite. One SNP was associated with 33 out of a total 53 metabolites. Bin size = 1.

4421 A total of 5 metabolites showed evidence of association for at least one outcome (Figure
 4422 6.6). Valine was the only metabolite found to reduce the risk of endometrial cancer (OR of
 4423 endometrial cancer per SD unit increase in valine = 0.82; 95% CI = 0.64–1; p-value = 0.03).
 4424 The largest effect was found for triglycerides in small VLDL, which increased the odds of
 4425 non-endometrioid cancer by 1.8 (95% CI = 1.4–2.1; p-value = 0.001). The largest effect on
 4426 endometrioid cancer was for phospholipids in very large HDL (OR 1.16; 95% CI = 1.02–1.31;
 4427 p-value = 0.04), while the largest effect on overall endometrial cancer was for triglycerides in
 4428 very large HDL (OR = 1.11; 95% CI = 1.00–1.22; p-value = 0.06).

4429 Of the 5 metabolites with evidence of association with endometrial cancer, endometrioid
 4430 cancer, and/or non-endometrioid cancer, only triglycerides in very large HDL was associated
 4431 with BF in the previous analysis conducted in this Chapter and in observational (Chapter 4)
 4432 and MR (Chapter 5) analyses. The remaining 4 metabolites were not associated with BF
 4433 in any analysis, but were all associated with BMI and WHR in the previous analysis here
 4434 and in observational (Chapter 4) and MR (Chapter 5) analyses. As all 5 metabolites showed
 4435 evidence of association with a measure of adiposity and endometrial cancer, all were taken
 4436 forward for MVMR analysis.

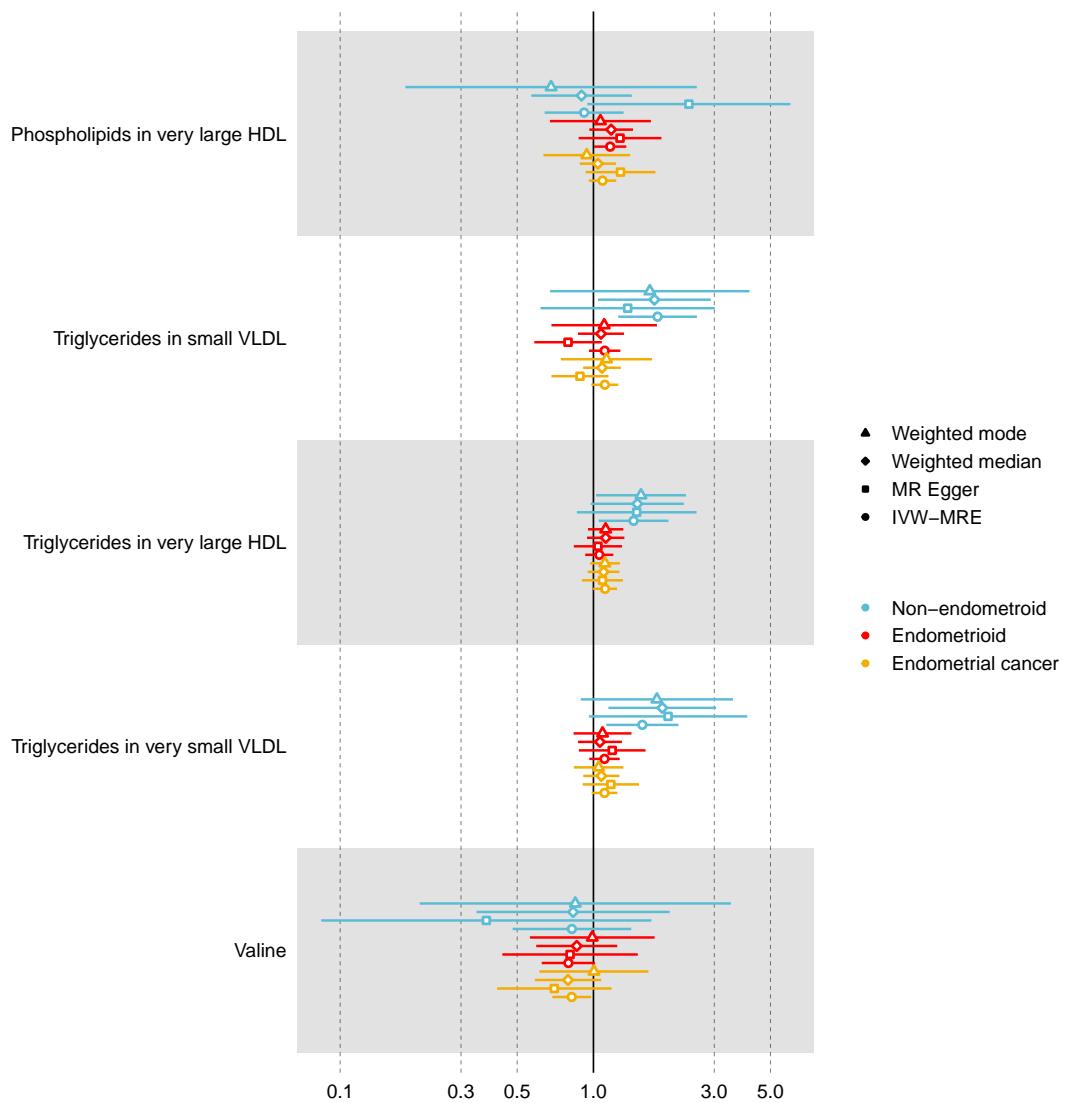


Figure 6.5: Two-sample MR: Effect of metabolites on endometrial cancer. Forest plot shows odds ratio (OR) and 95% confidence interval for metabolites associated with BMI, WHR, and/or BF on endometrial cancer, endometrioid cancer, and non-endometrioid cancer. The main analysis (IVW multiplicative random effects (MRE)) is presented alongside sensitivity analyses (weighted median, weighted mode, MR-Egger).

4437 **6.3.4 Two-sample multivariable Mendelian randomization: intermediate effects**
4438 **of adiposity associated metabolites on endometrial cancer**

4439 Of the 5 metabolites for which there was causal evidence of an effect of at least one
4440 measure of adiposity and an effect on endometrial cancer, two had consistent positive
4441 directions of effect with all measures of adiposity and endometrial cancer: triglycerides in
4442 small VLDL and triglycerides in very small VLDL. For triglycerides in small VLDL, the strongest
4443 effect of adiposity was found for WHR (SD unit increase in metabolite per SD increase in
4444 adiposity measure = 0.56; 95% CI = 0.26–0.86; p-value = 0); WHR also had the strongest
4445 effect on triglycerides in very small VLDL (beta = 0.45; 95% CI = 0.2–0.7; p-value = 0). BMI
4446 (effect on small VLDL = 0.07; effect on very small VLDL = 0.06) and BF (effect on small VLDL
4447 = 0.11; effect on very small VLDL = 0.05) both had positive effects on triglycerides in small
4448 VLDL and very small VLDL, however confidence intervals overlapped the null (Table 6.2).
4449 Of the remaining three metabolites, all three adiposity measures were found to decrease
4450 phospholipids in very large HDL and triglycerides in very large HDL, while an increase in
4451 these metabolites was associated with an increase in all three endometrial cancer outcomes.
4452 The remaining metabolite, valine, was found to be increased by all three adiposity measures,
4453 however increased valine was associated with a decreased risk of all three endometrial
4454 cancer outcomes. As such, triglycerides in small VLDL and triglycerides in very small VLDL
4455 were taken forward for MVMR analysis.

Table 6.2: Two-sample Mendelian randomization: Effect of adiposity measures on triglycerides in small and very small VLDL

		Effect	Lower CI	Upper CI	p-value
BMI	Triglycerides in small VLDL (mmol/l)	0.07	0.00	0.14	0.06
WHR	Triglycerides in small VLDL (mmol/l)	0.56	0.26	0.86	0.00
BF	Triglycerides in small VLDL (mmol/l)	0.11	-0.01	0.23	0.08
BMI	Triglycerides in very small VLDL (mmol/l)	0.06	0.00	0.13	0.06
WHR	Triglycerides in very small VLDL (mmol/l)	0.45	0.20	0.70	0.00
BF	Triglycerides in very small VLDL (mmol/l)	0.05	-0.12	0.23	0.55

Effect estimates are given as SD unit increase in metabolite per SD unit increase in adiposity measure. Metabolite labels are presented with their originally measured units

4456 The effect of both triglycerides in small (OR per SD unit increase in metabolite = 1.79;
4457 95% CI = 1.43–2.15; p-value = 0) and very small VLDL (OR = 1.56; 95% CI = 1.23–1.89;
4458 p-value = 0.01) was strongest for non-endometrioid cancer. Both metabolites had positive
4459 effects on endometrioid (OR small VLDL = 1.11; OR very small VLDL = 1.11) and overall
4460 endometrial cancer (OR small VLDL = 1.11; OR very small VLDL = 1.11) with highly similar
4461 effect sizes. However, effects were smaller than for non-endometrioid cancer and confidence
4462 intervals overlapped the null (Table 6.3).

Table 6.3: Two-sample Mendelian randomization: Effect of triglycerides in small and very small VLDL on endometrial cancer

		OR	Lower CI	Upper CI	p-value
Triglycerides in small VLDL (mmol/l)	Endometrial cancer	1.11	0.99	1.23	0.10
Triglycerides in small VLDL (mmol/l)	Endometrioid	1.11	0.96	1.25	0.16
Triglycerides in small VLDL (mmol/l)	Non-endometrioid	1.79	1.43	2.15	0.00
Triglycerides in very small VLDL (mmol/l)	Endometrial cancer	1.11	0.99	1.22	0.09
Triglycerides in very small VLDL (mmol/l)	Endometrioid	1.11	0.97	1.24	0.15
Triglycerides in very small VLDL (mmol/l)	Non-endometrioid	1.56	1.23	1.89	0.01

Effect estimates are given as the odds (OR) of endometrial cancer per SD unit increase in metabolite. Metabolite labels are presented with their originally measured units

4463 In MVMR analysis, the effect of BMI on overall endometrial cancer was highly similar after
 4464 adjustment for both metabolites: OR per SD unit increase in BMI adjusted for triglycerides in
 4465 small VLDL = 1.9; 95% CI = 1.71–2.08; p-value = 5.18×10^{-9} ; OR per SD unit increase in
 4466 BMI adjusted for triglycerides very small VLDL = 1.87; 95% CI = 1.67–2.07; p-value = 7.26×10^{-8} . These results were highly similar to the unadjusted (i.e. two-sample MR) effect of
 4467 BMI on endometrial cancer (OR per SD unit increase in BMI = 1.91; 95% CI = 1.75–2.07;
 4468 p-value = 7.47×10^{-15}). A similar pattern of association was found across endometrioid
 4469 and non-endometrioid cancer, with highly similar effect sizes and confidence intervals after
 4470 adjustment for each metabolite and when comparing effect size and confidence intervals
 4471 with the unadjusted effects (Figure 6.6). When looking at the effect of metabolites on all
 4472 endometrial cancer outcomes after adjustment for BMI, there was little difference with effect
 4473 size in comparison to the unadjusted effect, however confidence intervals were generally
 4474 wider in the adjusted estimates. For non-endometrioid cancer, however, the adjusted effect
 4475 size was larger than the unadjusted effect size for both metabolites, but confidence intervals
 4476 overlapped. Given the similarity between unadjusted and adjusted effects, it is unlikely
 4477 that either metabolite has a role in mediating the effects of BMI on overall endometrial,
 4478 endometrioid, and non-endometrioid cancer.

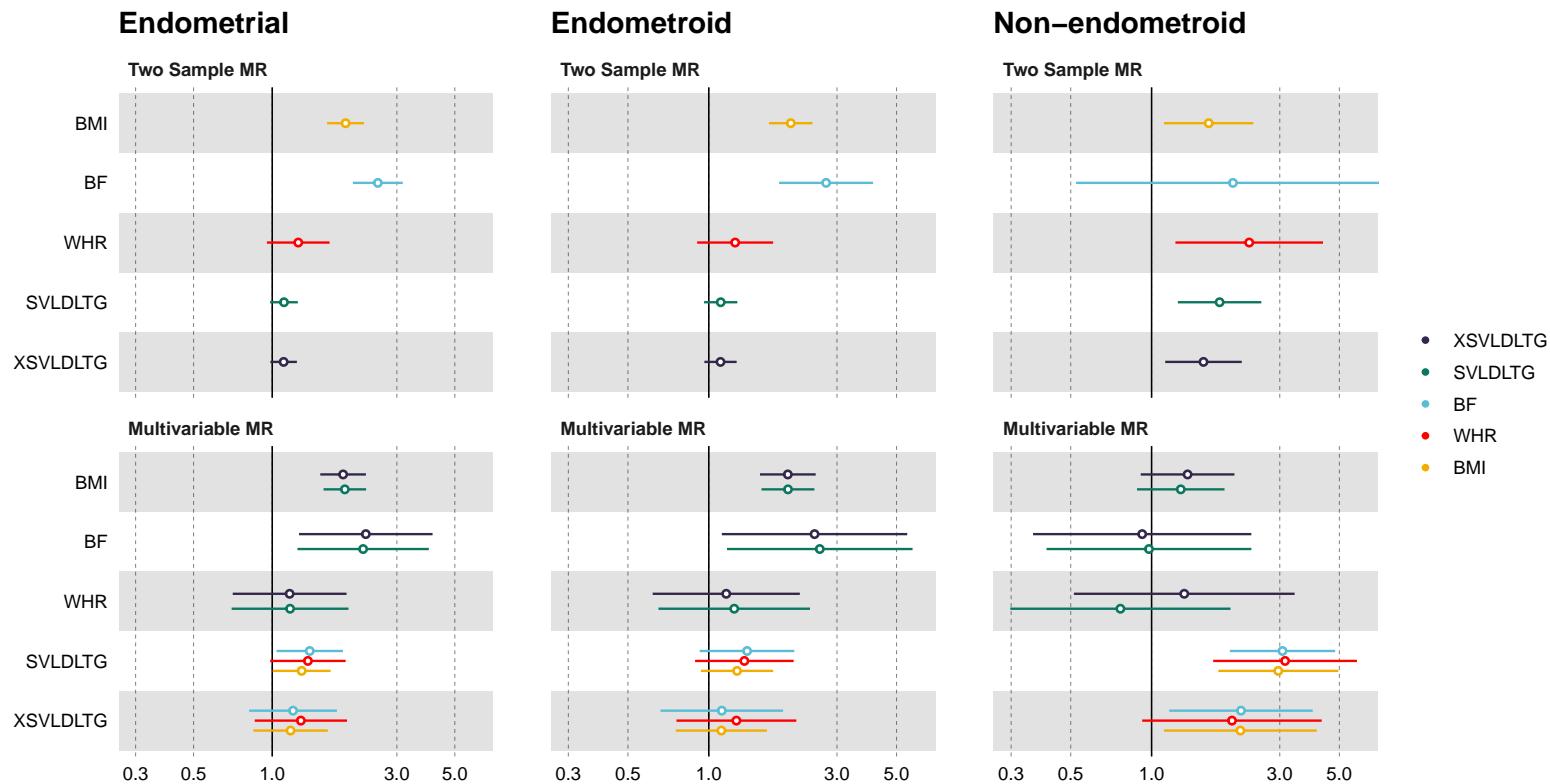


Figure 6.6: Multivariable MR: Effect of adiposity and metabolites on endometrial cancer.
 Forest plot shows odds ratio and 95% confidence interval for: the two-sample MR analyses described previously and the two-sample multivariable MR results for the effect of adiposity measures adjusted for metabolites and metabolites adjusted for adiposity measures on endometrial (left), endometrioid (middle) and non-endometrioid cancer (right). Points and confidence intervals are coloured for the variable they represent in the two-sample MR and for the adjusted variable in the multivariable MR, e.g. BMI is coloured orange and any metabolite effect adjusted for BMI is coloured orange also. An inverse variance weighted model was used. BMI = body mass index; BF = body fat percentage; WHR = waist hip ratio; SVLDLTG = triglycerides in small VLDL; XSVLDLTG = triglycerides in very small VLDL.

4480 For the effect of BF and WHR on endometrial and endometrioid cancer, a similar pattern
4481 of association to that of BMI was seen, with effect sizes and confidence intervals highly similar
4482 after adjustment with both metabolites. For example, the effect of WHR on endometrial cancer
4483 adjusted for triglycerides in small VLDL was 1.17 (95% CI = 0.66–1.69; p-value = 0.5547937)
4484 whereas after adjustment for triglycerides in very small VLDL the OR was 1.17 (95% CI =
4485 0.66–1.67; p-value = 0.5546007). For both BF and WHR, the adjusted effects had much
4486 wider confidence intervals than the unadjusted effects; for BF, adjustment for either metabolite
4487 resulted in confidence intervals which crossed the null. For WHR, confidence intervals
4488 crossed the null in unadjusted and adjusted analyses. When looking at the metabolites effects
4489 on endometrial and endometrioid cancer after adjustment for BF or WHR, there is a similar
4490 pattern of association to when adjusting for BMI, with similar sized effect sizes and confidence
4491 intervals which overlapped the null. The exception was for the effect of triglycerides in small
4492 VLDL on overall endometrial cancer, which, when adjusted for BF (OR = 1.39; 95% CI =
4493 1.1–1.68; p-value = 0.0168637) did not cross the null unlike in the unadjusted analysis (OR =
4494 1.11; 95% CI = 0.99–1.23; p-value = 0.0967917).

4495 For non-endometrioid cancer, the pattern of association for BF and WHR was different to
4496 that of BMI. Both BF and WHR adjusted for triglycerides in small VLDL resulted in directions
4497 of effect which were opposite to the unadjusted estimates; BF adjusted OR = 0.98 (95% CI =
4498 2.62–3.52; p-value = 0.9604153); BF unadjusted OR = 2.01 (95% CI = 0.66–3.35; p-value =
4499 0.3095126); WHR adjusted OR = 0.77 (95% CI = -0.18–1.71; p-value = 0.5839701); WHR
4500 unadjusted OR = 2.31 (95% CI = 1.68–2.94; p-value = 0.0095173). Confidence intervals
4501 overlapped the null and the unadjusted estimates for BF and WHR. When looking at the effect
4502 of triglycerides in small and very small VLDL on non-endometrioid cancer, adjustment for
4503 BF and WHR led to a greater increased risk of non-endometrioid cancer than was found in
4504 the unadjusted analysis. Confidence intervals for triglycerides in small VLDL were similar in
4505 width, and overlapped, to the unadjusted estimate. Confidence intervals for triglycerides in
4506 very small VLDL were much wider in adjusted analyses compared to unadjusted analysis.

4507 Overall, BMI, WHR, and BF adjusted for either metabolite led to an increased risk of
4508 overall endometrial and endometrioid cancer. While BMI adjusted for either metabolite
4509 increased the risk of non-endometrioid cancer, BF reduced the risk of non-endometrioid
4510 cancer as did WHR when adjusted for triglycerides in small VLDL. The effect of WHR on non-
4511 endometrioid cancer after adjusting for triglycerides in very small VLDL, although positive, was
4512 attenuated compared to the unadjusted effect. Confidence intervals for all adjusted analyses
4513 overlapped with confidence intervals for respective unadjusted analyses. For triglycerides in
4514 small VLDL, effect sizes were broadly consistent across adjusted and unadjusted analyses
4515 with overlapping confidence intervals; the exception being for the effect on non-endometrioid
4516 cancer, where effect sizes increased compared to the unadjusted effect size. A similar picture
4517 is apparent for triglycerides in very small VLDL.

4518 Of particular note is the effect of WHR adjusted for both metabolites and both metabolites
4519 adjusted for WHR on non-endometrioid cancer; WHR and both metabolites were associated
4520 with non-endometrioid cancer in the unadjusted analysis, however, after adjusting WHR for

4521 either metabolite, the effect of WHR on non-endometrioid cancer attenuated to the null while
 4522 the effect of the metabolites adjusted for WHR remained consistent with the unadjusted
 4523 analysis. A similar picture is seen for the effect of BF adjusted for both metabolites and
 4524 both metabolites adjusted for BF on non-endometrioid cancer; the effect of BMI adjusted for
 4525 both metabolites shows a smaller attenuation to the null. These results would suggest the
 4526 effect of WHR (and to a lesser extent BF) on non-endometrioid cancer is either mediated by
 4527 triglycerides in small and very small VLDL or there is considerable pleiotropy. A table of the
 4528 full results is available on [GitHub](#).

4529 **Sensitivity analysis**

4530 Instrument strength, estimated using the F-statistic, can give an estimate of weak instru-
 4531 ment bias in MR analyses. In general, an F-statistic of ≥ 10 is considered sufficiently strong
 4532 that there is not substantial weak instrument bias^{578,579}, in a single sample setting, the bias
 4533 is towards the observational estimate, whereas in a two-sample setting, the bias is towards
 4534 the null. Weak instrument bias was estimated using a generalized version of Cochran's Q⁵⁷⁶.
 4535 F-statistics were ≥ 10 for all but triglycerides in small VLDL adjusted for BMI ($F = 9$) and
 4536 WHR ($F = 7$) (Table 6.4). F-statistics in MVMR analysis were lower than those associated
 4537 with the instruments used in the two-sample MR analysis, for example BMI in two-sample MR
 4538 had an F-statistic of 66 while after adjustment for both metabolites this dropped to 54.

Table 6.4: F-statistics for adjusted exposures used in multivariable MR analyses

	BMI	WHR	BF	SVLDLTG	XSVLDLTG
SVLDLTG	54	17	12		
XSVLDLTG	54	21	17		
BMI				9	14
WHR				7	35
BF				14	23
Two-sample MR F	66	49	47	43	53

F-statistics are presented for each exposure (column) and the ad-
 justed (row) variable, e.g. BMI adjusted for SVLDLTG (column
 1 row 1) = 54. The mean F-statistic for instruments used in the
 two-sample MR are also presented for each exposure (column).
 BMI = body mass index; WHR = waist hip ratio; BF = body fat per-
 centage; SVLDLTG = Triglycerides in small VLDL; XSVLDLTG =
 Triglycerides in very small VLDL.

4539 Horizontal pleiotropy, whereby the effect of the instrument on the outcome is not exclusively
 4540 through the exposure, was estimated using a modified form of Cochran's Q⁵⁷⁶. Of the 18 tests,

4541 11 showed evidence of horizontal pleiotropy (p-value < 0.05). All 6 analyses investigating
 4542 associations with non-endometrioid cancer, and one analysis on endometrial cancer showed
 4543 weak evidence of pleiotropy (p-value > 0.05). On the whole, evidence of pleiotropy was
 4544 strongest when instrumenting triglycerides in very small VLDL (Table 6.5). These results
 4545 highlight the potential violation of the independence and exclusion restriction assumptions for
 4546 analyses investigating adjusted effects on overall endometrial and endometrioid cancer.

Table 6.5: Multivariable MR Q statistics

Instrument	Outcome	Q	P-value
WHR & XSVLDLTG	Endometrioid	56	0.0001
BMI & XSVLDLTG	Endometrioid	105	0.0007
WHR & XSVLDLTG	Endometrial cancer	49	0.0008
BMI & XSVLDLTG	Endometrial cancer	104	0.0010
BF & XSVLDLTG	Endometrioid	31	0.0012
WHR & SVLDLTG	Endometrioid	49	0.0014
BMI & SVLDLTG	Endometrioid	102	0.0037
BF & SVLDLTG	Endometrioid	29	0.0069
WHR & SVLDLTG	Endometrial cancer	42	0.0086
BMI & SVLDLTG	Endometrial cancer	97	0.0104
BF & XSVLDLTG	Endometrial cancer	23	0.0165
BF & SVLDLTG	Endometrial cancer	21	0.0732
WHR & XSVLDLTG	Non-endometrioid	21	0.5092
BF & XSVLDLTG	Non-endometrioid	7	0.7628
WHR & SVLDLTG	Non-endometrioid	17	0.8080
BMI & XSVLDLTG	Non-endometrioid	50	0.8799
BF & SVLDLTG	Non-endometrioid	6	0.9337
BMI & SVLDLTG	Non-endometrioid	47	0.9714

Table is ordered by smallest to largest p-value. BMI = body mass index; WHR = waist hip ratio; BF = body fat percentage; SVLDLTG = Triglycerides in small VLDL; XSVLDLTG = Triglycerides in very small VLDL; Q = Cochran's.

4547 6.4 Discussion

4548 In this Chapter, the potential intermediary effect of metabolites in the association between
 4549 adiposity and endometrial cancer was investigated. Two metabolites, triglycerides in small
 4550 and very small VLDL, with evidence for a causal relationship with adiposity measures and
 4551 endometrial cancer, were included in MVMR analyses. There was evidence of for an intermediate
 4552 effect of triglycerides in small VLDL and triglycerides in very small VLDL on the effect

4553 of WHR on non-endometrioid cancer. There was weaker evidence for a similar intermediate
4554 effect of triglycerides in small and very small VLDL on the effect of BF on non-endometrioid
4555 cancer. Whether the effect of WHR (and BF) on non-endometrioid cancer was mediated by
4556 triglycerides in small and very small VLDL was unclear; sensitivity analysis did not indicate
4557 that effects were due to horizontal pleiotropy, however weak instruments may have biased the
4558 results for the effect of triglycerides in small VLDL adjusted for WHR on non-endometrioid
4559 cancer towards the null. There was weak evidence for an intermediate effect of both metabo-
4560 lites on endometrioid and overall endometrial cancer across adiposity measures. Results
4561 here also provide strong evidence for a causal effect of BMI on endometrial cancer (OR per
4562 normalized SD unit increase in BMI = 1.91; 95% CI = 1.75–2.07; p-value = 1.74×10^{-15} ; (Table
4563 6.1). This was consistent with previous MR results^{431,447,457} presented as part of Chapter 2
4564 (OR = 1.57; 95% CI = 1.11 – 2.22; p-value = 0.01) and observational results reported in the
4565 literature by Bhaskaran et al. (2104; OR = 1.62; 95% CI = 1.56 – 1.69; p-value = 1×10^{-4})¹⁰⁸.

4566 Previous observational work has found increased triglycerides to be associated with an
4567 increased risk of endometrial cancer⁵⁸⁰. The effect increased in a dose response manner
4568 and persisted after adjusting for BMI and other potential confounders. Similar results have
4569 been found elsewhere^{581,582}, including associations between increasing triglycerides and
4570 advancing tumour stage⁵⁸³. However, results elsewhere have also shown weak evidence
4571 for an association between triglycerides and endometrial cancer after adjusting for BMI⁵⁸⁴.
4572 There is some evidence to suggest individual triglyceride metabolites, such as myristic
4573 acid (a component of the triglyceride trimyristin⁵⁸⁵), are associated with endometrial cancer
4574 development⁵⁸⁰. However, here are few studies focussing solely on the effects of specific
4575 triglycerides on endometrial cancer, as those discussed focussed primarily on triglycerides
4576 as a class rather than specific lipoprotein compositions. There is evidence that metabolites
4577 within a class may not follow the pattern of association as the class itself⁵⁸⁶, e.g. VLDL may
4578 be associated with a decreased risk of endometrial cancer, but triglycerides in small VLDL is
4579 associated with an increase in endometrial cancer risk. Additionally, it is reasonable to expect
4580 that the properties of specific metabolites are exhibited differently in different tissues and cells
4581 around the body^{587–590}. It is therefore hard to compare results here with those from previous
4582 analyses. Results here suggest there is an increasing effect of triglycerides on endometrial
4583 cancer risk, but that this is likely to be specific to non-endometrioid cancer, and specific to the
4584 alignment of triglycerides with small and very small VLDL, which has not been reported in the
4585 literature. Instead, associations have focussed on overall endometrial cancer risk.

4586 Alongside triglycerides in small and very small VLDL, 3 other metabolites showed evi-
4587 dence of association with endometrial cancer. Increased phospholipids in very large HDL
4588 and triglycerides in very large HDL were associated with an increased risk of endometrial,
4589 endometrioid, and non-endometrioid cancer. There is some evidence for an increasing effect
4590 of phospholipids and triglycerides on endometrial cancer risk⁵⁹¹ but whether this is specific to
4591 very large HDL composition is unclear as studies have not focussed on individual metabolites.
4592 Valine was also associated with endometrial, endometrioid, and non-endometrioid cancer risk.
4593 Previous studies have reported increased valine to be associated with endometrial cancer
4594 risk⁵⁹¹. In two-sample MR analysis here, BMI and WHR were associated with an increase

4595 in valine, which has been proposed as a potential intermediate in the association between
4596 adiposity and endometrial cancer⁵⁶⁴. However, in two-sample MR analysis here, valine was
4597 associated with a decreased risk of endometrial, endometrioid, and non-endometrioid cancer.
4598 This protective effect could be due to metabolic reprogramming of cancer cells which can
4599 utilise branched-chain amino acids, such as valine, as alternative fuel sources^{592,593}.

4600 When comparing two-sample MR (unadjusted) and two-sample MVMR (adjusted) effect
4601 estimates for the association of adiposity measures, there was little change in effect size
4602 and confidence intervals overlapped for effects on overall and endometrioid cancer risk. For
4603 non-endometrioid cancer risk, the effect of all three adiposity measures, for which confidence
4604 intervals did not cross the null in the unadjusted analysis (except for BF), crossed the null in
4605 adjusted analyses. This may be reflective of reduced power given the smaller number of cases
4606 for non-endometrioid cancer (N = 1,230) as opposed to endometrioid (N = 8,758) and overall
4607 endometrial cancer (N = 12,906). For metabolites, when comparing the unadjusted and
4608 adjusted effect estimates, effect sizes and confidence intervals were generally larger in the
4609 adjusted analysis for all endometrial cancer outcomes. Although in some instances, MVMR
4610 can increase power⁵⁷⁹, MVMR is generally considered to be less powered than two-sample
4611 MR. The observed decreased precision and larger effect sizes in adjusted analyses may be a
4612 result of weak instrument bias, especially for triglycerides in small VLDL adjusted for BMI and
4613 WHR. But it may also reflect a true causal association or a highly pleiotropic environment,
4614 especially for analyses of the effect on endometrioid and endometrial cancer, where evidence
4615 of horizontal pleiotropy was likely high.

4616 Previous observational and MR work has found similar effects to those found here.
4617 In one study, observational and MR results indicated an association between BMI and
4618 endometrial cancer but not WHR and endometrial cancer⁴⁴⁷. Additional studies using MR
4619 have found consistent effects^{431,457}. There is further observational evidence for associations
4620 between BMI and endometrial cancer, as well as WHR, waist circumference, and waist to
4621 height ratio⁹⁵. Observational investigations have highlighted links between BMI and both
4622 endometrioid and non-endometrioid^{569,594–599}, but there has been a lack of genetic data
4623 on both subtypes in large, well-powered studies to enable MR investigations. Additionally,
4624 studies have predominantly focussed on BMI with some focus on measures of deposition
4625 such as waist circumference (WC) and WHR, but as yet very little work using an absolute
4626 measure of adiposity such as BF. Here, investigation of both subtypes and multiple measures
4627 of adiposity was possible. MR results were consistent with those found in the aforementioned
4628 observational studies. This consistency was similar in the strength of effects, with a stronger
4629 effect observed for endometrioid cancer compared to non-endometrioid (though confidence
4630 intervals overlapped). Of note is the association found for WHR and non-endometrioid cancer,
4631 which has not previously been reported.

4632 A prominent strength of this work was the breadth of metabolomics data used here
4633 and in previous Chapters. Metabolites used in MVMR analyses here, were selected based
4634 on evidence from prior observational (Chapter 4) and MR analyses (Chapter 5) for causal
4635 relationships with measures of adiposity. In Chapter 4, the metabolic profile of BMI, WHR, and

4636 BF was investigated using the Avon Longitudinal Study of Parents and Children (ALSPAC).
4637 The effect of adiposity measures on numerous metabolites was robust after accounting for a
4638 number of potential confounders such as smoking status, diet, and physical activity level. In
4639 Chapter 5, the metabolic profile of BMI, WHR, and BF was investigated in a causal framework.
4640 Effects identified in the previous observational analysis, as well as in previously published
4641 studies, was replicated in two independent studies. Finally, in this Chapter, replication of
4642 the causal effects of adiposity on the metabolic profile was performed using the largest
4643 independent data set available, UK Biobank. Additionally, replication of the causal effect of
4644 BMI on endometrial cancer, as obtained from a meta-analysis in Chapter 2, was performed.

4645 **6.4.1 Instrumentation**

4646 Instrumentation, described briefly for measures of adiposity in Chapter 5, is primarily
4647 achieved by using either a single genetic instrumental variable or multiple genetic instrumental
4648 variables. The choice of instruments for measures of adiposity is generally to use the
4649 largest available GWAS and a genome wide p-value threshold of $p\text{-value} \leq 5 \times 10^{-8}$. This
4650 approach was utilised in the Chapter 5, however instruments for BMI and WHR were obtained
4651 from GWAS which included UK Biobank. UK Biobank is a unique resource with wide
4652 application, one of which is as a component of the most recent and largest metabolomics
4653 GWAS (unpublished) and endometrial cancer GWAS⁴⁶⁷, both used here. Overlap between
4654 the exposure and outcome in a two-sample MR can bias estimates towards the observational
4655 estimate in a manner proportional to the size of the overlap⁴¹⁸ (See limitations section below).
4656 As such, instruments for adiposity measures were obtained from GWAS which did not include
4657 UK Biobank. In Chapter 5, BMI^{36,41} and WHR^{37,42} were instrumented using the largest
4658 available GWAS (which included UK Biobank) and the second largest GWAS which was
4659 used in analyses in this Chapter, results were highly correlated. As endometrial cancer
4660 is a sex-specific condition, this work is limited by the use of sex combined instruments for
4661 adiposity, though this does have utility in identifying the general effects of adiposity, especially
4662 for metabolites. Instrumentation practices for metabolites are less well established, but have
4663 largely followed a similar principle to that described above. Here, a simple instrumentation
4664 approach was used for metabolites ($p\text{-value} \leq 5 \times 10^{-8}$ and LD $R^2 \leq 0.0001$ thresholds). This
4665 approach could have been strengthened by collapsing metabolites into broader categories³⁹²
4666 or by using composite instruments⁶⁰⁰.

4667 **6.4.2 Limitations**

4668 In MR, the three main assumptions (SNPs associate with the risk factor of interest,
4669 SNPs share no common cause with the outcome, and SNPs do not affect the outcome
4670 except through the risk factor) need to be satisfied in order to obtain robust causal estimates.
4671 Adiposity instruments used here have been used widely and are robustly associated with

4672 their respective phenotypic measures^{36,37,39}. For metabolites, previous work has used a
4673 standard approach of identifying SNPs reaching a genome wide p-value threshold (p-value
4674 $\leq 5 \times 10^{-8}$), which was employed here alongside a stringent LD clumping threshold ($R^2 =$
4675 0.001 and 10,000 base window). However, the strength of these instruments in two-sample
4676 MVMR was considerably weaker than in two-sample MR, and for triglycerides in small VLDL
4677 this was below the nominal threshold of 10. It is likely then that weak instrument bias⁵⁷⁸ is a
4678 present feature of the MVMR results and is likely to in part explain the decreased precision
4679 in confidence intervals. This reduced precision may for example explain why triglycerides in
4680 very small VLDL adjusted for WHR is no longer associated with non-endometrioid cancer in
4681 MVMR versus two-sample MR. In addition to this, instruments reflect a life time exposure and
4682 do not take into account change due to lifestyle for instance.

4683 It is possible that SNPs associated with metabolites do not conform to the third IV
4684 assumption. That is, SNPs associated with metabolites have an effect on the outcome which
4685 is not directly through the metabolites. This is of particular concern with metabolites given
4686 they are highly inter-correlated and have many have a shared genetic architecture. In this
4687 sense, the effect of a metabolite on an outcome could be through a highly correlated second
4688 metabolite. Of the 56 and 50 SNPs associated with triglycerides in small and very small
4689 VLDL respectively (90 unique SNPs), 16 were shared between them (18%). It is likely both
4690 metabolites share SNPs with many other metabolites given that just over half of the 934
4691 identified SNPs for 53 metabolites were associated with just one metabolite and the average
4692 number of SNPs associated with a metabolite was 49.

4693 In two-sample MR, sensitivity methods (MR-Egger, weighted median, and weighted mode)
4694 showed consistent directions of effect for both metabolites with the IVW-MRE method, though
4695 the majority of confidence intervals crossed the null. The exception to the consistent direction
4696 was triglycerides in small VLDL on endometrioid cancer in MR-Egger analysis which, although
4697 the confidence intervals crossed the null, was opposite to the directions of effect of other
4698 methods. In MVMR, sensitivity methods such as these have not been developed. Instead,
4699 unbalanced horizontal pleiotropy was evaluated using a modified form of Cochran's Q³⁴⁵,
4700 which found evidence (p-value < 0.05) of unbalanced horizontal pleiotropy for a majority of
4701 tests and may therefore bias estimates.

4702 An additional implication here is the bias introduced through sample overlap. As dis-
4703 cussed in the instrumentation section above, sample overlap can bias estimates towards
4704 the observational confounded estimate in a manner proportional to the overlap. In order to
4705 minimise overlap, and knowing that metabolite and outcome data included individuals from UK
4706 Biobank, exposure instruments were sought from GWAS which did not include UK Biobank.
4707 Overlap was therefore limited to the metabolite data endometrial cancer data, specifically the
4708 overall endometrial cancer data. For binary outcomes, this bias due to overlap is specific to
4709 the overlap in individuals in the exposure GWAS and those who are cases in the outcome
4710 GWAS. In this regard, the number of UK Biobank participants included in the case arm is
4711 at most 5% of the total cases. As such, the relative bias in analyses using metabolite and
4712 overall endometrial cancer data is at most 0.5%.

4713 A key limitation of this work is the use of non-sex-specific instruments for measures
4714 of adiposity. A major driver of endometrial cancer, specifically endometrioid cancer, is
4715 oestrogen⁵⁶⁵, which, as a sex hormone, is important in endometrial cell growth. Oestrogen is
4716 also involved in the distribution and accumulation of adipose tissue⁶⁵. Importantly, adipose
4717 tissue is a key source of oestrogen synthesis, particularly post-menopause⁶⁰¹. There is sexual
4718 dimorphism in the accumulation and distribution of adipose tissue that has been highlighted
4719 in genetic studies, which found sex-specific SNPs associated with adiposity^{36,37,41-43}. There
4720 is also evidence of sex differences in adiposity metabolite associations³⁶⁷. The use of non-
4721 sex-specific adiposity instruments may therefore have biased estimates. This is because,
4722 in a two-sample MR the assumption is that the underlying population of the two samples
4723 is the same. Violation of this assumption and therefore of the second IV assumption (that
4724 instruments do not share a common cause with the outcome) can lead to biased estimates
4725 for example from genetic confounding and population structure.

4726 In this Chapter, NMR-derived metabolites were used, as with work in Chapter 4 and 5.
4727 Although a relatively large number of metabolites were investigated, they are predominantly
4728 lipid based. This leaves a broad array of metabolites that have not been investigated. Mass
4729 spectrometry (MS) has investigated many hundreds of metabolites⁶⁰², but unlike NMR derived
4730 metabolites (see Wurtz et al. (2014)¹⁵⁷), few studies have focused exclusively on the metabolic
4731 profile of adiposity using MS-derived measurements. Those that have^{158,386,514,515}, have
4732 either focussed on case control analyses⁵¹⁴, on children⁵¹⁵, or included a small number of
4733 individuals with repeat measures¹⁵⁸. The study by Ho et al. (2016)³⁸⁶ focussed solely on
4734 the metabolic profile of adiposity in 2,383 adults and found 69 out of 217 metabolite to be
4735 associated with increased BMI, including triglycerides and a number of other associations (e.g.,
4736 amino acids) identified here. A number of studies have investigated associations between
4737 MS-derived metabolites and endometrial cancer. However, these studies have included fewer
4738 than 100 cases and have focussed on post-menopausal⁶⁰³ or obese women⁶⁰⁴ only. In
4739 the latter case, phospholipids were identified as a biomarker of significance. The study by
4740 Audet-Delage et al. (2018)⁶⁰³, which focussed on post-menopausal women, identified lipid
4741 pathways to be most affected by case status. Although numerous SNPs have been associated
4742 with MS-derived metabolites^{317,321,605}, to our knowledge no studies have employed MR to
4743 investigate metabolic associations with adiposity measures or endometrial cancer.

4744 6.4.3 Conclusion

4745 There were few metabolites that showed evidence of a causal relationship with adiposity
4746 measures and were causal for endometrial cancer. Of the few metabolites with evidence
4747 for a causal relationship with adiposity and endometrial cancer, there was evidence that
4748 triglycerides in small and very small VLDL may play an intermediary role between WHR and
4749 non-endometrioid cancer. There was some suggestive evidence for a similar intermediary role
4750 between BMI and BF and non-endometrioid cancer. Non-endometrioid cancer is the more
4751 aggressive of the two common subtypes of endometrial cancer, and is typically considered

4752 oestrogen independent but recent work has shown this is not necessarily the case. There
4753 was weak evidence of horizontal pleiotropy in MVMR analyses investigating effects on non-
4754 endometrioid cancer compared to endometrioid and overall endometrial cancer. However,
4755 this does not rule out the possibility of pleiotropic effects, especially as metabolites are
4756 highly intercorrelated and have a shared genetic architecture. Future work to elucidate
4757 the intermediary relationship of triglycerides in small and very small VLDL should have
4758 two focusses: (i) specific metabolites within the VLDL triglyceride subclass (ii) and wet lab
4759 studies, which, when informed by MR analyses, have been fruitful in identifying molecular
4760 mechanisms⁶⁰⁶.

4761 **Chapter 7**

4762 **Discussion**

4763 **Chapter summary**

4764 This thesis has focussed on the role of adiposity in disease development, with a specific
4765 focus on the potential intermediary role metabolites play in this relationship. At the end of
4766 each of Chapters 2, 4, 5, and 6, the main results, and strengths and limitations of each
4767 analysis are described. In this final discussion Chapter, a discussion and summary of the key
4768 findings and implications of this thesis as a whole are presented alongside suggestions for
4769 future directions that may aid in untangling the mechanisms leading from adiposity to disease.

4770 **7.1 Overview**

4771 Both observational and Mendelian randomization (MR) analyses provide evidence for
4772 an association between adiposity and many health outcomes and diseases. However, with
4773 the many limitations of singular adiposity measures (e.g., body mass index (BMI) is unable
4774 to differentiate lean and fat mass⁶⁴), complimentary assessments of body composition are
4775 recommended^{83,88}. This allows for a more detailed interrogation of potential effects as
4776 each measure provides a different assessment: BMI provides an overall assessment of
4777 body composition, body fat percentage (BF) provides an estimate of the specific make-up
4778 of the body, and waist hip ratio (WHR) provides an estimate of where fat mass is stored.
4779 Understanding how adiposity is associated with different outcomes can therefore help to
4780 establish a more accurate estimate of the impact of adiposity on health outcomes.

4781 The underlying mechanisms for many of the relationships between adiposity and
4782 diseases are not well understood. There is some evidence that the physical burden of

4783 fat mass is directly related to the development of some diseases. For many diseases,
4784 evidence points to pathway changes as the potential underlying mechanism of disease
4785 development^{158,160,162–165,175,176}. Many studies have highlighted associations between
4786 adiposity and changes to the metabolome^{157,158,328,363–373}, as well as associations between
4787 the metabolome and diseases such as type 2 diabetes³⁷⁴, coronary heart disease³⁷⁵,
4788 depression³⁷⁶, hypertension³⁷⁷, and more^{1,188,378–385}. However, whether metabolites explain
4789 the mechanisms of disease development has not been explored in depth. This thesis set out
4790 to identify metabolites that sit on the causal pathway from adiposity to disease.

4791 **7.1.1 Chapter 2: Adiposity is causally associated with numerous risk factors
4792 and diseases**

4793 The adverse effects of adiposity have been reviewed extensively in the observational
4794 literature^{61,227,607–613}. However, observational studies have a number of limitations, not least
4795 the difficulty in obtaining a causal estimate^{145,301,302,333–335}. These limitations, including
4796 confounding and reverse causation, can lead to biased results^{145,333–335}. MR, a form of
4797 instrumental variable analysis, is able to mitigate the effects of many of these limitations due
4798 to the fact that genetic variants, used as instrumental variables, are obtained randomly at
4799 conception, and are thus temporally prior to any subsequent outcome³³⁷. Although MR has its
4800 own set of assumptions and limitations it has shown promise in untangling causal effects^{145,334}.
4801 Recently, Riaz et al., (2018)⁴⁰⁷ attempted to gain an overview of the causal effects of adiposity
4802 on cardiovascular outcomes. They found consistent results with observational estimates, but
4803 did not perform quality assessment of the included studies and inappropriately assessed
4804 the independence and exclusion restriction assumptions (See Chapter 1 Section 1.8) as a
4805 combined assumption⁶¹⁴. There has not been a comprehensive investigation of the causal
4806 effect of adiposity.

4807 To gain a comprehensive overview of the causal effect of adiposity, a systematic review
4808 and meta-analysis was performed and presented in Chapter 2. A total of 173 studies met pre-
4809 published criteria for inclusion (see PROSPERO), with 34 studies included in meta-analyses
4810 of 31 adiposity-outcome pairs. Results from the meta-analyses, which were consistent
4811 with published observational studies, identified positive causal effects of adiposity on many
4812 cancers, cardiovascular traits, type 2 diabetes, and depression. A narrative synthesis of the
4813 studies not included in the meta-analyses revealed a literature that is broadly consistent with
4814 observational findings, showing that the causal effects of adiposity are wide reaching. Of note
4815 were results for endometrial cancer; in meta-analysis, higher BMI was associated with an
4816 increased risk of endometrial cancer (OR = 1.57; 95% CI = 1.11 – 2.22), while in the narrative
4817 synthesis there was inconsistent evidence, with some studies pointing to a decreasing effect
4818 and others showing little evidence of an effect in either direction. A similar picture was
4819 observed for colorectal cancer. This may be reflective of the differential association with
4820 subtypes.

4821 Although there was some conflicting evidence in the narrative synthesis for the different
4822 exposures, there was much evidence supporting the causal impact of adiposity on many
4823 diseases. There is a requirement to understand the mechanisms by which adiposity impacts
4824 these outcomes. Given the hypothesis of metabolic perturbations of adiposity, the thesis then
4825 focused on the causal role of adiposity on circulating metabolites, which was then expanded
4826 to understand the role of adiposity-related metabolites in endometrial cancer as an exemplar
4827 outcome.

4828 **7.1.2 Chapter 4: Adiposity is associated with individual and whole subclasses
4829 of metabolites across time points**

4830 There is considerable evidence that adiposity is observationally associated with a wide
4831 range of metabolites^{157,158,328,363–366,368–372}. However, the focus of many studies has been
4832 on BMI as a marker for overall adiposity. Additionally, few studies have looked at whether
4833 associations persist over time. To gain a greater understanding of the association between
4834 different measures of adiposity and metabolites, and to strengthen evidence from MR analyses
4835 to be conducted, linear regression analyses using data from the Avon Longitudinal Study of
4836 Parents and Children (ALSPAC) were conducted and presented in Chapter 4.

4837 Data on adiposity measures and up-to 230 predominantly lipid nuclear magnetic reso-
4838 nance (NMR) derived metabolites were available for children (N = 5,656; mean age = 7.56;
4839 SD = 0.36), adolescents (N = 4,489; mean age = 16.06; SD = 1.11), young adults (N =
4840 3,269; mean age = 24.03; SD = 0.85), and adults (N = 6,406; mean age = 49.53; SD = 5.32).
4841 Three models were used to estimate the effect of each adiposity measure on each metabolite:
4842 model 1 adjusted for age and sex; model 2 = model 1 + mothers/own education, smoking
4843 status, alcohol consumption, and diet; model 3 = model 2 + physical activity. Model 1 and 2
4844 were run on individuals with data on all confounders except for physical activity as there was
4845 substantially fewer individuals with physical activity data.

4846 The association between all adiposity measures and metabolites was far reaching, with
4847 effects seen across all subclasses of metabolites, and results similar to those previously
4848 reported¹⁵⁷. All three measures of adiposity were associated with a majority of the metabolites
4849 across all time points and all models, the exception being model 3 when investigating the
4850 effect of BMI and BF on metabolites in adolescents which may be due to reduced power
4851 given the smaller sample size available for physical activity. There was broad consistency
4852 in the directions of effect across models at each time point and adiposity measures at each
4853 time point. Generally the effect of adiposity was to increase metabolite concentrations,
4854 the exception being HDL metabolite measures which on the whole were associated with a
4855 decrease. Effect sizes tended to increase across all measures of adiposity and models with
4856 age.

4857 Work in this Chapter highlighted the wide reaching effect adiposity has on metabolites

4858 and shows that effects likely persist over time. Of particular note is the larger effect sizes
4859 and increasing number of associations seen as age increases. Given that many adiposity
4860 associated diseases occur later in life, exposure to an altered metabolic profile over time may
4861 be important in disease development. This is especially true as weight loss in overweight
4862 and obese individuals is associated with a normalizing of metabolite changes³⁷². This is
4863 noteworthy in regards to comparisons with MR analyses as exposures instrumented by
4864 genetic variants reflect a lifetime exposure, and evidence here suggests that adiposity exerts
4865 a greater effect as time progresses. Comparing evidence here with MR analyses strengthens
4866 this evidence for an association between adiposity measures and metabolites as observational
4867 and MR analyses have different assumptions and biases⁵¹⁸.

4868 **7.1.3 Chapter 5: The effect of adiposity on metabolites is consistent across**
4869 **observational and Mendelian randomization analyses**

4870 In order to obtain causal estimates of the effect of adiposity on metabolites identified as
4871 being associated with BMI, WHR, and/or BF in Chapter 4, parallel two-sample MR analyses in
4872 two datasets and subsequent meta-analysis were performed and reported in Chapter 5. Data
4873 on adiposity measures were available for: BMI from up to 795,624 individuals of European
4874 ancestries from GIANT⁴¹, WHR from up to 697,702 individuals of European ancestries from
4875 GIANT⁴², BF from up to 89,297 individuals European ancestries from Lu et al.,³⁹. Metabolite
4876 data were available from two independent studies: 123 NMR derived metabolites measured
4877 in up-to 24,925 individuals of European ancestries from Kettunen et al. (2016)³¹⁸, and 230
4878 NMR derived metabolites measured in up-to 40,905 individuals of European ancestries from
4879 INTERVAL (unpublished). Both metabolite datasets used the same NMR platform as that
4880 used in Chapter 4 and were therefore predominantly lipid metabolites. Meta-analysis of 110
4881 metabolites measured in both datasets and comparison with observational analyses in adults
4882 from Chapter 4 identified 56 associations between adiposity measures and metabolites that
4883 were consistent in their direction of effect across both analyses and which passed various
4884 multiple testing thresholds.

4885 Results highlighted the broad effect of adiposity across the metabolome, with evidence
4886 for an increasing effect on whole subclasses of metabolites such as VLDL (small, very small,
4887 medium, large, and very large), as well as for specific metabolites within the aromatic and
4888 branched chain amino acids subclasses. Negative effects were, similar to observational
4889 analyses, observed for HDL metabolite subclasses (medium, large and very large). The
4890 effects of BMI and WHR on metabolites in MR analyses were highly consistent with obser-
4891 vational analyses. However, where effects for BF were similar to those of BMI and WHR in
4892 observational analyses, effects of BF in MR analyses were generally opposite to BMI and
4893 WHR. Evidence for an association between BF and the metabolic profile was weak; effect
4894 sizes and confidence intervals were much larger than those for BMI and WHR, with the
4895 majority of confidence intervals crossing the null.

4896 Given the highly correlated nature of the exposures, and the consistency in observational
4897 estimates with metabolites, results for BF appear counter-intuitive. These differences may
4898 be due to a true causal negative effect of BF on metabolites, or a result of the difficulty
4899 in instrumenting complex traits such as adiposity. This complexity is reflected in the fact
4900 that 2 of the 7 SNPs used to instrument BF had previously been associated with 'favourable
4901 adiposity'^{35,539}, that is, they are associated with increased fat mass and a favourable metabolic
4902 profile. Removal of these SNPs and re-estimation of the effect of BF on metabolites resulted
4903 in a global tightening of confidence intervals and a number of effect estimates changing
4904 direction to be more consistent with BMI and WHR. A number of these effects subsequently
4905 reached the multiple testing threshold.

4906 Work in this Chapter, in combination with results from Chapter 4, highlights the effects of
4907 adiposity on whole subclasses of metabolites. The use of both MR and observational analyses,
4908 which hold different biases, strengthens evidence for an effect of adiposity on 56 metabolites.
4909 Predominantly, the effect of adiposity was to increase these metabolites, with a negative
4910 effect of adiposity associated with HDL metabolites. A number of these associations were
4911 consistent with previous observational and MR analyses^{157,392} and have been shown to be
4912 associated with adiposity-related diseases such as colorectal³⁹² and endometrial cancer⁵⁶⁴,
4913 meaning that they could play an intermediary role in these relationships.

4914 **7.1.4 Chapter 6: Intermediary metabolites may partly explain the association 4915 between adiposity and diseases identified in Chapter 2**

4916 To investigate whether the adiposity-related metabolites identified in Chapter 4 and 5
4917 played a role in the development of disease, multivariable MR (MVMR) was performed and
4918 reported in Chapter 6. Endometrial cancer was chosen as an exemplar for this analysis as it
4919 met four key requirements: there was strong evidence in Chapter 2 for an effect of one or more
4920 adiposity measures, there is consistent evidence across observational⁵⁶⁵ and MR^{431,447,457}
4921 analyses that adiposity is associated with an increased risk of endometrial cancer, there is a
4922 large and publicly available GWAS on endometrial cancer, and the extent to which circulating
4923 metabolites may play a role in the relationship between adiposity and endometrial cancer had
4924 not been published before and thus adds novelty to the literature.

4925 Data were available for three endometrial cancer outcomes: endometrioid cancer (cases
4926 = 8758; controls = 108,979), non-endometrioid cancer (cases = 1230; controls = 108,979),
4927 and overall endometrial cancer (cases = 12,906; controls = 108,979) which included cases
4928 from endometrioid and non-endometrioid cancer, as well as un-classified endometrial cancer
4929 cases. An independent metabolite dataset to those used in Chapter 4 and 5 was used to
4930 replicate the effect of adiposity on metabolites observed in Chapter 4 and 5 and to estimate
4931 the causal effect of metabolites on endometrial cancer risk. Two metabolites (triglycerides in
4932 small and very small VLDL) had evidence for a causal relationship with adiposity measures
4933 and endometrial cancer and were included in MVMR analyses. There was evidence for an

4934 intermediate effect of both metabolites on the effect of WHR on non-endometrioid cancer, and
4935 weaker evidence for an intermediate effect on the effect of BF on non-endometrioid cancer.
4936 Although sensitivity analysis did not indicate these effects were due to horizontal pleiotropy,
4937 they did indicate metabolite instruments were weak (F-statistic < 10), as a result it is unclear
4938 if these metabolites truly mediate the effect of adiposity on endometrial cancer.

4939 It is difficult to compare these results with previous work given few studies have looked at
4940 the intermediate role of metabolites in the relationship between adiposity and endometrial
4941 cancer, and those that have, have focussed solely on BMI and have used a clinical measure
4942 of triglycerides. These studies have found both an increasing⁵⁸⁰ and decreasing⁵⁸⁴ effect
4943 of triglycerides on endometrial cancer risk after adjustment for BMI. Given that there is
4944 evidence that metabolites within a class may not follow the same pattern of association as
4945 the class itself⁵⁸⁶ it is difficult to compare results for a clinical measure of triglycerides with
4946 the more specific metabolite measures used here. It is also reasonable to expect that the
4947 properties of specific metabolites are exhibited differently in different tissues and cells around
4948 the body⁵⁸⁷⁻⁵⁹⁰. As such, these results suggest that triglycerides are potential intermediates
4949 in the relationship between adiposity and endometrial cancer, but this is likely to be specific to
4950 non-endometrioid cancer, and specific to the alignment of triglycerides with small and very
4951 small VLDL.

4952 **7.2 Strengths and limitations**

4953 Within each chapter, the strengths and limitations of the techniques applied and data used
4954 are discussed. However, there are a number of strengths and limitations that are overarching
4955 and relevant to the interpretation of results. These overarching themes are discussed here in.

4956 **7.2.1 Replication, meta-analysis, and triangulation of evidence**

4957 The major strength of this work is the replication of the effect of adiposity on metabolites
4958 in observational (Chapter 4) and MR analyses (Chapter 5 and 6) across 4 independent
4959 datasets. In observational analyses (Chapter 4), a majority of metabolites tested were found
4960 to be associated with adiposity. In MR analyses (Chapter 5), a majority of the associations
4961 identified in observational analyses were replicated across two independent metabolite
4962 datasets. Subsequent meta-analysis of MR results, and comparison with observational
4963 results, highlighted 56 metabolites to be associated with adiposity. These associations were
4964 subsequently replicated in MR analyses using a much larger independent dataset (Chapter
4965 6).

4966 It was also possible to replicate previous MR analyses investigating the effect of BMI

4967 on endometrial cancer. In Chapter 2, three MR analyses were included in a meta-analysis
4968 of the effect of BMI on endometrial cancer resulting in an OR of 1.57 (95% CI = 1.11 –
4969 2.22). In Chapter 6, MR analysis of the effect of BMI on endometrial cancer resulted in an
4970 OR of 1.91 (95% CI = 1.75 – 2.07), similarly strong effects for BF (OR = 2.54; 95% CI =
4971 2.32 – 2.76) and WHR (OR = 1.26; 95% CI = 0.98 – 1.53) were also found. In addition,
4972 investigation of endometrial cancer subtypes (endometrioid and non-endometrioid) in Chapter
4973 6 provided evidence that BMI and BF increased the risk of endometrioid cancer while BMI
4974 and WHR increased the risk of non-endometrioid cancer. The effect of BMI and BF was
4975 strongest for endometrioid cancer, while the effect of WHR was strongest for non-endometrioid
4976 cancer. These results would suggest that body composition is more important in endometrioid
4977 development and fat deposition is more important in non-endometrioid cancer development.

4978 Replication of results using different datasets and techniques is a major goal of epidemiological
4979 analyses as different assumptions and biases are tested, and where evidence aligns, triangulation strengthens the evidence for a true effect of an exposure on an outcome⁵¹⁸.
4980 Analyses here used multiple independent datasets, this included individual level data across
4981 4 times points from ALSPAC, as well as summary level data from 3 independent metabolite
4982 GWAS and up-to 4 GWAS each for BMI, WHR, and BF. These datasets were analysed using
4983 linear regression analyses, adjusting for a multitude of confounders, and two-sample MR
4984 analyses using a number of methods to test multiple assumptions. As observational and MR
4985 analyses have different assumptions, which were tested extensively and satisfied, there is
4986 strong evidence for a causal relationship between adiposity and 56 metabolites. This is further
4987 strengthened by the fact that consistent results were obtained across multiple independent
4988 metabolite datasets: ALSPAC, Kettunen, INTERVAL, and UK Biobank.
4989

4990 Replication and triangulation of evidence was focussed primarily on the relationship
4991 between adiposity and metabolites. The effect of metabolites on endometrial cancer was not
4992 replicated, nor were MVMR analyses which investigated the intermediate effect of metabolites
4993 on endometrial cancer risk. The primary reason for a lack of replication was data availability
4994 and independence. Although, three independent endometrial cancer GWAS were included in
4995 the meta-analysis of the effect of BMI on endometrial cancer in Chapter 2, two of these were
4996 included in the GWAS used in MR analyses in Chapter 6. As the endometrial cancer GWAS
4997 used in Chapter 6 is an updated meta-analysis of these studies these datasets would not
4998 have been independent if used in replication analyses. Additional considerations include the
4999 fact that there are few publicly available cancer GWAS, limited funding, and time constraints.
5000 That being said, there is opportunity for replication using data from the European Prospective
5001 Investigation into Cancer and Nutrition (EPIC) where there is data from individuals who have
5002 undergone bariatric surgery.

5003 **7.2.2 Data, availability, and presentation**

5004 Data, its availability, and its independence to other datasets was a key component of this
5005 thesis. All data used through the thesis had previously been collected, this included individual
5006 level data from ALSPAC, which was used in Chapter 4, and summary level data for adiposity
5007 measures, metabolites, and endometrial cancer used in Chapter 5 and 6. The majority of
5008 the summary level data was publicly available; metabolite data from INTERVAL was provided
5009 by collaborators at the University of Cambridge (Adam Butterworth), and metabolite data
5010 from UK Biobank was provided by collaborators at the University of Bristol (Carolina Borges).
5011 These collaborations enabled the replication and meta-analysis of the effect of adiposity on
5012 metabolites. Although multiple testing thresholds were employed, without these additional
5013 datasets to perform replication, the chance of false positives is likely to have risen. Data
5014 on endometrial cancer was not as readily available. The lack of GWAS data availability is
5015 not restricted to endometrial cancer⁶¹⁵, but it is a common occurrence that, in comparison
5016 to other traits, there are few publicly available cancer GWAS, and even fewer GWAS of
5017 cancer subtypes? Platforms such as the OpenGWAS^{522,616} and the GWASCatalogue⁶¹⁷ make
5018 it easy for researchers to make their data available to the wider research community and
5019 requirements by funders and publishers, including Plan S (coalition-s.org/), to make data
5020 available should improve the availability of data going forward.

5021 In Chapter 2, BMI was found to be the most commonly used measure of adiposity when
5022 investigating the causal association with outcomes. A main reason for its commonality is
5023 that data on BMI is readily collected as a staple of prospective and cohort studies, and
5024 there is a history of making summary statistics publicly available^{33,36,41,42} which aids further
5025 analysis^{615,618}. BMI however has a number of limitations, not least its inability to differentiate
5026 fat and lean mass⁶⁴. Complimentary assessment of adiposity using a combination of body
5027 composition measures, may be beneficial when investigating associations with disease as
5028 each has different limitations^{83,88} (Chapter 1 Section 1.4). Measures of adiposity used in this
5029 thesis were chosen in order to encompass the key aspects of adiposity, body composition
5030 and fat deposition. BMI was chosen as it is the most commonly used measure and provides
5031 an overall estimate of body composition. BF was chosen as, unlike BMI, it is able to capture
5032 the bodies true composition. WHR was chosen as evidence has pointed to an important role
5033 of fat deposition in the development of many diseases. Importantly, all three measures have
5034 been shown to have different underlying biological pathways⁵⁵, meaning that when looked
5035 at together a deeper understanding of a relationship with an outcome can be achieved. An
5036 additional consideration for choosing WHR and BF, as opposed to other measures that may
5037 be more detailed (e.g., visceral adipose tissue), was the availability of data. Much like BMI,
5038 there is a history of publicly available data for WHR^{37,42,619} and an increasing number of
5039 publicly available BF datasets^{39,43}. There are comparatively fewer well powered and publicly
5040 available GWAS for more detailed body composition measures such as visceral adipose
5041 tissue⁶²⁰, making within exposure replication difficult.

5042 An additional challenge with the use of pre-collected and publicly available data is the

5043 lack of control over the data available and the methods of collection, which are likely to vary
5044 across studies. For instance, in ALSPAC, data on all three measures of adiposity was not
5045 available at all time points, data on WHR was not available for adolescents. In addition, data
5046 on BF was collected via bioelectrical impedance and dual-energy x-ray absorptiometry (DXA)
5047 in the different age groups. Although there was high correlation found between the measures
5048 of BF using impedance and DXA data in Chapter 4, subtle differences between the measures
5049 may lead to differential associations with metabolites and other traits. In order to ensure
5050 comparison of adiposity effects were appropriate in Chapter 4, adiposity measures were
5051 normalised and Z-scores were used in regression models. When using publicly available
5052 summary level data however, summary statistics can be interpreted differently depending on
5053 the adjustments and transformations applied prior to genome-wide analysis. Summary level
5054 data for all adiposity measures had undergone the same inverse rank normal transformation
5055 prior to genome wide analysis, but study specific adjustments were not the same across the
5056 datasets. As such, although estimates reflected a normalised standard deviation unit, that
5057 unit will have been adjusted for different confounders. This meant direct comparison with
5058 ALSPAC data was not possible, it also meant that obtaining clinically informative units was not
5059 possible as it is not appropriate to back-calculate and estimate an SD unit change without the
5060 underlying distributions of the data which was not available. The format of publicly available
5061 data used in two-sample MR is a common limitation and it is a strength of one-sample MR
5062 that the researcher can control the processing prior to analysis.

5063 Pre-analysis processing of data is particularly important for metabolomics data⁴⁹⁶, where
5064 researchers characterize and prepare data prior to analyses. This may include identifying
5065 poor quality samples and metabolites and/or a broad array of additional, and researcher
5066 specific processes, such as transformations. In an effort to perform transparent and informed
5067 processing of metabolomics data the metaboprep R package, which I was involved in the
5068 development of, was used in Chapter 4 to process ALSPAC metabolite data. In using
5069 metaboprep, there are a number of threshold settings that can be altered (default settings
5070 were used). The default settings however are arbitrarily assigned or have been set based on
5071 previous published work, for example, sample missingness is set to 20% by default based
5072 on work by Lotta et al. (2021)³²¹. A more stringent threshold (e.g., 5%) may have impacted
5073 on the number of samples excluded. Of particular note from this pre-analysis processing
5074 step performed in Chapter 4, is the fact that, although almost all metabolites were shared
5075 across age groups, the number of independent metabolites identified by metaboprep differed
5076 across age groups. As metabolite data used in MR analyses were publicly available or
5077 provided by collaborators it was not possible to use the same pre-analysis processing strategy.
5078 Comparison of effect sizes was not possible across observational and MR analyses (only
5079 comparison of the direction of effect) as different transformations and adjustments had been
5080 made to the GWAS data used in MR analyses. This also meant it was not possible to
5081 meta-analyse the MR analyses performed in Chapter 5 using effect estimates. Future work
5082 should look to use a standardised method for pre-analysis processing, such as metaboprep,
5083 while a centralised approach to analysis of metabolomics data across cohorts, such as that
5084 being employed by the Consortium of Metabolomics Studies⁶²¹, will allow for more efficient
5085 comparisons and meta-analyses of studies.

5086 A strength of the metabolomics data used in these analyses was however the fact that
5087 all of the metabolomics data came from the same NMR platform which has been described
5088 previously^{476,492–494} and in Chapter 4. As the same platform was used it was possible to
5089 compare directions of effect across methods and datasets for almost all metabolites. Of
5090 important note however is the fact that this NMR platform is predominantly lipid based.
5091 Although many lipids are larger than the traditionally defined size of a metabolite⁶²² they are
5092 detected reliably by the platform⁴⁹². The platform also provides many derived measures,
5093 such as ratios and number of bonds. Although this NMR platform measures a broad array
5094 of metabolites, including amino acids and glycoproteins, it is limited and is therefore not
5095 reflective of the metabolome as a whole. Rather it reflects the lipidome and a small number
5096 of non-lipid metabolites.

5097 Given the large number of datasets used and the need to compare across exposures and
5098 analyses, a particular challenge was how to effectively summarise and interpret the large
5099 number of association analyses performed. Throughout this thesis forest plots and Circos
5100 plots are used to summarise, interpret, and present results from the over 1,000 analyses
5101 conducted. A key aspect of this was EpiViz, presented in Chapter 3, which was developed
5102 in order to improve the efficiency and reproducibility of producing Circos plots and to make
5103 them more accessible to a wider audience. EpiViz has been successfully used in published
5104 work^{1,468,469} and as the availability and interest in molecular data grows, it is likely more
5105 researchers will look to similar visualisation tools to interpret and communicate results. As it
5106 stands, EpiViz successfully creates Circos plots for summarising and communicating large
5107 association analyses. Future development should focus first on converting the current code
5108 style to the *tidy style* to improve readability and consistency. This will also improve the use
5109 of the package in the future enabling more contributors. Outside of this, future development
5110 will be through requests from the community to meet their needs.

5111 7.2.3 Methodology, instrumentation, and assumptions

5112 A key aspect of the work in this thesis was the use of a systematic review and meta-
5113 analyses (Chapter 2, and observational (Chapter 4) and two-sample MR (Chapter 5) anal-
5114 yses to inform downstream analyses of the investigation of the effect of adiposity-related
5115 metabolites on adiposity-related diseases. These analyses were performed using multiple
5116 independent datasets. This strategy of triangulation strengthens evidence for observed effects
5117 as each method has its own limitations and sources of bias. For example, observational
5118 epidemiology is limited by confounding and reverse causation. Prospective studies, such as
5119 ALSPAC used in Chapter 4, are able to provide some separation between the measurement
5120 of an exposure and outcome, and thus mitigate the potential effects of reverse causation.
5121 However, as data on adiposity measures and metabolites were collected at the same time
5122 reverse causation remains a limitation. It was however possible to account for the potential
5123 effects of confounding in Chapter 4 by adjusting linear models for potential confounders.
5124 Although many potential confounders were included across the models, it is likely these

5125 analyses will not have fully accounted for confounding, either due to measurement error or
5126 unmeasured confounding. For example, adults were asked 'do you take part in physical
5127 activity (e.g. running, swimming, dancing, golf, tennis, squash, jogging, bowls)?', with possible
5128 answers of 'no', 'occasionally (less than monthly)' and 'frequently (once a month or more)'.
5129 Broad categories such as these are unlikely to capture the full impact of physical activity given
5130 that 'frequently' will encompass individuals who exercise once a month as well as every day.

5131 Mendelian randomization analyses, which are robust to limitations of observational anal-
5132 yses such as confounding and reverse causation³³⁶, has its own set of assumptions and
5133 limitations, such as genetic confounding and horizontal pleiotropy (Chapter 1 Section 1.8).
5134 MR has three main assumptions and a number of additional assumptions, all discussed in
5135 detail in Chapter 1 Section 1.8, that must be satisfied in order to obtain reliable results. The
5136 first of these assumptions, that instruments are robustly associated with the exposure, can be
5137 satisfied by using instruments obtained from well powered GWAS, which are independent of
5138 one another, and which meet a robust genome-wide significance threshold (e.g., p-value <
5139 5×10^{-8}). All instruments used in MR analyses in this thesis were obtained from well pow-
5140 ered GWAS which used strict linkage disequilibrium (LD) R^2 thresholds and a genome-wide
5141 significance threshold of $< 5 \times 10^{-8}$. In additional analyses performed in Chapter 5 a more
5142 conservative LD R^2 threshold was used, the effect of these instruments on metabolites was
5143 consistent with the less conservative LD R^2 threshold used in the main analyses.

5144 This instrument selection strategy, of selecting the largest most recent GWAS with a
5145 strong independence and genome-wide significance threshold, was the most commonly used
5146 approach in studies included in the systematic review in Chapter 2. However, it is not clear
5147 whether this strategy is the most appropriate for all traits. In this regard, consideration needs
5148 to be given to the potential that SNPs associated with the measures of adiposity may also
5149 be associated directly with metabolites or metabolic pathways. The exclusion restriction
5150 assumption states that the instrument must act on the outcome only via the exposure, where
5151 violation of this assumption can bias results. Horizontal pleiotropy such as this is difficult to test
5152 for directly. Sensitivity analyses, using models sensitive to the effects of horizontal pleiotropy
5153 (MR-Egger, weighted median, and weight mode), were consistent with the main analysis in
5154 Chapter 5 and 6 suggesting weak evidence of horizontal pleiotropy. The Steiger directionality
5155 test can also be used to test whether the "true" causal direction is the one under investigation,
5156 i.e., the effect of adiposity measures on metabolites. Results from Steiger directionality tests
5157 in Chapter 5 suggested that the effects of horizontal pleiotropy were most likely apparent
5158 when using instruments with large SNP lists such as the 941 SNP BMI instrument. However,
5159 given that the directions of effect for these large SNP lists were highly consistent with the
5160 smaller SNP lists instrumenting the same exposures (which showed a majority of "true" causal
5161 directions) the interpretation of the directions of effect are likely the same across the SNP lists.
5162 This does not fully address the potential for direct associations between SNPs associated
5163 with the exposure and the outcome however, and future work should look at exploring the
5164 genetic correlation between adiposity measures and metabolites.

5165 For analyses in Chapter 6 investigating the effect of metabolites on endometrial cancer,

5166 instruments were identified using the same strategy described above and applying the
5167 stringent LD R^2 threshold used in additional analyses in Chapter 5. However, given the
5168 strong inter-correlated nature of metabolites and the fact many metabolites are products of
5169 one-another, this instrumentation strategy may not be appropriate. For instance, it is possible
5170 that SNPs associated with metabolites do not conform to the third IV assumption, that the
5171 SNPs do not affect the outcome except through the exposure. In this sense, the effect of
5172 a metabolite on an outcome could be through a highly correlated second metabolite. For
5173 the two metabolites identified in Chapter 6 with evidence of an intermediary role on the
5174 effect of adiposity on endometrial cancer, a total of 56 and 50 SNPs were associated at a
5175 genome-wide significance p-value threshold of 5×10^{-8} and an LD R^2 threshold of 0.0001.
5176 Of these SNPs, 16 were shared across the two metabolites. It is likely both metabolites also
5177 share SNPs with many other metabolites given that just over half of the 934 identified SNPs
5178 for the 53 metabolites investigated in Chapter 6 were associated with just one metabolite and
5179 the average number of SNPs associated with a metabolite was 49.

5180 When investigating the effect of adiposity measures on metabolites in Chapter 5, the
5181 exposure and outcome GWAS were sex-combined. In addition, adiposity instruments used
5182 in Chapter 6 were obtained from sex-combined GWAS. A key limitation of this work is
5183 therefore the use of non-sex specific instruments for measures of adiposity. A major driver
5184 of endometrial cancer, specifically endometrioid cancer, is oestrogen⁵⁶⁵, which, as a sex
5185 hormone, is important in endometrial cell growth. Oestrogen is also involved in the distribution
5186 and accumulation of adipose tissue⁶⁵. Importantly, adipose tissue is a key source of oestrogen
5187 synthesis, particularly post-menopause⁶⁰¹. There is sexual dimorphism in the accumulation
5188 and distribution of adipose tissue which is highlighted in genetic studies, which have found sex
5189 specific SNPs associated with adiposity^{36,37,41-43}. There is also evidence of sex differences
5190 in adiposity metabolite associations³⁶⁷. There is however utility in the use of sex-combined
5191 instruments as they provide an overarching view of the effect of adiposity that is independent
5192 of the sexual dimorphism associated with adiposity. That being said, replication using sex-
5193 specific instruments will only add to the evidence found in this thesis.

5194 A key assumption of two-sample MR is the independence of the two samples. Overlap
5195 between the exposure and outcome in a two-sample MR can bias estimates towards the
5196 observational estimate in a manner proportional to the size of the overlap⁴¹⁸. Given that many
5197 publicly available summary statistics are meta-analyses which include many different cohorts
5198 there is always the possibility of overlap when using GWAS with hundreds of thousands of
5199 participants. However, for analyses performed in Chapter 5, there was no overlap between the
5200 cohorts included in the adiposity GWAS with cohorts used in the metabolite GWAS. BMI and
5201 WHR GWAS used in these main analyses included individuals from UK Biobank, additional
5202 analyses used GWAS which did not include UK Biobank and resulted in highly consistent
5203 results. In Chapter 6, metabolite and endometrial cancer data were available from GWAS
5204 which included individuals from UK Biobank. In order to minimise the amount of overlap given
5205 metabolite and endometrial cancer data used included individuals from UK Biobank, BMI
5206 and WHR instruments were obtained from GWAS used in the additional analyses in Chapter
5207 5 which did not include individuals from UK Biobank. For binary outcomes, the bias due

5208 to overlap is specific to the overlap between individuals in the exposure GWAS and those
5209 who are cases in the outcome GWAS. As 5% of the endometrial cancer cases were UK
5210 Biobank participants, the relative bias for the effect of metabolites on endometrial cancer was
5211 at most 0.5%. Specifically, this bias is applicable to the effect of metabolites on endometrial
5212 cancer, and the effect of metabolites adjusted for adiposity measures on endometrial cancer.
5213 The bias is not applicable to the effect of adiposity measures adjusted for metabolites on
5214 endometrial cancer as the instruments for these analyses were obtained from the adiposity
5215 GWAS which did not include individuals from UK Biobank. Recent work has shown that the
5216 use of strong instruments can mitigate the bias induced by sample overlap⁶²³, although there
5217 was evidence for weak instruments for some metabolites, including when used in MVMR
5218 analysis, the majority of those instruments used in univariable MR were strong (F-statistic >
5219 10).

5220 Issues of sample overlap were not limited to the MR analyses performed in Chapter 5
5221 and 6, it was also an issue in Chapter 2. Of the over 2,000 MR analyses identified in the
5222 systematic review, a total of 71 MR analyses were included in 31 meta-analyses. The majority
5223 of these 31 meta-analyses included only two MR analyses. This meant that interpretation
5224 of heterogeneity statistics^{458,466} and the use of funnel plots to assess publication bias, was
5225 difficult. One of the main reasons for these small numbers was the fact that studies were
5226 excluded from the meta-analyses if there was: (i) sample overlap between outcomes across
5227 MR analyses, (ii) and/or if there was sample overlap between exposures and outcomes
5228 across MR analyses. As many of the largest and most recent MR analyses used samples
5229 from prior smaller MR analyses, these larger analyses were preferred for inclusion as they
5230 encompassed estimates from previous analyses. However, given that many MR analyses use
5231 data from large GWAS consortia, and given that reporting quality was not high for a majority
5232 of studies, it is possible that there is overlap in the meta-analysed MR analyses that was not
5233 identified during screening. It is possible to estimate the overlap between samples using LD
5234 score regression (LDSR)⁶²⁴, but this can only be achieved with summary data which was not
5235 readily available when extracting data for studies included in the Systematic Review. With the
5236 drive for GWAS to use ever larger sample sizes and for consortium to join forces to perform
5237 meta-analyses of meta-analyses it may become more challenging to identify truly independent
5238 samples in MR analyses. This will be a particular concern in regards to MR meta-analyses
5239 where there is more opportunity for sample overlap given the need to have independence
5240 between outcomes across studies and exposures and outcomes across studies.

5241 **7.2.4 General assumptions of MR analyses**

5242 Generally, MR analyses use unrelated individuals and assume that genotypes are inde-
5243 pendent of the environment. Where this assumption is not satisfied biased SNP-phenotype
5244 associations, due to population structure, selection bias, assortative mating, dynastic effects,
5245 and canalization^{336,625}. Canalization, whereby what would otherwise be developmentally
5246 deleterious genetic effects which are nullified by compensatory mechanisms, results in effect

estimates attenuating to the null³³⁷. Though being aware of the underlying biology can inform analyses, accounting for canalization is difficult as it is currently difficult to test for its presence⁶²⁶. Confounding by population structure on the other hand would bias estimates towards the observational confounded estimate. Generally this is accounted for in GWAS using principle components or linear mixed models^{533,627}. However, there is evidence of population structure in large and small GWAS of apparently homogeneous samples using both methods^{346,351,523-525}. In Chapter 5, additional analyses aimed to obtain consistent effects using a number of different instrumentation practices and different GWAS to instrument BMI, WHR, and BF. Consistent effects using GWAS of different populations suggests any effect of population structure is limited and did not drastically alter results. That being said, there was considerable difference in directions of effect when using the two BF GWAS. It is possible that these conflicting results are however due to the difficulty in instrumenting BF as opposed to population structure, given that when removing two 'favourable adiposity' SNPs from one SNP list resulted in a number of effect estimates changing direction.

Selection bias, whereby study participation is non-random, can induce an association between genotypes associated with study participation and other traits related to study participation. For example, a study aims to recruit individuals with endometrial cancer, but individuals with a high BMI are more willing to participate in the study than those with a lower BMI. In this example, an association between high BMI and endometrial cancer would have been induced as a result of study participation. As many of the cohort studies used in GWAS used in this thesis were prospective, individuals would not have been selected based on disease status. However, there is some evidence that individuals with certain cancers, e.g., prostate cancer, are diagnosed at a later stage as a result of a higher BMI¹¹¹. This could mean that individuals with higher adiposity during early cancer stages are less likely to be diagnosed or that higher adiposity is associated with a faster progression to advanced prostate cancer.

An additional consideration is that individuals do not mate at random. Assortative mating, where parents are more alike than one would expect at random, is associated with adiposity⁶²⁸. This can be because of a direct relationship, that is partners select because of a phenotype, or an indirect relationship, partners select because of background (e.g., ancestry or culture). Additionally there is potential for partner phenotypes to converge. These effects can be either single (e.g., adiposity) or multi-trait (e.g., adiposity and level of education). Bias in MR results is only induced by multi-trait assortative mating, as it induces associations between the two exposures through their associated instruments, that is SNPs associated with adiposity will become associated with level of education and vice versa. It is not possible to evaluate or test for the effects of assortative mating using summary level data⁶²⁵. However, as many independent datasets were used, resulting in highly consistent results, it is unlikely that patterns of assortative mating applicable to one study or cohort are replicated in all of the others.

Dynastic effects are a form of confounding as a result of generational traits which influence offspring phenotypes^{344,345}. Inherited parental alleles can influence the offspring phenotype

5288 through a direct path. In addition, the inherited and non-inherited parental alleles influence the
5289 parental phenotype which in-turn influences offspring phenotype through an environmental
5290 path. In this regard, the second MR assumption would be violated as there would be an
5291 association between offspring phenotype and parental genotype that was not directly through
5292 the transmission of alleles. As with all of these limitations, dynastic effects are not easily
5293 tested when using summary level data. Nor is it possible to evaluate their effects outside of
5294 family based studies^{344,345}.

5295 The issues of dynastic effects, assortative mating, and population structure (familial
5296 biases) can be mitigated through the use of family based studies^{344,625}. This is because
5297 alleles, inherited from parents, if conditioned on the parents genotype will be independent
5298 of the familial biases. Importantly, an association between non-inherited alleles and traits
5299 in the offspring can only be through parental phenotypes or via the aforementioned familial
5300 biases. However, conditioning on the parents genotype does not however account for the
5301 effect of the parental phenotype. It is possible to estimate the effect of the parental phenotype
5302 using MVMR, but it is not possible to assign a specific bias to any estimated effect. Although
5303 family, specifically trios, based studies can account for familial biases, the scarcity of genetic
5304 and phenotypic data on parents and offspring makes these analyses difficult^{344,625}. ALSPAC
5305 is one such study with trios data, however this data is limited in size and, as offspring are
5306 comparatively young, investigations of disease such as cancer are not possible. Prospective
5307 studies such as the Norwegian Mother, Father and Child Cohort Study (MoBa) and The
5308 Nord-Trøndelag Health Study (HUNT) include a larger number of trios but are also limited in
5309 their use to investigate disease outcomes given that data is needed on trios and outcomes.
5310 Twin-based and sibship studies, which are larger in number, can also be used, though not to
5311 the same level, to test for familial biases⁶²⁵. Though analyses may include small sample sizes,
5312 used in conjunction with traditional MR analyses using unrelated individuals, if evidence is
5313 consistent across analyses this will strengthen evidence⁵¹⁸ and suggest the effects of familial
5314 bias are limited.

5315 7.2.5 Additional considerations

5316 This thesis hypothesised that the direction of effect was true for the effect of adiposity
5317 on metabolites and endometrial cancer, and for metabolites on endometrial cancer. It did
5318 not however investigate the reverse of this, that the direction of effect was from metabolites
5319 to adiposity, and from endometrial cancer to metabolites and adiposity. Although there was
5320 evidence for the former, this does not rule out an effect of the latter. This is especially true
5321 given: (i) adipose tissue is a complex signalling organ, (ii) cancerous cells are complex
5322 signallers that re-wire metabolic processes, (iii) and metabolites are multi-functional with
5323 roles in energy, signalling, transportation, and as structural components. In observational
5324 and MR analyses there is considerable evidence that the relationship between adiposity and
5325 endometrial cancer is in the tested direction. As such, the reverse was not investigated here.
5326 In regards to the effect of metabolites on adiposity, the Steiger directionality tests performed

5327 in Chapter 5, as discussed in Section 7.2.3, suggested the majority of tests of the effect of
5328 adiposity on metabolites were 'true' causal directions. Further investigation of the effects
5329 of those analyses which did not reflect a 'true' causal direction is warranted, as there may
5330 be evidence that those metabolites are either highly pleiotropic with adiposity or influence
5331 adiposity. For the effect of metabolites on endometrial cancer and its reverse, the picture
5332 is complex. There is evidence that concentrations of metabolites are altered as a result of
5333 cancer⁵⁹⁰. However, it is not clear how these changes manifest, are they a result of cancer
5334 cells responding to a nutrient poor environment (nutrients may be available but they may not
5335 benefit the cancer cell) and their subsequent metabolic reprogramming, are they a result of
5336 metabolic reprogramming of cancer cells and the subsequent survival of those cancerous
5337 cells that are able to utilise the available nutrients, or is it they are a result of both selective
5338 pressures at once. These questions can not be asked in an epidemiological context. Instead,
5339 as was done here, we can inform wet lab studies by providing potential candidates that can
5340 be used to investigate these questions.

5341 Finally, a major component of this work was the comparison of observational and MR
5342 estimates to obtain evidence for an effect of adiposity measures on metabolites. A key
5343 limitation of this comparison is the fact that MR studies may represent different underlying
5344 processes to that of observational studies. Exposures instrumented by genetic variants reflect
5345 a lifetime exposure, while for observational studies, the exposures are determined by genetic
5346 and non-genetic factors at that point in time. The fact that consistent results were observed
5347 across observational (Chapter 4) and MR analyses (Chapter 5) using independent datasets,
5348 and that these results were replicated in a further MR analysis using an additional independent
5349 dataset (Chapter 6) provides robust evidence for an effect of adiposity on metabolites even if
5350 the underlying processes are different.

5351 7.3 Future work

5352 Work in this thesis identified two metabolites that may play intermediary roles in the
5353 development of endometrial cancer as a consequence of adiposity. The primary focus of this
5354 work was the use of complimentary assessment of adiposity, NMR based metabolites, and
5355 use of MR methodology. The aim and objectives of this thesis (Chapter 1 Section 1.10) were
5356 designed to better understand the underlying mechanisms of adiposity associated disease
5357 development. Although this aim has largely been achieved, there remain some unanswered
5358 questions and some new ones stemming from the presented results.

5359 A key question that remains is the replication of the effect of metabolites on endometrial
5360 cancer. Availability, funding, and time constraints limited the ability to perform replication
5361 using independent endometrial cancer datasets. There is however opportunity within the
5362 EPIC collaboration to investigate these associations. Data is available on individuals who
5363 have undergone bariatric surgery, including endometrial tissue, that could enable the tracking

5364 of metabolite changes through the surgery process, which would include investigating the
5365 effect of weight reduction. This would act as a reverse analysis to analyses performed in this
5366 thesis, investigating whether weight reduction leads to a reduction in metabolites identified in
5367 this thesis as up-regulated by adiposity and associated with endometrial cancer. Changes in
5368 these metabolites, within endometrial tissue, as a result of weight loss would add weight to
5369 the evidence presented in this thesis for potential intermediary effects on endometrial cancer
5370 as a result of adiposity.

5371 Endometrial cancer was used as an exemplar and a natural follow-on would be to in-
5372 vestigate the intermediary role of metabolites with other outcomes identified in Chapter
5373 ??systematic-review) as associated with adiposity. Of particular interest would be inves-
5374 tigations of breast and prostate cancer, both of which exhibit hormone dependence and
5375 independence, as the intermediary role of metabolites here appeared to be specific to hor-
5376 mone independent (non-endometrioid cancer) cancer. Follow-up analyses, especially of
5377 those relating to cancer, could be to examine the effect of increased triglycerides in small
5378 and very small VLDL on cell lines. There is evidence that dosing with specific metabolites is
5379 shown to increase proliferation rates⁶²⁹. A candidate metabolite could be that of myristic acid
5380 (a component of the triglyceride trimyristin⁵⁸⁵), which is associated with endometrial cancer
5381 development⁵⁸⁰. Follow-up experiments such as these, which use epidemiology and causal
5382 methods to identify candidates for molecular investigation, have shown promise fruitful in
5383 identifying molecular mechanisms of disease development⁶⁰⁶.

5384 Follow-up wet lab studies would benefit greatly from a more comprehensive assessment
5385 of the metabolome. As discussed earlier, metabolites here were limited to an NMR platform
5386 that was predominantly composed of lipids. Many of these metabolites are large transporters
5387 of smaller metabolites. Expanding the metabolites available for analysis to include these
5388 smaller metabolites such as those identified in un-targeted mass spectrometry analyses,
5389 where there is available genetic data³¹⁷, will enable a broader assessment of the effects of
5390 adiposity.

5391 The key take away from this thesis in regards to the effect of adiposity-related metabolites
5392 on endometrial cancer is that there is some evidence for an intermediate effect, that is
5393 not driven by horizontal pleiotropy but may be a result of weak instrument bias. As such,
5394 future work should look to improve on the instrumentation strategy used here. Closely
5395 related metabolites within the same subclass as those tested here could be combined and
5396 collapsed to create a new instrument that represents the broad effect of these metabolites³⁹².
5397 Alternatively, as both metabolites tested here share similar attributes and have similar genetic
5398 associations, a composite measure could be used to instrument their combined effect⁶⁰⁰.

5399 Conclusion

5400 Within this thesis, I have demonstrated: (i) a literature with strong evidence for a causal
5401 effect of adiposity on many outcomes, (ii) that adiposity is observationally and causally
5402 associated with a large number of predominantly lipid metabolites, (iii) and that a number
5403 of these metabolites may partly explain the relationship between adiposity and endometrial
5404 cancer risk. I have highlighted the complex issues that surround these findings, notably the
5405 difficulty instrumenting adiposity and metabolites, and have used a broad array of techniques
5406 to strengthen the evidence for an effect of adiposity on metabolites and of adiposity-related
5407 metabolites on endometrial cancer risk. In this discussion chapter I have summarised the
5408 findings of this thesis and the overarching strengths and limitations. I have made suggestions
5409 for future work that could extend and strengthen results here, and have provided a basis from
5410 which hypotheses involving large molecular datasets can be investigated in the future.

5411 **Appendix A**

5412 All tables and figures presented here are also available on [GitHub](#) within each chapters
5413 respective folder.

5414 **A.1 Chapter 1: Introduction**

5415 **Literature search**

5416 Manual literature searching is prone to bias. Literature mining tools, though susceptible
5417 to publication and other biases, provide an alternative approach enabling a large number of
5418 articles to be assessed in a semi-systematic way. MELODI⁶³⁰, a literature mining tool, was
5419 used to identify intermediate diseases between BMI and mortality.

5420 Briefly, MELODI creates individual article sets based on search terms ('body mass
5421 index' and 'mortality') and looks for enriched overlapping terms. MELODI uses PubMed and
5422 SemMedDB to identify enriched terms. PubMed is a data base of health and biomedical
5423 research literature, and SemMedDB is a semantic predication repository built from PubMed
5424 citations. Identifying enriched overlapping terms is a two-step process. First, overlapping
5425 terms are identified. Second, the degree of overlap given the observed and expected
5426 frequency of terms across all articles not just those included in the article sets is quantified.

5427 All articles published from 01/01/2000–09/12/2019 (maximum number of articles per
5428 article set is 1,000,000) using the terms 'body mass index' as the source and 'mortality' as
5429 the outcome were included. Raw results are available on [GitHub](#). A total of 187,951 and
5430 787,451 articles were retrieved and included in the source and outcome article sets. Using
5431 the SemMedDB Triple results a total of 10828 enriched overlapping terms were identified.
5432 This included similar terms to 'body mass index' as the source, which were removed (n =
5433 424). A Bonferroni corrected p-value (1.2×10^{-4}) removed 0 terms. Terms were filtered
5434 for uniqueness (n = 156) and presence in the following categories: Age Group (n = 152),
5435 Bacterium (included because of confounding with pneumonia; n = 4), finding (not an official
5436 MELODI category; n = 1), general (not an official MELODI category; n = 1), Fungus (included
5437 because of confounding with pneumonia; n = 2), Health Care Activity (n = 5), Human (n = 14),
5438 Injury or Poisoning (n = 1), Mammal (animal studies; n = 1), Patient or Disabled Group (n = 8),
5439 Population Group (n = 14), Sign or Symptom (n = 1), Virus (included because of confounding;
5440 n = 3).

5441 The following terms were removed and merged into immunocompromised host: infections,
5442 hospital and opportunistic infections. The following terms were removed because of duplication
5443 under different names: cardiac death, sudden death, and sudden cardiac death (merged
5444 into cessation of life); coronary arteriosclerosis (atherosclerosis); cardiovascular morbidity,
5445 cardiac event, heart diseases, vascular diseases (cardiovascular diseases); depressed mood
5446 (depressive disorder); diabetes mellitus and insulin-dependent (diabetic); metabolic diseases
5447 (metabolic syndrome). The following terms were removed because they were top-level cate-
5448 gories that could either include a wide variety of terms already included or were duplicated by
5449 other terms: chronic disease, critical illness, disability, pregnancy complications and perinatal
5450 morbidity (included in pregnancy category), pathogenesis. As a result of filtering a total of

5451 77 terms remained. These terms were combined into 0 categories: . The 'other' category
5452 included traits that did not fit into one of any of the other categories and which did not have
5453 aligned traits to form a separate category (Table A.1). The majority of intermediates were
5454 cardiovascular related.

5455 It should be noted that the search did not include articles prior to the year 2000 and
5456 focussed only on those archived by PubMed. Enrichment aims to reduce the noise introduced
5457 when searching hundreds of thousands of articles, however manual curation, which will hold
5458 its own biases, is needed in order to obtain an informative list of enriched terms.

Table A.1: MELDOI analysis: intermediates between "body mass index" and "mortality"

Intermediate	Category
Primary carcinoma of the liver cells	Cancer
Malignant neoplasm of stomach	Cancer
Malignant neoplasm of prostate	Cancer
Malignant neoplasm of lung	Cancer
Common Neoplasm	Cancer
Liver neoplasms	Cancer
Malignant disease	Cancer
Carcinoma of the Large Intestine	Cancer
Pancreatic carcinoma	Cancer
Heart failure	Cardiovascular
Anemia	Cardiovascular
Dyslipidemias	Cardiovascular
Cerebrovascular accident	Cardiovascular
Cardiovascular Diseases	Cardiovascular
Atherosclerosis	Cardiovascular
Myocardial Infarction	Cardiovascular
Ischemic stroke	Cardiovascular
Acute coronary syndrome	Cardiovascular
Atrial Fibrillation	Cardiovascular
Coronary heart disease	Cardiovascular
Systemic arterial pressure	Cardiovascular
Thrombosis	Cardiovascular
Cerebrovascular Disorders	Cardiovascular
Acute myocardial infarction	Cardiovascular
Sinus rhythm	Cardiovascular
Cardiomyopathies	Cardiovascular
Myocardial Ischemia	Cardiovascular
Peripheral Vascular Diseases	Cardiovascular

Table A.1: MELDOI analysis: intermediates between "body mass index" and "mortality"
(continued)

Intermediate	Category
Vascular calcification	Cardiovascular
Heart Arrest	Cardiovascular
Myocardial rupture	Cardiovascular
Shock, Cardiogenic	Cardiovascular
Hemorrhage	Cardiovascular
Ischemia	Cardiovascular
Congestive heart failure	Cardiovascular
Ventricular Dysfunction, Left	Cardiovascular
Mitral Valve Insufficiency	Cardiovascular
Hyperglycaemia	Cardiovascular
Pancreatitis	immune
Inflammatory disorder	immune
Immunocompromised Host	immune
Bacteremia	immune
Septicemia	immune
Lupus Erythematosus, Systemic	immune
Sepsis Syndrome	immune
End stage renal failure	Kidney
Kidney Failure, Chronic	Kidney
Glomerular Filtration Rate	Kidney
Kidney Diseases	Kidney
Kidney Failure	Kidney
Renal function	Kidney
Liver diseases	Liver
Non-alcoholic fatty liver	Liver
Liver and Intrahepatic Biliary Tract	Liver
Carcinoma	
Chronic liver disease	Liver
Depressive disorder	Neurological/behavioural
Dementia	Neurological/behavioural
Metabolic syndrome	Other
Cessation of life	Other
Malnutrition	Other
Diabetic	Other
Multiple Organ Failure	Other
Fibrosis	Other
Deglutition Disorders	Other

Table A.1: MELDOI analysis: intermediates between "body mass index" and "mortality"
(continued)

Intermediate	Category
Vitamin D Deficiency	Other
Pre-Eclampsia	Pregnancy
Pregnancy	Pregnancy
Hypertension induced by pregnancy	Pregnancy
Tuberculosis	Respiratory
Sleep Apnea, Obstructive	Respiratory
Pneumonia	Respiratory
Chronic Obstructive Airway Disease	Respiratory
Respiration Disorders	Respiratory
Respiratory Distress Syndrome, Adult	Respiratory
Respiratory Tract Infections	Respiratory
Respiratory Failure	Respiratory
Acute respiratory failure	Respiratory

5459 **A.2 Chapter 2: Systematic Review**

Table A.2: Quality assessment tool

		Weak instrument bias	Genetic confounding bias	Other Confounding bias	Additional direct effects between IV and outcome (exclusion restriction assumption)	Bias due to selection of participants (exposure)	Bias due to selection of participants (outcome)	IV association	Sample overlap	Sensitivity analyses	Descriptive data	Data availability	Statistical parameters
Question	Strength of association between instrument and exposure	Reported test on association between confounders and IV	Included confounders in the IV analysis	Presence of pleiotropy for IV tested	Homogenous population or similar ancestry?	Homogenous population or similar ancestry?	Provide information on the similarity of IV in exposure/outcome. Two-Sample MR: same allele frequency, same ancestry etc.	Provide information on sample overlap	Have they performed sensitivity analyses.	Is information on the methodology missing/incorrect.	Is the methodology reproducible?	Are the statistical parameters for the analysis given	
High	F less than 10	Yes AND there is an obvious association	Yes (lifestyle factors)	No test/investigation performed or not discussed	non-homogenous population (e.g. black and white together, etc.)	non-homogenous population (e.g. black and white together, etc.)	Different ancestry	Not reported/assumed overlap	Not reported / not performed	Majority missing or incorrect - key info such as MR estimator and/or N SNPs missing	No code provided and no software/packages referenced	No	
Moderate	F missing	Not presented or presented and there is some degree of association	Not discussed (one-sample/meta-analysis)	Post-hoc sensitivity analysis	population described as homogenous BUT no ancestry covariate included	population described as homogenous BUT no ancestry covariate included	Not reported/Assumed same underlying ancestry	Not reported/assumed no overlap	Reported	Some information missing such as N for exposure/outcome	Software/packages referenced	Some	
Low	F greater than 10	Presented and no obvious association	No	Excluded/re-ran after exclusion of pleiotropic SNPs / sensitivity analysis consistent with main analysis	population described as homogenous AND ancestry covariate included	population described as homogenous AND ancestry covariate included	Tested and same ancestry	Reported/no overlap	Reported in the context of main analysis / can't do it (e.g. one SNP used)	Majority of information presented and correct	Code is provided and all software/packages are referenced	Yes/NA because 1 SNP/directly genotyped	

Table A.3: Quality assessment results for studies included in meta-analyses

Author	Year	doi	1	2	3	4	5	6	7	8	9	10	11	12	Total
Kivimaki	2008	10.1093/eurheartj/ehn252	1	2	2	3	3	3	2	1	3	1	3	1	25
Palmer	2011	10.1093/aje/kwr026	2	1	2	1	3	3	2	1	1	1	3	1	21
Fall	2013	10.1371/journal.pmed.1001474	2	2	3	3	3	3	2	2	3	1	3	1	28
Fall	2013	10.1371/journal.pmed.1001474	2	2	3	3	3	3	2	2	3	1	3	1	28
Fall	2013	10.1371/journal.pmed.1001474	2	2	3	3	3	3	2	2	3	1	3	1	28
Fall	2013	10.1371/journal.pmed.1001474	2	2	3	3	3	3	2	2	3	1	3	1	28
Fall	2013	10.1371/journal.pmed.1001474	2	2	3	3	3	3	2	2	3	1	3	1	28
Fall	2013	10.1371/journal.pmed.1001474	2	2	3	3	3	3	2	2	3	1	3	1	28
Holmes	2014	10.1016/j.ajhg.2013.12.014	1	1	2	2	3	3	2	1	2	1	2	3	23
Holmes	2014	10.1016/j.ajhg.2013.12.014	1	1	2	2	3	3	2	1	2	1	2	3	23
Holmes	2014	10.1016/j.ajhg.2013.12.014	1	1	2	2	3	3	2	1	2	1	2	3	23
Wurtz	2014	10.1371/journal.pmed.1001765	1	2	2	2	3	3	2	1	2	1	3	2	24
Wurtz	2014	10.1371/journal.pmed.1001765	1	2	2	2	3	3	2	1	2	1	3	2	24
Østergaard	2015	10.1371/journal.pmed.1001841	2	2	1	1	3	3	2	2	1	1	3	2	23
Gao	2016	10.1093/ije/dyw129	2	2	1	1	3	3	2	1	1	1	3	2	22
Gao	2016	10.1093/ije/dyw129	2	2	1	1	3	3	2	1	1	1	3	2	22
Gao	2016	10.1093/ije/dyw129	2	2	1	1	3	3	2	1	1	1	3	2	22
Gao	2016	10.1093/ije/dyw129	2	2	1	1	3	3	2	1	1	2	3	2	23
Gao	2016	10.1093/ije/dyw129	2	2	1	1	3	3	2	1	1	2	3	2	23
Gao	2016	10.1093/ije/dyw129	2	2	1	1	3	3	2	1	1	1	3	2	22
Gao	2016	10.1093/ije/dyw129	2	2	1	1	3	3	2	1	1	1	2	3	22
Gao	2016	10.1093/ije/dyw129	2	2	1	1	3	3	2	1	1	1	2	3	22
Gao	2016	10.1093/ije/dyw129	2	2	1	1	3	3	2	1	1	1	2	3	22
Gao	2016	10.1093/ije/dyw129	2	2	1	1	3	3	2	1	1	1	2	3	22
Gao	2016	10.1093/ije/dyw129	2	2	1	1	3	3	2	1	1	1	2	3	22
Gao	2016	10.1093/ije/dyw129	2	2	1	1	3	3	2	1	1	1	2	3	22
Gao	2016	10.1093/ije/dyw129	2	2	1	1	3	3	2	1	1	1	2	3	22
Jarvis	2016	10.1038/bjc.2016.188	2	2	1	2	3	3	2	1	2	1	3	3	25
Jarvis	2016	10.1038/bjc.2016.188	2	2	1	2	3	3	2	1	2	1	3	3	25
Lyall	2016	10.1001/jamocardio.2016.5804	2	2	3	1	3	2	2	1	1	1	2	3	23
Painter	2016	10.1158/1055-9965.EPI-16-0147	2	2	2	1	3	3	2	1	1	1	3	3	24
Wang	2016	10.1210/jc.2017-02789	1	1	3	1	3	3	2	1	1	1	3	3	23
Wang	2016	10.1210/jc.2017-02789	1	1	3	1	3	3	2	1	1	1	3	3	23
Dale	2017	10.1161/CIRCULATIONAHA.116.026560	1	1	1	1	2	3	2	2	1	1	2	2	19
Dale	2017	10.1161/CIRCULATIONAHA.116.026560	1	1	2	1	2	3	2	2	1	2	2	2	21
Klarin	2017	10.1161/CIRGENETICS.116.001643	2	2	3	3	3	3	2	1	3	2	3	3	30
Larsson	2017	10.1212/WNL.0000000000004173	2	2	1	1	3	3	2	2	1	2	3	3	25
Lindstrom	2017	10.1007/s00439-017-1811-x	2	2	1	2	3	3	2	2	2	1	3	2	25
Nordestgaard	2017	10.1210/jc.2017-00195	2	2	3	3	3	3	2	2	3	1	3	3	30
Xu	2017	10.1007/s00125-017-4396-y	1	2	1	1	3	3	2	2	1	1	3	2	22
Xu	2017	10.1007/s00125-017-4396-y	1	2	1	1	3	3	2	2	1	1	3	2	22
Xu	2017	10.1007/s00125-017-4396-y	1	2	1	1	3	3	2	2	1	1	3	2	22
Xu	2017	10.1007/s00125-017-4396-y	1	2	1	3	3	3	2	2	3	1	3	2	26
Brower	2018	10.1093/humrep/dey343	1	2	1	3	3	3	2	1	3	1	2	2	24
Day	2018	10.1371/journal.pgen.1007813	2	2	2	1	3	3	2	1	1	2	3	1	23
Kar	2018	10.1007/s10654-019-00485-7	1	2	1	1	3	3	3	1	1	1	2	2	21
Kar	2018	10.1007/s10654-019-00485-7	1	2	1	1	3	3	3	1	1	1	2	2	21
Larsson	2018	10.1093/rheumatology/key229	2	2	1	1	3	3	2	2	1	1	3	3	24
Lv	2018	10.1038/s41431-018-0180-9	2	2	1	1	3	3	2	3	1	2	3	3	26
Shapland	2018	10.1002/sim.8029	3	2	2	3	3	3	2	1	3	1	3	3	29
Shu	2018	10.1093/ije/dyy201	2	2	1	3	3	3	3	2	3	1	3	3	29

Table A.3: Quality assessment results for studies included in meta-analyses (*continued*)

Author	Year	doi	1	2	3	4	5	6	7	8	9	10	11	12	Total
Shu	2018	10.1093/ije/dyy201	2	2	1	3	3	3	3	2	3	1	3	3	29
Skaaby	2018	10.1111/all.13242	1	2	1	2	3	3	2	1	2	1	2	3	23
Tyrrell	2018	10.1093/ije/dyy223	2	2	1	1	2	2	1	1	1	1	3	3	20
van den Broek	2018	10.1136/jech-2017-210000	2	2	1	1	3	3	3	2	1	1	2	2	23
Wade	2018	10.1161/CIRCULATIONAHA.117.033278	2	2	1	1	2	2	1	1	1	1	3	3	20
Wang	2018	10.1002/oby.22167	2	2	3	3	3	3	2	2	3	1	3	3	30
Censin	2019	10.1371/journal.pgen.1008405	1	2	3	1	3	3	2	1	1	2	2	3	24
Censin	2019	10.1371/journal.pgen.1008405	1	2	3	1	3	3	2	1	1	2	2	3	24
Censin	2019	10.1371/journal.pgen.1008405	1	2	3	1	3	3	2	1	1	1	2	3	23
Censin	2019	10.1371/journal.pgen.1008405	1	2	3	1	3	3	2	1	1	1	2	3	23
Censin	2019	10.1371/journal.pgen.1008405	1	2	3	1	3	3	2	1	1	1	2	3	24
Censin	2019	10.1371/journal.pgen.1008405	1	2	3	1	3	3	2	1	1	2	2	3	24
Censin	2019	10.1371/journal.pgen.1008405	1	2	3	1	3	3	2	2	1	2	2	3	25
Gharahkhani	2019	10.1038/s41416-019-0386-9	1	1	3	1	2	2	1	1	1	1	3	3	20
Gharahkhani	2019	10.1038/s41416-019-0386-9	1	1	3	1	2	2	1	1	1	1	3	3	20
Gharahkhani	2019	10.1038/s41416-019-0386-9	1	1	3	1	2	2	1	1	1	1	3	3	20
Gharahkhani	2019	10.1038/s41416-019-0386-9	1	1	3	1	2	2	1	1	1	1	3	3	20
Gharahkhani	2019	10.1038/s41416-019-0386-9	1	1	3	1	2	2	1	1	1	1	3	3	20
Richardson	2019	10.7554/eLife.43657	2	2	1	3	3	3	2	2	3	1	3	3	28
Speed	2019	10.1101/539601	1	2	1	1	2	3	2	2	1	1	2	2	20
Yarmolinsky	2019	10.1101/472696	1	2	1	1	3	3	2	1	1	1	3	2	21

Table A.4: Meta-analyses results from MR analyses using adiposity as an exposure

Exposure on outcome	Studies	B/OR	Lower CI	Upper CI	p-value	q	q_df	q_p	tau2	tau2_setau	h	i1
BMI (SD) on Alzheimers	436,446	1.01	0.88	1.15	0.92	0.05	1	0.82	0.00	0.00	0.00	1.00
BMI (SD) on hemorrhagic stroke	389,428	1.08	0.73	1.59	0.71	1.19	1	0.28	0.04	0.07	0.20	1.09
BMI (SD) on ischemic stroke	430,441	1.24	1.00	1.53	0.05	4.98	1	0.03	0.02	0.02	0.13	2.23
Birthweight (SD) on ER- breast cancer	429,434	0.94	0.77	1.14	0.51	0.15	1	0.70	0.00	0.00	0.02	1.00
Birthweight (SD) on breast cancer	429,434	1.01	0.72	1.42	0.97	4.70	1	0.03	0.04	0.06	0.21	2.17
Birthweight (SD) on colon cancer	429,432	0.94	0.54	1.64	0.83	4.04	1	0.04	0.11	0.14	0.33	2.01
BMI (SD) on colorectal cancer	429,431,432	1.18	1.01	1.37	0.03	4.14	2	0.13	0.01	0.01	0.10	1.44
WHR (SD) on colorectal cancer	429,432	1.48	1.08	2.03	0.01	0.38	1	0.54	0.00	0.01	0.06	1.00
BMI (SD) on endometrial cancer	431,447,448	1.57	1.11	2.22	0.01	25.52	2	0.00	0.08	0.07	0.28	3.57
BMI (SD) on lung cancer	429,430	1.30	1.17	1.45	0.00	0.17	1	0.68	0.00	0.00	0.01	1.00
BMI (SD) on ovarian cancer	429,431	1.39	1.20	1.61	0.00	0.08	1	0.78	0.00	0.00	0.01	1.00
BMI (SD) on prostate cancer	429,431	1.08	0.91	1.28	0.37	1.29	1	0.26	0.01	0.01	0.08	1.13
WHR (SD) on CAD	430,444	1.63	1.40	1.91	0.00	3.09	1	0.08	0.01	0.01	0.09	1.76

Exposure on outcome	Studies	B/OR	Lower CI	Upper CI	p-value	q	q_df	q_p	tau2	tau2_setau	h	i1
WHRadjBMI (SD) on CAD	428,430	1.40	1.33	1.47	0.00	0.63	1	0.43	0.00	0.00	0.02	1.00 0.00
BMI (SD) on hypertension	389,445	1.36	0.94	1.98	0.11	41.65	1	0.00	0.07	0.10	0.26	6.45 0.98
BMI (SD) on venous thromboembolism	440,443	1.58	1.33	1.87	0.00	0.27	2	0.87	0.00	0.00	0.02	1.00 0.00
BMI (SD) on depression	452–454	1.11	1.04	1.19	0.00	0.00	1	0.97	0.00	0.00	0.00	1.00 0.00
BMI (SD) on type 2 diabetes	430,435	2.48	1.52	4.07	0.00	11.23	2	0.00	0.00	0.00	0.05	2.37 0.82
WHRadjBMI (SD) on type 2 diabetes	430,435	2.06	1.90	2.24	0.00	0.04	1	0.85	0.00	0.00	0.01	1.00 0.00
BMI (SD) on PCOS	437,438	2.55	1.22	5.34	0.01	29.71	3	0.00	0.00	0.00	0.06	3.15 0.90
BMI (SD) on asthma	448,451	1.06	1.03	1.10	0.00	4.37	1	0.04	0.00	0.00	0.03	2.09 0.77
BMI (SD) on arthritis	442,449	1.48	0.68	3.25	0.33	7.09	2	0.03	0.04	0.04	0.19	1.88 0.72
BMI (SD) on SBP (mm/Hg)	390,450,455	0.79	0.54	1.05	0.00	22.35	1	0.00	0.02	0.03	0.14	4.73 0.96
BMI (SD) on cholesterol (mmol/L)	389,433	0.01	-0.02	0.03	0.63	1.72	1	0.19	0.17	0.25	0.41	1.31 0.42
BMI (SD) on fasting glucose (mmol/L)	357,389,390,430	0.08	0.02	0.15	0.02	7.90	2	0.02	0.01	0.02	0.11	1.99 0.75
BMI (SD) on HbA1c (%)	357,389	0.03	-0.02	0.07	0.22	1.30	1	0.25	0.00	0.00	0.04	1.14 0.23
BMI (SD) on HDL (mmol/L)	389,390,433	-0.11	-0.34	0.12	0.34	33.29	1	0.00	0.12	0.17	0.35	5.77 0.97
BMI (SD) on HDL (SD)	157,357	-0.12	-0.33	0.09	0.25	1.13	1	0.29	0.00	0.00	0.05	1.06 0.12

Exposure on outcome	Studies	B/OR	Lower CI	Upper CI	p- value	q	q_df	q_p	tau2	tau2_setau	h	i1
BMI (SD) on HOMA IR (SD)	439,456	0.26	-0.42	0.95	0.45	1.96	1	0.16	0.19	0.26	0.43	1.40
BMI (SD) on LDL (mmol/L)	389,390,433	-	-0.06	0.04	0.65	0.64	1	0.42	0.01	0.03	0.11	1.00
BMI (SD) on LDL (SD)	157,357	0.01 0.02	-0.03	0.07	0.43	32.48	1	0.00	0.30	0.42	0.55	5.70
												0.97

5460 **A.3 Chapter 4: Observational**

5461 **A.3.1 Methods**

5462 In Children, a measure for BF was not available. Instead, impedance values were used
5463 to calculate estimates of BF using equation (4.1). The calculated BF resulted in negative
5464 BF estimates for some children (Figure A.1). However, the calculated BF was positively
5465 correlated with weight, height and BMI in children. In adolescents, the calculated BF did
5466 not produce negative estimates and was positively correlated with impedance and DXA
5467 derived BF estimates (Figure ??), as well as with weight, height, and BMI (Figure A.2). Child
5468 calculated BF positively correlated with adolescent measures of BF (Figure ??).

5469 Given that equation (4.1) was derived using adult data and the volumetric difference be-
5470 tween adults and children (i.e., BMI cubed rather than BMI squared may be more appropriate
5471 in children⁶³¹), differences in the range of estimates is unsurprising. There is however strong
5472 correlation between calculated BF and impedance derived and DXA derived estimates of BF
5473 in adolescents. Given that in a linear model the estimate is based on the per-unit increase, the
5474 absolute value of the exposure therefore does not need to be positive. As such, BF calculated
5475 using equation (4.1) was used in subsequent analyses as a measure of BF in children.

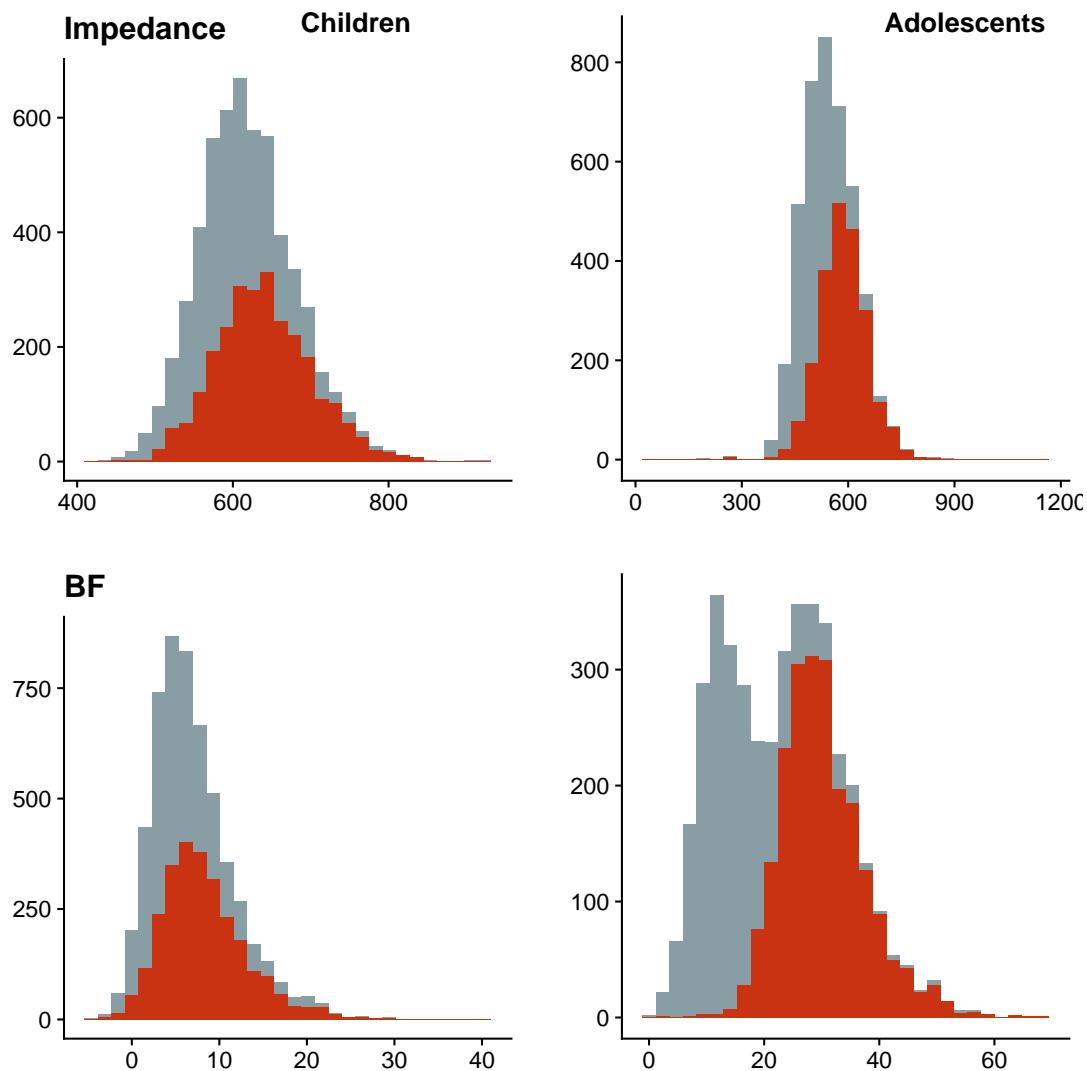


Figure A.1: Distribution of raw impedance and calculated body fat percentage in children and adolescents. Raincloud plots give the distribution of the raw impedance value measured in ohms and body fat percentage (BF) derived using equation (4.1) for children and adolescents. Data is presented for complete cases across: raw metabolomics, body mass index, waist hip ratio (children only), impedance, height, weight, age, sex, and BF derived by impedance and dual energy x-ray absorptiometry in adolescents. Red is female data, grey is male data.

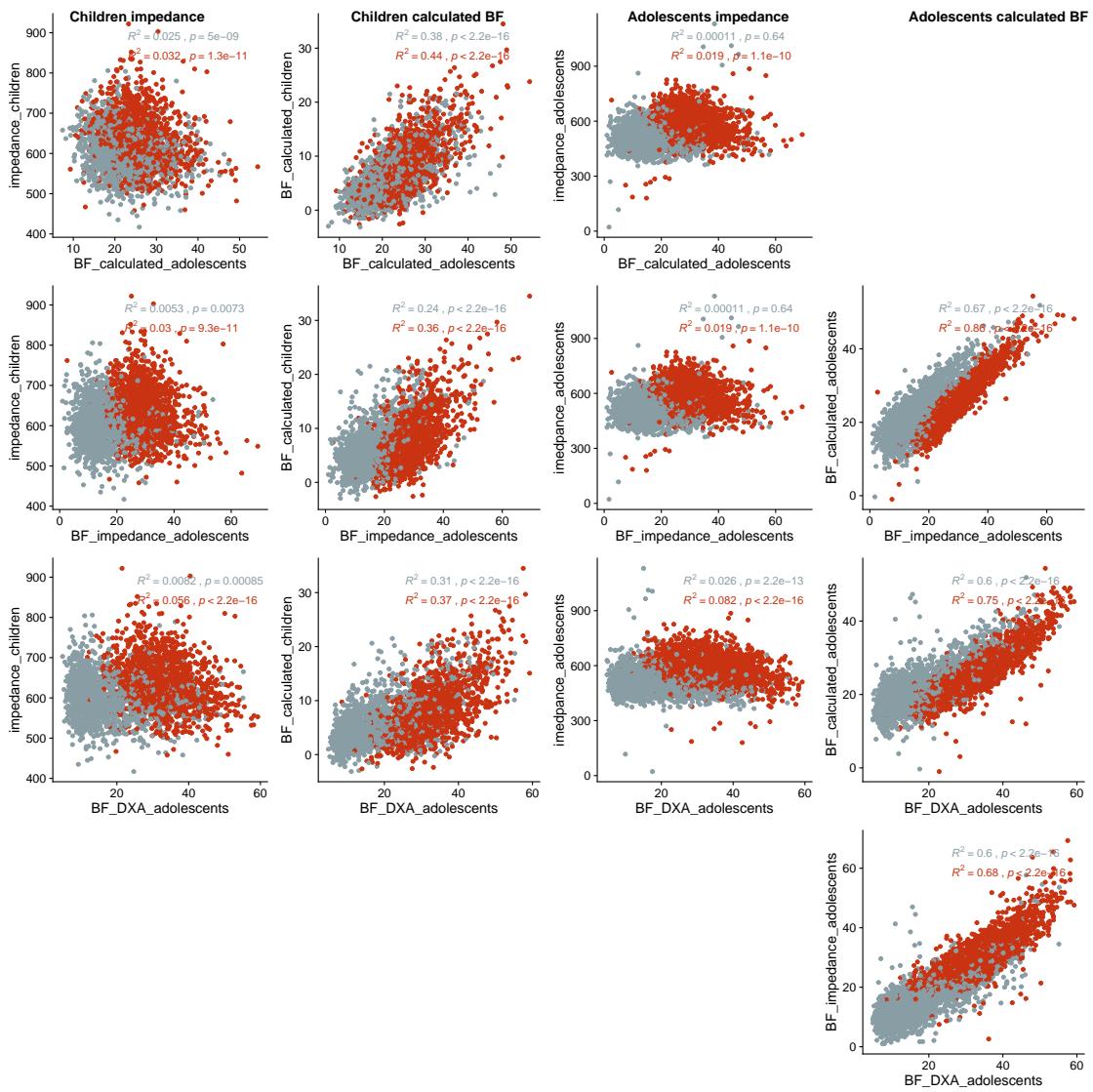


Figure A.2: Correlation between different body fat percentage measures in children and adolescents. Scatter plots are presented alongside Pearson's correlation coefficients for each sex; data for males are in grey and females in red. Plots are arranged in 4 columns and 4 rows. Column 1 gives raw impedance with body fat percentage (BF) calculated in children and adolescents using equation (4.1), and BF derived from dual energy x-ray absorptiometry (DXA) in adolescents. Column 2 gives child calculated BF using equation (4.1) with adolescent calculated BF using equation (4.1) and DXA derived BF. Column 3 gives adolescent raw impedance with adolescent BF calculated using equation (4.1) and DXA derived BF. Column 4 gives Adolescent calculated BF using equation (4.1) with adolescent raw impedance, DXA derived BF, and adolescent raw impedance with DXA derived BF. Data is presented for complete cases across: QC'd metabolomics, body mass index, waist hip ratio (children only), impedance, height, weight, age, sex, and BF derived by impedance and DXA in adolescents.

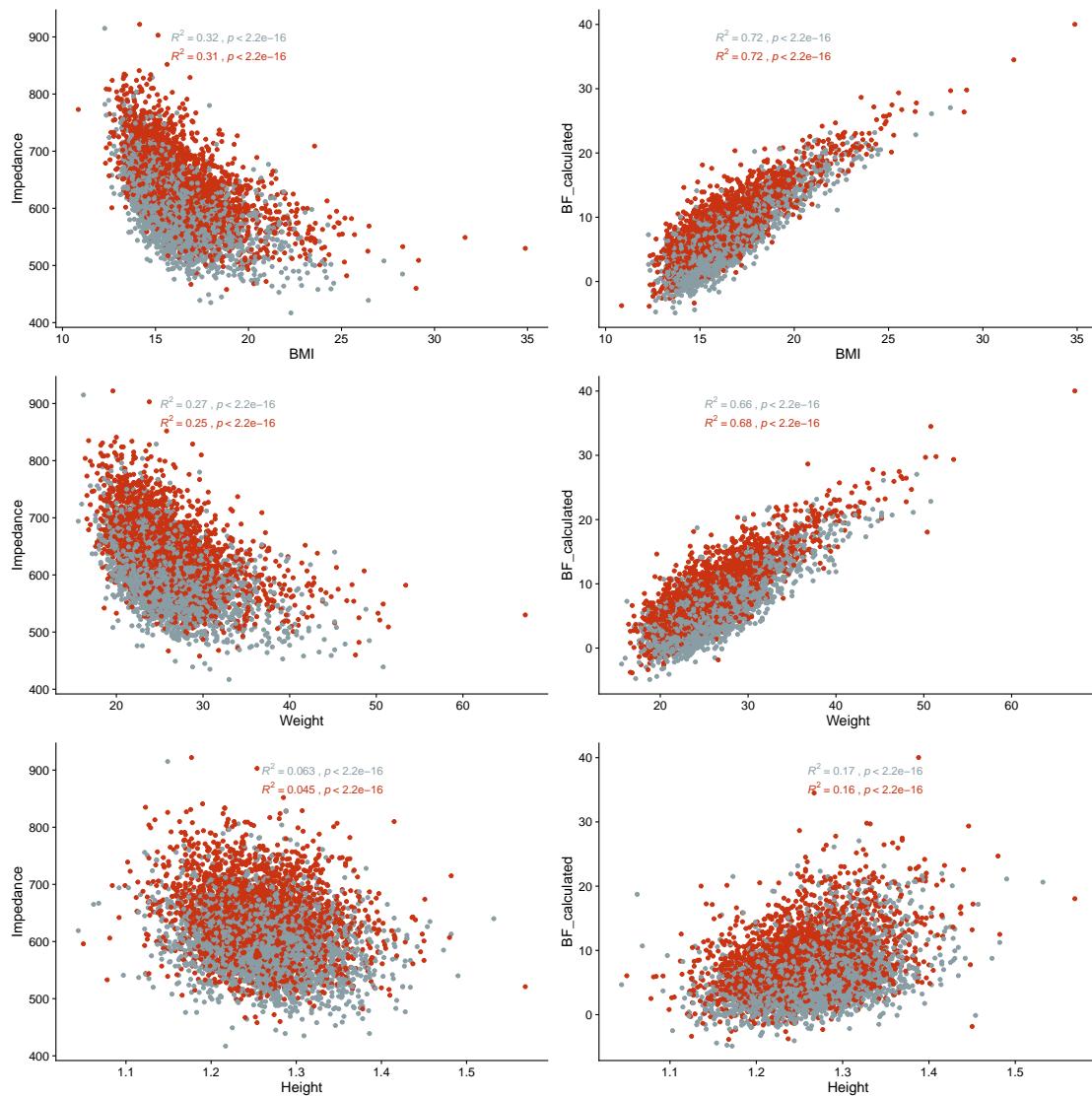


Figure A.3: Correlation between body fat percentage and anthropometric measures in children. Scatter plots are presented alongside Pearson's correlation coefficients for each sex; data for males are in grey and females in red. Column 1 = impedance; column 2 = body fat percentage (BF). Impedance is given as ohms. Data is presented for complete cases across: QC'd metabolomics, body mass index, waist hip ratio (children only), impedance, height, weight, age, sex, and BF derived by impedance and DXA in adolescents.

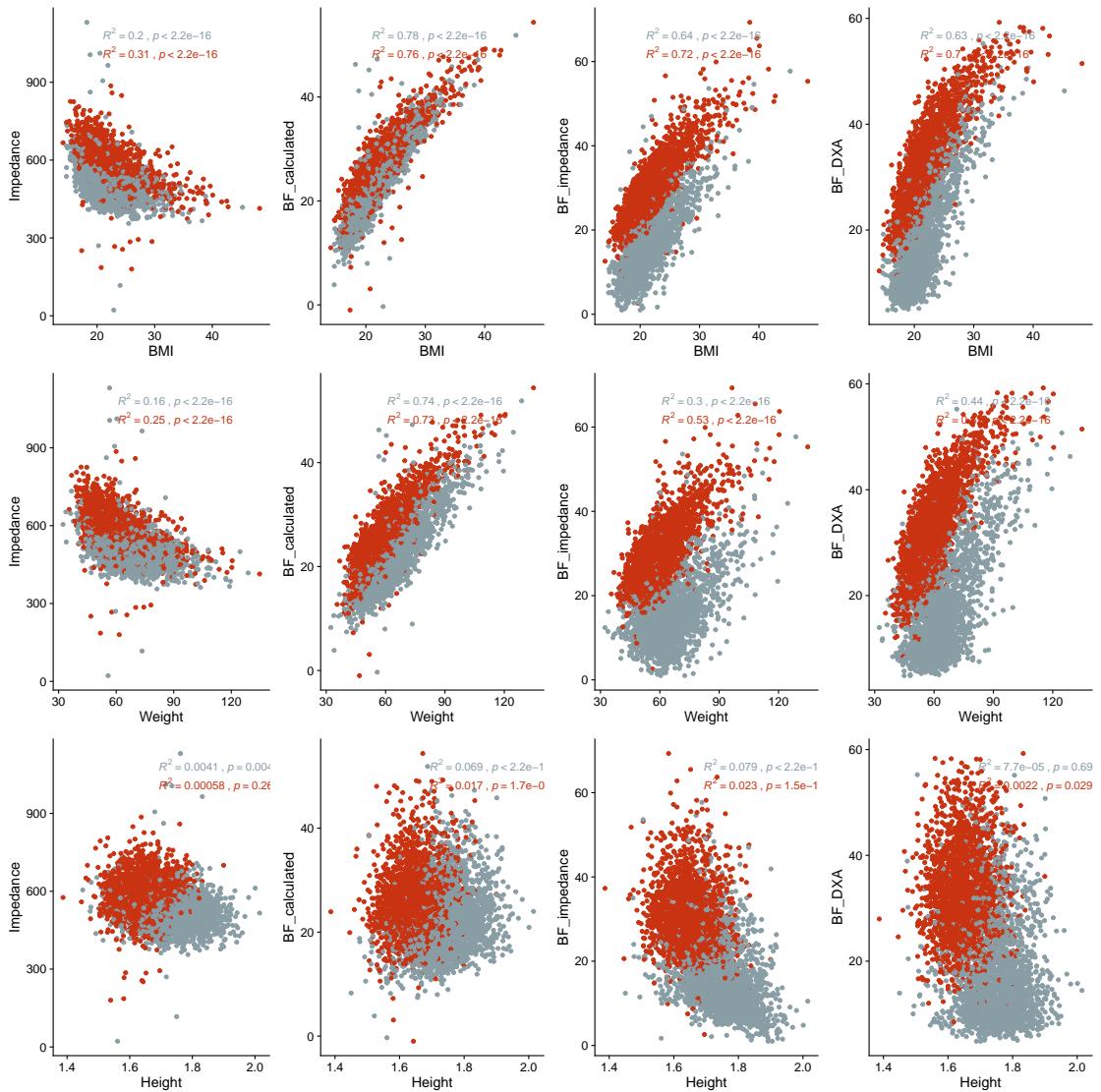


Figure A.4: Correlation between body fat percentage and anthropometric measures in adolescents. Scatter plots are presented alongside Pearson's correlation coefficients for each sex; data for males are in grey and females in red. Column 1 = impedance; column 2 = BF calculated using equation (4.1); column 3 = BF derived from impedance device; column 4 = BF derived from dual energy x-ray absorptiometry (DXA). Impedance is given as ohms. Data is presented for complete cases across: QC'd metabolomics, body mass index, waist hip ratio (children only), impedance, height, weight, age, sex, and BF derived by impedance and DXA in adolescents.

5476 **A.3.2 Results**

5477 **Correlations**

5478 Due to the large number of tests performed, concordance between models, exposures,
5479 and age groups were investigated using Spearman's correlations.

```
[1] "Correlations across models within exposures and age groups"
```

```
[1] "Children BMI"
```

	model1	model2
model1	1.0000000	0.9552825
model2	0.9552825	1.0000000

```
[1] "Children WHR"
```

	model1	model2
model1	1.000000	0.999278
model2	0.999278	1.000000

```
[1] "Children BF"
```

	model1	model2
model1	1.0000000	0.9407415
model2	0.9407415	1.0000000

```
[1] "Adolescents BMI"
```

	model1	model2	model3
model1	1.0000000	0.9795601	0.9521372
model2	0.9795601	1.0000000	0.9741246
model3	0.9521372	0.9741246	1.0000000

```
[1] "Adolescents BF"
```

```
      model1     model2     model3
model1 1.0000000 0.9571041 0.9459728
model2 0.9571041 1.0000000 0.9418077
model3 0.9459728 0.9418077 1.0000000
```

```
[1] "Young adults BMI"
```

```
      model1     model2     model3
model1 1.0000000 0.9918482 0.8245953
model2 0.9918482 1.0000000 0.8143402
model3 0.8245953 0.8143402 1.0000000
```

```
[1] "Young adults WHR"
```

```
      model1     model2     model3
model1 1.0000000 0.9992526 0.9606961
model2 0.9992526 1.0000000 0.9594533
model3 0.9606961 0.9594533 1.0000000
```

```
[1] "Young adults BF"
```

```
      model1     model2     model3
model1 1.0000000 0.9923692 0.8796946
model2 0.9923692 1.0000000 0.8642943
model3 0.8796946 0.8642943 1.0000000
```

```
[1] "Adults BMI"
```

```
      model1     model2     model3
model1 1.0000000 0.9760175 0.9650574
model2 0.9760175 1.0000000 0.9947999
model3 0.9650574 0.9947999 1.0000000
```

```
[1] "Adults WHR"
```

```
      model1      model2      model3
model1 1.0000000 0.9982737 0.9517970
model2 0.9982737 1.0000000 0.9554885
model3 0.9517970 0.9554885 1.0000000
```

```
[1] "Adults BF"
```

```
      model1      model2      model3
model1 1.0000000 0.9945195 0.9771555
model2 0.9945195 1.0000000 0.9850295
model3 0.9771555 0.9850295 1.0000000
```

```
[1] "Correlations across exposures within age groups for model 2"
```

```
[1] "Children"
```

```
      bf      bmi      whr
bf 1.0000000 0.8914358 0.6066604
bmi 0.8914358 1.0000000 0.7711406
whr 0.6066604 0.7711406 1.0000000
```

```
[1] "Adolescents"
```

```
      bf      bmi
bf 1.0000000 0.9377422
bmi 0.9377422 1.0000000
```

```
[1] "Young adults"
```

```
      bf      bmi      whr
bf 1.0000000 0.9659641 0.8012407
bmi 0.9659641 1.0000000 0.8123532
whr 0.8012407 0.8123532 1.0000000
```

```
[1] "Adults"
```

```

          bf      bmi      whr
bf  1.0000000 0.9179145 0.9542118
bmi 0.9179145 1.0000000 0.9366665
whr 0.9542118 0.9366665 1.0000000

```

[1] "Correlations within exposures across age groups for model 2"

[1] "BMI"

	adolescents	adults	children	young_adults
adolescents	1.0000000	0.4808187	0.7695565	0.7356921
adults	0.4808187	1.0000000	0.5592282	0.6902187
children	0.7695565	0.5592282	1.0000000	0.7299576
young_adults	0.7356921	0.6902187	0.7299576	1.0000000

[1] "WHR"

	adults	children	young_adults
adults	1.0000000	0.5473523	0.7357566
children	0.5473523	1.0000000	0.5023414
young_adults	0.7357566	0.5023414	1.0000000

[1] "BF"

	adolescents	adults	children	young_adults
adolescents	1.0000000	0.5328382	0.8151367	0.6930515
adults	0.5328382	1.0000000	0.5135602	0.7160939
children	0.8151367	0.5135602	1.0000000	0.7088497
young_adults	0.6930515	0.7160939	0.7088497	1.0000000

5480 **Figures**

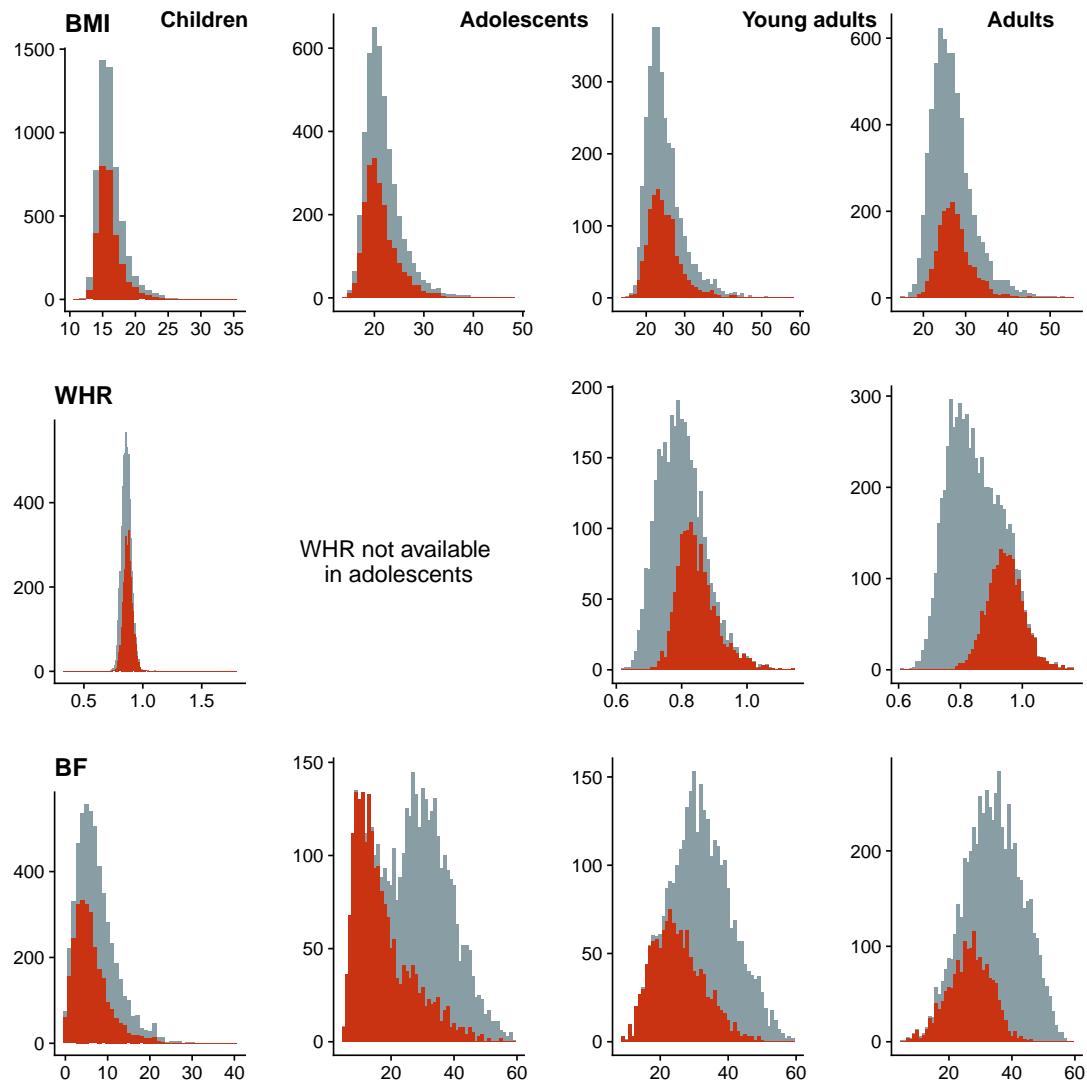


Figure A.5: Distribution of adiposity measures across all age groups. Raincloud plots give the distribution of all adiposity measures across age groups by sex. Body fat percentage (BF) in children is derived from equation (4.1). BMI = body mass index; WHR = waist hip ratio. Data is presented for all individuals given in Table 4.3. The interquartile range and median are also shown. Raincloud plots produced using the RainCloudPlots R package⁶³².

5481 **Forestplots**

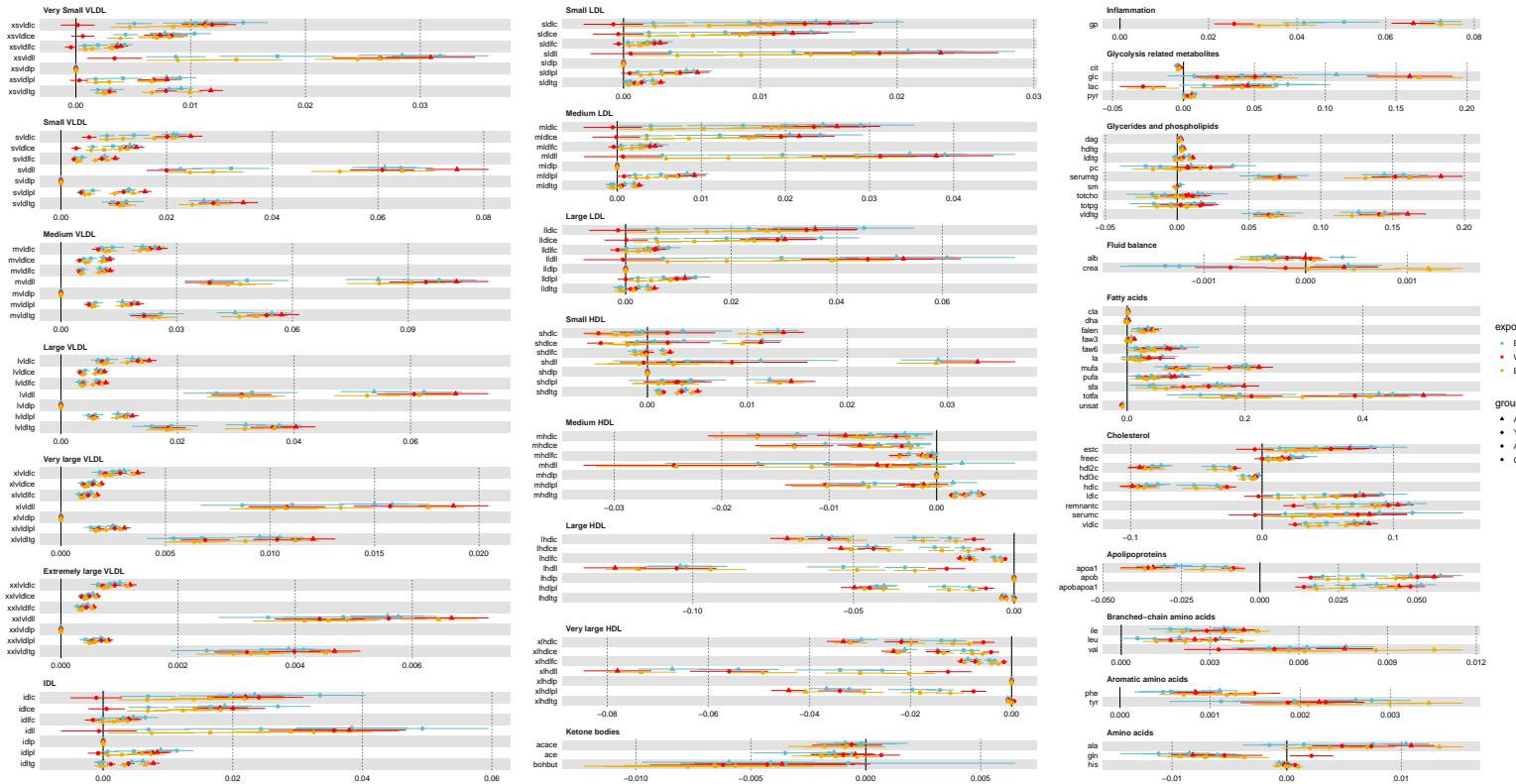


Figure A.6: Effect estimates from linear regression of adiposity measures on metabolites. Forestplot shows effect estimates and 95% confidence intervals from model 2 for all exposures and age groups. Effect estimates are absolute change in metabolite per-standard deviation increase in exposure.



Figure A.7: Effect estimates from linear regression of adiposity measures on metabolites. Forestplot shows effect estimates and 95% confidence intervals from model 2 for all exposures and age groups. Effect estimates are absolute change in metabolite per-standard deviation increase in exposure.

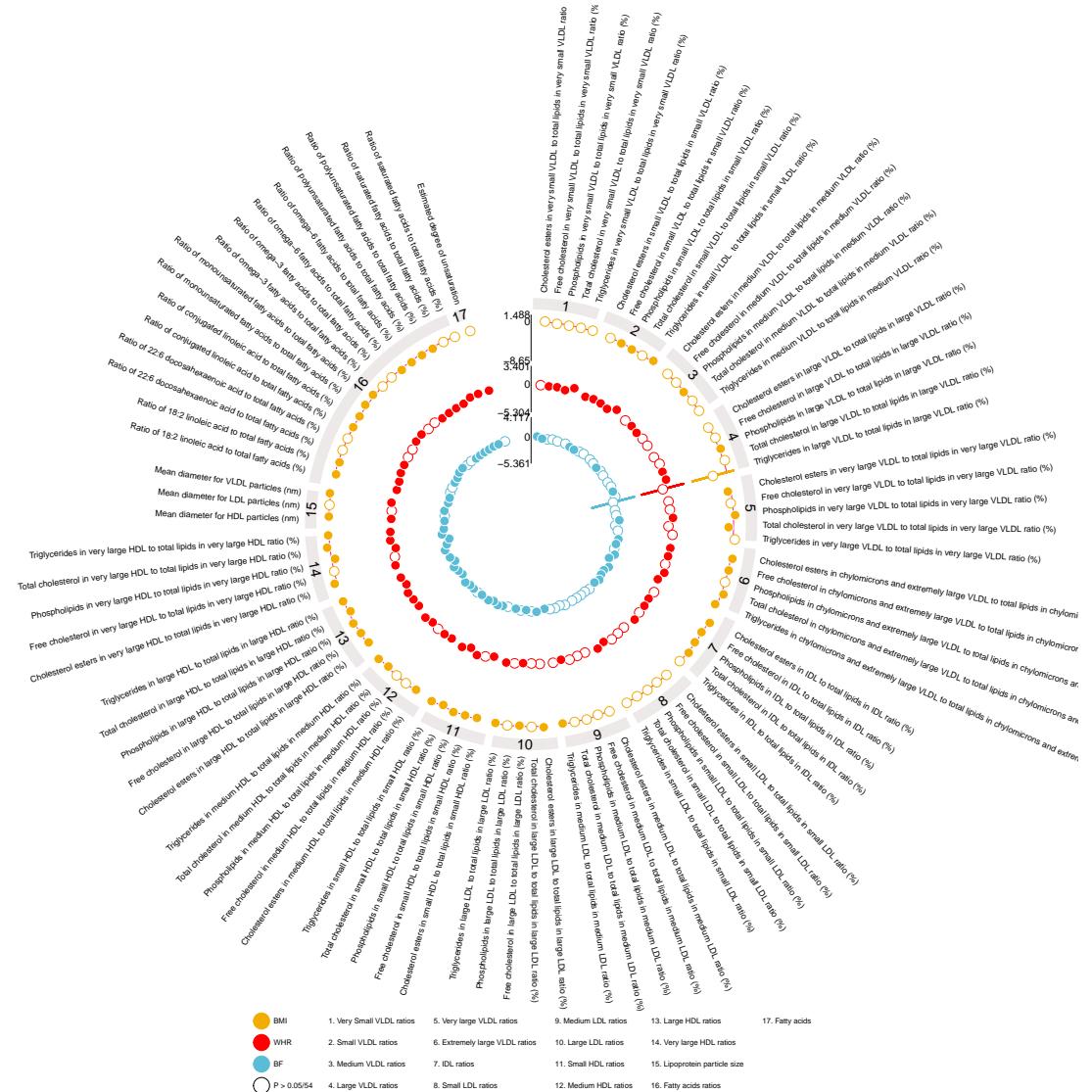


Figure A.8: Effect estimates from linear regression of adiposity measures on derived supplementary metabolites in children. Circos plot shows effect estimates and 95% confidence intervals from model 2 for all exposures in children. Effect estimates are absolute change in metabolite per-standard deviation increase in exposure. The outer track is body mass index (BMI), the middle track is waist hip ratio (WHR), the inner track is body fat percentage (BF). Solid points indicate a multiple testing threshold has been met; multiple testing threshold (0.05/54) set as the number of independent metabolite features as calculated by metaboprep.

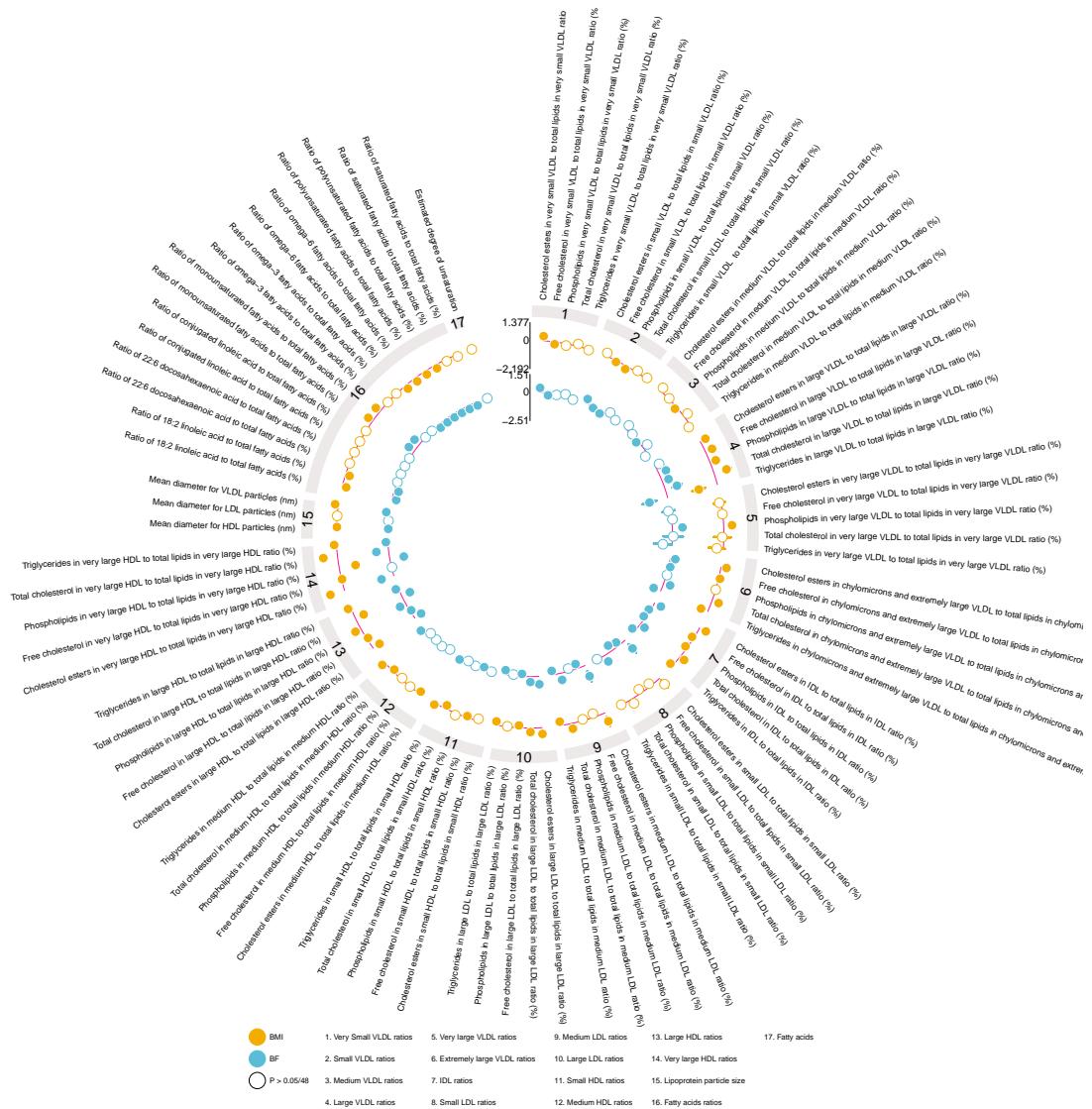


Figure A.9: Effect estimates from linear regression of adiposity measures on derived supplementary metabolites in adolescents. Circos plot shows effect estimates and 95% confidence intervals from model 2 for all exposures in adolescents. Effect estimates are absolute change in metabolite per-standard deviation increase in exposure. The outer track is body mass index (BMI), the inner track is body fat percentage (BF). Solid points indicate a multiple testing threshold has been met; multiple testing threshold (0.05/48) set as the number of independent metabolite features as calculated by metaboprep.

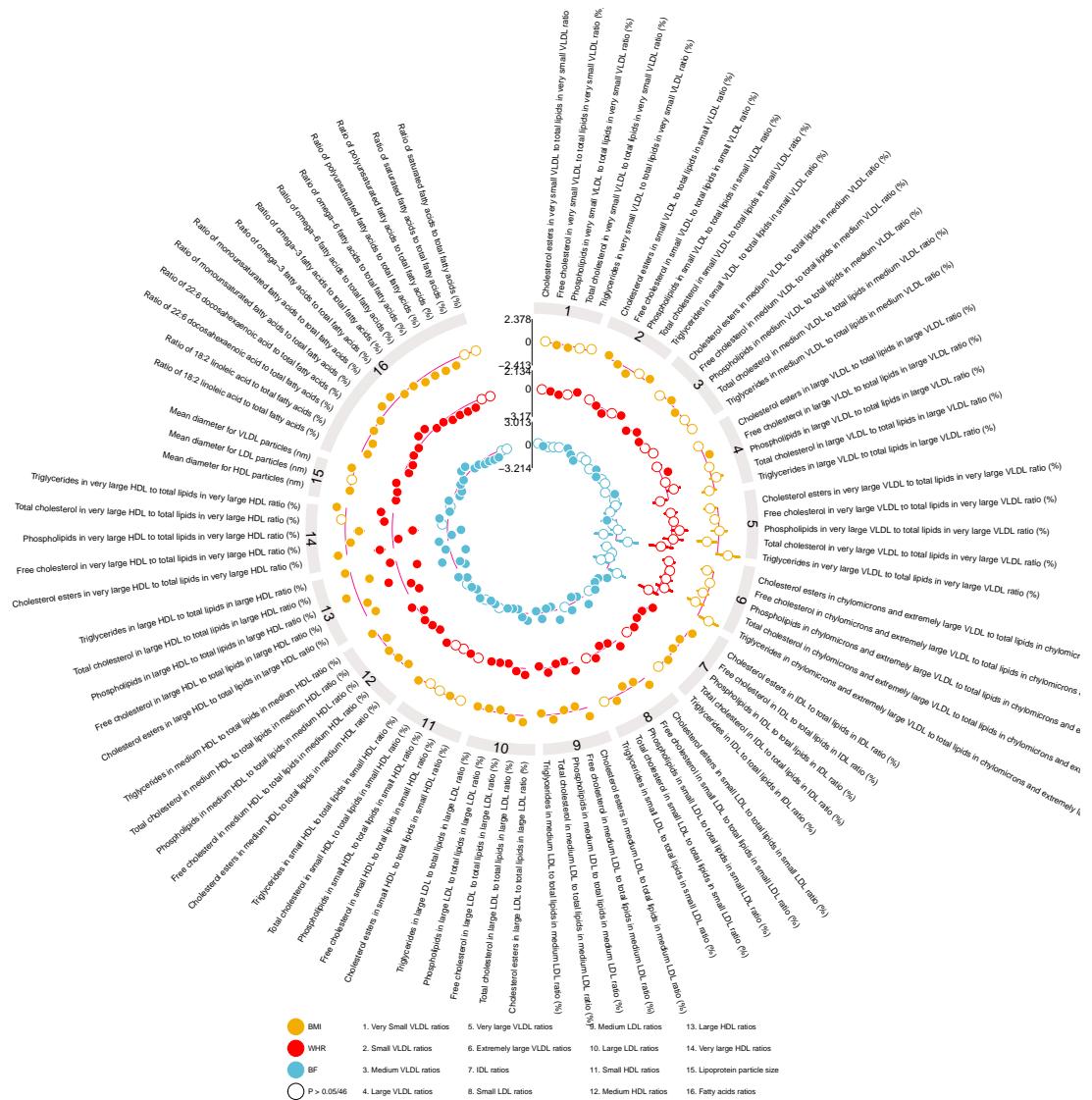


Figure A.10: Effect estimates from linear regression of adiposity measures on derived supplementary metabolites in young adults. Circos plot shows effect estimates and 95% confidence intervals from model 2 for all exposures in young adults. Effect estimates are absolute change in metabolite per-standard deviation increase in exposure. The outer track is body mass index (BMI), the middle track is waist hip ratio (WHR), the inner track is body fat percentage (BF). Solid points indicate a multiple testing threshold has been met; multiple testing threshold (0.05/46) set as the number of independent metabolite features as calculated by metaboprep.

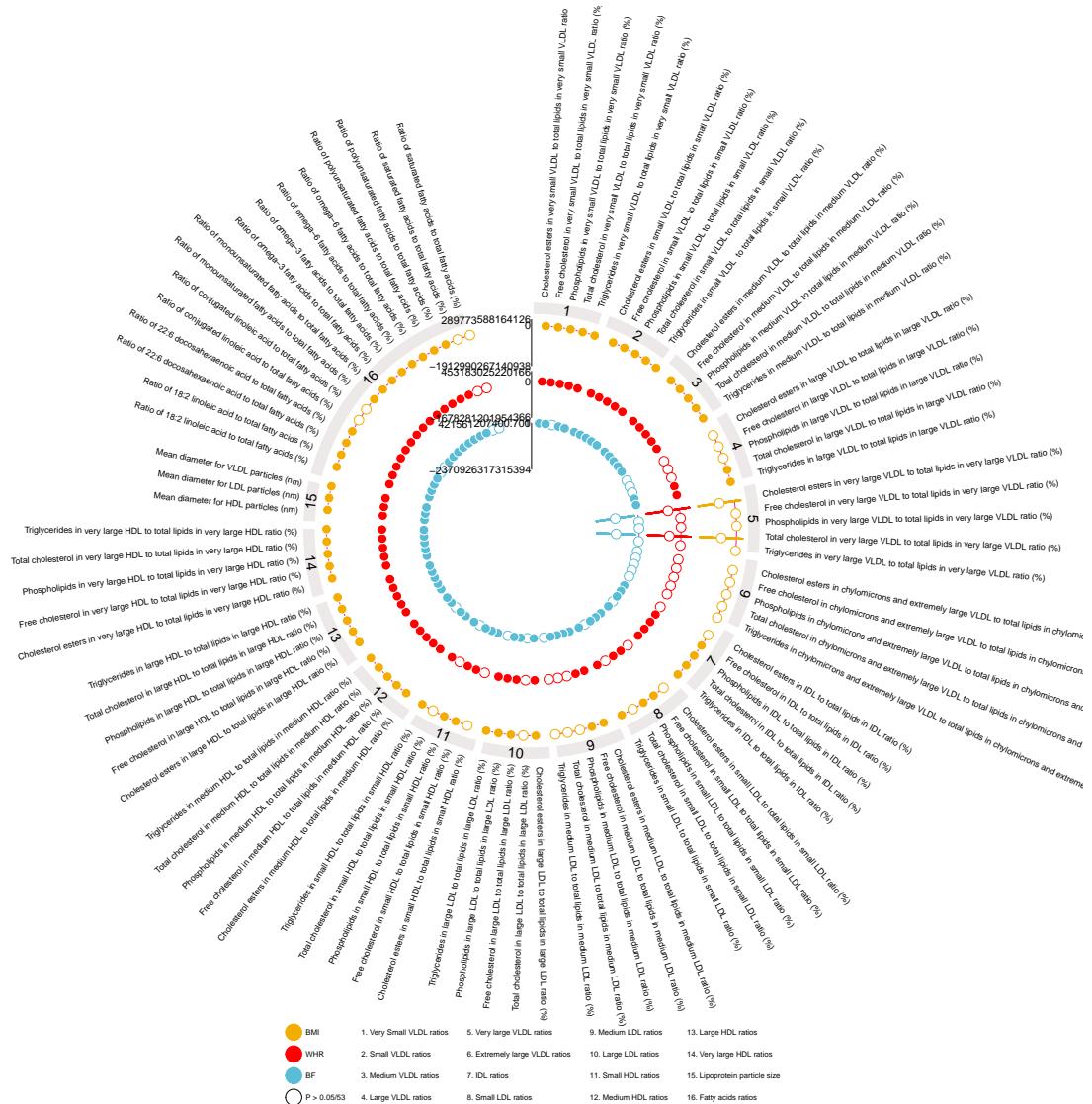


Figure A.11: Effect estimates from linear regression of adiposity measures on derived supplementary metabolites in adults. Circos plot shows effect estimates and 95% confidence intervals from model 2 for all exposures in adults. Effect estimates are absolute change in metabolite per-standard deviation increase in exposure. The outer track is body mass index (BMI), the middle track is waist hip ratio (WHR), the inner track is body fat percentage (BF). Solid points indicate a multiple testing threshold has been met; multiple testing threshold (0.05/53) set as the number of independent metabolite features as calculated by `metaboprep`.

5483 **A.4 Chapter 5: MR**

5484 **A.4.1 Methods**

5485 **Exposure instruments**

5486 In the main analysis BMI, WHR, and BF were instrumented using SNPs from Yengo et
5487 al. (2018)⁴¹, Pulit et al. (2019)⁴², and Lu et al. (206)³⁹ respectively. Additional instrument lists
5488 were obtained for each of the adiposity measures. A list of all instruments used is available
5489 on [GitHub](#).

5490 **Body mass index** Two additional instrument lists were obtained for BMI. The first came
5491 from the same study as used in the main analysis (Yengo et al. (2019)) but did not use the
5492 COJO GWAS summary statistics. As such, the details for this GWAS are presented in the
5493 main text; a total of 656 primary associations reaching a genome wide significance threshold
5494 of 5×10^{-8} were identified and used for the additional analysis. The second instrument list
5495 came from Locke et al. (2016), in which 322,154 individuals of European ancestries were
5496 included in an inverse variance weighted fixed effects meta-analysis. A total of 82 GWAS and
5497 43 Metabochip studies were included in the meta-analysis. Individual GWAS were adjusted
5498 for age, age², and study specific covariates with residuals inverse normally transformed.
5499 Imputations was performed using HapMap phase II CEU reference panel. Each study used
5500 a linear regression model assuming an additive genetic model with quality control following
5501 procedures outlined previously⁵⁷¹. A fixed effect inverse variance weighted meta-analysis
5502 was performed using METAL for the 82 GWAS and 43 Metabochip studies separately. Study
5503 GWAS and meta-analysis results were corrected for genomic control using all SNPs; study
5504 Metabochip and meta-analysis of Metabochip results were corrected for genomic control
5505 using 4,425 SNPs included on Metabochip for replication of associations with QT-interval, a
5506 phenotype not correlated with BMI, after pruning SNPs within 500kb of a BMI replication SNP.
5507 The final meta-analysis combined the genomic control corrected GWAS and Metabochip
5508 meta-analysis results. A total of 77 loci reaching genome-wide significance (5×10^{-8}) and
5509 separated by at least 500 kilobases were identified.

5510 **Waist hip ratio** The main analysis used summary statistics for waist hip ratio (WHR) from
5511 Pulit et al. (2019). This included the meta-analysis of published summary statistics from
5512 Shungin et al. (2016) which was used as an additional analysis of WHR here. A total of
5513 210,088 individuals of European ancestries were included in an inverse variance weighted
5514 fixed effects meta-analysis. A total of 57 GWAS and 44 Metabochip studies were included in
5515 teh meta-analysis. WHR was adjusted for age, age², study-specific covariates if necessary
5516 with residuals inverse normally transformed. Imputations was performed using HapMap phase

5517 II CEU reference panel. Each study used a linear regression model assuming an additive
5518 genetic model. Study GWAS and meta-analysis results were corrected for genomic control
5519 using all SNPs; study Metabochip and meta-analysis of Metabochip results were corrected
5520 for genomic control using 4,425 SNPs included on Metabochip for replication of associations
5521 with QT-interval, a phenotype not correlated with WHR, after pruning SNPs within 500kb of a
5522 BMI replication SNP. The final meta-analysis combined the genomic control corrected GWAS
5523 and Metabochip meta-analysis results. A total of 26 loci reaching genome-wide significance
5524 (5×10^{-8}) and separated by at least 500 kilobases were identified.

5525 **Body fat percentage** The main analysis for body fat percentage (BF) used summary
5526 statistics from Lu et al. (2016) which utilised 7 SNPs. Additional analyses included the
5527 same summary statistics but after removal of two SNPs which had been associated with
5528 'favourable adiposity'^{35,539} as well as a 76 SNPs identified by Hubel et al. (2019)⁴³ who
5529 performed a GWAS of BF in UK Biobank. In UK Biobank, body composition was assessed
5530 using bioelectrical impedance (Tanita BC-418 MA; Tanita Corporation, Arlington Heights, IL).
5531 Fat mass and fat free mass was calculated from raw impedance data adjusting for age, sex,
5532 height and athleticism. Assuming an additive genetic model, sex-specific GWASs (female =
5533 70,700; male = 85,261) of BF were conducted using BGENIE (V1.2)3 in healthy and drug
5534 free individuals of European ancestry descent from UK Biobank. European ancestry was
5535 defined by k-means clustering of the first two principal components from the genetic data.
5536 This equated to 32% of the genotyped UK Biobank population. In total 7,794,483 SNPs with
5537 a MAF > 1%, imputation quality score > 0.8 were included. BF residuals were adjusted for
5538 age, socioeconomic status (measured by Townsend deprivation index), assessment centre,
5539 genotyping batch, smoking status, alcohol consumption, menopause and the first 6 principal
5540 components. Genotypes were imputed to the HRC. A fixed effect inverse variance weighted
5541 meta-analysis of the sex-specific GWASs was performed using METAL6. Associations for
5542 the sex-combined analysis were considered if reaching genome wide significance (5×10^{-8}).
5543 Linkage disequilibrium score regression (LDSC intercept = 1.05; SE = 0.01) suggested no
5544 effect of population stratification. They identified 76 associations reaching a genome wide
5545 significance threshold of 5×10^{-8} .

Table A.5: Metabolites used in MR and meta-analyses

Metabolite	Label	Class	Subclass	Study
ala	Alanine (mmol/l)	Amino acids	Amino acids	Both
gln	Glutamine (mmol/l)	Amino acids	Amino acids	Both
gly	Glycine (mmol/l)	Amino acids	Amino acids	Interval
his	Histidine (mmol/l)	Amino acids	Amino acids	Both
phe	Phenylalanine (mmol/l)	Amino acids	Aromatic amino acids	Both
tyr	Tyrosine (mmol/l)	Amino acids	Aromatic amino acids	Both
ile	Isoleucine (mmol/l)	Amino acids	Branched-chain amino acids	Both
leu	Leucine (mmol/l)	Amino acids	Branched-chain amino acids	Both
val	Valine (mmol/l)	Amino acids	Branched-chain amino acids	Both
apoA1	Apolipoprotein A-I (g/l)	Apolipoproteins	Apolipoproteins	Both
apoB	Apolipoprotein B (g/l)	Apolipoproteins	Apolipoproteins	Both
estc	Esterified cholesterol (mmol/l)	Cholesterol	Cholesterol	Both
freec	Free cholesterol (mmol/l)	Cholesterol	Cholesterol	Both
hdl2c	Total cholesterol in HDL2 (mmol/l)	Cholesterol	Cholesterol	Interval
hdl3c	Total cholesterol in HDL3 (mmol/l)	Cholesterol	Cholesterol	Interval
hdlc	Total cholesterol in HDL (mmol/l)	Cholesterol	Cholesterol	Both
ldlc	Total cholesterol in LDL (mmol/l)	Cholesterol	Cholesterol	Both
remnanc	Remnant cholesterol (non-HDL, non-LDL -cholesterol) (mmol/l)	Cholesterol	Cholesterol	Interval
serumc	Serum total cholesterol (mmol/l)	Cholesterol	Cholesterol	Both
vldlc	Total cholesterol in VLDL (mmol/l)	Cholesterol	Cholesterol	Interval
cla	Conjugated linoleic acid (mmol/l)	Fatty acids	Fatty acids	Interval
dha	22:6, docosahexaenoic acid (mmol/l)	Fatty acids	Fatty acids	Both
falen	Estimated description of fatty acid chain length, not actual carbon number	Fatty acids	Fatty acids	Both
faw3	Omega-3 fatty acids (mmol/l)	Fatty acids	Fatty acids	Both
faw6	Omega-6 fatty acids (mmol/l)	Fatty acids	Fatty acids	Both
la	18:2, linoleic acid (mmol/l)	Fatty acids	Fatty acids	Both
mufa	Monounsaturated fatty acids; 16:1, 18:1 (mmol/l)	Fatty acids	Fatty acids	Both
pufa	Polyunsaturated fatty acids (mmol/l)	Fatty acids	Fatty acids	Interval
sfa	Saturated fatty acids (mmol/l)	Fatty acids	Fatty acids	Interval
totfa	Total fatty acids (mmol/l)	Fatty acids	Fatty acids	Both
alb	Albumin (signal area)	Fluid balance	Fluid balance	Both
crea	Creatinine (mmol/l)	Fluid balance	Fluid balance	Both
dag	Diacylglycerol (mmol/l)	Glycerides and phospholipids	Glycerides and phospholipids	Interval
hdltg	Triglycerides in HDL (mmol/l)	Glycerides and phospholipids	Glycerides and phospholipids	Interval
ldltg	Triglycerides in LDL (mmol/l)	Glycerides and phospholipids	Glycerides and phospholipids	Interval
pc	Phosphatidylcholine and other cholines (mmol/l)	Glycerides and phospholipids	Glycerides and phospholipids	Both
serumtg	Serum total triglycerides (mmol/l)	Glycerides and phospholipids	Glycerides and phospholipids	Both
sm	Sphingomyelins (mmol/l)	Glycerides and phospholipids	Glycerides and phospholipids	Both

Table A.5: Metabolites used in MR and meta-analyses (*continued*)

Metabolite	Label	Class	Subclass	Study
totcho	Total cholines (mmol/l)	Glycerides and phospholipids	Glycerides and phospholipids	Interval
totpg	Total phosphoglycerides (mmol/l)	Glycerides and phospholipids	Glycerides and phospholipids	Both
vldltg	Triglycerides in VLDL (mmol/l)	Glycerides and phospholipids	Glycerides and phospholipids	Interval
cit	Citrate (mmol/l)	Glycolysis related metabolites	Glycolysis related metabolites	Kettunen
glc	Glucose (mmol/l)	Glycolysis related metabolites	Glycolysis related metabolites	Both
glol	Glycerol (mmol/l)	Glycolysis related metabolites	Glycolysis related metabolites	Kettunen
lac	Lactate (mmol/l)	Glycolysis related metabolites	Glycolysis related metabolites	Both
pyr	Pyruvate (mmol/l)	Glycolysis related metabolites	Glycolysis related metabolites	Both
gp	Glycoprotein acetyls, mainly α1-acid glycoprotein (mmol/l)	Inflammation	Inflammation	Kettunen
acace	Acetoacetate (mmol/l)	Ketone bodies	Ketone bodies	Both
ace	Acetate (mmol/l)	Ketone bodies	Ketone bodies	Both
bohbut	3-hydroxybutyrate (mmol/l)	Ketone bodies	Ketone bodies	Kettunen
hdld	Mean diameter for HDL particles (nm)	Lipoprotein particle size	Lipoprotein particle size	Both
ldld	Mean diameter for LDL particles (nm)	Lipoprotein particle size	Lipoprotein particle size	Both
vldld	Mean diameter for VLDL particles (nm)	Lipoprotein particle size	Lipoprotein particle size	Both
xxlvldlc	Total cholesterol in chylomicrons and extremely large VLDL (mmol/l)	Lipoprotein subclasses	Extremely large VLDL	Interval
xxvldlce	Cholesterol esters in chylomicrons and extremely large VLDL (mmol/l)	Lipoprotein subclasses	Extremely large VLDL	Interval
xxvldlfc	Free cholesterol in chylomicrons and extremely large VLDL (mmol/l)	Lipoprotein subclasses	Extremely large VLDL	Interval
xxvldll	Total lipids in chylomicrons and extremely large VLDL (mmol/l)	Lipoprotein subclasses	Extremely large VLDL	Both
xxvldlp	Concentration of chylomicrons and extremely large VLDL particles (mol/l)	Lipoprotein subclasses	Extremely large VLDL	Both
xxvldlpl	Phospholipids in chylomicrons and extremely large VLDL (mmol/l)	Lipoprotein subclasses	Extremely large VLDL	Both
xxvldltg	Triglycerides in chylomicrons and extremely large VLDL (mmol/l)	Lipoprotein subclasses	Extremely large VLDL	Both
idlc	Total cholesterol in IDL (mmol/l)	Lipoprotein subclasses	IDL	Both
idlce	Cholesterol esters in IDL (mmol/l)	Lipoprotein subclasses	IDL	Interval
idlfc	Free cholesterol in IDL (mmol/l)	Lipoprotein subclasses	IDL	Both
idll	Total lipids in IDL (mmol/l)	Lipoprotein subclasses	IDL	Both
idlpl	Concentration of IDL particles (mol/l)	Lipoprotein subclasses	IDL	Both
idlpl	Phospholipids in IDL (mmol/l)	Lipoprotein subclasses	IDL	Both
idlgt	Triglycerides in IDL (mmol/l)	Lipoprotein subclasses	IDL	Both
lhdlc	Total cholesterol in large HDL (mmol/l)	Lipoprotein subclasses	Large HDL	Both
lhdlce	Cholesterol esters in large HDL (mmol/l)	Lipoprotein subclasses	Large HDL	Both
lhdlfc	Free cholesterol in large HDL (mmol/l)	Lipoprotein subclasses	Large HDL	Both
lhdlpl	Total lipids in large HDL (mmol/l)	Lipoprotein subclasses	Large HDL	Both
lhdlgt	Concentration of large HDL particles (mol/l)	Lipoprotein subclasses	Large HDL	Both
lhdlpl	Phospholipids in large HDL (mmol/l)	Lipoprotein subclasses	Large HDL	Both
lhdlgt	Triglycerides in large HDL (mmol/l)	Lipoprotein subclasses	Large HDL	Interval
lldlc	Total cholesterol in large LDL (mmol/l)	Lipoprotein subclasses	Large LDL	Both
lldlce	Cholesterol esters in large LDL (mmol/l)	Lipoprotein subclasses	Large LDL	Both

Table A.5: Metabolites used in MR and meta-analyses (*continued*)

Metabolite	Label	Class	Subclass	Study
lldfc	Free cholesterol in large LDL (mmol/l)	Lipoprotein subclasses	Large LDL	Both
lldll	Total lipids in large LDL (mmol/l)	Lipoprotein subclasses	Large LDL	Both
lldlp	Concentration of large LDL particles (mol/l)	Lipoprotein subclasses	Large LDL	Both
lldpl	Phospholipids in large LDL (mmol/l)	Lipoprotein subclasses	Large LDL	Both
lldtg	Triglycerides in large LDL (mmol/l)	Lipoprotein subclasses	Large LDL	Interval
lvldc	Total cholesterol in large VLDL (mmol/l)	Lipoprotein subclasses	Large VLDL	Both
lvldce	Cholesterol esters in large VLDL (mmol/l)	Lipoprotein subclasses	Large VLDL	Both
lvldfc	Free cholesterol in large VLDL (mmol/l)	Lipoprotein subclasses	Large VLDL	Both
lvldll	Total lipids in large VLDL (mmol/l)	Lipoprotein subclasses	Large VLDL	Both
lvldlp	Concentration of large VLDL particles (mol/l)	Lipoprotein subclasses	Large VLDL	Both
lvldpl	Phospholipids in large VLDL (mmol/l)	Lipoprotein subclasses	Large VLDL	Both
lvldtg	Triglycerides in large VLDL (mmol/l)	Lipoprotein subclasses	Large VLDL	Both
mhdlc	Total cholesterol in medium HDL (mmol/l)	Lipoprotein subclasses	Medium HDL	Both
mhdlc	Cholesterol esters in medium HDL (mmol/l)	Lipoprotein subclasses	Medium HDL	Both
mhdlfc	Free cholesterol in medium HDL (mmol/l)	Lipoprotein subclasses	Medium HDL	Both
mhdll	Total lipids in medium HDL (mmol/l)	Lipoprotein subclasses	Medium HDL	Both
mhdlp	Concentration of medium HDL particles (mol/l)	Lipoprotein subclasses	Medium HDL	Both
mhdpl	Phospholipids in medium HDL (mmol/l)	Lipoprotein subclasses	Medium HDL	Both
mhdltg	Triglycerides in medium HDL (mmol/l)	Lipoprotein subclasses	Medium HDL	Interval
mldlc	Total cholesterol in medium LDL (mmol/l)	Lipoprotein subclasses	Medium LDL	Both
mldlc	Cholesterol esters in medium LDL (mmol/l)	Lipoprotein subclasses	Medium LDL	Both
mldfc	Free cholesterol in medium LDL (mmol/l)	Lipoprotein subclasses	Medium LDL	Interval
mldll	Total lipids in medium LDL (mmol/l)	Lipoprotein subclasses	Medium LDL	Both
mldlp	Concentration of medium LDL particles (mol/l)	Lipoprotein subclasses	Medium LDL	Both
mldpl	Phospholipids in medium LDL (mmol/l)	Lipoprotein subclasses	Medium LDL	Both
mldtg	Triglycerides in medium LDL (mmol/l)	Lipoprotein subclasses	Medium LDL	Interval
mvldlc	Total cholesterol in medium VLDL (mmol/l)	Lipoprotein subclasses	Medium VLDL	Both
mvldlc	Cholesterol esters in medium VLDL (mmol/l)	Lipoprotein subclasses	Medium VLDL	Both
mvldfc	Free cholesterol in medium VLDL (mmol/l)	Lipoprotein subclasses	Medium VLDL	Both
mvldll	Total lipids in medium VLDL (mmol/l)	Lipoprotein subclasses	Medium VLDL	Both
mvldlp	Concentration of medium VLDL particles (mol/l)	Lipoprotein subclasses	Medium VLDL	Both
mvldpl	Phospholipids in medium VLDL (mmol/l)	Lipoprotein subclasses	Medium VLDL	Both
mvldtg	Triglycerides in medium VLDL (mmol/l)	Lipoprotein subclasses	Medium VLDL	Both
shdlc	Total cholesterol in small HDL (mmol/l)	Lipoprotein subclasses	Small HDL	Interval
shdlce	Cholesterol esters in small HDL (mmol/l)	Lipoprotein subclasses	Small HDL	Interval
shdlfc	Free cholesterol in small HDL (mmol/l)	Lipoprotein subclasses	Small HDL	Interval
shdlll	Total lipids in small HDL (mmol/l)	Lipoprotein subclasses	Small HDL	Both
shdlp	Concentration of small HDL particles (mol/l)	Lipoprotein subclasses	Small HDL	Both
shdlpl	Phospholipids in small HDL (mmol/l)	Lipoprotein subclasses	Small HDL	Interval
shdltg	Triglycerides in small HDL (mmol/l)	Lipoprotein subclasses	Small HDL	Both
sldlc	Total cholesterol in small LDL (mmol/l)	Lipoprotein subclasses	Small LDL	Both
sldlc	Cholesterol esters in small LDL (mmol/l)	Lipoprotein subclasses	Small LDL	Interval
sldfc	Free cholesterol in small LDL (mmol/l)	Lipoprotein subclasses	Small LDL	Interval
sldll	Total lipids in small LDL (mmol/l)	Lipoprotein subclasses	Small LDL	Both
sldlp	Concentration of small LDL particles (mol/l)	Lipoprotein subclasses	Small LDL	Both
sldpl	Phospholipids in small LDL (mmol/l)	Lipoprotein subclasses	Small LDL	Interval
sldtg	Triglycerides in small LDL (mmol/l)	Lipoprotein subclasses	Small LDL	Interval
svldlc	Total cholesterol in small VLDL (mmol/l)	Lipoprotein subclasses	Small VLDL	Both
svldlc	Cholesterol esters in small VLDL (mmol/l)	Lipoprotein subclasses	Small VLDL	Interval

Table A.5: Metabolites used in MR and meta-analyses (*continued*)

Metabolite	Label	Class	Subclass	Study
svldfc	Free cholesterol in small VLDL (mmol/l)	Lipoprotein subclasses	Small VLDL	Both
svldll	Total lipids in small VLDL (mmol/l)	Lipoprotein subclasses	Small VLDL	Both
svldlp	Concentration of small VLDL particles (mol/l)	Lipoprotein subclasses	Small VLDL	Both
svldlpl	Phospholipids in small VLDL (mmol/l)	Lipoprotein subclasses	Small VLDL	Both
svldtg	Triglycerides in small VLDL (mmol/l)	Lipoprotein subclasses	Small VLDL	Both
xlhdlc	Total cholesterol in very large HDL (mmol/l)	Lipoprotein subclasses	Very large HDL	Both
xlhdlce	Cholesterol esters in very large HDL (mmol/l)	Lipoprotein subclasses	Very large HDL	Both
xlhdlfc	Free cholesterol in very large HDL (mmol/l)	Lipoprotein subclasses	Very large HDL	Both
xlhndl	Total lipids in very large HDL (mmol/l)	Lipoprotein subclasses	Very large HDL	Both
xlhdlp	Concentration of very large HDL particles (mol/l)	Lipoprotein subclasses	Very large HDL	Both
xlhdlpl	Phospholipids in very large HDL (mmol/l)	Lipoprotein subclasses	Very large HDL	Both
xlhdtg	Triglycerides in very large HDL (mmol/l)	Lipoprotein subclasses	Very large HDL	Both
xlvldlc	Total cholesterol in very large VLDL (mmol/l)	Lipoprotein subclasses	Very large VLDL	Interval
xlvldlce	Cholesterol esters in very large VLDL (mmol/l)	Lipoprotein subclasses	Very large VLDL	Interval
xlvldlfc	Free cholesterol in very large VLDL (mmol/l)	Lipoprotein subclasses	Very large VLDL	Interval
xlvldll	Total lipids in very large VLDL (mmol/l)	Lipoprotein subclasses	Very large VLDL	Both
xlvldlp	Concentration of very large VLDL particles (mol/l)	Lipoprotein subclasses	Very large VLDL	Both
xlvldlpl	Phospholipids in very large VLDL (mmol/l)	Lipoprotein subclasses	Very large VLDL	Both
xlvldtg	Triglycerides in very large VLDL (mmol/l)	Lipoprotein subclasses	Very large VLDL	Both
xsvldlc	Total cholesterol in very small VLDL (mmol/l)	Lipoprotein subclasses	Very Small VLDL	Interval
xsvldlce	Cholesterol esters in very small VLDL (mmol/l)	Lipoprotein subclasses	Very Small VLDL	Interval
xsvldlfc	Free cholesterol in very small VLDL (mmol/l)	Lipoprotein subclasses	Very Small VLDL	Interval
xsvldll	Total lipids in very small VLDL (mmol/l)	Lipoprotein subclasses	Very Small VLDL	Both
xsvldlp	Concentration of very small VLDL particles (mol/l)	Lipoprotein subclasses	Very Small VLDL	Both
xsvldlpl	Phospholipids in very small VLDL (mmol/l)	Lipoprotein subclasses	Very Small VLDL	Both
xsvldtg	Triglycerides in very small VLDL (mmol/l)	Lipoprotein subclasses	Very Small VLDL	Both
apobbyapoa1				Interval
clabyfa				Interval
dagbytg				Interval
dhabyfa				Interval
faw3byfa				Interval
faw6byfa				Interval
glyca				Interval
labyfa				Interval
mufabyfa				Interval
pufabyfa				Interval
sfabyfa				Interval
tgbypg				Interval
unsatdeg				Interval
	Ratio of bisallylic groups to double bonds			Kettunen
	Ratio of bisallylic groups to total fatty acids			Kettunen
	Average number of methylene groups per double bond			Kettunen
	Average number of methylene groups in a fatty acid chain			Kettunen
	Average number of double bonds in a fatty acid chain			Kettunen
	Omega-7, omega-9 and saturated fatty acids			Kettunen

Table A.5: Metabolites used in MR and meta-analyses (*continued*)

Metabolite	Label	Class	Subclass	Study
	Glycoproteins			Kettunen
	Other polyunsaturated fatty acids than 18:2			Kettunen
	Urea			Kettunen

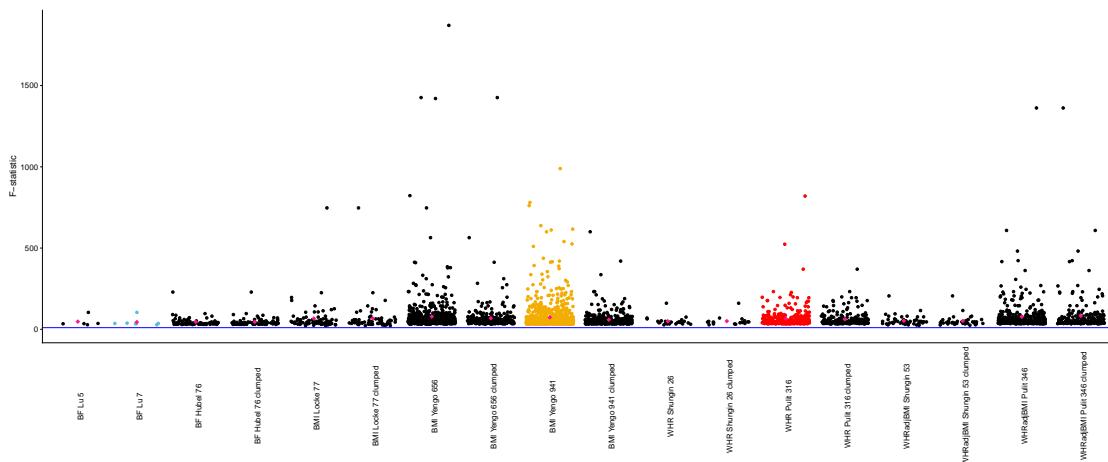
F-statistics

Figure A.12: F-statistics for individual genetic instrumental variables and mean F-statistic for each exposure used in Mendelian randomization analyses. Mean f-statistic for each exposure indicated by the pink diamond. The blue line indicates a nominal threshold of 10. Exposures used in the main analysis are highlighted with coloured points. The name after each exposure trait represents the first authors last name of the original GWAS publication for which each exposure was obtained. The number following the first authors last name represents the number of SNPs obtained from the original GWAS; clumped refers to this original number of SNPs having been pruned based on an LD R^2 of 0.001. BMI = body mass index, WHR = waist hip ratio, WHRadjBMI = waist hip ratio adjusted for BMI, BF = body fat percentage.

5548 **A.4.2 Results**

5549 **Sensitivity analyses**

5550 Assumptions of no pleiotropy were explored using MR-Egger⁵³⁵, weighted median⁵³⁶ and
5551 weighted mode⁵³⁷ based estimators. Globally, sensitivity analysis was visually reflective of
5552 the main analysis for each exposure, though with wider confidence intervals. Confidence
5553 intervals for sensitivity analyses tended to cross the null and were widest for MR Egger,
5554 which is unsurprising given the lower power afforded with this model. Sensitivity results for
5555 WHR appeared to show most consistency with the main analysis for both the Kettunen and
5556 INTERVAL analyses; confidence intervals for weighted median and mode models did not
5557 cross the null in a majority of results for subclasses.

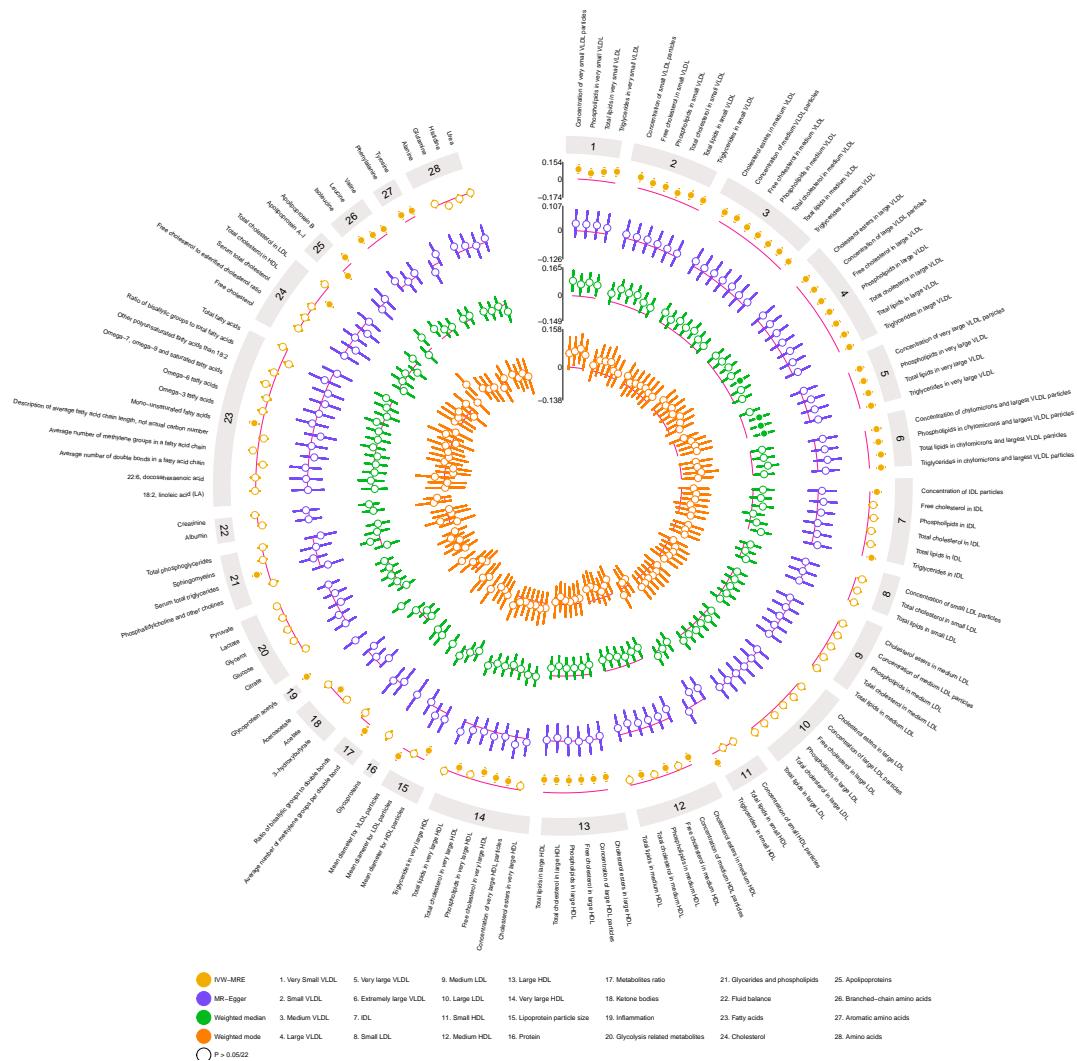


Figure A.13: Global metabolic profile of BMI on 123 NMR derived metabolites from Kettunen et al. (2016)³¹⁸: sensitivity analysis. Circos plot shows each track as one MR model; the outer track is inverse variance weighted multiplicative random effects (IVW-MRE), the second track is MR Egger, the third track is Weighted median, the inner track is Weighted mode. Solid points indicate a multiple testing threshold (0.0023) has been reached. Effect estimates represent the change in the inverse rank position of each metabolite per change in the inverse rank position of the exposure. 95% confidence intervals shown. Metabolites are grouped by subclass and arranged alphabetically within each subclass.

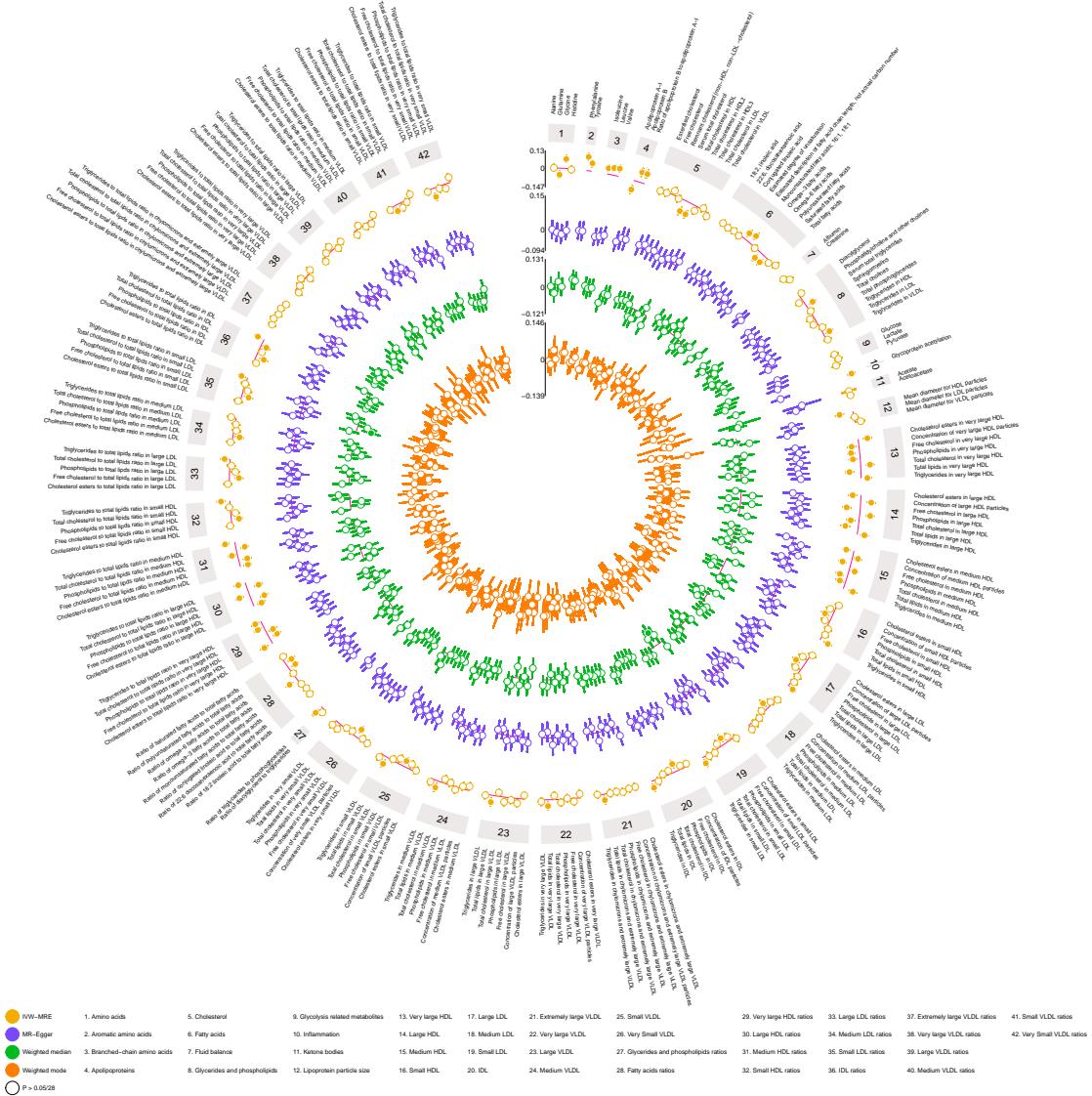


Figure A.14: Global metabolic profile of BMI on 230 NMR derived metabolites from INTERVAL: sensitivity analysis. Circos plot shows each track as one MR model; the outer track is inverse variance weighted multiplicative random effects (IVW-MRE), the second track is MR Egger, the third track is Weighted median, the inner track is Weighted mode. Solid points indicate a multiple testing threshold (2×10^{-4}) has been reached. Effect estimates represent the change in the inverse rank position of each log transformed metabolite per change in the inverse rank position of the exposure. 95% confidence intervals shown. Metabolites are grouped by subclass and arranged alphabetically within each subclass.

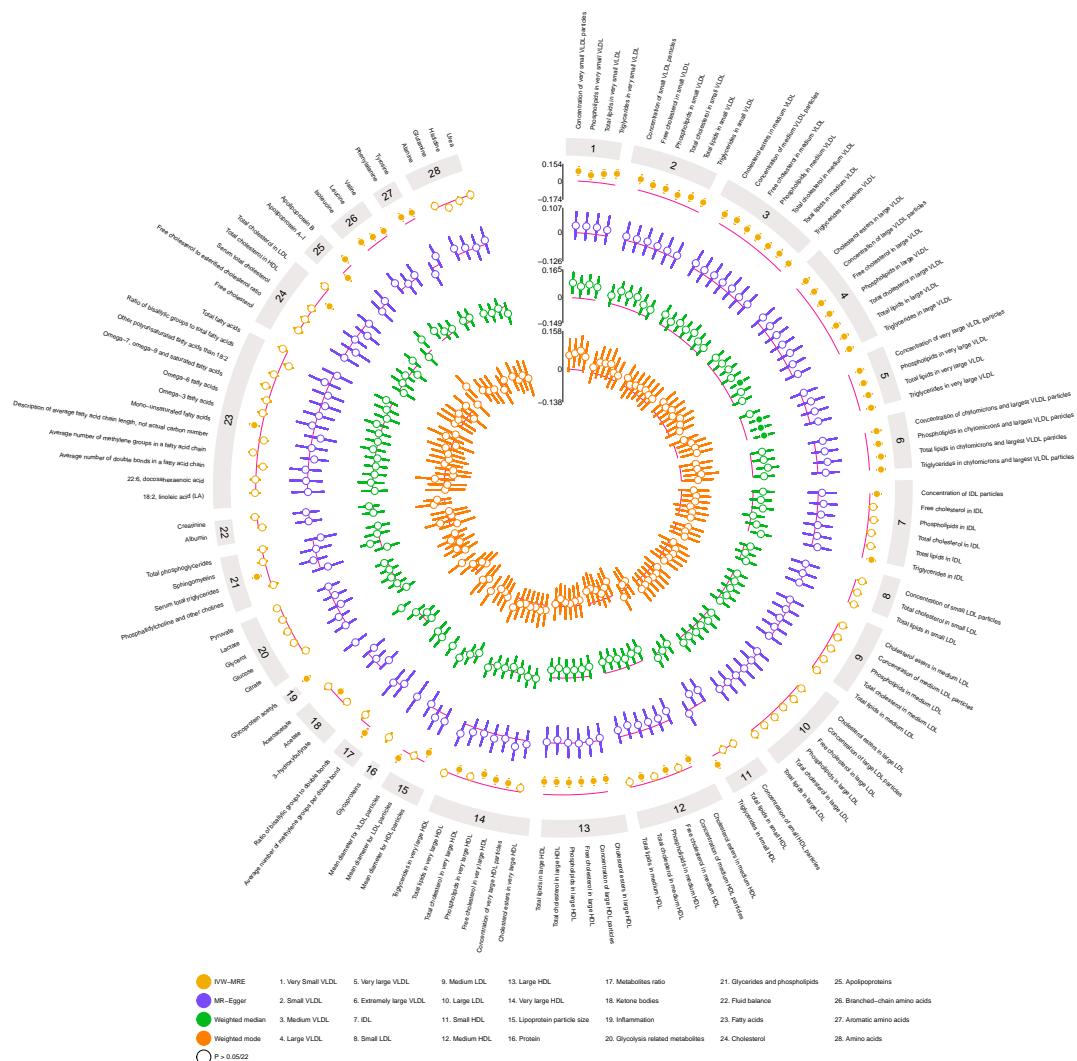


Figure A.15: Global metabolic profile of WHR on 123 NMR derived metabolites from Kettunen et al. (2016)³¹⁸: sensitivity analysis. Circos plot shows each track as one MR model; the outer track is inverse variance weighted multiplicative random effects (IVW-MRE), the second track is MR Egger, the third track is Weighted median, the inner track is Weighted mode. Solid points indicate a multiple testing threshold (0.0023) has been reached. Effect estimates represent the change in the inverse rank position of each metabolite per change in the inverse rank position of the exposure. 95% confidence intervals shown. Metabolites are grouped by subclass and arranged alphabetically within each subclass.

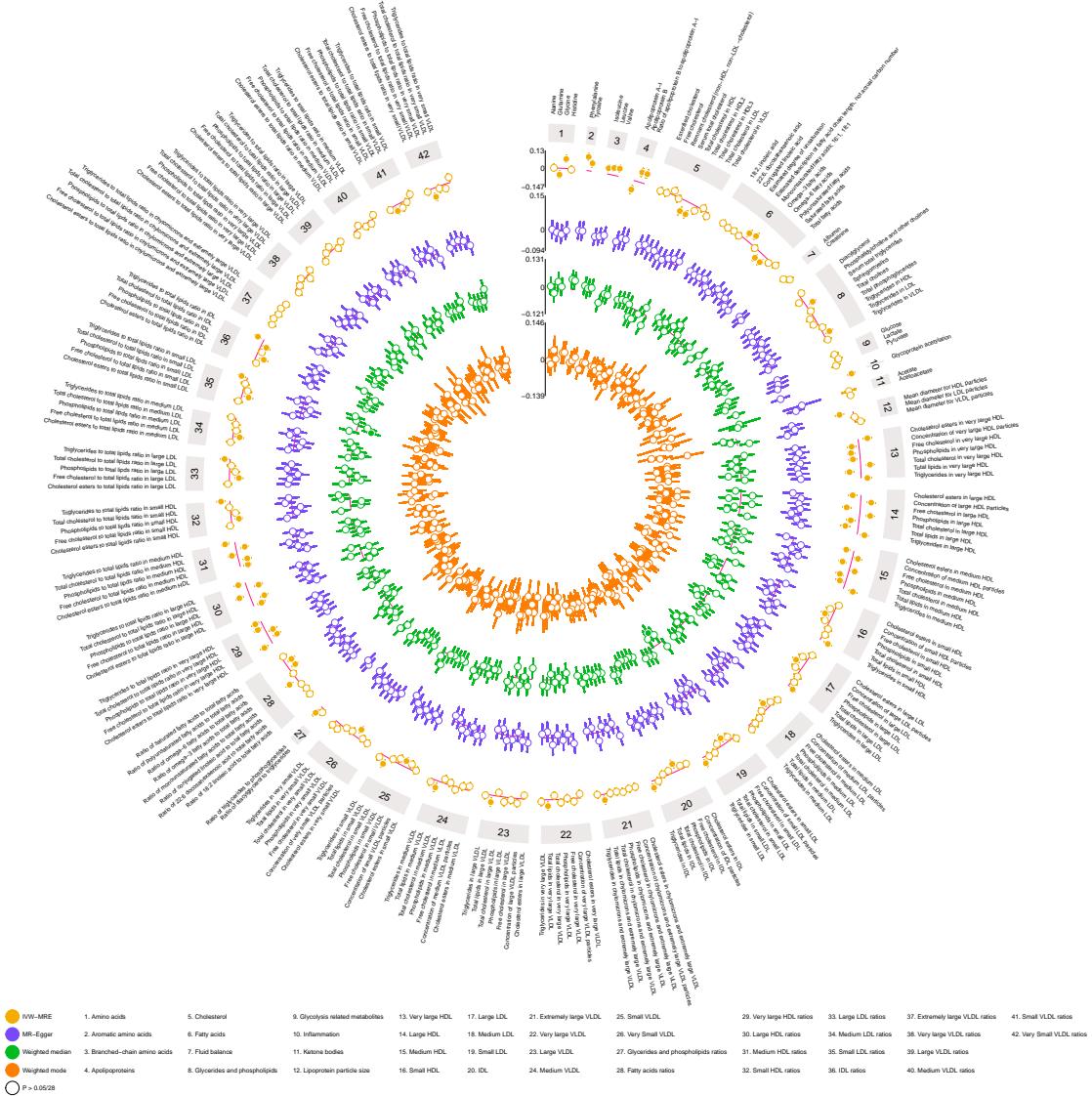


Figure A.16: Global metabolic profile of WHR on 230 NMR derived metabolites from INTERVAL: sensitivity analysis. Circos plot shows each track as one MR model; the outer track is inverse variance weighted multiplicative random effects (IVW-MRE), the second track is MR Egger, the third track is Weighted median, the inner track is Weighted mode. Solid points indicate a multiple testing threshold (2×10^{-4}) has been reached. Effect estimates represent the change in the inverse rank position of each log transformed metabolite per change in the inverse rank position of the exposure. 95% confidence intervals shown. Metabolites are grouped by subclass and arranged alphabetically within each subclass.



Figure A.17: Global metabolic profile of BF on 123 NMR derived metabolites from Kettunen et al. (2016)³¹⁸: sensitivity analysis. Circos plot shows each track as one MR model; the outer track is inverse variance weighted multiplicative random effects (IVW-MRE), the second track is MR Egger, the third track is Weighted median, the inner track is Weighted mode. Solid points indicate a multiple testing threshold (0.0023) has been reached. Effect estimates represent the change in the inverse rank position of each metabolite per change in the inverse rank position of the exposure. 95% confidence intervals shown. Metabolites are grouped by subclass and arranged alphabetically within each subclass.

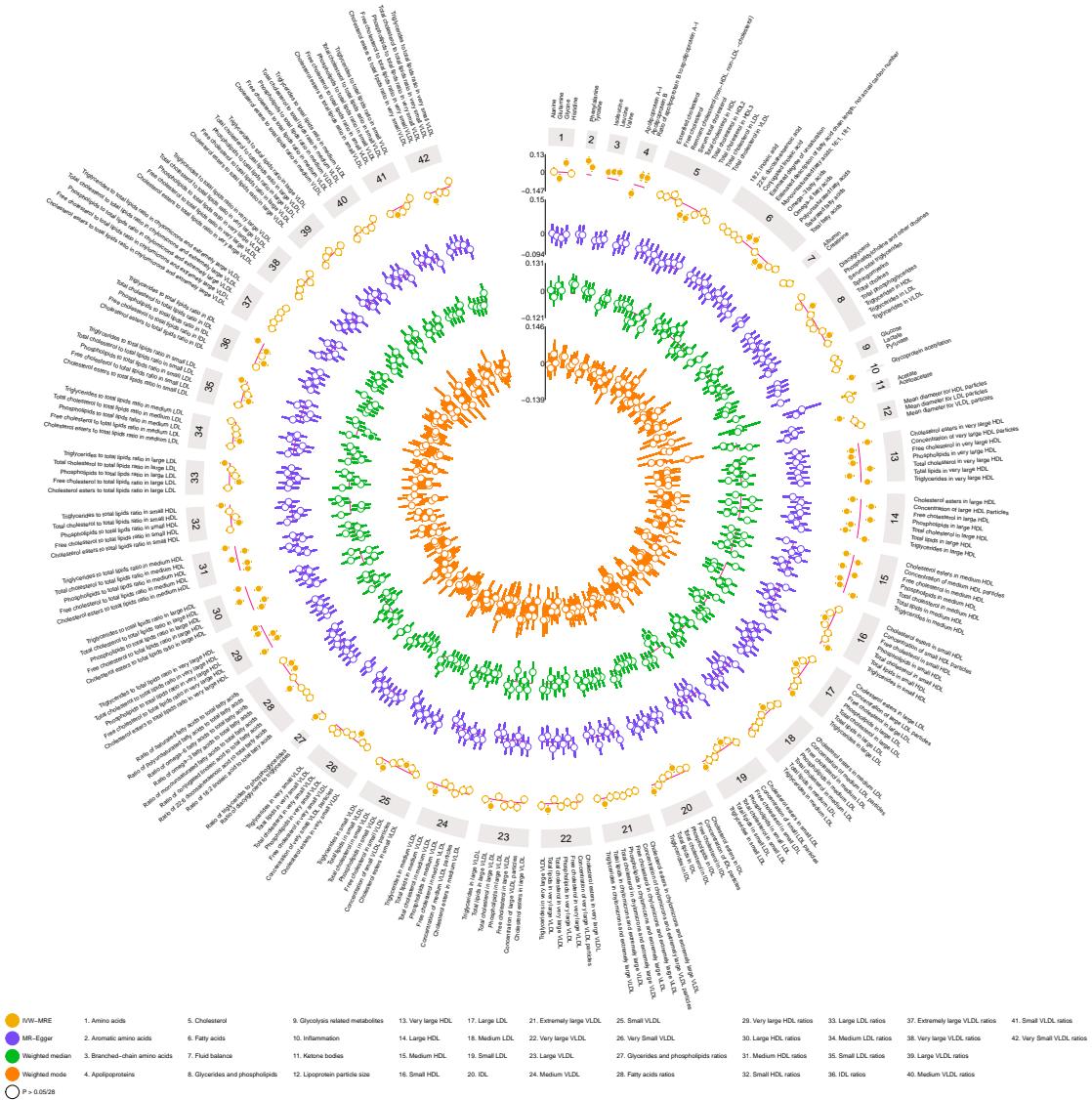


Figure A.18: Global metabolic profile of BF on 230 NMR derived metabolites from INTERVAL: sensitivity analysis. Circos plot shows each track as one MR model; the outer track is inverse variance weighted multiplicative random effects (IVW-MRE), the second track is MR Egger, the third track is Weighted median, the inner track is Weighted mode. Solid points indicate a multiple testing threshold (2×10^{-4}) has been reached. Effect estimates represent the change in the inverse rank position of each log transformed metabolite per change in the inverse rank position of the exposure. 95% confidence intervals shown. Metabolites are grouped by subclass and arranged alphabetically within each subclass.

5561 **Additional sensitivity analyses**

5562 **Kettunen** In single SNP MR for the Kettunen analysis, visual inspection of forest plots
5563 showed *S* shaped distributions of effect estimates for all tests (Representative Figure A.19).
5564 Effect estimates for some SNPs in the single SNP MR analysis appeared to be outliers. For
5565 example, for BMI on Glycoproteins, rs4673553 showed a disproportionately larger effect
5566 estimate of 22 (standard error = 0.85; *p*-value = 5.66×10^{-148}) when compared to other SNPs.
5567 Additionally, rs7777102 showed a disproportionately larger effect estimate of -9 for Mean
5568 diameter for VLDL particles (standard error = 1.14; *p*-value = 6.15×10^{-17}). When looking
5569 at the median effect size across all metabolites for each SNP, a number of SNPs (including
5570 rs7777102 for BMI) showed larger median effect sizes across the board. Funnel plots did not
5571 highlight outlying SNPs, but did reflect some SNPs having larger effect estimates across the
5572 board (Representative Figure A.20. The low number of SNPs used for BF did not result in
5573 meaningfully interpretable funnel plots (Representative Figure ??).

5574 Although a number of SNPs showed disproportionately larger effect sizes, in leave-one-
5575 out analysis, visual inspection of forest plots showed that no single SNP altered the direction
5576 of effect for any metabolite across exposures. For BF, confidence intervals for one or more
5577 SNPs crossed the null for every metabolite tested (Representative Figure A.21). This was not
5578 the case for BMI and WHR, where for many metabolites confidence intervals did not cross
5579 the null for any SNPs (Representative Figure A.22).

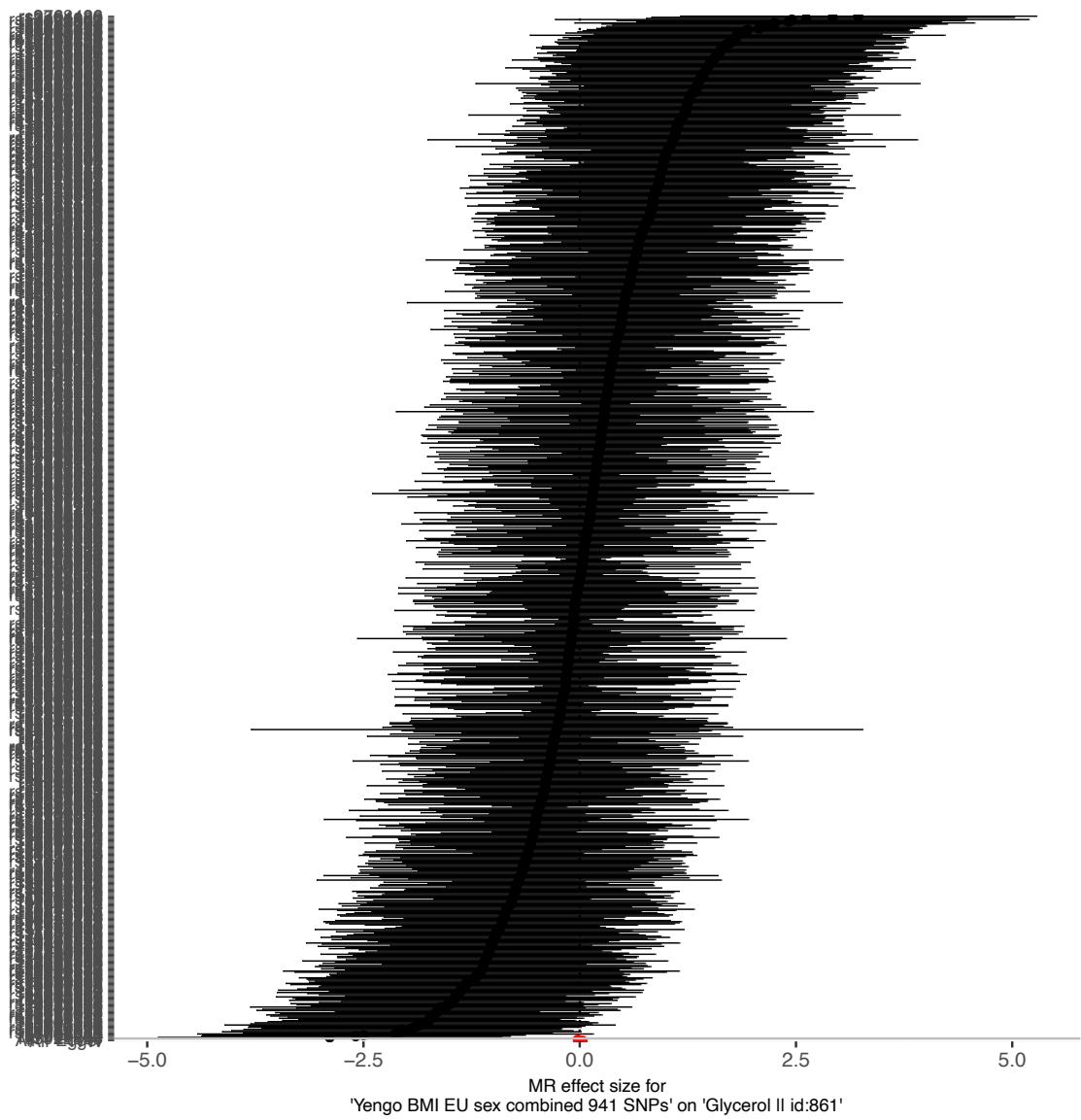


Figure A.19: Representative figure: Single SNP MR analysis of BMI on glycerol from Kettunen et al. (2016)³¹⁸. Forest plot shows the effect estimate and 95% confidence interval for each individual SNPs effect on glycerol. Effect estimates represent the change in the inverse rank position of each metabolite per change in the inverse rank position of the exposure.

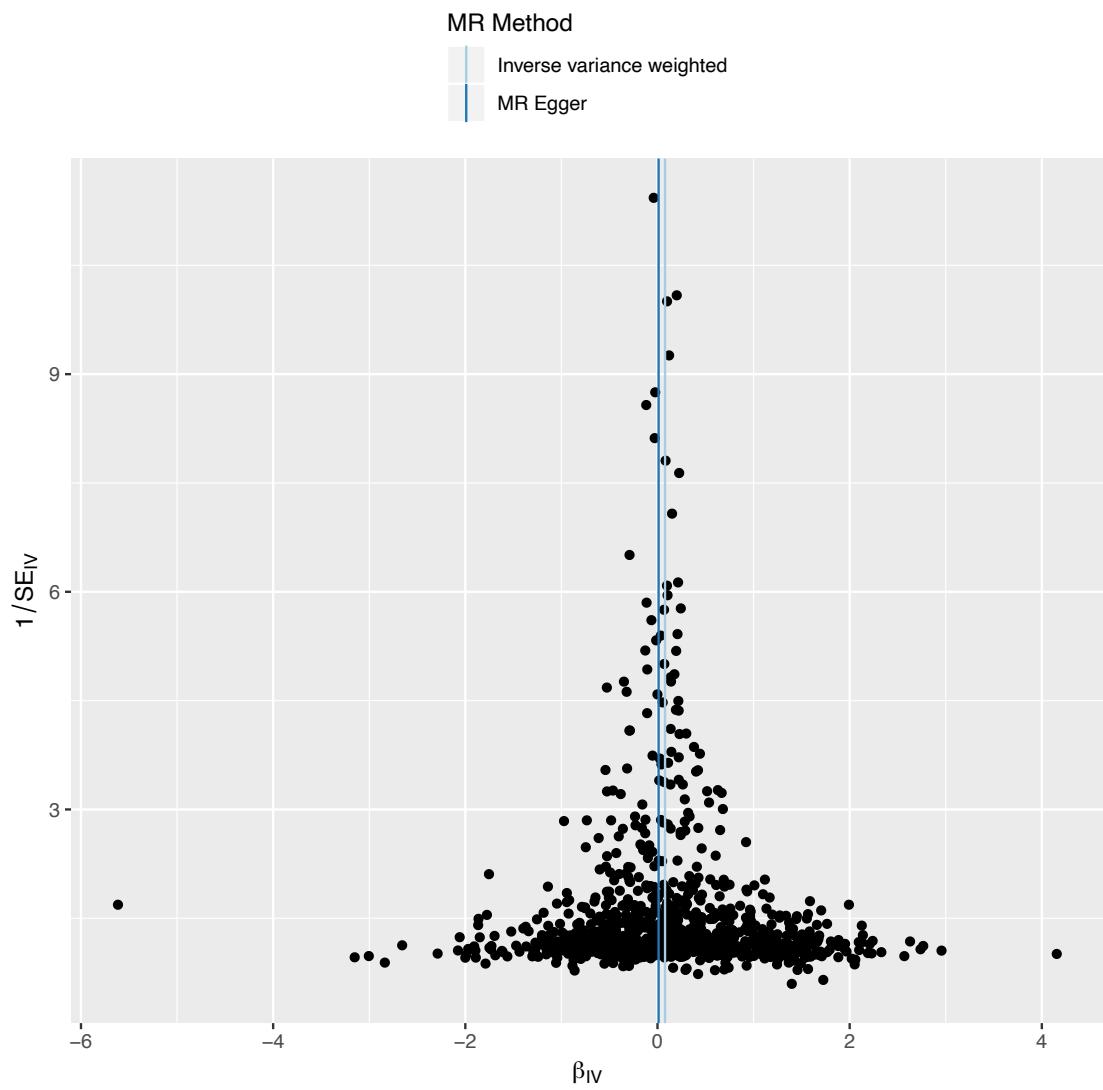


Figure A.20: **Representative figure: Funnel plot of BMI on one metabolite from Kettunen et al. (2016)**³¹⁸. Funnel plot shows the results of a single SNP MR with effect estimate and standard error. Asymmetry in the funnel indicates unreliable effect estimates.

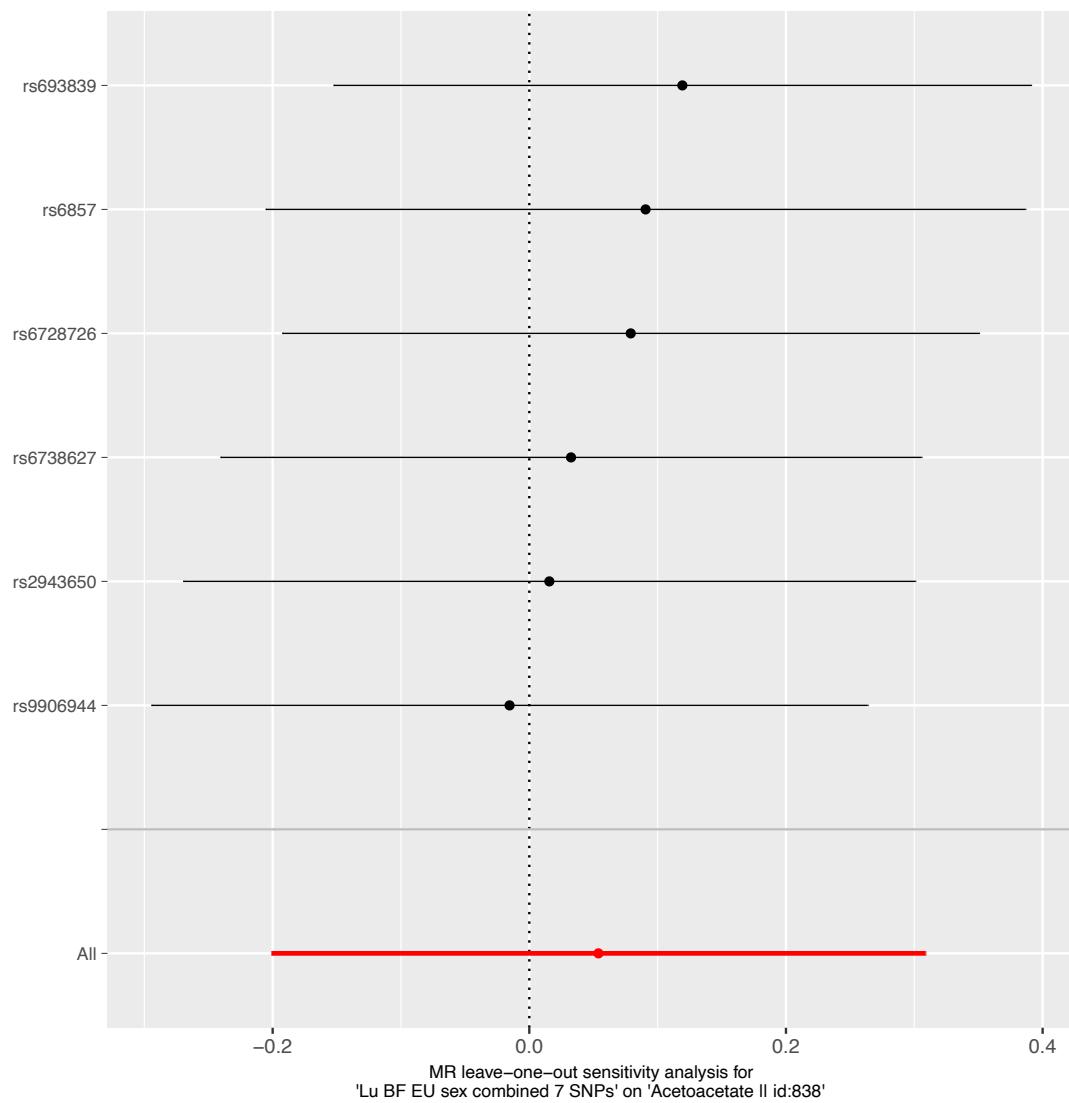


Figure A.21: Representative figure: Leave-one-out MR analysis of BF on Acetoacetate from Kettunen et al. (2016)³¹⁸. A leave-one-out analysis performs an MR of exposure on outcome for all SNPs excluding a different SNP each time. Forest plot shows the effect estimate and 95% confidence interval for each SNP exclusion on acetoacetate. Effect estimates represent the change in the inverse rank position of each metabolite per change in the inverse rank position of the exposure.



Figure A.22: Representative figure: Leave-one-out MR analysis of BMI on Acetate from Kettunen et al. (2016)³¹⁸. A leave-one-out analysis performs an MR of exposure on outcome for all SNPs excluding a different SNP each time. Forest plot shows the effect estimate and 95% confidence interval for each SNP exclusion on acetate. Effect estimates represent the change in the inverse rank position of each log transformed metabolite per change in the inverse rank position of the exposure.

5580 **INTERVAL** Broadly speaking, INTERVAL additional sensitivity analyses were similar to
5581 that of the Kettunen additional sensitivity analyses. In single SNP MR visual inspection of
5582 forest plots showed *S* shaped distributions of effect estimates for all tests (Representative
5583 Figure A.23, figure also shows outlier SNP with effect estimate close to -6). As with the
5584 Kettunen analysis, effect estimates for some SNPs in the single SNP MR analysis were much
5585 greater than others (See Representative Figure A.23). As an example of this in practice, for
5586 BF, more often than not, rs6857 showed an effect estimate greater than the 6 other SNPs
5587 and did not overlap confidence intervals with them or the null. rs6857 was found in both the
5588 Kettunen and INTERVAL analyses to have a disproportionately larger effect estimate than
5589 other SNPs. For BMI and WHR these SNPs generally did have confidence intervals which
5590 overlapped other SNPs, however overlap was minimal.

5591 Funnel plots did not highlight outlying SNPs, but did reflect some SNPs having larger
5592 effect estimates across the board (Representative Figure A.24. The low number of SNPs used
5593 for BF did not result in meaningfully interpretable funnel plots (Representative Figure ??). In
5594 leave-one-out analysis, visual inspection of forest plots showed that no single SNP altered the
5595 direction of effect for any metabolite across exposures. For BF, confidence intervals for one or
5596 more SNPs crossed the null for a majority of metabolite tested (Representative Figure A.25).
5597 This was not the case for BMI and WHR, where for many metabolites confidence intervals did
5598 not cross the null for any SNPs (Representative Figure A.26).

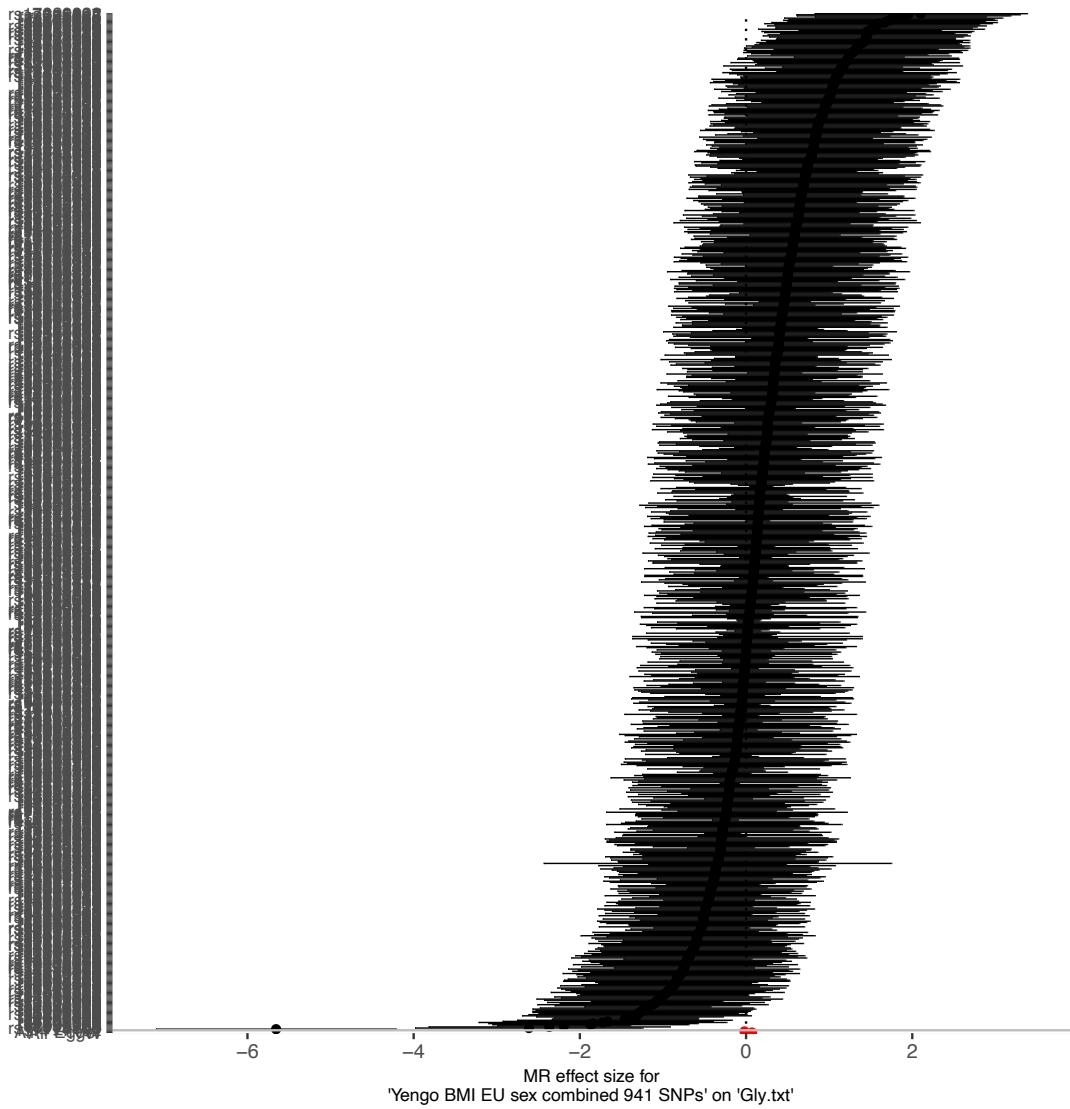


Figure A.23: Representative figure: Single SNP MR analysis of BMI on glycine from INTERVAL. Forest plot shows the effect estimate and 95% confidence interval for each individual SNPs effect on glycine. Effect estimates represent the change in the inverse rank position of each log transformed metabolite per change in the inverse rank position of the exposure.

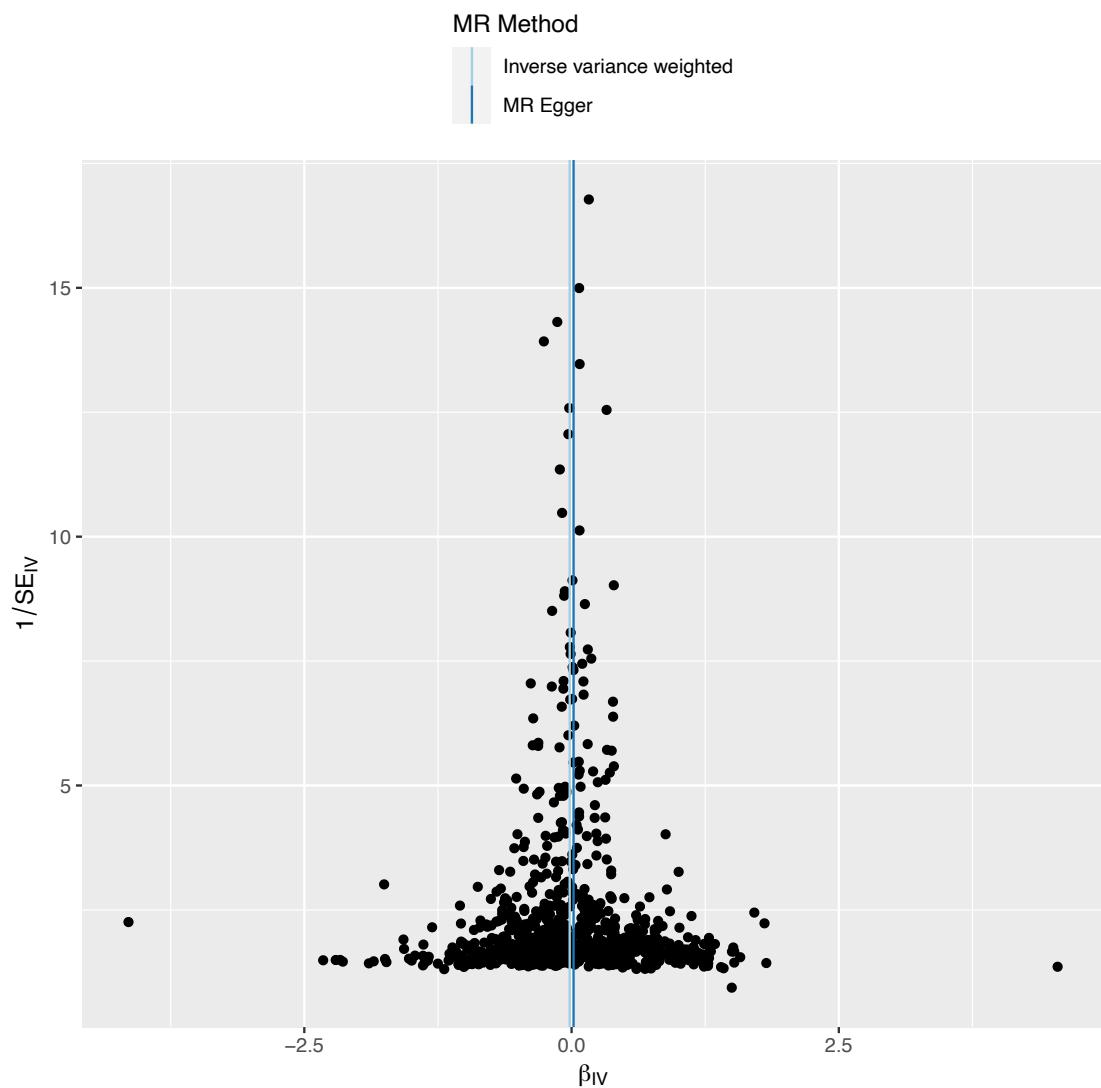


Figure A.24: **Representative figure: Funnel plot of BMI on one metabolite from INTERVAL.** Funnel plot shows the results of a single SNP MR with effect estimate and standard error. Asymmetry in the funnel indicates unreliable effect estimates.

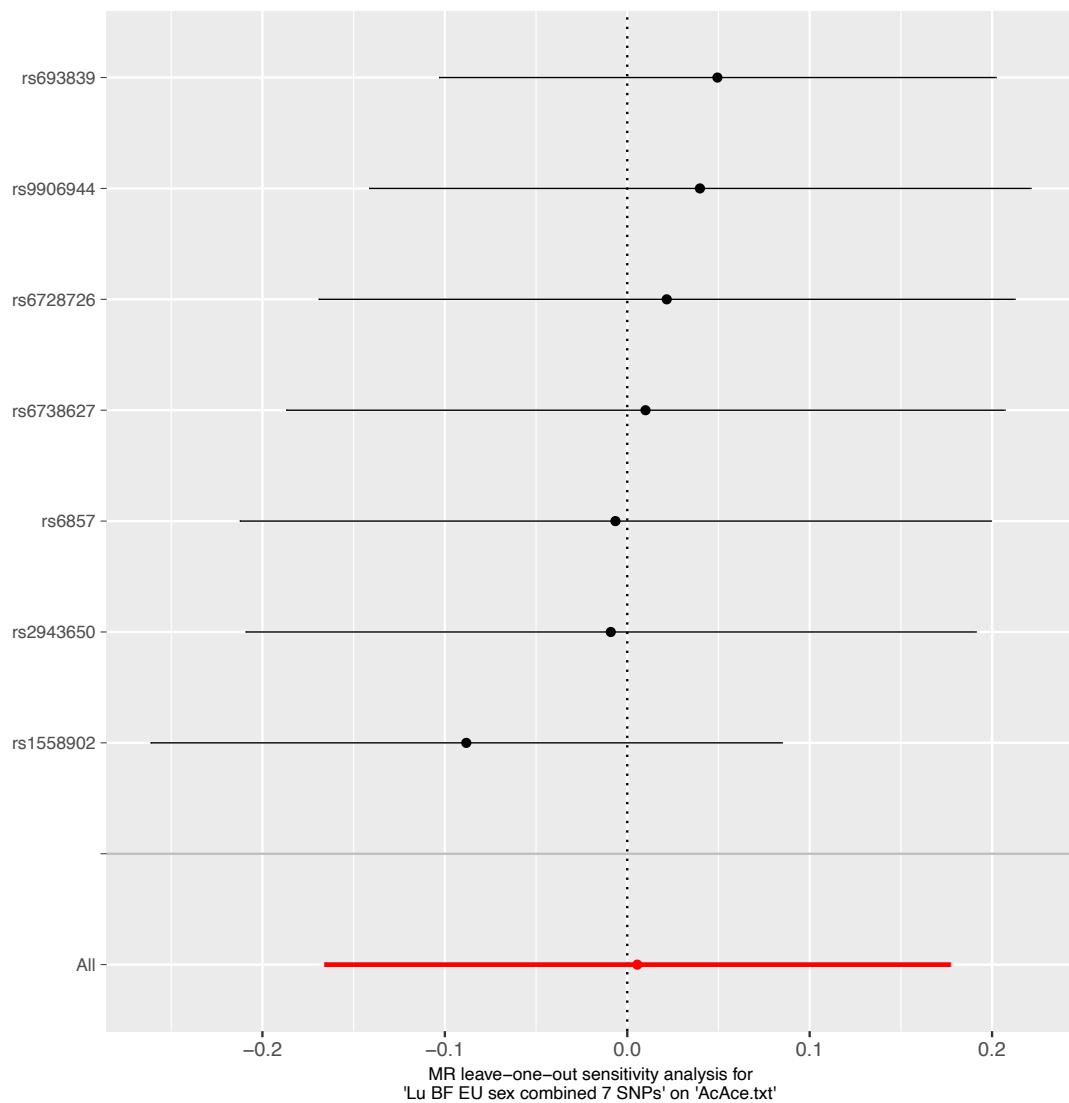


Figure A.25: Representative figure: Leave-one-out MR analysis of BF on Acetoacetate from Kettunen et al. (2016)³¹⁸. A leave-one-out analysis performs an MR of exposure on outcome for all SNPs excluding a different SNP each time. Forest plot shows the effect estimate and 95% confidence interval for each SNP exclusion on acetoacetate Effect estimates represent the change in the inverse rank position of each log transformed metabolite per change in the inverse rank position of the exposure.

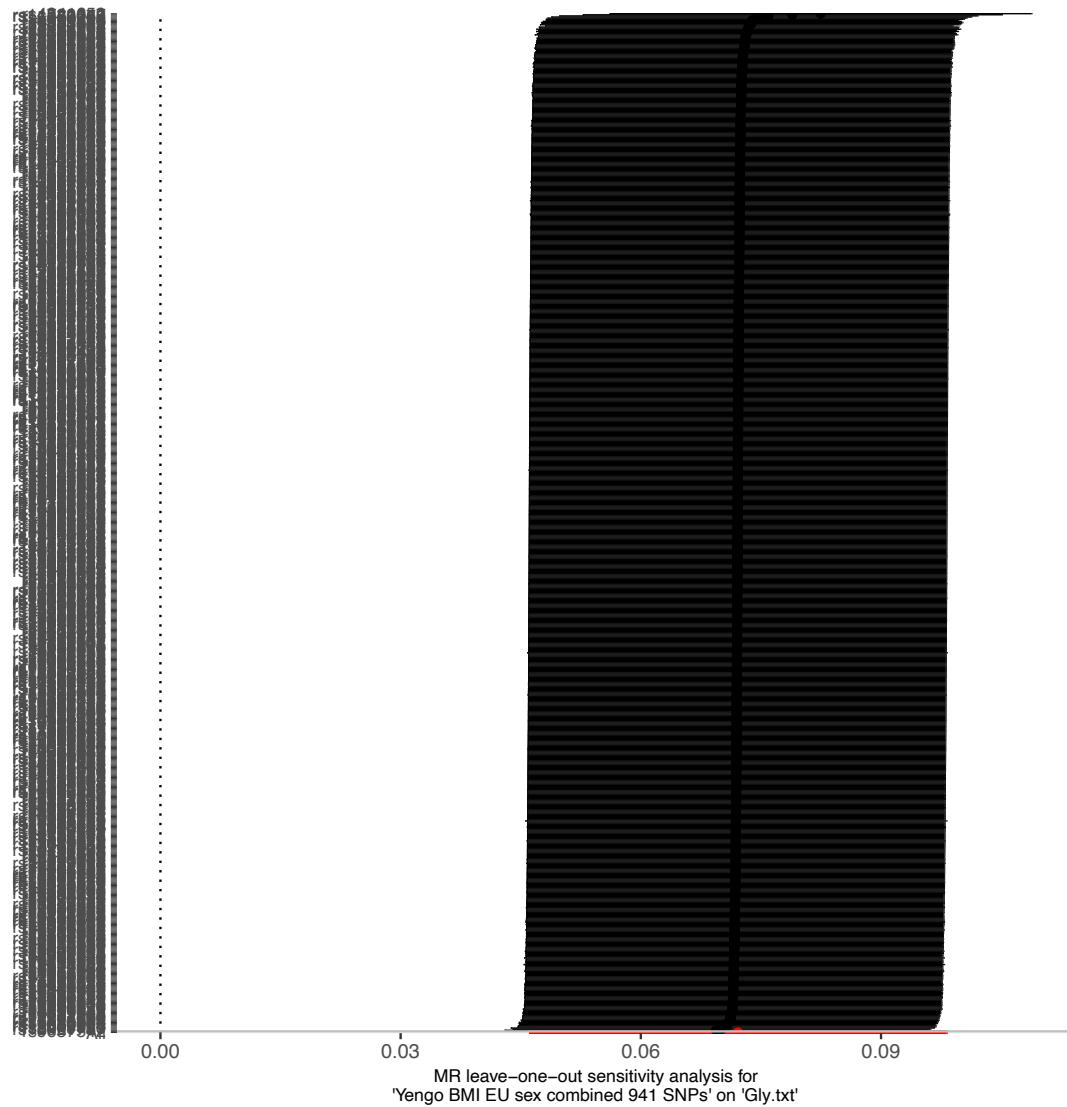


Figure A.26: Representative figure: Leave-one-out MR analysis of BMI on Glycine from INTERVAL. A leave-one-out analysis performs an MR of exposure on outcome for all SNPs excluding a different SNP each time. Forest plot shows the effect estimate and 95% confidence interval for each SNP exclusion on glycine. Effect estimates represent the change in the inverse rank position of each log transformed metabolite per change in the inverse rank position of the exposure.

5599 **Post-hoc additional analyses**

5600 A number of additional SNP lists were used to instrument BMI, WHR, and BF due to
5601 potential limitations of instrument lists used in the main analysis. All analysis, including
5602 sensitivity and additional sensitivity analyses, were repeated for these additional SNP lists.

5603 Results from additional exposures for BMI showed broadly larger effect estimates but
5604 consistent directions of effect across metabolites. For the BMI SNPs obtained from a non-UK
5605 Biobank GWAS, effect estimates had much wider confidence intervals. Spearman's Rho
5606 correlation of MR results was highest between the two SNP lists from Yengo et al. (2018)
5607 (0.98) – correlation between the Locke et al (2014) SNP list and the COJO SNP list from
5608 Yengo et al (2018) (0.9) and the non-COJO SNP list from Yengo et al. (2018) (0.93) were
5609 also high. For WHR the global pattern of association was similar between both the main
5610 and additional exposure (Figure ??) with high correlation between MR results (0.9). Effect
5611 estimates were larger for the additional exposure from Shungin et al (2014), for which fewer
5612 results reached the multiple testing threshold with wider confidence intervals which crossed
5613 the null more often than with the main analysis.

5614 For BF there was considerable similarity between the main analysis and the additional
5615 analysis from Lu et al (2016) which did not include two SNPs previously identified as being
5616 associated with favourable adiposity (Figure ??). More metabolites reached the multiple
5617 testing threshold when using the 5 SNPs from Lu et al (2016) as opposed to the full 7
5618 SNPs, this included associations with *Apolipoprotein A1*, *Phenylalanine*, *Tyrosine*, *Glucose*,
5619 and *Cholesterol esters in very large HDL*. For the additional analysis which used 76 SNPs
5620 from Hubel et al (2016), MR results were considerably smaller and appeared to show
5621 conflicting directions of effect with that of the Lu et al. (2016) SNPs (both using 7 and 5 SNPs).
5622 Confidence intervals were much tighter and two metabolites (*Phenylalanine* and *Glycoprotein*
5623 *acetyls*) reached the multiple testing threshold. Correlation between the two Lu et al (2016)
5624 SNP lists was high (0.93), however both the 5 (-0.64) and 7 (-0.52) SNP lists from Lu et
5625 al. (2016) showed weaker inverse correlations with the SNP list from Hubel et al (2016).

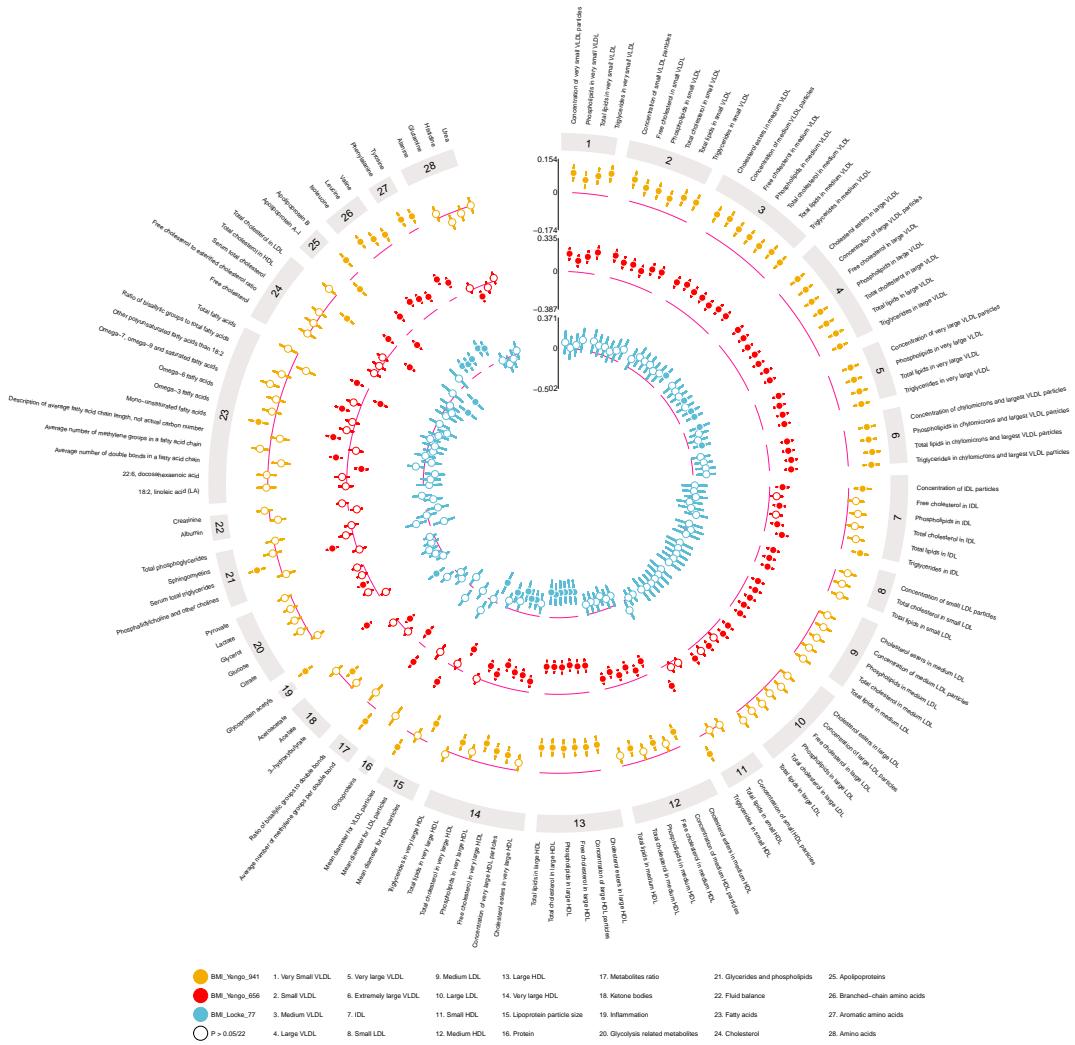


Figure A.27: Global metabolic profile of BMI on 123 NMR derived metabolites from Kettunen et al. (2016)³¹⁸: additional analysis. Circos plot shows each track as different BMI instrumentation: the outer track is Yengo et al. (2018)⁴¹ using 941 COJO SNPs; the middle track is Yengo et al. (2018)⁴¹ using 656 non-COJO SNPs; the inner track is Locke et al. (2016)³⁶ 77 SNPs. Solid points indicate a multiple testing threshold (0.0023) has been reached. Effect estimates represent the change in the inverse rank position of each metabolite per change in the inverse rank position of the exposure. 95% confidence intervals shown. Metabolites are grouped by subclass and arranged alphabetically within each subclass.

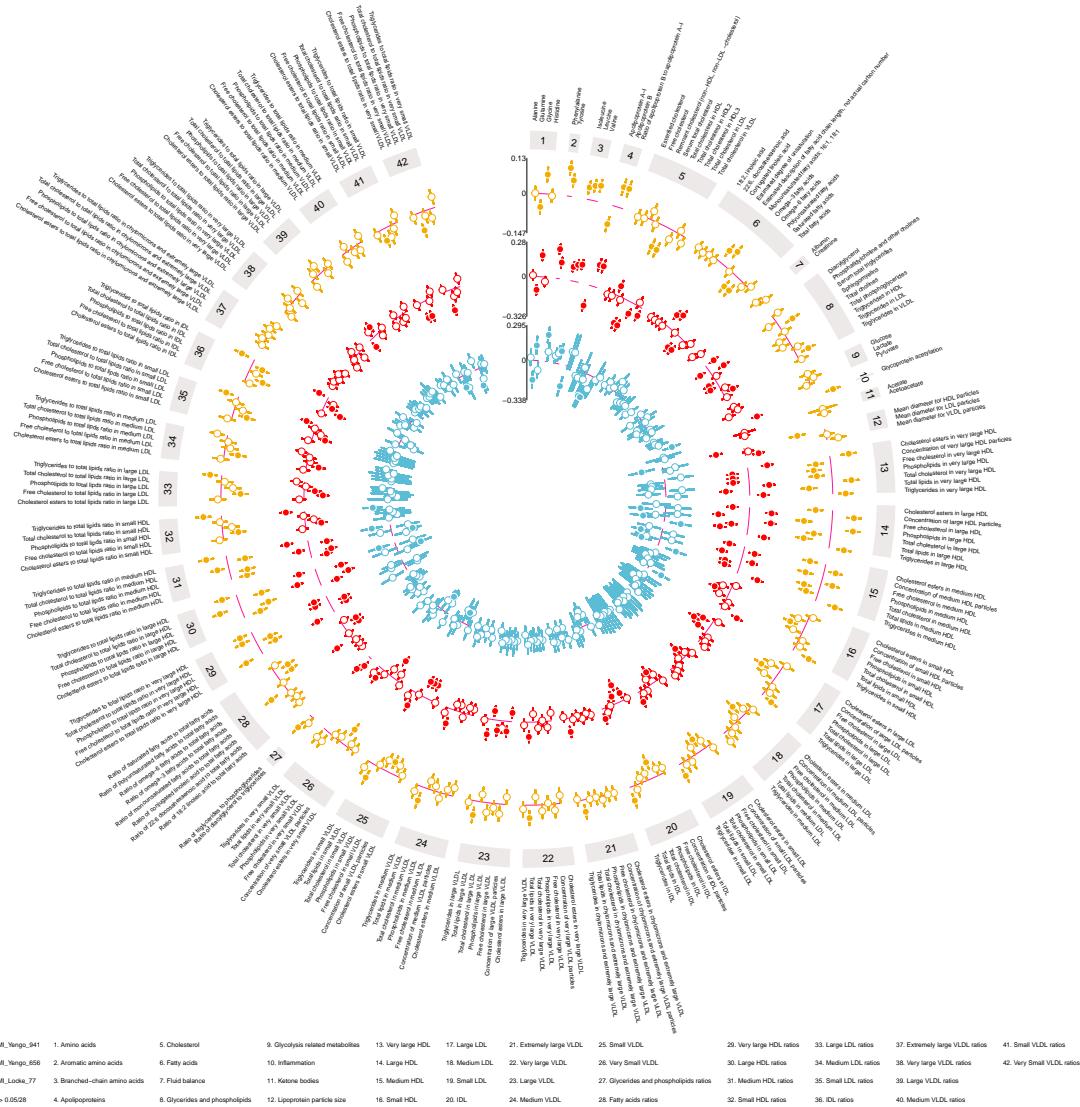


Figure A.28: Global metabolic profile of BMI on 230 NMR derived metabolites from INTERVAL: additional analysis. Circos plot shows each track as different BMI instrumentation: the outer track is Yengo et al. (2018)⁴¹ using 941 COJO SNPs; the middle track is Yengo et al. (2018)⁴¹ using 656 non-COJO SNPs; the inner track is Locke et al. (2016)³⁶ 77 SNPs. Solid points indicate a multiple testing threshold (2×10^{-4}) has been reached. Effect estimates represent the change in the inverse rank position of each log transformed metabolite per change in the inverse rank position of the exposure. 95% confidence intervals shown. Metabolites are grouped by subclass and arranged alphabetically within each subclass.

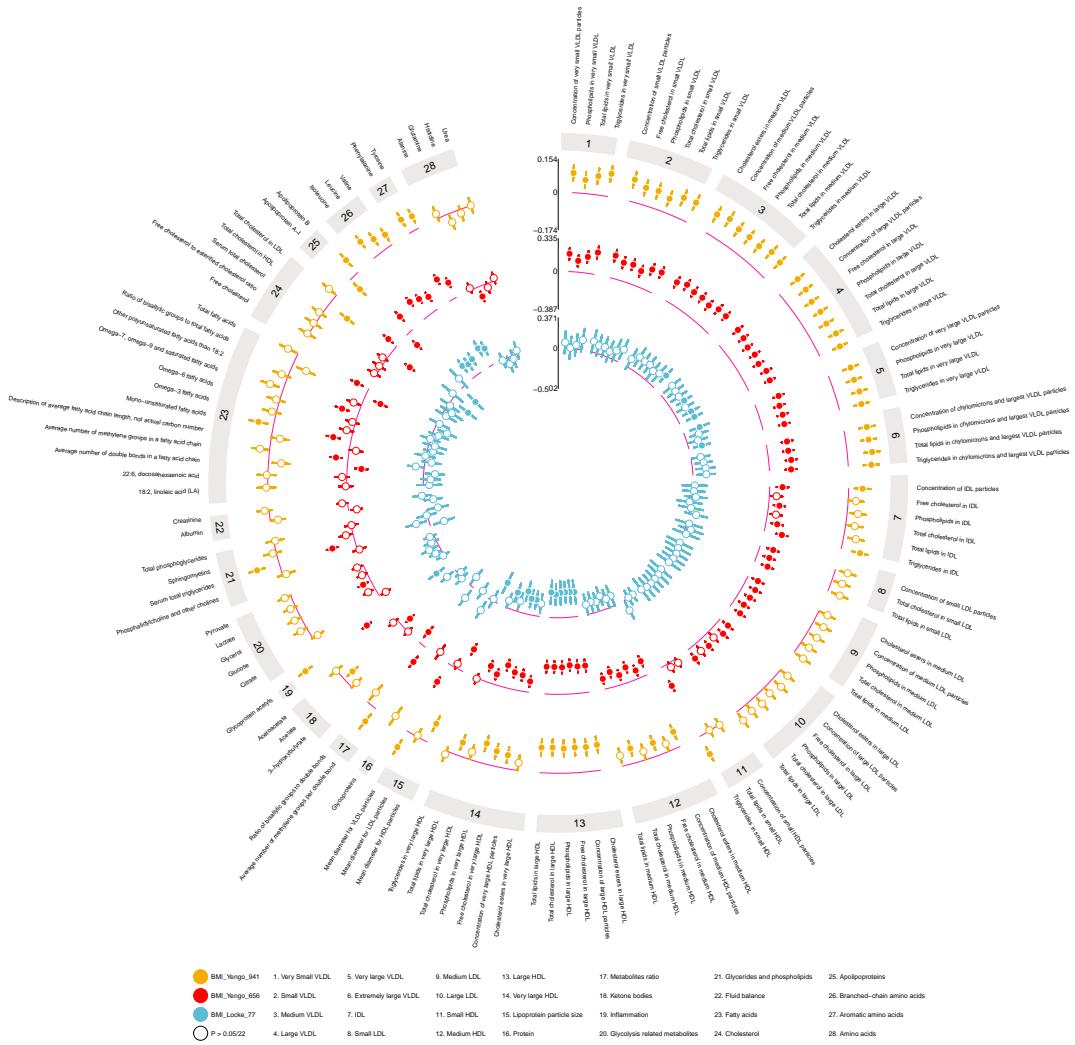


Figure A.29: Global metabolic profile of WHR on 123 NMR derived metabolites from Kettunen et al. (2016)³¹⁸: additional analysis. Circos plot shows each track as different WHR instrumentation: the outer track is Pulin et al. (2020)⁴² using 316 SNPs; the inner track is Shungin et al. (2015)³⁷ using 26 SNPs. Solid points indicate a multiple testing threshold (0.0023) has been reached. Effect estimates represent the change in the inverse rank position of each metabolite per change in the inverse rank position of the exposure. 95% confidence intervals shown. Metabolites are grouped by subclass and arranged alphabetically within each subclass.

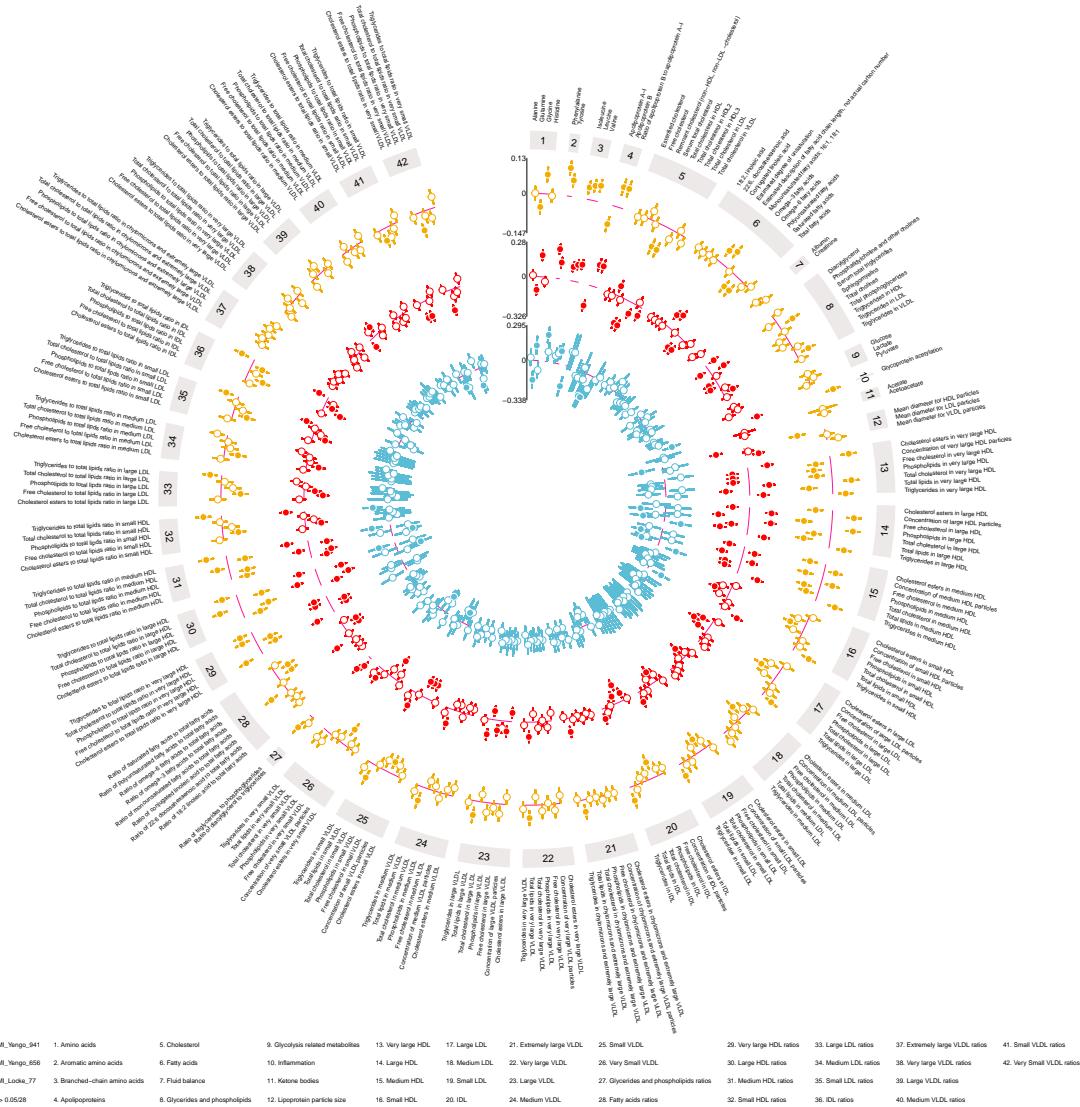


Figure A.30: Global metabolic profile of WHR on 230 NMR derived metabolites from INTERVAL: additional analysis. Circos plot shows each track as different WHR instrumentation: the outer track is Pulit et al. (2020)⁴² using 316 SNPs; the inner track is Shungin et al. (2015)³⁷ using 26 SNPs. Solid points indicate a multiple testing threshold (2×10^{-4}) has been reached. Effect estimates represent the change in the inverse rank position of each log transformed metabolite per change in the inverse rank position of the exposure. 95% confidence intervals shown. Metabolites are grouped by subclass and arranged alphabetically within each subclass.

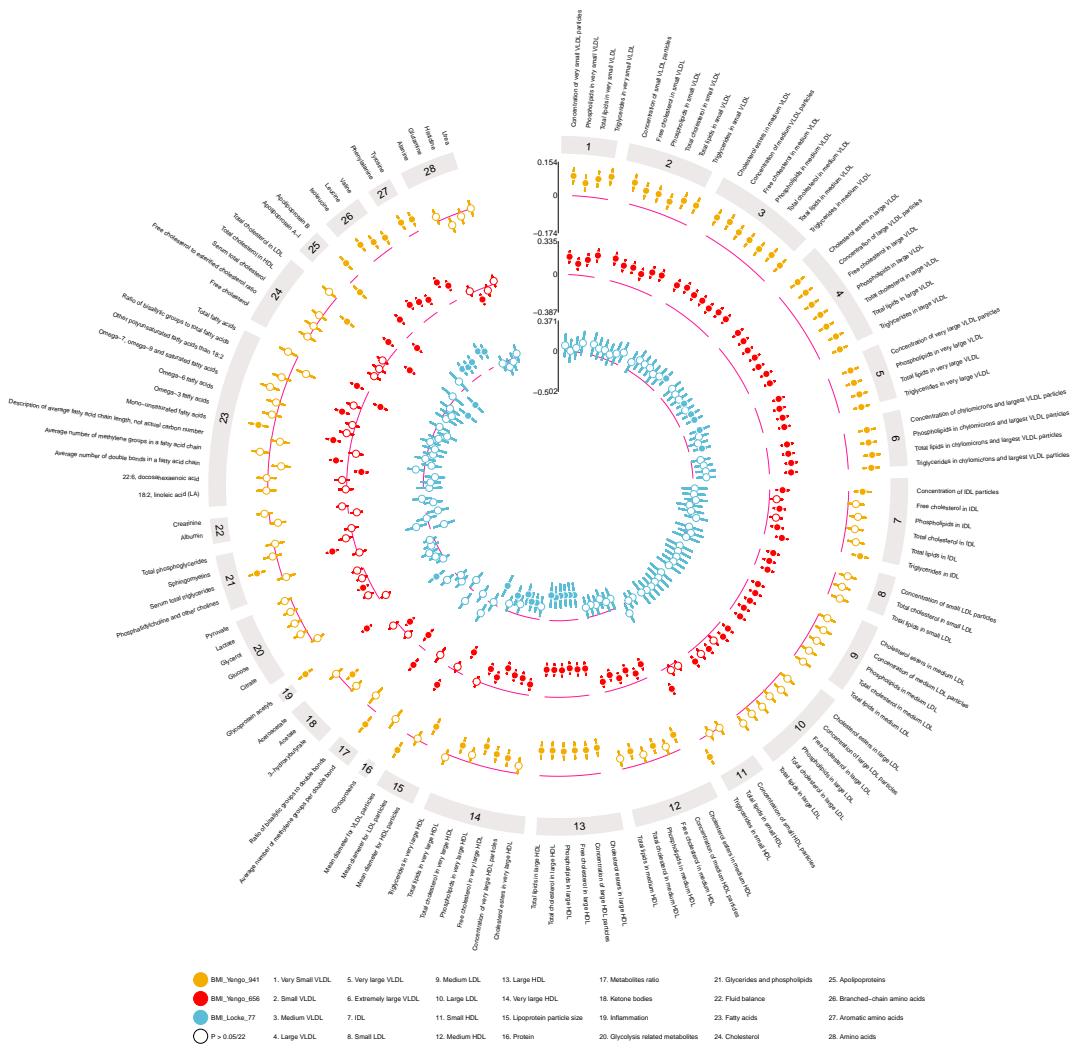


Figure A.31: Global metabolic profile of BF on 123 NMR derived metabolites from Kettunen et al. (2016)³¹⁸: additional analysis. Circos plot shows each track as different BF instrumentation: the outer track is Lu et al. (2016)³⁹ using 7 SNPs; the middle track is Lu et al. (2016)³⁹ using 5 SNPs having removed two for being associated with favourable adiposity; the inner track is Hubel et al. (2019)⁴³ using 76 SNPs. Solid points indicate a multiple testing threshold (0.0023) has been reached. Effect estimates represent the change in the inverse rank position of each metabolite per change in the inverse rank position of the exposure. 95% confidence intervals shown. Metabolites are grouped by subclass and arranged alphabetically within each subclass.

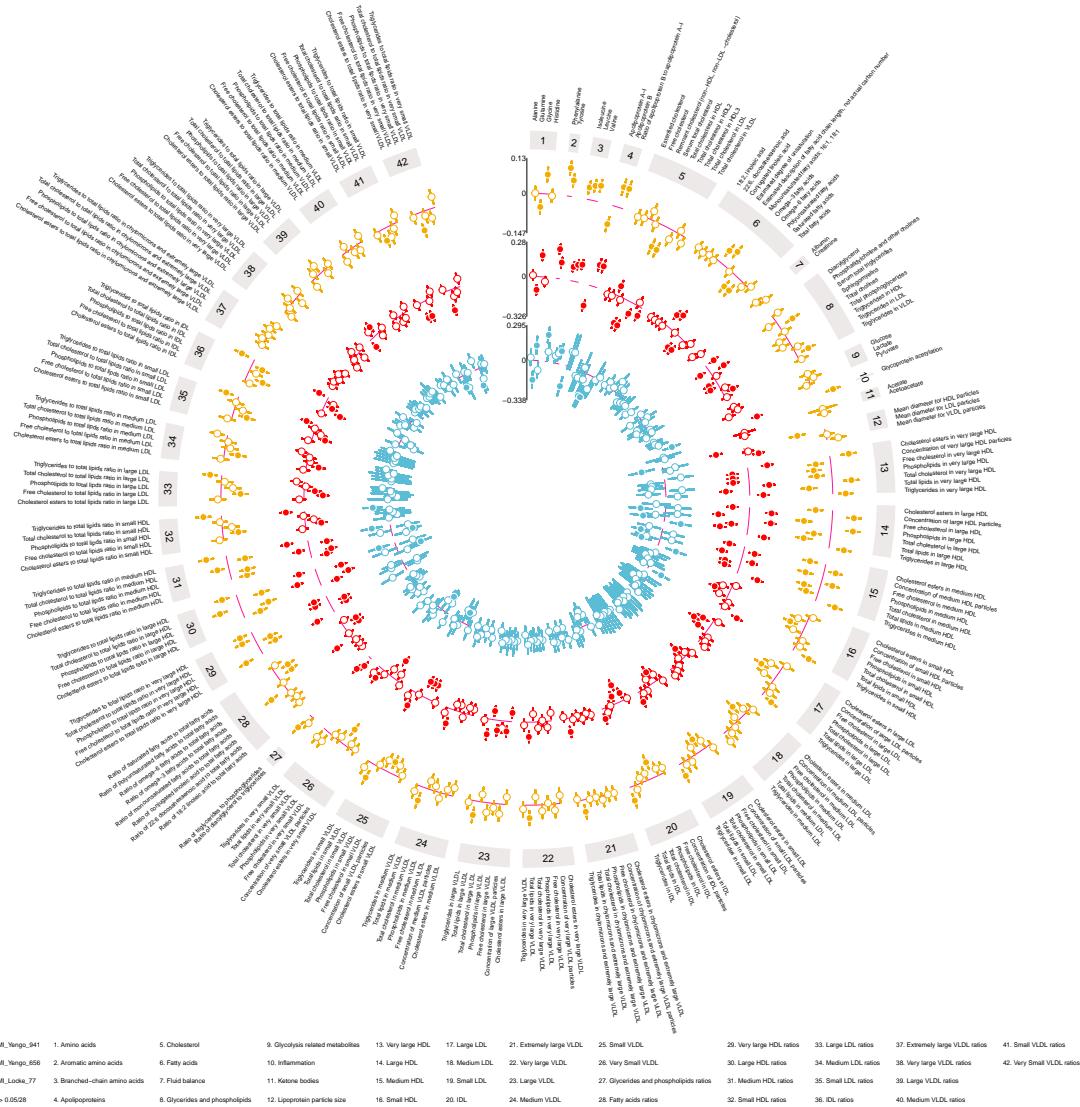


Figure A.32: Global metabolic profile of BF on 230 NMR derived metabolites from INTERVAL: additional analysis. Circos plot shows each track as different BF instrumentation: the outer track is Lu et al. (2016)³⁹ using 7 SNPs; the middle track is Lu et al. (2016)³⁹ using 5 SNPs having removed two for being associated with favourable adiposity; the inner track is Hubel et al. (2019)⁴³ using 76 SNPs. Solid points indicate a multiple testing threshold (2×10^{-4}) has been reached. Effect estimates represent the change in the inverse rank position of each log transformed metabolite per change in the inverse rank position of the exposure. 95% confidence intervals shown. Metabolites are grouped by subclass and arranged alphabetically within each subclass.

5629 **A.5 Chapter 6: Mediation**

5630 **A.5.1 Methods**

Table A.6: Metabolites used in MR analyses

Metabolite	Label	Class	Subclass
gln	Glutamine (mmol/l)	Amino acids	Amino acids
his	Histidine (mmol/l)	Amino acids	Amino acids
phe	Phenylalanine (mmol/l)	Amino acids	Aromatic amino acids
tyr	Tyrosine (mmol/l)	Amino acids	Aromatic amino acids
ile	Isoleucine (mmol/l)	Amino acids	Branched-chain amino acids
leu	Leucine (mmol/l)	Amino acids	Branched-chain amino acids
val	Valine (mmol/l)	Amino acids	Branched-chain amino acids
apoaa1	Apolipoprotein A-I (g/l)	Apolipoproteins	Apolipoproteins
apob	Apolipoprotein B (g/l)	Apolipoproteins	Apolipoproteins
hdc	Total cholesterol in HDL (mmol/l)	Cholesterol	Cholesterol
mufa	Monounsaturated fatty acids; 16:1, 18:1 (mmol/l)	Fatty acids	Fatty acids
totalfa	Total fatty acids (mmol/l)	Fatty acids	Fatty acids
albumin	Albumin (signal area)	Fluid balance	Fluid balance
lactate	Lactate (mmol/l)	Glycolysis related metabolites	Glycolysis related metabolites
pyruvate	Pyruvate (mmol/l)	Glycolysis related metabolites	Glycolysis related metabolites
hdlsize	Mean diameter for HDL particles (nm)	Lipoprotein particle size	Lipoprotein particle size
vldlsize	Mean diameter for VLDL particles (nm)	Lipoprotein particle size	Lipoprotein particle size
xxlvldll	Total lipids in chylomicrons and extremely large VLDL (mmol/l)	Lipoprotein subclasses	Extremely large VLDL
xxlvldlp	Concentration of chylomicrons and extremely large VLDL particles (mol/l)	Lipoprotein subclasses	Extremely large VLDL
xxvldlpl	Phospholipids in chylomicrons and extremely large VLDL (mmol/l)	Lipoprotein subclasses	Extremely large VLDL
idlgt	Triglycerides in IDL (mmol/l)	Lipoprotein subclasses	IDL
lhdc	Total cholesterol in large HDL (mmol/l)	Lipoprotein subclasses	Large HDL
lhdcce	Cholesterol esters in large HDL (mmol/l)	Lipoprotein subclasses	Large HDL
lhdfc	Free cholesterol in large HDL (mmol/l)	Lipoprotein subclasses	Large HDL
lhdl	Total lipids in large HDL (mmol/l)	Lipoprotein subclasses	Large HDL
lvldfc	Free cholesterol in large VLDL (mmol/l)	Lipoprotein subclasses	Large VLDL
lvldll	Total lipids in large VLDL (mmol/l)	Lipoprotein subclasses	Large VLDL
lvldlp	Concentration of large VLDL particles (mol/l)	Lipoprotein subclasses	Large VLDL
lvldlpl	Phospholipids in large VLDL (mmol/l)	Lipoprotein subclasses	Large VLDL
lvldtg	Triglycerides in large VLDL (mmol/l)	Lipoprotein subclasses	Large VLDL
mhdc	Total cholesterol in medium HDL (mmol/l)	Lipoprotein subclasses	Medium HDL
mhdcce	Cholesterol esters in medium HDL (mmol/l)	Lipoprotein subclasses	Medium HDL
mhdfc	Free cholesterol in medium HDL (mmol/l)	Lipoprotein subclasses	Medium HDL
mhdl	Total lipids in medium HDL (mmol/l)	Lipoprotein subclasses	Medium HDL
mvldlfc	Free cholesterol in medium VLDL (mmol/l)	Lipoprotein subclasses	Medium VLDL
mvldll	Total lipids in medium VLDL (mmol/l)	Lipoprotein subclasses	Medium VLDL
mvldtg	Triglycerides in medium VLDL (mmol/l)	Lipoprotein subclasses	Medium VLDL
shdl	Total lipids in small HDL (mmol/l)	Lipoprotein subclasses	Small HDL

Table A.6: Metabolites used in MR analyses (*continued*)

Metabolite	Label	Class	Subclass
shdltg	Triglycerides in small HDL (mmol/l)	Lipoprotein subclasses	Small HDL
svldll	Total lipids in small VLDL (mmol/l)	Lipoprotein subclasses	Small VLDL
svldltg	Triglycerides in small VLDL (mmol/l)	Lipoprotein subclasses	Small VLDL
xhdcce	Cholesterol esters in very large HDL (mmol/l)	Lipoprotein subclasses	Very large HDL
xhdfc	Free cholesterol in very large HDL (mmol/l)	Lipoprotein subclasses	Very large HDL
xhndl	Total lipids in very large HDL (mmol/l)	Lipoprotein subclasses	Very large HDL
xhdp	Concentration of very large HDL particles (mol/l)	Lipoprotein subclasses	Very large HDL
xhdp1	Phospholipids in very large HDL (mmol/l)	Lipoprotein subclasses	Very large HDL
xhdtg	Triglycerides in very large HDL (mmol/l)	Lipoprotein subclasses	Very large HDL
xvldll	Total lipids in very large VLDL (mmol/l)	Lipoprotein subclasses	Very large VLDL
xvldlp	Concentration of very large VLDL particles (mol/l)	Lipoprotein subclasses	Very large VLDL
xvldpl	Phospholipids in very large VLDL (mmol/l)	Lipoprotein subclasses	Very large VLDL
xvldtg	Triglycerides in very large VLDL (mmol/l)	Lipoprotein subclasses	Very large VLDL
xsvldll	Total lipids in very small VLDL (mmol/l)	Lipoprotein subclasses	Very Small VLDL
xsvldtg	Triglycerides in very small VLDL (mmol/l)	Lipoprotein subclasses	Very Small VLDL

5631 **A.5.2 Results**

5632 **Two Sample MR: adiposity-endometrial cancer**

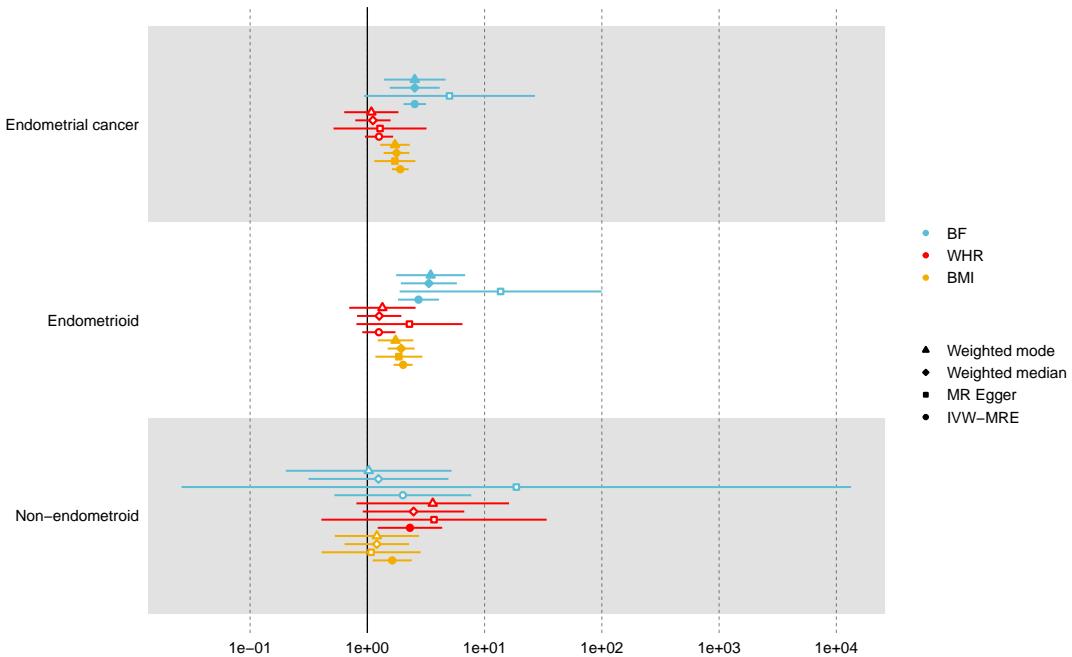


Figure A.33: **Two-sample MR: Effect of adiposity on endometrial cancer.** Forest plot shows odds ratios and associated 95% confidence intervals per normalized standard deviation unit increase for the effects of BMI, WHR, and BF on endometrial cancer, endometrioid cancer, and non-endometrioid cancer. The main analysis (IVW-MRE; multiplicative random effects) is presented alongside sensitivity analyses (weighted median, weighted mode, and MR Egger). Solid points indicate a nominal p-value threshold of 0.05 has been reached.

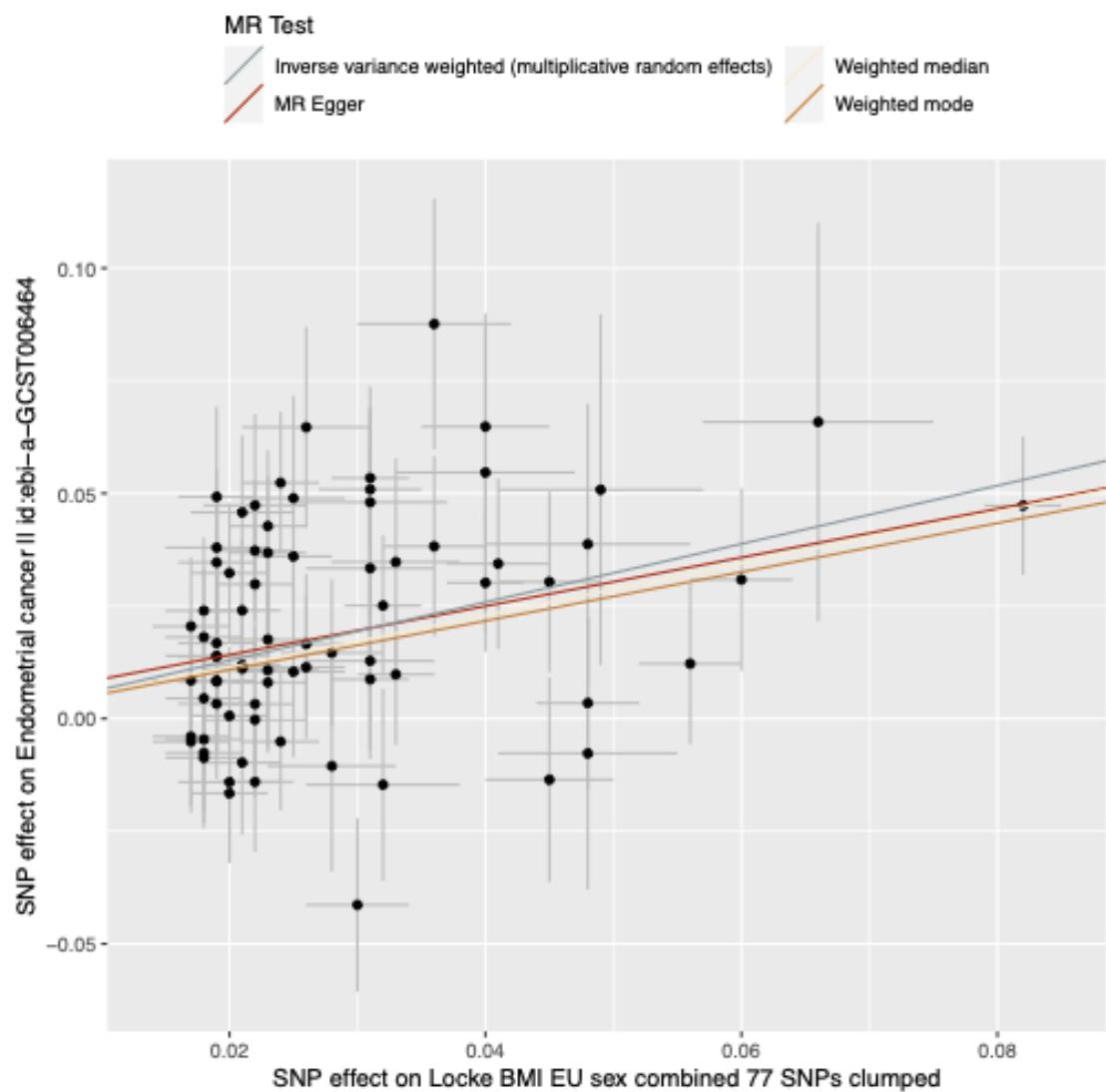


Figure A.34: Representativ figure: Scatter plot of MR effect estimates from two-sample MR of adiposity on endometrial cancer.

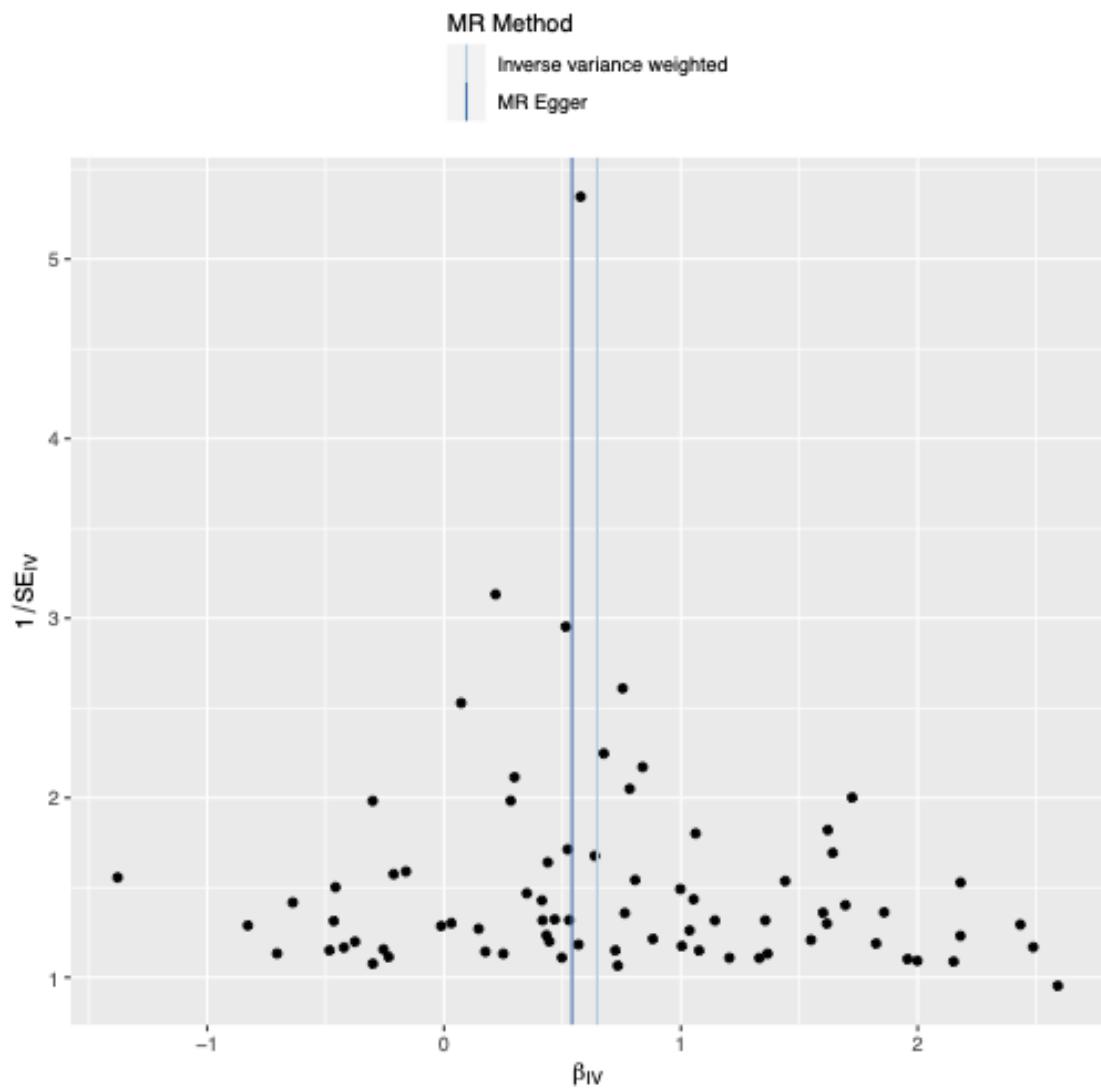


Figure A.35: Representativ figure: Funnel plot of MR effect estimates from two-sample MR of adiposity on endometrial cancer.

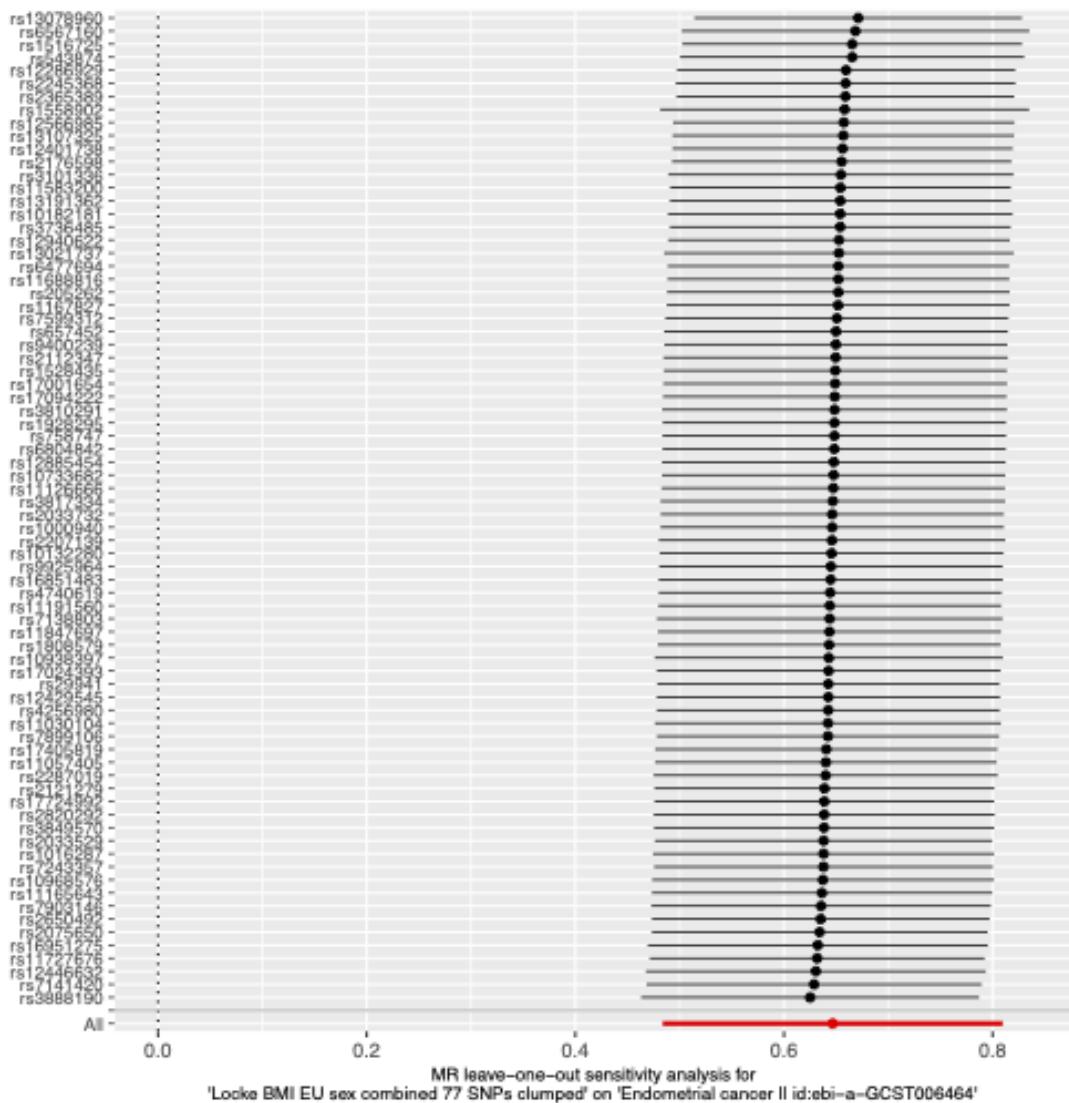


Figure A.36: **Representativ figure: Leave one out plot of MR effect estimates from two-sample MR of adiposity on endometrial cancer.**

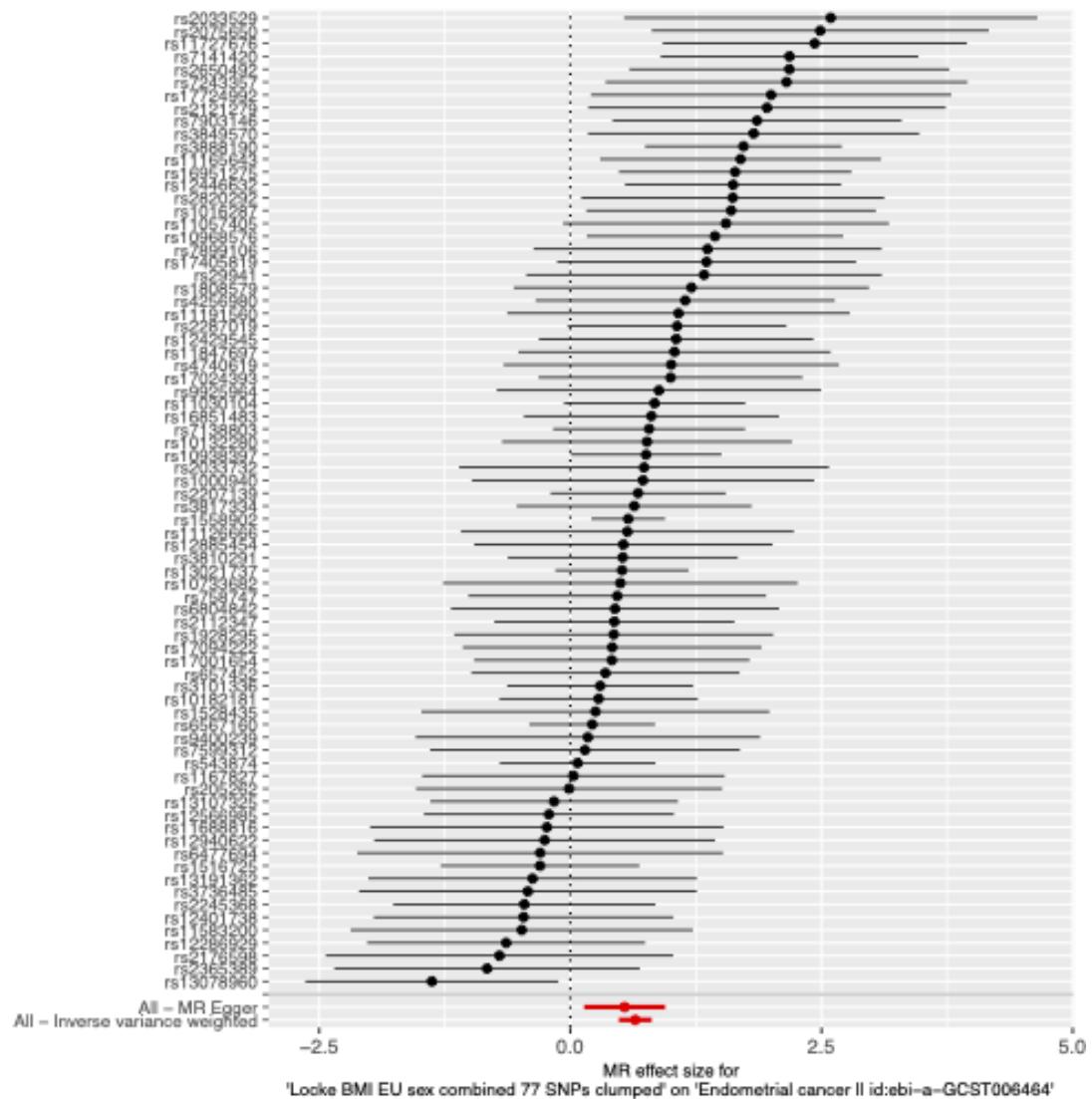


Figure A.37: **Representativ figure: Single SNP plot of MR effect estimates from two-sample MR of adiposity on endometrial cancer.**

5633 Two Sample MR: adiposity-metabolites

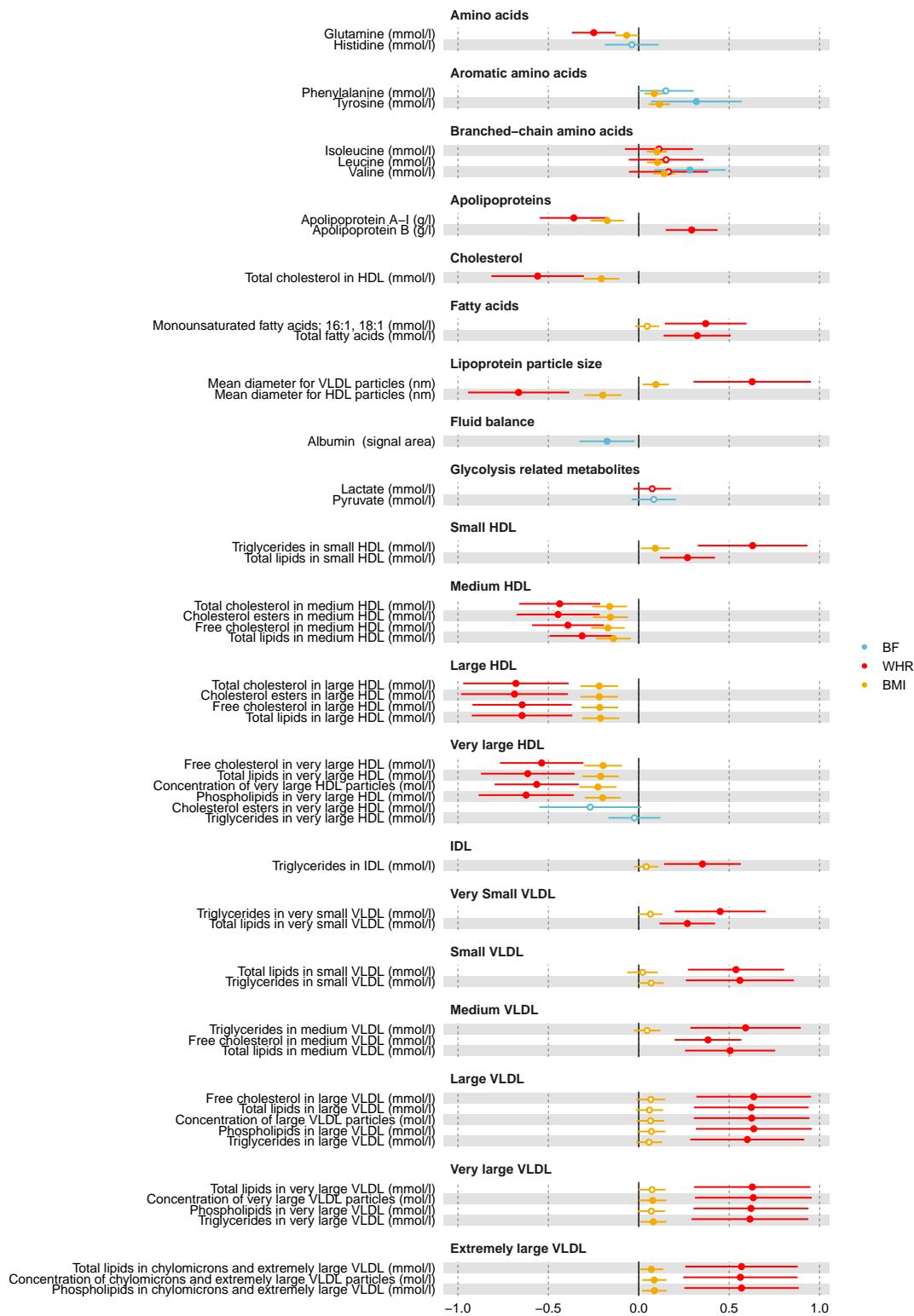


Figure A.38: **Two-sample MR: Effect of adiposity on metabolites.** Forest plot shows odds ratios and associated 95% confidence intervals per normalized standard deviation unit increase for the effects of BMI, WHR, and BF on 45, 46, and 8 metabolites respectively. The inverse variance weighted multiplicative random effects model is presented. Solid points indicate a nominal p-value threshold of 0.05 has been reached.

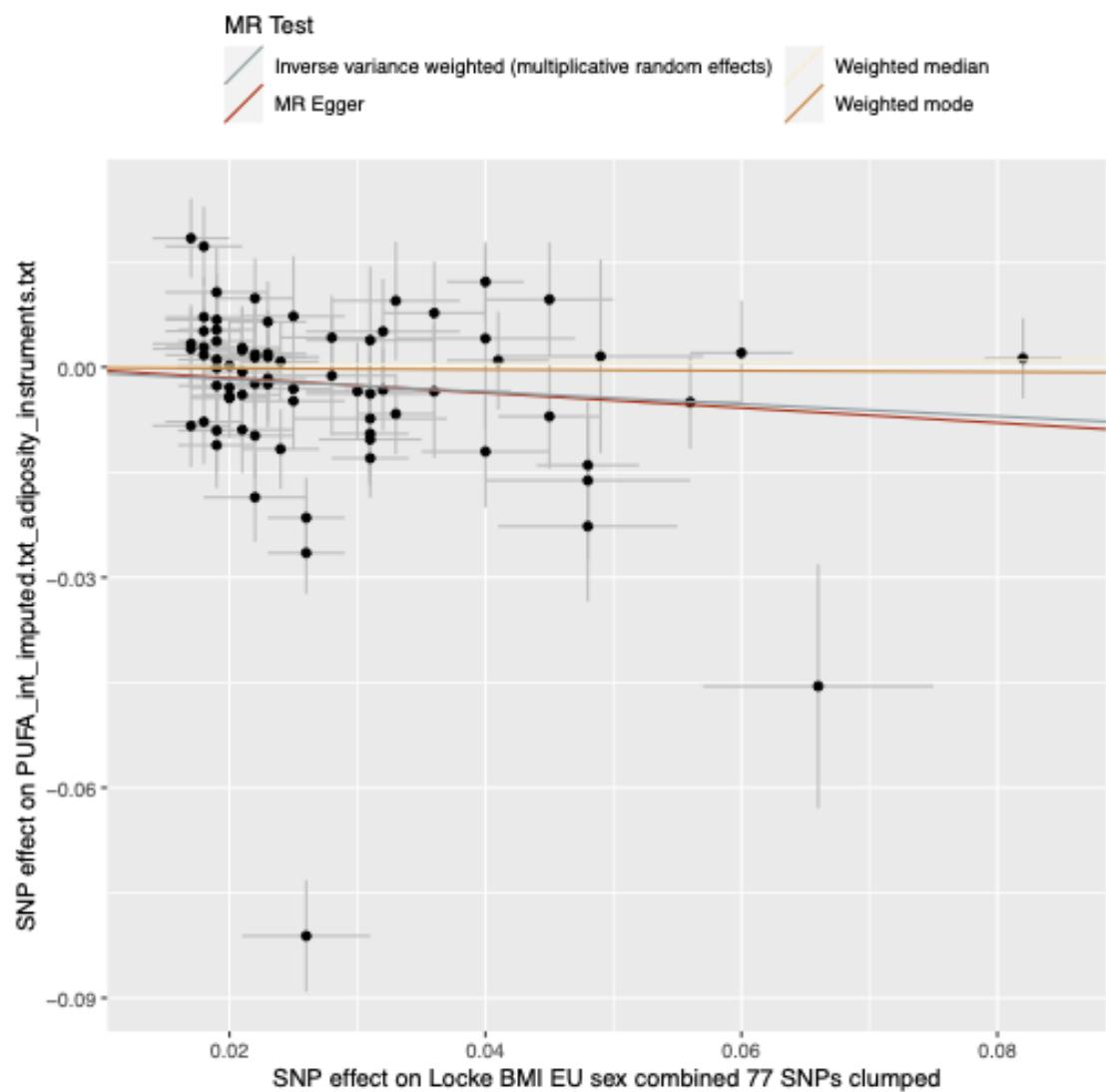


Figure A.39: Representativ figure: Scatter plot of MR effect estimates from two-sample MR of adiposity on metabolites.

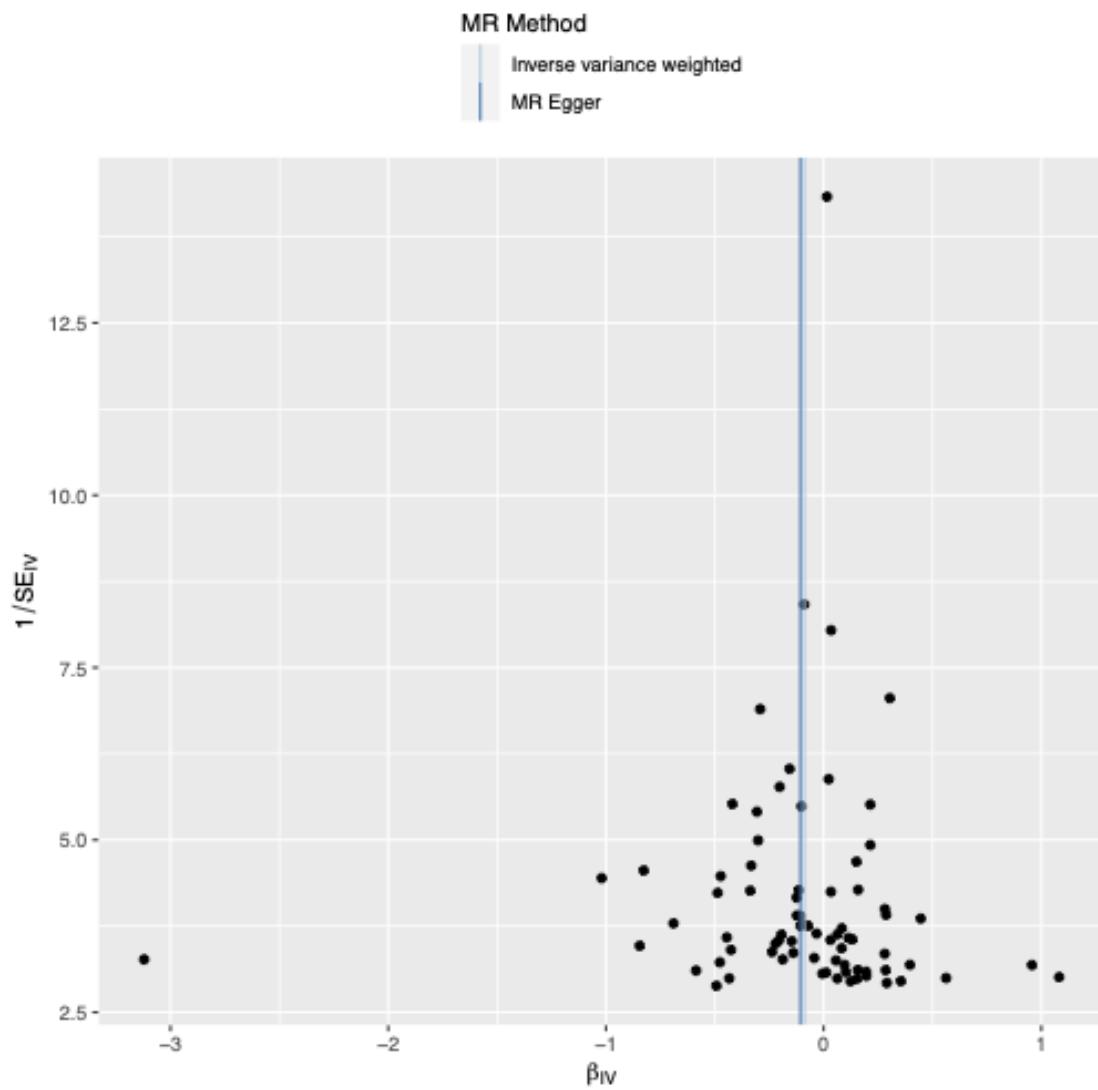


Figure A.40: Representativ figure: Funnel plot of MR effect estimates from two-sample MR of adiposity on metabolites.

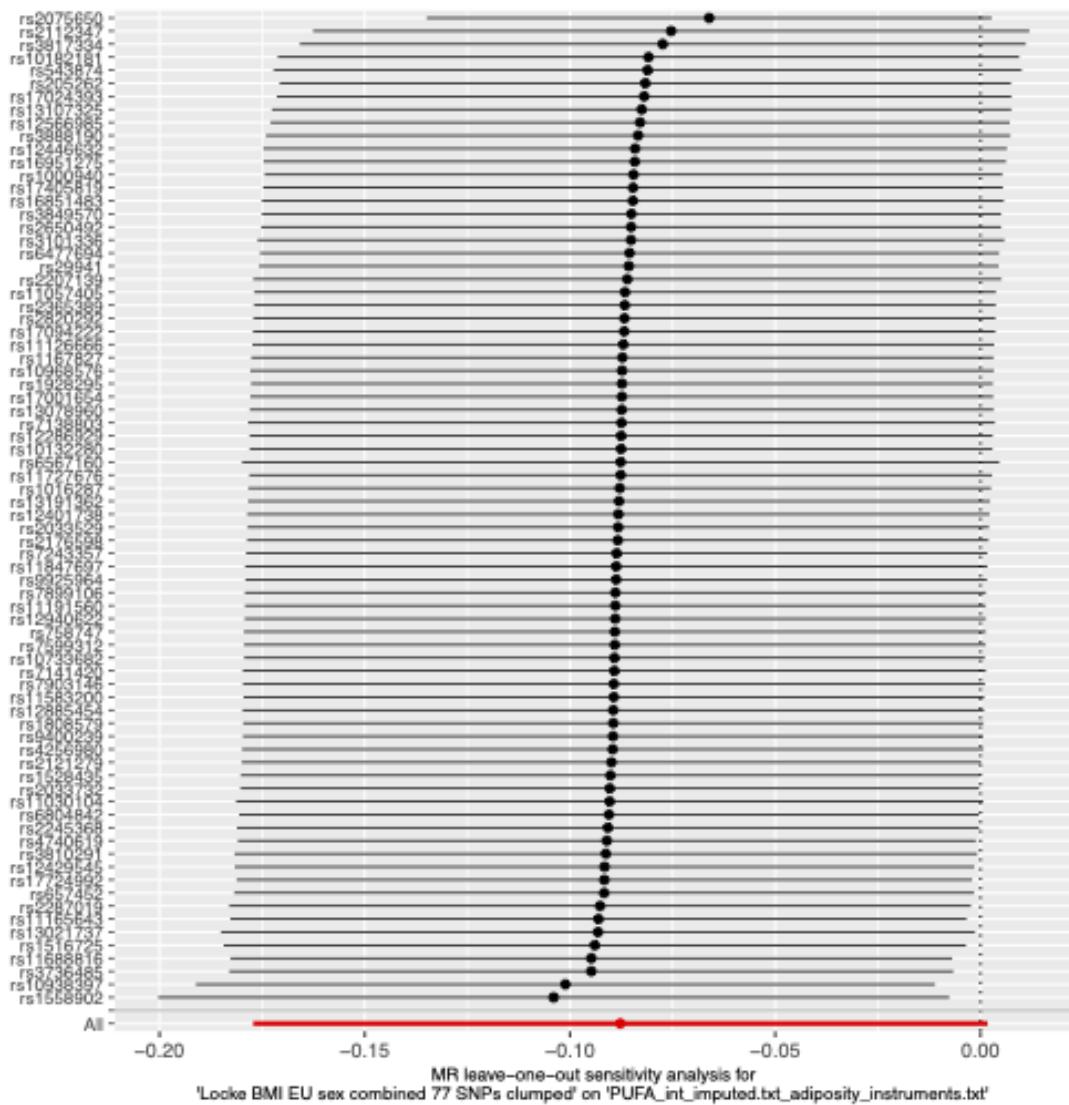


Figure A.41: **Representativ figure: Leave one out plot of MR effect estimates from two-sample MR of adiposity on metabolites.**

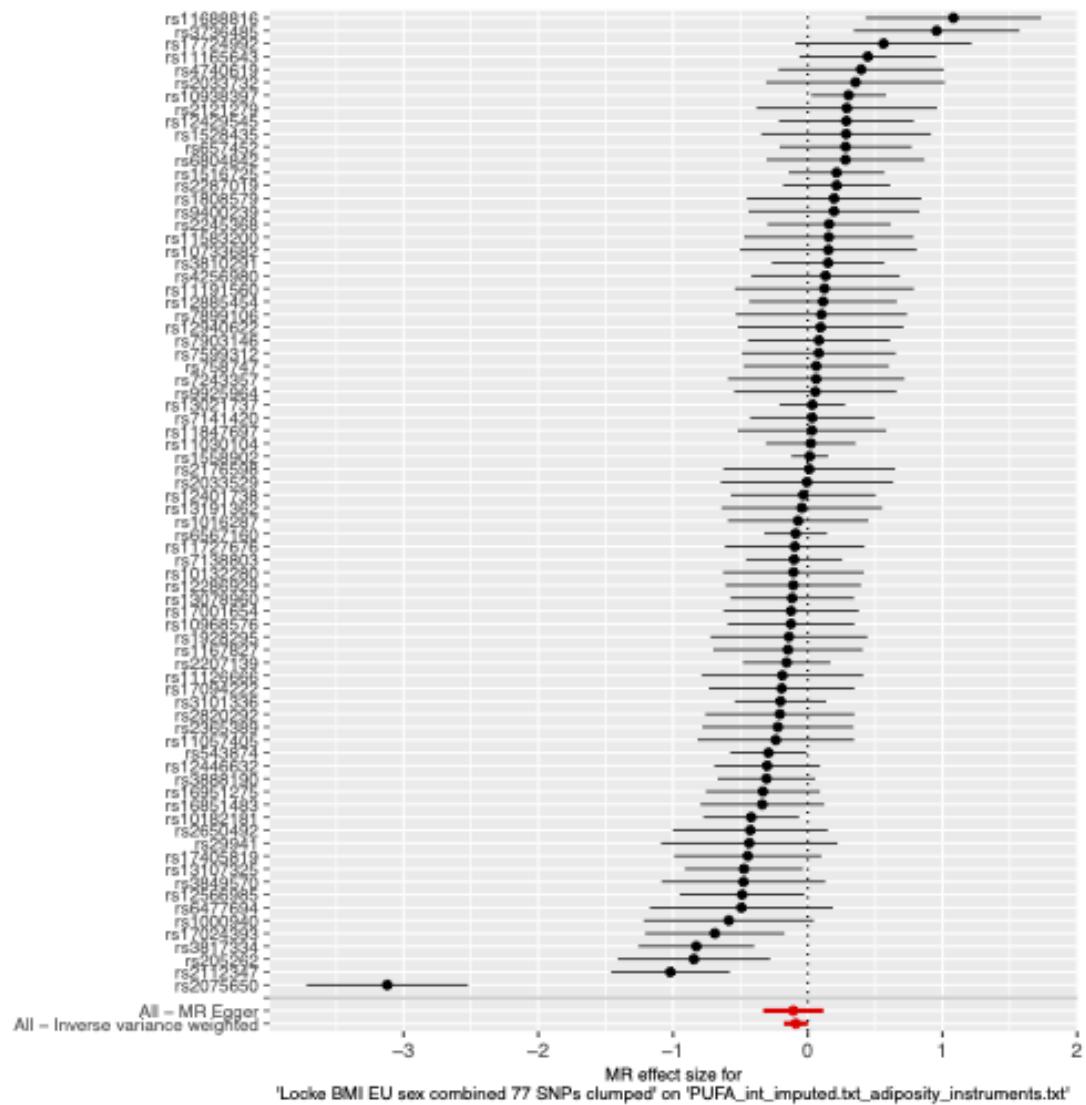


Figure A.42: **Representativ figure: Single SNP plot of MR effect estimates from two-sample MR of adiposity on metabolites.**

5634 Two Sample MR: metabolites-endometrial cancer

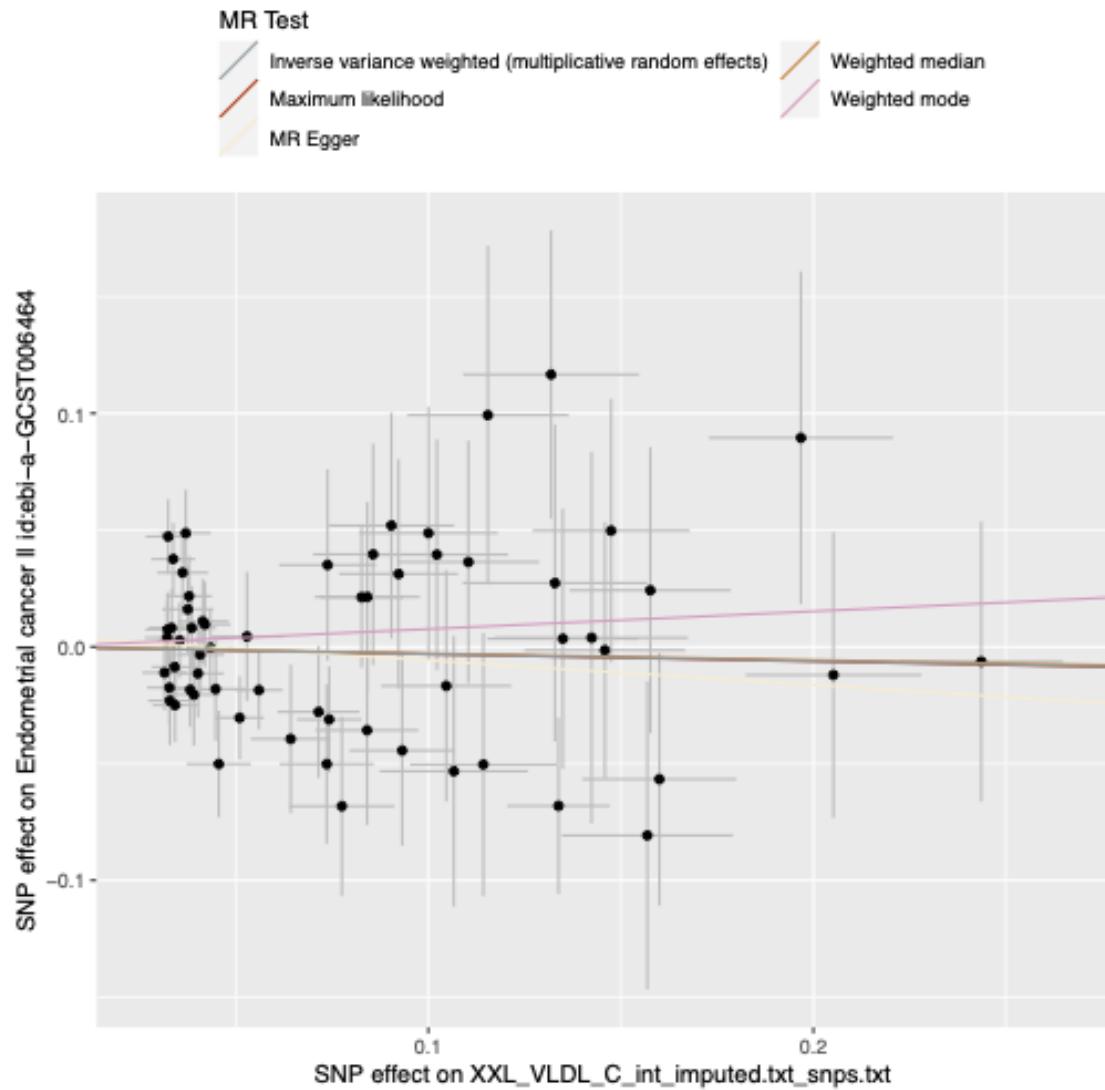


Figure A.43: Representativ figure: Scatter plot of MR effect estimates from two-sample MR of metabolites on endometrial cancer.

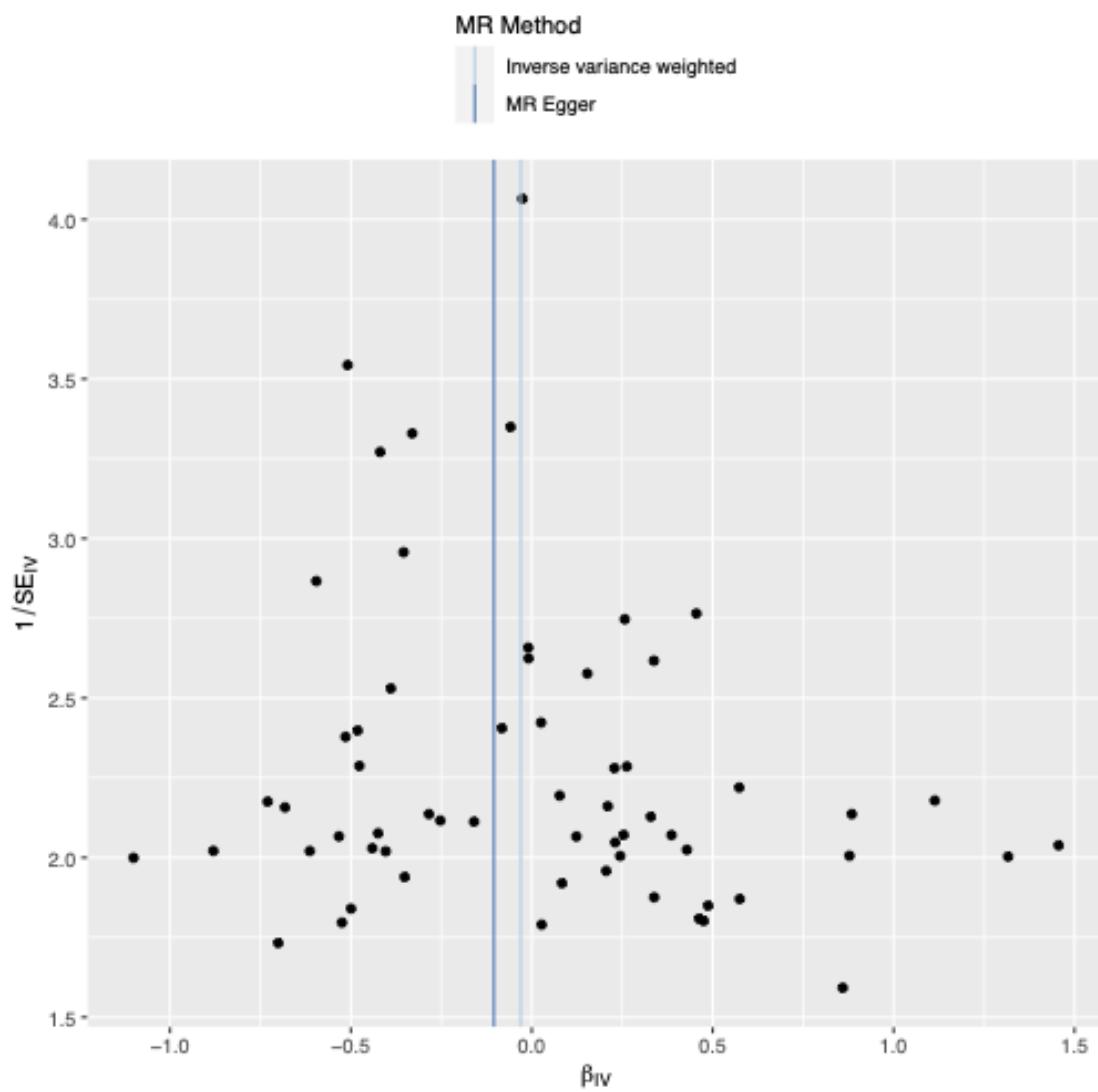


Figure A.44: Representativ figure: Funnel plot of MR effect estimates from two-sample MR of metabolites on endometrial cancer.

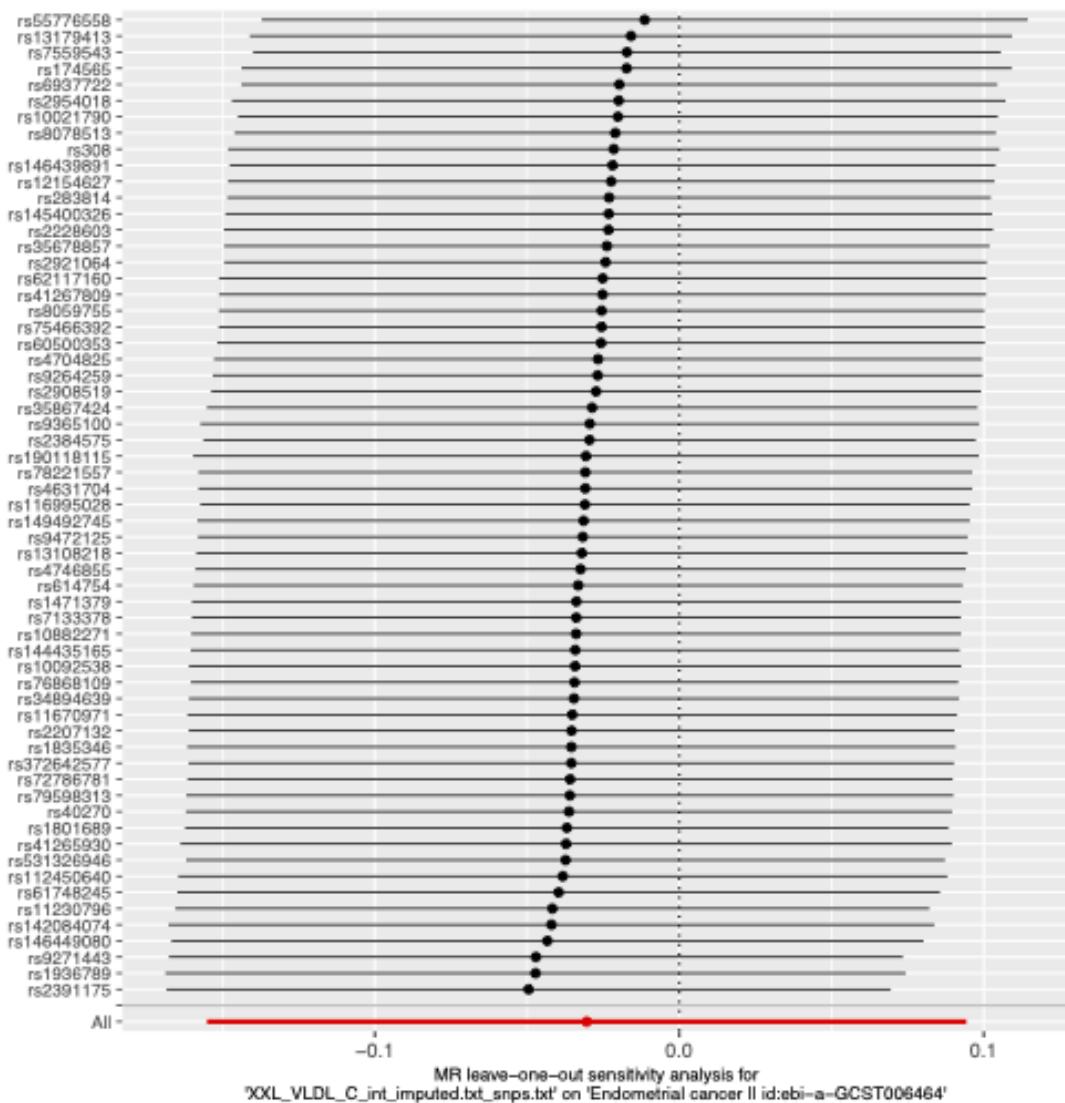


Figure A.45: **Representativ figure: Leave one out plot of MR effect estimates from two-sample MR of metabolites on endometrial cancer.**

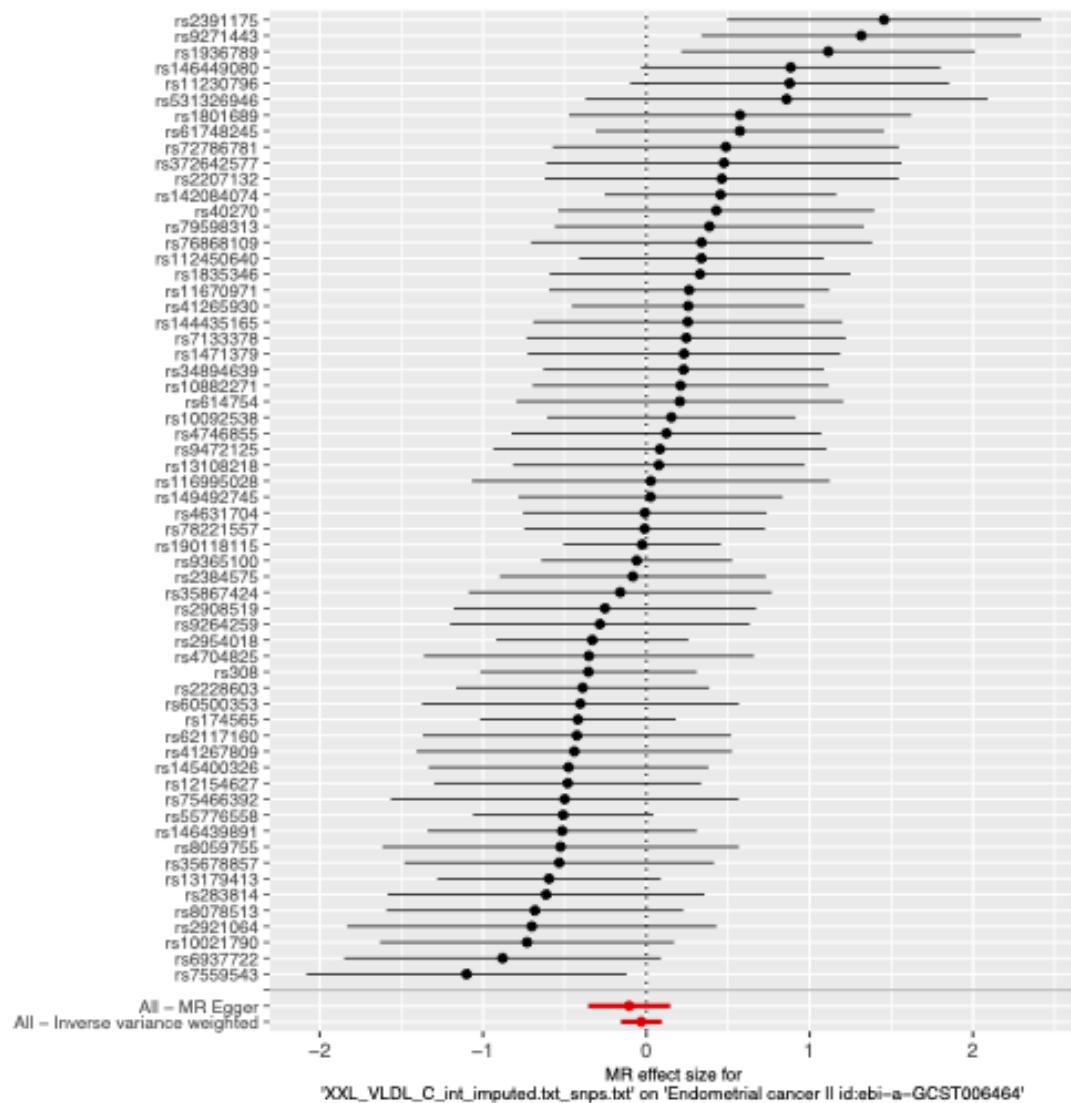


Figure A.46: **Representativ figure: Single SNP plot of MR effect estimates from two-sample MR of metabolites on endometrial cancer.**

5635

References

- 5636 1. Bos, M. M. *et al.* Investigating the relationships between unfavourable habitual
5637 sleep and metabolomic traits: evidence from multi-cohort multivariable regression
and Mendelian randomization analyses. *BMC medicine* **19**, 69 (2021).
- 5638 2. World Health Organisation. *Factsheet: Obesity and overweight*. (2018).
- 5639
- 5640 3. Ritchie, H. & Roser, M. *Obesity*. (2019).
- 5641
- 5642 4. Stanaway, J. D. *et al.* Global, regional, and national comparative risk assessment of
84 behavioural, environmental and occupational, and metabolic risks or clusters of
risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global
5643 Burden of Disease Stu. *The Lancet* **392**, 1923–1994 (2018).
- 5644 5. Ng, M. *et al.* Global, regional, and national prevalence of overweight and obesity in
children and adults during 1980–2013: a systematic analysis for the Global Burden of
5645 Disease Study 2013. *Lancet (London, England)* **384**, 766–781 (2014).
- 5646 6. (NCD-RisC), N. R. F. C. Trends in adult body-mass index in 200 countries from 1975
5647 to 2014: a pooled analysis of 1698 population-based measurement studies with 19·2
million participants. *The Lancet* **387**, 1377–1396 (2016).
- 5648 7. Abarca-Gómez, L. *et al.* Worldwide trends in body-mass index, underweight, over-
weight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based
5649 measurement studies in 128·9 million children, adolescents, and adults. *The
Lancet* **390**, 2627–2642 (2017).
- 5650 8. Kim, C. H. Measurements of Adiposity and Body Composition. *The Korean Journal of
Obesity* **25**, 115–120 (2016).

- 5652 9. Frayn, K. N., Karpe, F., Fielding, B. A., Macdonald, I. A. & Coppack, S. W. Integrative
5653 physiology of human adipose tissue. *International Journal of Obesity* **27**, 875–888
(2003).
- 5654 10. Cohen, P. & Spiegelman, B. M. Cell biology of fat storage. *Molecular biology of the cell*
5655 **27**, 2523–2527 (2016).
- 5656 11. Luo, L. & Liu, M. Adipose tissue in control of metabolism. *Journal of Endocrinology*
5657 **231**, R77–R99 (2016).
- 5658 12. Kershaw, E. E. & Flier, J. S. Adipose Tissue as an Endocrine Organ. *The Journal of*
5659 *Clinical Endocrinology & Metabolism* **89**, 2548–2556 (2004).
- 5660 13. Dahlman, I. & Arner, P. Chapter 3 - Genetics of Adipose Tissue Biology. in *Genes*
5661 and *obesity* (eds. Bouchard, C. B. T.-. P. in M. B. & Science, T.) **94**, 39–74 (Academic
5662 Press, 2010).
- 5663 14. Galgani, J. & Ravussin, E. Energy metabolism, fuel selection and body weight regulation.
5664 *International journal of obesity* (2005) **32 Suppl 7**, S109–S119 (2008).
- 5665 15. Amitani, M., Asakawa, A., Amitani, H. & Inui, A. The role of leptin in the control of
5666 insulin-glucose axis . **7**, 51 (2013).
- 5667 16. Gray, S. L. & Vidal-Puig, A. J. Adipose Tissue Expandability in the Maintenance of
5668 Metabolic Homeostasis. *Nutrition Reviews* **65**, S7–S12 (2007).
- 5669 17. Cannon, B. & Nedergaard, J. A. N. Brown Adipose Tissue: Function and Physiological
5670 Significance. *Physiological Reviews* **84**, 277–359 (2004).
- 5671 18. Lee, J.-E. & Ge, K. Transcriptional and epigenetic regulation of PPAR γ expression
5672 during adipogenesis. *Cell & Bioscience* **4**, 29 (2014).
- 5673 19. Schreiber, R., Xie, H. & Schweiger, M. Of mice and men: The physiological role of
5674 adipose triglyceride lipase (ATGL). *Biochimica et Biophysica Acta (BBA) - Molecular*
and Cell Biology of Lipids **1864**, 880–899 (2019).
- 5675 20. Lehr, S., Hartwig, S. & Sell, H. Adipokines: A treasure trove for the discovery of
5676 biomarkers for metabolic disorders. *PROTEOMICS – Clinical Applications* **6**, 91–101
(2012).

- 5676 21. Fasshauer, M. & Blüher, M. Adipokines in health and disease. *Trends in Pharmacological Sciences* **36**, 461–470 (2015).
- 5677
- 5678 22. Blüher, M. Adipose tissue dysfunction contributes to obesity related metabolic diseases. *Best Practice & Research Clinical Endocrinology & Metabolism* **27**, 163–177 (2013).
- 5679
- 5680 23. Blüher, M. & Mantzoros, C. S. From leptin to other adipokines in health and disease: Facts and expectations at the beginning of the 21st century. *Metabolism* **64**, 131–145 (2015).
- 5681
- 5682 24. Yaghoodeh, H. *et al.* Mendelian Randomization Studies Do Not Support a Causal Role for Reduced Circulating Adiponectin Levels in Insulin Resistance and Type 2 Diabetes. *Diabetes* **62**, 3589 LP–3598 (2013).
- 5683
- 5684 25. Borges, M. C. *et al.* Metabolic Profiling of Adiponectin Levels in Adults: Mendelian Randomization Analysis. *Circulation. Cardiovascular genetics* **10**, e001837–e001837 (2017).
- 5685
- 5686 26. Pan, W. W. & Myers, M. G. Leptin and the maintenance of elevated body weight. *Nature Reviews Neuroscience* **19**, 95–105 (2018).
- 5687
- 5688 27. Ahn, J., Wu, H. & Lee, K. Integrative Analysis Revealing Human Adipose-Specific Genes and Consolidating Obesity Loci. *Scientific Reports* **9**, 3087 (2019).
- 5689
- 5690 28. Achari, E. A. & Jain, K. S. Adiponectin, a Therapeutic Target for Obesity, Diabetes, and Endothelial Dysfunction. **18**, (2017).
- 5691
- 5692 29. Lefterova, M. I. & Lazar, M. A. New developments in adipogenesis. *Trends in Endocrinology & Metabolism* **20**, 107–114 (2009).
- 5693
- 5694 30. Huang, L. O. *et al.* Genome-wide discovery of genetic loci that uncouple excess adiposity from its comorbidities. *Nature Metabolism* **3**, 228–243 (2021).
- 5695
- 5696 31. Maguire, D., Talwar, D., Shiels, P. G. & McMillan, D. The role of thiamine dependent enzymes in obesity and obesity related chronic disease states: A systematic review. *Clinical Nutrition ESPEN* **25**, 8–17 (2018).
- 5697
- 5698 32. Ma, R. C. W. & Chan, J. C. N. Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States. *Annals of the New York Academy of Sciences* **1281**, 64–91 (2013).
- 5699

- 5700 33. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci
5701 associated with body mass index. *Nat Genet* **42**, 937–948 (2010).
- 5702 34. Kilpeläinen, T. O. *et al.* Genetic variation near IRS1 associates with reduced adiposity
5703 and an impaired metabolic profile. *Nature genetics* **43**, 753–760 (2011).
- 5704 35. Yaghootkar, H. *et al.* Genetic evidence for a normal-weight ‘metabolically obese’
5705 phenotype linking insulin resistance, hypertension, coronary artery disease and type 2
diabetes. *Diabetes* (2014). doi:10.2337/db14-0318
- 5706 36. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity
5707 biology. *Nature* **518**, 197–206 (2015).
- 5708 37. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat
5709 distribution. *Nature* **518**, 187–196 (2015).
- 5710 38. Sung, Y. J. *et al.* Genome-wide association studies suggest sex-specific loci associated
5711 with abdominal and visceral fat. *International Journal of Obesity* **40**, 662–674 (2016).
- 5712 39. Lu, Y. *et al.* New loci for body fat percentage reveal link between adiposity and
5713 cardiometabolic disease risk. *Nature Communications* **7**, 10495 (2016).
- 5714 40. Akiyama, M. *et al.* Genome-wide association study identifies 112 new loci for body
5715 mass index in the Japanese population. *Nature Genetics* **49**, 1458–1467 (2017).
- 5716 41. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body
5717 mass index in 700000 individuals of European ancestry. *Human molecular genetics*
27, 3641–3649 (2018).
- 5718 42. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat
5719 distribution in 694 649 individuals of European ancestry. *Human molecular genetics*
28, 166–174 (2019).
- 5720 43. Hübel, C. *et al.* Genomics of body fat percentage may contribute to sex bias in anorexia
5721 nervosa. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the
official publication of the International Society of Psychiatric Genetics* **180**, 428–438
(2019).
- 5722 44. Elks, C. E. *et al.* Variability in the heritability of body mass index: a systematic review
5723 and meta-regression. *Frontiers in endocrinology* **3**, 29 (2012).

- 5724 45. Silventoinen, K. *et al.* Differences in genetic and environmental variation in adult BMI
5725 by sex, age, time period, and region: an individual-based pooled analysis of 40 twin
cohorts. *The American journal of clinical nutrition* **106**, 457–466 (2017).
- 5726 46. Robinson, M. R. *et al.* Genotype–covariate interaction effects and the heritability of
5727 adult body mass index. *Nature Genetics* **49**, 1174–1181 (2017).
- 5728 47. Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible
5729 missing heritability for human height and body mass index. *Nature genetics* **47**,
1114–1120 (2015).
- 5730 48. Rose, K. M., Newman, B., Mayer-Davis, E. J. & Selby, J. V. Genetic and Behavioral
5731 Determinants of Waist-Hip Ratio and Waist Circumference in Women Twins. *Obesity
Research* **6**, 383–392 (1998).
- 5732 49. Henneman, P. *et al.* Prevalence and heritability of the metabolic syndrome and its
5733 individual components in a Dutch isolate: the Erasmus Rucphen Family study. *Journal
of Medical Genetics* **45**, 572 LP–577 (2008).
- 5734 50. Kutalik, Z. *et al.* Novel method to estimate the phenotypic variation explained by
5735 genome-wide association studies reveals large fraction of the missing heritability.
Genetic Epidemiology **35**, 341–349 (2011).
- 5736 51. Hsu, F.-C. *et al.* Heritability of Body Composition Measured by DXA in the Diabetes
5737 Heart Study. *Obesity Research* **13**, 312–319 (2005).
- 5738 52. Frayling, T. M. *et al.* A common variant in the FTO gene is associated with body mass
5739 index and predisposes to childhood and adult obesity. *Science (New York, N.Y.)* **316**,
889–894 (2007).
- 5740 53. Scuteri, A. *et al.* Genome-Wide Association Scan Shows Genetic Variants in the FTO
5741 Gene Are Associated with Obesity-Related Traits. *PLOS Genetics* **3**, e115 (2007).
- 5742 54. Ng, M. C. Y. *et al.* Discovery and fine-mapping of adiposity loci using high density
5743 imputation of genome-wide association studies in individuals of African ancestry:
African Ancestry Anthropometry Genetics Consortium. *PLoS genetics* **13**, e1006719–
e1006719 (2017).
- 5744 55. Loos, R. J. The genetics of adiposity. *Current opinion in genetics & development* **50**,
5745 86–95 (2018).

- 5746 56. Liu, X., Li, Y. I. & Pritchard, J. K. Trans Effects on Gene Expression Can Drive
5747 Omnipotent Inheritance. *Cell* **177**, 1022–1034.e6 (2019).
- 5748 57. Consortium, G. The Genotype-Tissue Expression (GTEx) project. *Nature genetics* **45**,
5749 580–585 (2013).
- 5750 58. Ghosh, S. & Bouchard, C. Convergence between biological, behavioural and genetic
5751 determinants of obesity. *Nature Reviews Genetics* **18**, 731–748 (2017).
- 5752 59. WHO. Physical Status: The Use and Interpretation of Anthropometry: Report of a
5753 World Health Organization (WHO) Expert Committee. *Geneva, Switzerland: World
Health Organization* (1995).
- 5754 60. World Health Organisation. *Obesity : preventing and managing the global epidemic :
5755 report of a WHO consultation.* (WHO, 2000).
- 5756 61. Donato, K. A. Executive summary of the clinical guidelines on the identification, evalua-
5757 tion, and treatment of overweight and obesity in adults. *Archives of Internal Medicine*
158, 1855–1867 (1998).
- 5758 62. U.S. Department of Health and Human Services. Managing Overweight and Obesity
5759 in Adults: Systematic Evidence Review from the Obesity Expert Panel. *National Heart,
Lung, and Blood Institute* 1–501 (2013).
- 5760 63. Flegal, K. M. & Graubard, B. I. Estimates of excess deaths associated with body mass
5761 index and other anthropometric variables. *The American journal of clinical nutrition* **89**,
1213–1219 (2009).
- 5762 64. Flegal, K. M. *et al.* Comparisons of percentage body fat, body mass index, waist
5763 circumference, and waist-stature ratio in adults. *The American journal of clinical
nutrition* **89**, 500–508 (2009).
- 5764 65. Wells, J. C. K. Sexual dimorphism of body composition. *Best Practice & Research
5765 Clinical Endocrinology & Metabolism* **21**, 415–430 (2007).
- 5766 66. Wagner, D. R. & Heyward, V. H. Measures of body composition in blacks and whites: a
5767 comparative review. *The American Journal of Clinical Nutrition* **71**, 1392–1402 (2000).
- 5768 67. Flegal, K. M. *et al.* High adiposity and high body mass index-for-age in US children
5769 and adolescents overall and by race-ethnic group. *The American journal of clinical
nutrition* **91**, 1020–1026 (2010).

- 5770 68. Wang, J. *et al.* Asians have lower body mass index (BMI) but higher percent body fat
5771 than do whites: comparisons of anthropometric measurements. *The American Journal
of Clinical Nutrition* **60**, 23–28 (1994).
- 5772 69. Consultation, W. expert. Appropriate body-mass index for Asian populations and its
5773 implications for policy and intervention strategies. *The Lancet* **363**, 157–163 (2004).
- 5774 70. Prentice, A. M. & Jebb, S. A. Beyond body mass index. *Obesity Reviews* **2**, 141–147
5775 (2001).
- 5776 71. Pischeddu, T. *et al.* General and Abdominal Adiposity and Risk of Death in Europe. *New
5777 England Journal of Medicine* **359**, 2105–2120 (2008).
- 5778 72. Romero-Corral, A. *et al.* Accuracy of body mass index in diagnosing obesity in the
5779 adult general population. *International Journal of Obesity* **32**, 959–966 (2008).
- 5780 73. Okorodudu, D. O. *et al.* Diagnostic performance of body mass index to identify obesity
5781 as defined by body adiposity: a systematic review and meta-analysis. *International
Journal of Obesity* **34**, 791–799 (2010).
- 5782 74. Pasco, J. A., Nicholson, G. C., Brennan, S. L. & Kotowicz, M. A. Prevalence of obesity
5783 and the relationship between the body mass index and body fat: Cross-sectional,
population-based data. *PLoS ONE* **7**, (2012).
- 5784 75. Flegal, K. M., Kit, B. K., Orpana, H. & Graubard, B. I. Association of all-cause mortality
5785 with overweight and obesity using standard body mass index categories: a systematic
review and meta-analysis. *JAMA* **309**, 71–82 (2013).
- 5786 76. Lee, J., Meyerhardt, J. A., Giovannucci, E. & Jeon, J. Y. Association between body
5787 mass index and prognosis of colorectal cancer: a meta-analysis of prospective cohort
studies. *PloS one* **10**, e0120706–e0120706 (2015).
- 5788 77. Elagizi, A. *et al.* An Overview and Update on Obesity and the Obesity Paradox in
5789 Cardiovascular Diseases. *Progress in Cardiovascular Diseases* **61**, 142–150 (2018).
- 5790 78. Banack, H. R. & Kaufman, J. S. The ‘Obesity Paradox’ Explained. *Epidemiology* **24**,
5791 (2013).
- 5792 79. Willett, W. C., Hu, F. B. & Thun, M. Overweight, Obesity, and All-Cause Mortality. *JAMA*
5793 **309**, 1681–1682 (2013).

- 5794 80. Khan, S. S. *et al.* Association of Body Mass Index With Lifetime Risk of Cardiovascular
5795 Disease and Compression of Morbidity. *JAMA Cardiology* **3**, 280–287 (2018).
- 5796 81. Sahakyan, K. R. *et al.* Normal-Weight Central Obesity: Implications for Total and
5797 Cardiovascular Mortality. *Annals of internal medicine* **163**, 827–835 (2015).
- 5798 82. Hamer, M., O'Donovan, G., Stensel, D. & Stamatakis, E. Normal-Weight Central
5799 Obesity and Risk for Mortality. *Annals of Internal Medicine* **166**, 917–918 (2017).
- 5800 83. World Health Organisation. *Waist Circumference and Waist-Hip Ratio: Report of a*
5801 *WHO Expert Consultation.* (WHO, 2008).
- 5802 84. World Health Organisation. *Obesity : preventing and managing the global epidemic : report of a WHO consultation.* (WHO, 2000).
- 5803
- 5804 85. Jebb, S. A., Cole, T. J., Doman, D., Murgatroyd, P. R. & Prentice, A. M. Evaluation of
5805 the novel Tanita body-fat analyser to measure body composition by comparison with a
four-compartment model. *British Journal of Nutrition* **83**, 115–122 (2000).
- 5806 86. Chouinard, L. E. *et al.* Bioelectrical Impedance vs. Four-compartment Model to Assess
5807 Body Fat Change in Overweight Adults. *Obesity* **15**, 85–92 (2007).
- 5808 87. Tallroth, K., Kettunen, J. A. & Kujala, U. M. Reproducibility of Regional DEXA Examinations
5809 of Abdominal Fat and Lean Tissue. *Obesity Facts* **6**, 203–210 (2013).
- 5810 88. Prospective Studies Collaboration *et al.* Body-mass index and cause-specific mortality
5811 in 900 000 adults: collaborative analyses of 57 prospective studies. *Lancet (London, England)* **373**, 1083–1096 (2009).
- 5812 89. Jenkins, D. A. *et al.* Adiposity-Mortality Relationships in Type 2 Diabetes, Coronary
5813 Heart Disease, and Cancer Subgroups in the UK Biobank, and Their Modification by
Smoking. *Diabetes Care* **41**, 1878 LP–1886 (2018).
- 5814 90. Rost, S. *et al.* New indexes of body fat distribution and sex-specific risk of total and
5815 cause-specific mortality: a prospective cohort study. *BMC Public Health* **18**, 427 (2018).
- 5816 91. Mørkedal, B., Romundstad, P. R. & Vatten, L. J. Informativeness of indices of blood
5817 pressure, obesity and serum lipids in relation to ischaemic heart disease mortality: the
HUNT-II study. *European journal of epidemiology* **26**, 457–461 (2011).

- 5818 92. Dong, B. *et al.* Joint association between body fat and its distribution with all-cause
5819 mortality: A data linkage cohort study based on NHANES (1988-2011). *PLoS one* **13**,
5819 e0193368–e0193368 (2018).
- 5820 93. Lee, D. H. *et al.* Predicted lean body mass, fat mass, and all cause and cause specific
5821 mortality in men: prospective US cohort study. *BMJ* **362**, k2575 (2018).
- 5822 94. Bigaard, J. *et al.* Body Fat and Fat-Free Mass and All-Cause Mortality. *Obesity
5823 Research* **12**, 1042–1049 (2004).
- 5824 95. Barberio, A. M. *et al.* Central body fatness is a stronger predictor of cancer risk than
5825 overall body size. *Nature Communications* **10**, 383 (2019).
- 5826 96. Dobbelsteyn, C. J., Joffres, M. R., MacLean, D. R., Flowerdew, G. & Group, T. C. H.
5827 H. S. R. A comparative evaluation of waist circumference, waist-to-hip ratio and body
mass index as indicators of cardiovascular risk factors. The Canadian Heart Health
Surveys. *International Journal of Obesity* **25**, 652–661 (2001).
- 5828 97. Yusuf, S. *et al.* Obesity and the risk of myocardial infarction in 27000 participants from
5829 52 countries: a case-control study. *The Lancet* **366**, 1640–1649 (2005).
- 5830 98. Elsayed, E. F. *et al.* Waist-to-hip ratio, body mass index, and subsequent kidney
5831 disease and death. *American journal of kidney diseases : the official journal of the
National Kidney Foundation* **52**, 29–38 (2008).
- 5832 99. Sahlman, P. *et al.* Genetic and lifestyle risk factors for advanced liver disease among
5833 men and women. *Journal of gastroenterology and hepatology* **35**, 291–298 (2020).
- 5834 100. Wiltink, J. *et al.* Associations between depression and different measures of obesity
5835 (BMI, WC, WHtR, WHR). *BMC psychiatry* **13**, 223 (2013).
- 5836 101. Dye, L., Boyle, N. B., Champ, C. & Lawton, C. The relationship between obesity and
5837 cognitive health and decline. *The Proceedings of the Nutrition Society* **76**, 443–454
(2017).
- 5838 102. Basraon, S. K. *et al.* Relationship of Early Pregnancy Waist-to-Hip Ratio versus Body
5839 Mass Index with Gestational Diabetes Mellitus and Insulin Resistance. *American
journal of perinatology* **33**, 114–121 (2016).

- 5840 103. Price, G. M., Uauy, R., Breeze, E., Bulpitt, C. J. & Fletcher, A. E. Weight, shape, and
5841 mortality risk in older persons: elevated waist-hip ratio, not high body mass index, is
5842 associated with a greater risk of death. *The American journal of clinical nutrition* **84**,
5843 449–460 (2006).
- 5842 104. Qiao, Q. & Nyamdjorj, R. Is the association of type II diabetes with waist circumference
5843 or waist-to-hip ratio stronger than that with body mass index?. *European journal of
5844 clinical nutrition* **64**, 30–34 (2010).
- 5844 105. Romero-Corral, A. *et al.* Normal weight obesity: a risk factor for cardiometabolic
5845 dysregulation and cardiovascular mortality. *European heart journal* **31**, 737–746
(2010).
- 5846 106. Oh, S. W. *et al.* Relationship between changes in body fat and a decline of renal
5847 function in the elderly. *PLoS one* **9**, e84052–e84052 (2014).
- 5848 107. Jo, A. & Mainous III, A. G. Informational value of percent body fat with body mass index
5849 for the risk of abnormal blood glucose: a nationally representative cross-sectional
5850 study. *BMJ Open* **8**, e019200 (2018).
- 5850 108. Bhaskaran, K. *et al.* Body-mass index and risk of 22 specific cancers: a population-
5851 based cohort study of 5·24 million UK adults. *Lancet (London, England)* **384**, 755–765
(2014).
- 5852 109. Choi, J. B. *et al.* Does increased body mass index lead to elevated prostate cancer
5853 risk? It depends on waist circumference. *BMC Cancer* **20**, 589 (2020).
- 5854 110. Kyrgiou, M. *et al.* Adiposity and cancer at major anatomical sites: umbrella review of
5855 the literature. *BMJ (Clinical research ed.)* **356**, j477–j477 (2017).
- 5856 111. Hu, M.-B., Liu, S.-H., Jiang, H.-W., Bai, P.-D. & Ding, Q. Obesity affects the biopsy-
5857 mediated detection of prostate cancer, particularly high-grade prostate cancer: a
5858 dose-response meta-analysis of 29,464 patients. *PLoS one* **9**, e106677–e106677
(2014).
- 5858 112. Hajiebrahimi, M., Bahmanyar, S., Öberg, S., Iliadou, A. N. & Cnattingius, S. Breast
5859 Cancer Risk in Opposite-Sexed Twins: Influence of Birth Weight and Co-Twin Birth
Weight. *JNCI: Journal of the National Cancer Institute* **105**, 1833–1836 (2013).

- 5860 113. Munsell, M. F., Sprague, B. L., Berry, D. A., Chisholm, G. & Trentham-Dietz, A. Body
5861 mass index and breast cancer risk according to postmenopausal estrogen-progestin
use and hormone receptor status. *Epidemiologic reviews* **36**, 114–136 (2014).
- 5862 114. Suzuki, R., Orsini, N., Saji, S., Key, T. J. & Wolk, A. Body weight and incidence of
5863 breast cancer defined by estrogen and progesterone receptor status—A meta-analysis.
International Journal of Cancer **124**, 698–712 (2009).
- 5864 115. Harvie, M., Hooper, L. & Howell, A. H. Central obesity and breast cancer risk: a
5865 systematic review. *Obesity Reviews* **4**, 157–173 (2003).
- 5866 116. Gaudet, M. M. *et al.* Waist circumference, body mass index, and postmenopausal
5867 breast cancer incidence in the Cancer Prevention Study-II Nutrition Cohort. *Cancer
Causes & Control* **25**, 737–745 (2014).
- 5868 117. Lauby-Secretan, B. *et al.* Body Fatness and Cancer — Viewpoint of the IARC Working
5869 Group. *New England Journal of Medicine* **375**, 794–798 (2016).
- 5870 118. Freisling, H. *et al.* Comparison of general obesity and measures of body fat distribution
5871 in older adults in relation to cancer risk: meta-analysis of individual participant data of
seven prospective cohorts in Europe. *British journal of cancer* **116**, 1486–1497 (2017).
- 5872 119. Parra-Soto, S. *et al.* Associations of six adiposity-related markers with incidence and
5873 mortality from 24 cancers—findings from the UK Biobank prospective cohort study.
BMC medicine **19**, 7 (2021).
- 5874 120. Barrington, W. E. *et al.* Difference in Association of Obesity With Prostate Cancer
5875 Risk Between US African American and Non-Hispanic White Men in the Selenium and
Vitamin E Cancer Prevention Trial (SELECT). *JAMA oncology* **1**, 342–349 (2015).
- 5876 121. Heo, J.-W., Kim, S.-E. & Sung, M.-K. Sex Differences in the Incidence of Obesity-
5877 Related Gastrointestinal Cancer. *International journal of molecular sciences* **22**, 1253
(2021).
- 5878 122. Paul, P. *et al.* Obesity and Cardiovascular Disease: Pathophysiology, Evaluation, and
5879 Effect of Weight Loss. *Circulation* **113**, 898–918 (2006).
- 5880 123. Koliaki, C., Liatis, S. & Kokkinos, A. Obesity and cardiovascular disease: revisiting an
5881 old relationship. *Metabolism* **92**, 98–107 (2019).

- 5882 124. H., E. R. & M., K. R. American Heart Association Call to Action: Obesity as a Major
5883 Risk Factor for Coronary Heart Disease. *Circulation* **97**, 2099–2100 (1998).
- 5884 125. Cepeda-Lopez, A. C., Aeberli, I. & Zimmermann, M. B. Does Obesity Increase Risk for
5885 Iron Deficiency? A Review of the Literature and the Potential Mechanisms. *International Journal for Vitamin and Nutrition Research* **80**, 263–270 (2010).
- 5886 126. Blokhin, I. O. & Lentz, S. R. Mechanisms of thrombosis in obesity. *Current opinion in
5887 hematology* **20**, 437–444 (2013).
- 5888 127. Thomsen, M. & Nordestgaard, B. G. Myocardial Infarction and Ischemic Heart Disease
5889 in Overweight and Obesity With and Without Metabolic Syndrome. *JAMA Internal
Medicine* **174**, 15–22 (2014).
- 5890 128. Klop, B., Elte, J. W. F. & Cabezas, M. C. Dyslipidemia in obesity: mechanisms and
5891 potential targets. *Nutrients* **5**, 1218–1240 (2013).
- 5892 129. Mathew, B., Francis, L., Kayalar, A. & Cone, J. Obesity: Effects on Cardiovascular
5893 Disease and its Diagnosis. *The Journal of the American Board of Family Medicine* **21**,
562 LP–568 (2008).
- 5894 130. Hubert, H. B., Feinleib, M., McNamara, P. M. & Castelli, W. P. Obesity as an indepen-
5895 dent risk factor for cardiovascular disease: a 26-year follow-up of participants in the
Framingham Heart Study. *Circulation* **67**, 968–977 (1983).
- 5896 131. Kenchaiah, S. *et al.* Obesity and the Risk of Heart Failure. *New England Journal of
5897 Medicine* **347**, 305–313 (2002).
- 5898 132. Yusuf, S. *et al.* Effect of potentially modifiable risk factors associated with myocardial
5899 infarction in 52 countries (the INTERHEART study): case-control study. *The Lancet*
364, 937–952 (2004).
- 5900 133. Cuillin, Z., M., R. K., M., van D. R., Y., L. T. & B., H. F. Abdominal Obesity and the
5901 Risk of All-Cause, Cardiovascular, and Cancer Mortality. *Circulation* **117**, 1658–1667
(2008).
- 5902 134. Gelber, R. P. *et al.* Measures of Obesity and Cardiovascular Risk Among Men and
5903 Women. *Journal of the American College of Cardiology* **52**, 605–615 (2008).

- 5904 135. Litwin, S. E. Which Measures of Obesity Best Predict Cardiovascular Risk?**Editorials
5905 published in the Journal of the American College of Cardiology reflect the views of the
authors and do not necessarily represent the views of JACC or the American College
of Cardiolog. *Journal of the American College of Cardiology* **52**, 616–619 (2008).
- 5906 136. C., M. H. *et al.* Obesity Accelerates the Progression of Coronary Atherosclerosis in
5907 Young Men. *Circulation* **105**, 2712–2718 (2002).
- 5908 137. Ades, P. A. & Savage, P. D. Obesity in coronary heart disease: An unaddressed
5909 behavioral risk factor. *Preventive medicine* **104**, 117–119 (2017).
- 5910 138. Plourde, B., Sarrazin, J.-F., Nault, I. & Poirier, P. Sudden cardiac death and obesity.
5911 *Expert Review of Cardiovascular Therapy* **12**, 1099–1110 (2014).
- 5912 139. Pathak, R. K., Mahajan, R., Lau, D. H. & Sanders, P. The Implications of Obesity for
5913 Cardiac Arrhythmia Mechanisms and Management. *Canadian Journal of Cardiology*
31, 203–210 (2015).
- 5914 140. Lavie, C. J., Pandey, A., Lau, D. H., Alpert, M. A. & Sanders, P. Obesity and Atrial
5915 Fibrillation Prevalence, Pathogenesis, and Prognosis: Effects of Weight Loss and
Exercise. *Journal of the American College of Cardiology* **70**, 2022–2035 (2017).
- 5916 141. X., W. C. *et al.* Associations of Epicardial, Abdominal, and Overall Adiposity With Atrial
5917 Fibrillation. *Circulation: Arrhythmia and Electrophysiology* **9**, e004378 (2016).
- 5918 142. Kurth, T. *et al.* Body Mass Index and the Risk of Stroke in Men. *Archives of Internal
5919 Medicine* **162**, 2557–2562 (2002).
- 5920 143. N., K. W., E., I. S., Carla, S., F., M. R. & L., F. K. Obesity. *Stroke* **44**, 278–286 (2013).
- 5921
- 5922 144. S., R. N. *et al.* Plasma Concentration of C-Reactive Protein and Risk of Ischemic
5923 Stroke and Transient Ischemic Attack. *Stroke* **32**, 2575–2579 (2001).
- 5924 145. (CCGC), C. R. P. C. H. D. G. C. *et al.* Association between C reactive protein and coro-
5925 nary heart disease: mendelian randomisation analysis based on individual participant
data. *BMJ* **342**, d548 (2011).
- 5926 146. Cornier, M.-A. *et al.* The metabolic syndrome. *Endocrine reviews* **29**, 777–822 (2008).
- 5927

- 5928 147. Martyn, J. A. J., Kaneki, M. & Yasuhara, S. Obesity-induced insulin resistance and
5929 hyperglycemia: etiologic factors and molecular mechanisms. *Anesthesiology* **109**,
137–148 (2008).
- 5930 148. Cohn, G., Valdes, G. & Capuzzi, D. M. Pathophysiology and treatment of the dyslipi-
5931 demia of insulin resistance. *Current Cardiology Reports* **3**, 416–423 (2001).
- 5932 149. TOMÁS, E. V. A. *et al.* Hyperglycemia and Insulin Resistance: Possible Mechanisms.
5933 *Annals of the New York Academy of Sciences* **967**, 43–51 (2002).
- 5934 150. Ormazabal, V. *et al.* Association between insulin resistance and the development of
5935 cardiovascular disease. *Cardiovascular Diabetology* **17**, 122 (2018).
- 5936 151. Ye, J. Mechanisms of insulin resistance in obesity. *Frontiers of medicine* **7**, 14–24
5937 (2013).
- 5938 152. National Collaborating Centre for Chronic Conditions 2008. *National clinical guideline
5939 for early identification and management in adults in primary and secondary care.* (2008).
- 5940 153. Al-Goblan, A. S., Al-Alfi, M. A. & Khan, M. Z. Mechanism linking diabetes mellitus and
5941 obesity. *Diabetes, metabolic syndrome and obesity : targets and therapy* **7**, 587–591
(2014).
- 5942 154. Wilkin, T. J. The accelerator hypothesis: weight gain as the missing link between Type
5943 I and Type II diabetes. *Diabetologia* **44**, 914–922 (2001).
- 5944 155. Corbin, K. D. *et al.* Obesity in Type 1 Diabetes: Pathophysiology, Clinical Impact, and
5945 Mechanisms. *Endocrine Reviews* **39**, 629–663 (2018).
- 5946 156. Kahn, S. E. The Importance of β -Cell Failure in the Development and Progression of
5947 Type 2 Diabetes. *The Journal of Clinical Endocrinology & Metabolism* **86**, 4047–4058
(2001).
- 5948 157. Würtz, P. *et al.* Metabolic Signatures of Adiposity in Young Adults: Mendelian Ran-
5949 domization Analysis and Effects of Weight Change. *PLOS Medicine* **11**, e1001765
(2014).
- 5950 158. Cirulli, E. T. *et al.* Profound Perturbation of the Metabolome in Obesity Is Associated
5951 with Health Risk. *Cell Metab* **29**, 488–500 e2 (2019).

- 5952 159. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity
5953 for energy harvest. *Nature* **444**, 1027–1031 (2006).
- 5954 160. Castaner, O. *et al.* The Gut Microbiome Profile in Obesity: A Systematic Review.
5955 *International journal of endocrinology* **2018**, 4095789 (2018).
- 5956 161. Durack, J. & Lynch, S. V. The gut microbiome: Relationships with disease and oppor-
5957 tunities for therapy. *The Journal of experimental medicine* **216**, 20–40 (2019).
- 5958 162. Cominetti, O. *et al.* Obesity shows preserved plasma proteome in large independent
5959 clinical cohorts. *Scientific Reports* **8**, 16981 (2018).
- 5960 163. Figarska, S. M. *et al.* Proteomic profiles before and during weight loss: Results from
5961 randomized trial of dietary intervention. *Scientific Reports* **10**, 7913 (2020).
- 5962 164. Fernandes Vilegas, D. & Cicogna, A. C. Effects of obesity on the cardiac proteome.
5963 *Endocrine and Metabolic Science* **2**, 100076 (2021).
- 5964 165. De Heredia, F. P., Gómez-Martínez, S. & Marcos, A. Chronic and degenerative dis-
5965 eases: Obesity, inflammation and the immune system. *Proceedings of the Nutrition
Society* **71**, 332–338 (2012).
- 5966 166. Festa, A. *et al.* The relation of body fat mass and distribution to markers of chronic
5967 inflammation. *International Journal of Obesity* **25**, 1407–1415 (2001).
- 5968 167. Park, H. S., Park, J. Y. & Yu, R. Relationship of obesity and visceral adiposity with
5969 serum concentrations of CRP, TNF-alpha and IL-6. *Diabetes Research and Clinical
Practice* **69**, 29–35 (2005).
- 5970 168. Bastard, J.-P. *et al.* Recent advances in the relationship between obesity, inflammation,
5971 and insulin resistance. *European Cytokine Network* **17**, 4–12 (2006).
- 5972 169. Lapice, E. *et al.* Abdominal adiposity is associated with elevated C-reactive protein
5973 independent of BMI in healthy nonobese people. *Diabetes care* **32**, 1734–1736 (2009).
- 5974 170. Hermsdorff, H. H. M., Zulet, M. Á., Puchau, B. & Martínez, J. A. Central Adiposity
5975 Rather Than Total Adiposity Measurements Are Specifically Involved in the Inflamma-
tory Status from Healthy Young Adults. *Inflammation* **34**, 161–170 (2011).
- 5976 171. Wärnberg, J. *et al.* Lifestyle-related determinants of inflammation in adolescence.
5977 *British Journal of Nutrition* **98**, 116–120 (2007).

- 5978 172. Quijada, Z. *et al.* The triglyceride/HDL-cholesterol ratio as a marker of cardiovascular
5979 risk in obese children; association with traditional and emergent risk factors. *Pediatric
Diabetes* **9**, 464–471 (2008).
- 5980 173. Martí, A., Marcos, A. & Martínez, J. A. Obesity and immune function relationships.
5981 *Obesity Reviews* **2**, 131–140 (2001).
- 5982 174. Tilg, H. & Moschen, A. R. Adipocytokines: mediators linking adipose tissue, inflamma-
5983 tion and immunity. *Nature Reviews Immunology* **6**, 772–783 (2006).
- 5984 175. Fernandez-Real, J.-M., Vendrell, J. & Ricart, W. Circulating Adiponectin and Plasma
5985 Fatty Acid Profile. *Clinical Chemistry* **51**, 603–609 (2005).
- 5986 176. Stryjecki, C. & Mutch, D. M. Fatty acid–gene interactions, adipokines and obesity.
5987 *European Journal of Clinical Nutrition* **65**, 285–297 (2011).
- 5988 177. Shu, C. J., Benoist, C. & Mathis, D. The immune system's involvement in obesity-driven
5989 type 2 diabetes. *Seminars in immunology* **24**, 436–442 (2012).
- 5990 178. Wang, Z. & Nakayama, T. Inflammation, a link between obesity and cardiovascular
5991 disease. *Mediators of inflammation* **2010**, 535918 (2010).
- 5992 179. Yudkin, J. S., Kumari, M., Humphries, S. E. & Mohamed-Ali, V. Inflammation, obesity,
5993 stress and coronary heart disease: is interleukin-6 the link? *Atherosclerosis* **148**,
209–214 (2000).
- 5994 180. Pepys, M. B. & Hirschfield, G. M. C-reactive protein: a critical update. *The Journal of
5995 clinical investigation* **111**, 1805–1812 (2003).
- 5996 181. Visser, M., Bouter, L. M., McQuillan, G. M., Wener, M. H. & Harris, T. B. Elevated
5997 C-Reactive Protein Levels in Overweight and Obese Adults. *JAMA* **282**, 2131–2135
(1999).
- 5998 182. Collaboration, E. R. F. *et al.* C-reactive protein concentration and risk of coronary heart
5999 disease, stroke, and mortality: an individual participant meta-analysis. *Lancet (London,
England)* **375**, 132–140 (2010).

- 6000 183. Momiyama, Y. *et al.* Associations Between Plasma C-reactive Protein Levels and the
6001 Severities of Coronary and Aortic Atherosclerosis. *Journal of Atherosclerosis and*
Thrombosis **17**, 460–467 (2010).
- 6002 184. Radmard, A. R. *et al.* Gallstone disease and obesity: a population-based study on
6003 abdominal fat distribution and gender differences. *Annals of Hepatology* **14**, 702–709
(2015).
- 6004 185. Khatua, B., El-Kurdi, B. & Singh, V. P. Obesity and pancreatitis. *Current Opinion in*
6005 *Gastroenterology* **33**, (2017).
- 6006 186. Lee, J. Y. J., Keane, M. G. & Pereira, S. Diagnosis and treatment of gallstone disease.
6007 *The Practitioner* **259**, 2, 15–19 (2015).
- 6008 187. Çoban, G. *et al.* Body Mass Index, Cholecystitis, Cholelithiasis, Pancreatitis and
6009 Imaging of Common Bile Duct Stones. *The American Journal of the Medical Sciences*
347, 364–369 (2014).
- 6010 188. Kovesdy, C. P., Furth, S., Zoccali, C. & Committee, W. K. D. S. Obesity and Kidney
6011 Disease: Hidden Consequences of the Epidemic. *Iranian journal of kidney diseases*
11, 99–108 (2017).
- 6012 189. Foster, M. C. *et al.* Overweight, obesity, and the development of stage 3 CKD: the
6013 Framingham Heart Study. *American journal of kidney diseases : the official journal of*
the National Kidney Foundation **52**, 39–48 (2008).
- 6014 190. Pinto-Sietsma, S.-J. *et al.* A central body fat distribution is related to renal function
6015 impairment, even in lean subjects. *American Journal of Kidney Diseases* **41**, 733–741
(2003).
- 6016 191. Kramer, H. *et al.* Obesity and Prevalent and Incident CKD: The Hypertension Detection
6017 and Follow-Up Program. *American Journal of Kidney Diseases* **46**, 587–594 (2005).
- 6018 192. Lu, J. L. *et al.* Association of age and BMI with kidney function and mortality: a cohort
6019 study. *The lancet. Diabetes & endocrinology* **3**, 704–714 (2015).
- 6020 193. Hsu, C., McCulloch, C. E., Iribarren, C., Darbinian, J. & Go, A. S. Body Mass Index and
6021 Risk for End-Stage Renal Disease. *Annals of Internal Medicine* **144**, 21–28 (2006).

- 6022 194. Munkhaugen, J., Lydersen, S., Widerøe, T.-E. & Hallan, S. Prehypertension, Obesity, and Risk of Kidney Disease: 20-Year Follow-up of the HUNT I Study in Norway. *American Journal of Kidney Diseases* **54**, 638–646 (2009).
- 6023
- 6024 195. Thoenes, M. *et al.* Abdominal obesity is associated with microalbuminuria and an elevated cardiovascular risk profile in patients with hypertension. *Vascular health and risk management* **5**, 577–585 (2009).
- 6025
- 6026 196. Kramer, H. *et al.* Waist Circumference, Body Mass Index, and ESRD in the REGARDS (Reasons for Geographic and Racial Differences in Stroke) Study. *American journal of kidney diseases : the official journal of the National Kidney Foundation* **67**, 62–69 (2016).
- 6027
- 6028 197. Sharma, K. The link between obesity and albuminuria: adiponectin and podocyte dysfunction. *Kidney International* **76**, 145–148 (2009).
- 6029
- 6030 198. Wolf, G. & Ziyadeh, F. Leptin and Renal Fibrosis. in *Contributions to nephrology* **151**, 175–183 (2006).
- 6031
- 6032 199. Ruan, X. Z., Varghese, Z. & Moorhead, J. F. An update on the lipid nephrotoxicity hypothesis. *Nature Reviews Nephrology* **5**, 713–721 (2009).
- 6033
- 6034 200. Fabbrini, E., Sullivan, S. & Klein, S. Obesity and nonalcoholic fatty liver disease: biochemical, metabolic, and clinical implications. *Hepatology (Baltimore, Md.)* **51**, 679–689 (2010).
- 6035
- 6036 201. Sarwar, R., Pierce, N. & Koppe, S. Obesity and nonalcoholic fatty liver disease: current perspectives. *Diabetes, metabolic syndrome and obesity : targets and therapy* **11**, 533–542 (2018).
- 6037
- 6038 202. Browning, J. D. *et al.* Prevalence of hepatic steatosis in an urban population in the United States: Impact of ethnicity. *Hepatology* **40**, 1387–1395 (2004).
- 6039
- 6040 203. Petersen, K. F. *et al.* Increased prevalence of insulin resistance and nonalcoholic fatty liver disease in Asian-Indian men. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 18273–18277 (2006).
- 6041
- 6042 204. Romeo, S. *et al.* Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nature genetics* **40**, 1461–1465 (2008).
- 6043

- 6044 205. Donnelly, K. L. *et al.* Sources of fatty acids stored in liver and secreted via lipoproteins
6045 in patients with nonalcoholic fatty liver disease. *The Journal of clinical investigation* **115**, 1343–1351 (2005).
- 6046 206. O'Brien, P. D., Hinder, L. M., Callaghan, B. C. & Feldman, E. L. Neurological conse-
6047 quences of obesity. *The Lancet. Neurology* **16**, 465–477 (2017).
- 6048 207. Qizilbash, N. *et al.* BMI and risk of dementia in two million people over two decades: a
6049 retrospective cohort study. *The Lancet Diabetes & Endocrinology* **3**, 431–436 (2015).
- 6050 208. Elias, M. F., Elias, P. K., Sullivan, L. M., Wolf, P. A. & D'Agostino, R. B. Obesity, diabetes
6051 and cognitive deficit: The Framingham Heart Study. *Neurobiology of Aging* **26**, 11–16 (2005).
- 6052 209. Anstey, K. J., Cherbuin, N., Budge, M. & Young, J. Body mass index in midlife and
6053 late-life as a risk factor for dementia: a meta-analysis of prospective studies. *Obesity
Reviews* **12**, e426–e437 (2011).
- 6054 210. Pedditizi, E., Peters, R. & Beckett, N. The risk of overweight/obesity in mid-life and
6055 late life for the development of dementia: a systematic review and meta-analysis of
longitudinal studies. *Age and Ageing* **45**, 14–21 (2016).
- 6056 211. Mrak, R. E. Alzheimer-type neuropathological changes in morbidly obese elderly
6057 individuals. *Clinical Neuropathology* **28**, 40–45 (2009).
- 6058 212. Gustafson, D., Lissner, L., Bengtsson, C., Björkelund, C. & Skoog, I. A 24-year follow-up
6059 of body mass index and cerebral atrophy. *Neurology* **63**, 1876 LP–1881 (2004).
- 6060 213. Jagust, W., Harvey, D., Mungas, D. & Haan, M. Central Obesity and the Aging Brain.
6061 *Archives of Neurology* **62**, 1545–1548 (2005).
- 6062 214. Ward, M. A., Carlsson, C. M., Trivedi, M. A., Sager, M. A. & Johnson, S. C. The effect
6063 of body mass index on global brain volume in middle-aged adults: a cross sectional
study. *BMC neurology* **5**, 23 (2005).
- 6064 215. Taki, Y. *et al.* Relationship Between Body Mass Index and Gray Matter Volume in 1,428
6065 Healthy Individuals. *Obesity* **16**, 119–124 (2008).
- 6066 216. Gazdzinski, S., Kornak, J., Weiner, M. W. & Meyerhoff, D. J. Body mass index and
6067 magnetic resonance markers of brain integrity in adults. *Annals of neurology* **63**,
652–657 (2008).

- 6068 217. Raji, C. A. *et al.* Brain structure and obesity. *Human Brain Mapping* **31**, 353–364
6069 (2010).
- 6070 218. Yau, P. L., Castro, M. G., Tagani, A., Tsui, W. H. & Convit, A. Obesity and metabolic
6071 syndrome and functional and structural brain impairments in adolescence. *Pediatrics*
130, e856–e864 (2012).
- 6072 219. Alosco, M. L. *et al.* Body mass index and brain structure in healthy children and
6073 adolescents. *International Journal of Neuroscience* **124**, 49–55 (2014).
- 6074 220. Molteni, R., Barnard, R. J., Ying, Z., Roberts, C. K. & Gómez-Pinilla, F. A high-fat,
6075 refined sugar diet reduces hippocampal brain-derived neurotrophic factor, neuronal
plasticity, and learning. *Neuroscience* **112**, 803–814 (2002).
- 6076 221. Ledreux, A., Wang, X., Schultzberg, M., Granholm, A.-C. & Freeman, L. R. Detimental
6077 effects of a high fat/high cholesterol diet on memory and hippocampal markers in aged
rats. *Behavioural brain research* **312**, 294–304 (2016).
- 6078 222. Thaler, J. P. *et al.* Obesity is associated with hypothalamic injury in rodents and
6079 humans. *The Journal of clinical investigation* **122**, 153–162 (2012).
- 6080 223. Thaler, J. P. & Schwartz, M. W. Minireview: Inflammation and obesity pathogenesis:
6081 the hypothalamus heats up. *Endocrinology* **151**, 4109–4115 (2010).
- 6082 224. Myers Jr, M. G., Leibel, R. L., Seeley, R. J. & Schwartz, M. W. Obesity and leptin
6083 resistance: distinguishing cause from effect. *Trends in endocrinology and metabolism: TEM* **21**, 643–651 (2010).
- 6084 225. Gelaye, B. *et al.* Body composition status and the risk of migraine: A meta-analysis.
6085 *Neurology* **88**, 1795–1804 (2017).
- 6086 226. Peterlin, B. L., Sacco, S., Bernecker, C. & Scher, A. I. Adipokines and Migraine: A
6087 Systematic Review. *Headache* **56**, 622–644 (2016).
- 6088 227. Luppino, F. S. *et al.* Overweight, obesity, and depression: a systematic review and
6089 meta-analysis of longitudinal studies. *Archives of general psychiatry* **67**, 220–229
(2010).
- 6090 228. Carey, M. *et al.* Prevalence of comorbid depression and obesity in general practice: a
6091 cross-sectional survey. *British Journal of General Practice* **64**, e122 LP–e127 (2014).

- 6092 229. Puhl, R. M. & Heuer, C. A. Obesity stigma: important considerations for public health. *American journal of public health* **100**, 1019–1028 (2010).
- 6093
- 6094 230. Vallis, M. Quality of life and psychological well-being in obesity management: improving the odds of success by managing distress. *International Journal of Clinical Practice* **70**, 196–205 (2016).
- 6095
- 6096 231. Zammit, C., Liddicoat, H., Moonsie, I. & Makker, H. Obesity and respiratory diseases. *International journal of general medicine* **3**, 335–343 (2010).
- 6097
- 6098 232. Gibson, G. J. Obesity, respiratory function and breathlessness. *Thorax* **55 Suppl 1**, S41–S44 (2000).
- 6099
- 6100 233. Kress, J. P., Pohlman, A. S., Alverdy, John. & Hall, J. B. The Impact of Morbid Obesity on Oxygen Cost of Breathing (V02RESP) at Rest. *American Journal of Respiratory and Critical Care Medicine* **160**, 883–886 (1999).
- 6101
- 6102 234. Laghi, F. & Tobin, M. J. Disorders of the Respiratory Muscles. *American Journal of Respiratory and Critical Care Medicine* **168**, 10–48 (2003).
- 6103
- 6104 235. Beuther, D. A. & Sutherland, E. R. Overweight, obesity, and incident asthma: a meta-analysis of prospective epidemiologic studies. *American journal of respiratory and critical care medicine* **175**, 661–666 (2007).
- 6105
- 6106 236. Rodrigo, G. J. & Plaza, V. Body Mass Index and Response to Emergency Department Treatment In Adults With Severe Asthma Exacerbations: A Prospective Cohort Study. *CHEST* **132**, 1513–1519 (2007).
- 6107
- 6108 237. Aaron, S. D. *et al.* Overdiagnosis of asthma in obese and nonobese adults. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne* **179**, 1121–1131 (2008).
- 6109
- 6110 238. Sin, D. D. & Sutherland, E. R. Obesity and the lung: 4 · Obesity and asthma. *Thorax* **63**, 1018 LP–1023 (2008).
- 6111
- 6112 239. Crummy, F., Piper, A. J. & Naughton, M. T. Obesity and the lung: 2 · Obesity and sleep-disordered breathing. *Thorax* **63**, 738 LP–746 (2008).
- 6113

- 6114 240. Mortimore, I. L., Marshall, I., Wraith, P. K., Sellar, R. J. & Douglas, N. J. Neck and Total
6115 Body Fat Deposition in Nonobese and Obese Patients with Sleep Apnea Compared
with That in Control Subjects. *American Journal of Respiratory and Critical Care
Medicine* **157**, 280–283 (1998).
- 6116 241. Eisner, M. D. *et al.* Body composition and functional limitation in COPD. *Respiratory
6117 research* **8**, 7 (2007).
- 6118 242. Franssen, F. M. E., O'Donnell, D. E., Goossens, G. H., Blaak, E. E. & Schols, A. M. W.
6119 J. Obesity and the lung: 5 · Obesity and COPD. *Thorax* **63**, 1110 LP–1117 (2008).
- 6120 243. McClean, K. M., Kee, F., Young, I. S. & Elborn, J. S. Obesity and the lung: 1 ·
6121 Epidemiology. *Thorax* **63**, 649 LP–654 (2008).
- 6122 244. Nakajima, K. *et al.* A Possible Association Between Suspected Restrictive Pattern as
6123 Assessed by Ordinary Pulmonary Function Test and the Metabolic Syndrome. *CHEST*
134, 712–718 (2008).
- 6124 245. Jones, R. L. & Nzekwu, M.-M. U. The Effects of Body Mass Index on Lung Volumes.
6125 *CHEST* **130**, 827–833 (2006).
- 6126 246. Peppard, P. E., Ward, N. R. & Morrell, M. J. The impact of obesity on oxygen desatu-
6127 ration during sleep-disordered breathing. *American journal of respiratory and critical
care medicine* **180**, 788–793 (2009).
- 6128 247. Chen, Y., Rennie, D., Cormier, Y. F. & Dosman, J. Waist circumference is associated
6129 with pulmonary function in normal-weight, overweight, and obese subjects. *The
American Journal of Clinical Nutrition* **85**, 35–39 (2007).
- 6130 248. Stang, J. & Huffman, L. G. Position of the Academy of Nutrition and Dietetics: Obesity,
6131 Reproduction, and Pregnancy Outcomes. *Journal of the Academy of Nutrition and
Dietetics* **116**, 677–691 (2016).
- 6132 249. Barker, D. J. P. The origins of the developmental origins theory. *Journal of Internal
6133 Medicine* **261**, 412–417 (2007).
- 6134 250. Sharp, G. C. & Lawlor, D. A. Paternal impact on the life course development of obesity
6135 and type 2 diabetes in the offspring. *Diabetologia* **62**, 1802–1810 (2019).

- 6136 251. Robker, R. L. *et al.* Obese Women Exhibit Differences in Ovarian Metabolites, Hormones, and Gene Expression Compared with Moderate-Weight Women. *The Journal of Clinical Endocrinology & Metabolism* **94**, 1533–1540 (2009).
- 6137
- 6138 252. Wise, L. A. *et al.* An internet-based prospective study of body size and time-to-pregnancy. *Human reproduction (Oxford, England)* **25**, 253–264 (2010).
- 6139
- 6140 253. Chandra, A., Copen, C. E. & Stephen, E. H. *Infertility and impaired fecundity in the United States, 1982-2010: data from the National Survey of Family Growth.* (US Department of Health; Human Services, Centers for Disease Control and . . . , 2013).
- 6141
- 6142 254. Álvarez-Blasco, F., Botella-Carretero, J. I., San Millán, J. L. & Escobar-Morreale, H. F. Prevalence and Characteristics of the Polycystic Ovary Syndrome in Overweight and Obese Women. *Archives of Internal Medicine* **166**, 2081–2086 (2006).
- 6143
- 6144 255. Beatriz Motta, A. The role of obesity in the development of polycystic ovary syndrome. *Current pharmaceutical design* **18**, 2482–2491 (2012).
- 6145
- 6146 256. Ehrmann, D. A. *et al.* Prevalence and Predictors of the Metabolic Syndrome in Women with Polycystic Ovary Syndrome. *The Journal of Clinical Endocrinology & Metabolism* **91**, 48–53 (2006).
- 6147
- 6148 257. Ehrmann, D. A., Barnes, R. B., Rosenfield, R. L., Cavaghan, M. K. & Imperial, J. Prevalence of impaired glucose tolerance and diabetes in women with polycystic ovary syndrome. *Diabetes care* **22**, 141–146 (1999).
- 6149
- 6150 258. Orio Jr, F. *et al.* The cardiovascular risk of young women with polycystic ovary syndrome: an observational, analytical, prospective case-control study. *The Journal of Clinical Endocrinology & Metabolism* **89**, 3696–3701 (2004).
- 6151
- 6152 259. Hammoud, A. O., Gibson, M., Peterson, C. M., Meikle, A. W. & Carrell, D. T. Impact of male obesity on infertility: a critical review of the current literature. *Fertility and sterility* **90**, 897–904 (2008).
- 6153
- 6154 260. Chen, R. Y. T., Wittert, G. A. & Andrews, G. R. Relative androgen deficiency in relation to obesity and metabolic status in older men. *Diabetes, Obesity and Metabolism* **8**, 429–435 (2006).
- 6155
- 6156 261. Heidler, S. *et al.* Is the metabolic syndrome an independent risk factor for erectile dysfunction? *The Journal of urology* **177**, 651–654 (2007).
- 6157

- 6158 262. Schummers, L., Hutcheon, J. A., Bodnar, L. M., Lieberman, E. & Himes, K. P. Risk of
adverse pregnancy outcomes by prepregnancy body mass index: a population-based
study to inform prepregnancy weight loss counseling. *Obstetrics and gynecology* **125**,
133 (2015).
- 6159
- 6160 263. Bautista-Castaño, I. *et al.* Maternal obesity in early pregnancy and risk of adverse
outcomes. *PLoS one* **8**, e80410 (2013).
- 6161
- 6162 264. Lo, J. O., Mission, J. F. & Caughey, A. B. Hypertensive disease of pregnancy and
maternal mortality. *Current Opinion in Obstetrics and Gynecology* **25**, 124–132 (2013).
- 6163
- 6164 265. Gaillard, R. *et al.* Risk factors and outcomes of maternal obesity and excessive weight
gain during pregnancy. *Obesity* **21**, 1046–1055 (2013).
- 6165
- 6166 266. Feig, D. S. *et al.* Preeclampsia as a risk factor for diabetes: a population-based cohort
study. *PLoS Med* **10**, e1001425 (2013).
- 6167
- 6168 267. Athukorala, C., Rumbold, A. R., Willson, K. J. & Crowther, C. A. The risk of adverse
pregnancy outcomes in women who are overweight or obese. *BMC pregnancy and
childbirth* **10**, 1–8 (2010).
- 6169
- 6170 268. Shin, D. & Song, W. O. Prepregnancy body mass index is an independent risk factor
for gestational hypertension, gestational diabetes, preterm labor, and small-and large-
for-gestational-age infants. *The Journal of Maternal-Fetal & Neonatal Medicine* **28**,
1679–1686 (2015).
- 6171
- 6172 269. Li, N. *et al.* Maternal prepregnancy body mass index and gestational weight gain on
pregnancy outcomes. *PLoS one* **8**, e82310 (2013).
- 6173
- 6174 270. Belbasis, L., Savvidou, M. D., Kanu, C., Evangelou, E. & Tzoulaki, I. Birth weight in
relation to health and disease in later life: an umbrella review of systematic reviews
and meta-analyses. *BMC Medicine* **14**, 147 (2016).
- 6175
- 6176 271. Stothard, K. J., Tennant, P. W. G., Bell, R. & Rankin, J. Maternal overweight and obesity
and the risk of congenital anomalies: a systematic review and meta-analysis. *Jama*
301, 636–650 (2009).
- 6177
- 6178 272. Carmichael, S. L., Rasmussen, S. A. & Shaw, G. M. Prepregnancy obesity: a complex
risk factor for selected birth defects. *Birth Defects Research Part A: Clinical and
Molecular Teratology* **88**, 804–810 (2010).
- 6179

- 6180 273. Aune, D., Saugstad, O. D., Henriksen, T. & Tonstad, S. Maternal body mass index and
6181 the risk of fetal death, stillbirth, and infant death: a systematic review and meta-analysis.
Jama **311**, 1536–1546 (2014).
- 6182 274. Bray, G. A. Medical Consequences of Obesity. *The Journal of Clinical Endocrinology*
6183 & *Metabolism* **89**, 2583–2589 (2004).
- 6184 275. Haslam, D. W. & James, W. P. T. Obesity. *The Lancet* **366**, 1197–1209 (2005).
- 6185
- 6186 276. Poulain, M. *et al.* The effect of obesity on chronic respiratory diseases: pathophysiology
6187 and therapeutic strategies. *CMAJ : Canadian Medical Association journal = journal de
l'Association medicale canadienne* **174**, 1293–1299 (2006).
- 6188 277. Carreras-Torres, R. *et al.* Role of obesity in smoking behaviour: Mendelian randomisa-
6189 tion study in UK Biobank. *BMJ* **361**, k1767 (2018).
- 6190 278. Jayedi, A., Rashidy-Pour, A., Khorshidi, M. & Shab-Bidar, S. Body mass index, abdomi-
6191 nal adiposity, weight gain and risk of developing hypertension: a systematic review and
dose–response meta-analysis of more than 2.3 million participants. *Obesity Reviews*
19, 654–667 (2018).
- 6192 279. Marie-Eve, P., André, T. & Jean-Pierre, D. Obesity Phenotypes, Diabetes, and Cardio-
6193 vascular Diseases. *Circulation Research* **126**, 1477–1500 (2021).
- 6194 280. Bulun, S. E., Chen, D., Moy, I., Brooks, D. C. & Zhao, H. Aromatase, breast cancer
6195 and obesity: a complex interaction. *Trends in endocrinology and metabolism: TEM* **23**,
83–89 (2012).
- 6196 281. Macciò, A. *et al.* Correlation of body mass index and leptin with tumor size and stage
6197 of disease in hormone-dependent postmenopausal breast cancer: preliminary results
and therapeutic implications. *Journal of molecular medicine (Berlin, Germany)* **88**,
677–686 (2010).
- 6198 282. Rose, D. P. & Vona-Davis, L. Interaction between menopausal status and obesity in
6199 affecting breast cancer risk. *Maturitas* **66**, 33–38 (2010).
- 6200 283. Kerlikowske, K. *et al.* Obesity, mammography use and accuracy, and advanced breast
6201 cancer risk. *Journal of the National Cancer Institute* **100**, 1724–1733 (2008).

- 6202 284. Lahmann, P. H. *et al.* Body size and breast cancer risk: findings from the European
6203 Prospective Investigation into Cancer And Nutrition (EPIC). *International journal of
cancer* **111**, 762–771 (2004).
- 6204 285. Dagfinn, A. *et al.* Body Mass Index, Abdominal Fatness, and Heart Failure Incidence
6205 and Mortality. *Circulation* **133**, 639–649 (2016).
- 6206 286. Ma, C. *et al.* Effects of weight loss interventions for adults who are obese on mortality,
6207 cardiovascular disease, and cancer: systematic review and meta-analysis. *BMJ* **359**,
j4849 (2017).
- 6208 287. Force, U. S. P. S. T. Behavioral Weight Loss Interventions to Prevent Obesity-Related
6209 Morbidity and Mortality in Adults: US Preventive Services Task Force Recommendation
Statement. *JAMA* **320**, 1163–1171 (2018).
- 6210 288. Arnold, M., Freisling, H. & Gunter, M. J. Intentional Weight Loss and Cancer Risk:
6211 Never Too Late to Lose Weight. *JNCI Cancer Spectrum* **3**, (2019).
- 6212 289. Luo, J., Hendryx, M. & Chlebowski, R. T. Intentional weight loss and cancer risk.
6213 *Oncotarget* **8**, 81719–81720 (2017).
- 6214 290. Clifton, P. M. & Keogh, J. B. Effects of Different Weight Loss Approaches on CVD Risk.
6215 *Current Atherosclerosis Reports* **20**, 27 (2018).
- 6216 291. Brown, J. D., Buscemi, J., Milsom, V., Malcolm, R. & O'Neil, P. M. Effects on cardiovascular
6217 risk factors of weight losses limited to 5-10. *Translational behavioral medicine* **6**,
339–346 (2016).
- 6218 292. Cercato, C. & Fonseca, F. A. Cardiovascular risk and obesity. *Diabetology & Metabolic
6219 Syndrome* **11**, 74 (2019).
- 6220 293. McVay, M. A., Yancy, W. S., Bennett, G. G., Jung, S.-H. & Voils, C. I. Perceived barriers
6221 and facilitators of initiation of behavioral weight loss interventions among adults with
obesity: a qualitative study. *BMC Public Health* **18**, 854 (2018).
- 6222 294. Picot, J., Jones, J., Colquitt, J. L., Loveman, E. & Clegg, A. J. Weight Loss Surgery for
6223 Mild to Moderate Obesity: A Systematic Review and Economic Evaluation. *Obesity
Surgery* **22**, 1496–1506 (2012).
- 6224 295. Madura 2nd, J. A. & Dibaise, J. K. Quick fix or long-term cure? Pros and cons of
6225 bariatric surgery. *F1000 medicine reports* **4**, 19 (2012).

- 6226 296. Corbin, L. J. & Timpson, N. J. Body mass index: Has epidemiology started to break
6227 down causal contributions to health and disease? *Obesity (Silver Spring, Md.)* **24**,
1630–1638 (2016).
- 6228 297. Gray, L. J. *et al.* A systematic review and mixed treatment comparison of pharma-
6229 cological interventions for the treatment of obesity. *Obesity Reviews* **13**, 483–498
(2012).
- 6230 298. Farr, O. M., Gavrieli, A. & Mantzoros, C. S. Leptin applications in 2015: what have
6231 we learned about leptin and obesity? *Current opinion in endocrinology, diabetes, and*
obesity **22**, 353–359 (2015).
- 6232 299. Lee, J. *et al.* Withaferin A is a leptin sensitizer with strong antidiabetic properties in
6233 mice. *Nature Medicine* **22**, 1023–1032 (2016).
- 6234 300. Okada-Iwabu, M., Iwabu, M., Yamauchi, T. & Kadowaki, T. Drug development re-
6235 search for novel adiponectin receptor-targeted antidiabetic drugs contributing to healthy
longevity. *Diabetology International* **10**, 237–244 (2019).
- 6236 301. Batty, D. Epidemiology Kept Simple: An Introduction to Classic and Modern Epidemiol-
6237 ogy. *BMJ* **318**, 470 (1999).
- 6238 302. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal
6239 inference in epidemiological studies. *Human molecular genetics* **23**, R89–R98 (2014).
- 6240 303. Schisterman, E. F., Cole, S. R. & Platt, R. W. Overadjustment bias and unnecessary
6241 adjustment in epidemiologic studies. *Epidemiology (Cambridge, Mass.)* **20**, 488–495
(2009).
- 6242 304. Richiardi, L., Bellocchio, R. & Zugna, D. Mediation analysis in epidemiology: methods,
6243 interpretation and bias. *International journal of epidemiology* **42**, 1511–1519 (2013).
- 6244 305. Prabhu, S. *et al.* Nanocarriers targeting adipose macrophages increase glucocorticoid
6245 anti-inflammatory potency to ameliorate metabolic dysfunction. *Biomaterials Science*
9, 506–518 (2021).
- 6246 306. Vangaveti, V., Shashidhar, V., Jarrod, G., Baune, B. T. & Kennedy, R. L. Free fatty
6247 acid receptors: emerging targets for treatment of diabetes and its complications.
Therapeutic advances in endocrinology and metabolism **1**, 165–175 (2010).

- 6248 307. Achari, E. A. & Jain, K. S. Adiponectin, a Therapeutic Target for Obesity, Diabetes, and
6249 Endothelial Dysfunction. **18**, (2017).
- 6250 308. Griffin, J. L. The Cinderella story of metabolic profiling: does metabolomics get to go
6251 to the functional genomics ball? *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**, 147–161 (2006).
- 6252 309. Su, L. J. *et al.* The use of metabolomics in population-based research. *Advances in nutrition (Bethesda, Md.)* **5**, 785–788 (2014).
- 6254 310. Chu, S. H. *et al.* Integration of Metabolomic and Other Omics Data in Population-Based
6255 Study Designs: An Epidemiological Perspective. *Metabolites* **9**, 117 (2019).
- 6256 311. Wishart, D. S. Metabolomics for Investigating Physiological and Pathophysiological
6257 Processes. *Physiological Reviews* **99**, 1819–1875 (2019).
- 6258 312. Johnson, C. H., Ivanisevic, J. & Siuzdak, G. Metabolomics: beyond biomarkers and
6259 towards mechanisms. *Nature Reviews Molecular Cell Biology* **17**, 451–459 (2016).
- 6260 313. Fearnley, L. G. & Inouye, M. Metabolomics in epidemiology: from metabolite con-
6261 centrations to integrative reaction networks. *International journal of epidemiology* **45**,
1319–1328 (2016).
- 6262 314. Roberts, L. D., Souza, A. L., Gerszten, R. E. & Clish, C. B. Targeted metabolomics.
6263 *Current protocols in molecular biology* **Chapter 30**, Unit30.2–30.2.24 (2012).
- 6264 315. Liu, X. & Locasale, J. W. Metabolomics: A Primer. *Trends in biochemical sciences* **42**,
6265 274–284 (2017).
- 6266 316. Vinayavekhin, N. & Saghatelian, A. Untargeted metabolomics. *Current protocols in
6267 molecular biology* **Chapter 30**, Unit 30.1.1–24 (2010).
- 6268 317. Shin, S.-Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nature
6269 Genetics* **46**, 543 (2014).
- 6270 318. Kettunen, J. *et al.* Genome-wide study for circulating metabolites identifies 62 loci and
6271 reveals novel systemic effects of LPA. *Nature Communications* **7**, 11122 (2016).
- 6272 319. Long, T. *et al.* Whole-genome sequencing identifies common-to-rare variants associ-
6273 ated with human blood metabolites. *Nature Genetics* **49**, 568–578 (2017).

- 6274 320. Gallois, A. *et al.* A comprehensive study of metabolite genetics reveals strong pleiotropy
6275 and heterogeneity across time and context. *Nature Communications* **10**, 4788 (2019).
- 6276 321. Lotta, L. A. *et al.* A cross-platform approach identifies genetic regulators of human
6277 metabolism and health. *Nature Genetics* **53**, 54–64 (2021).
- 6278 322. Bell, J. A. *et al.* Sex differences in systemic metabolites at four life stages: cohort study
6279 with repeated metabolomics. *BMC Medicine* **19**, 58 (2021).
- 6280 323. Carayol, M. *et al.* Reliability of Serum Metabolites over a Two-Year Period: A Targeted
6281 Metabolomic Approach in Fasting and Non-Fasting Samples from EPIC. *PLoS one* **10**,
e0135437–e0135437 (2015).
- 6282 324. Sedlmeier, A. *et al.* The human metabolic profile reflects macro- and micronutrient
6283 intake distinctly according to fasting time. *Scientific Reports* **8**, 12262 (2018).
- 6284 325. Teruya, T., Chaleckis, R., Takada, J., Yanagida, M. & Kondoh, H. Diverse metabolic
6285 reactions activated during 58-hr fasting are revealed by non-targeted metabolomic
analysis of human blood. *Scientific Reports* **9**, 854 (2019).
- 6286 326. Guasch-Ferré, M. *et al.* Metabolomics in Prediabetes and Diabetes: A Systematic
6287 Review and Meta-analysis. *Diabetes care* **39**, 833–846 (2016).
- 6288 327. Liesenfeld, D. B., Habermann, N., Owen, R. W., Scalbert, A. & Ulrich, C. M. Review of
6289 mass spectrometry-based metabolomics in cancer research. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **22**, 2182–
2201 (2013).
- 6290 328. Moore, S. C. *et al.* Human metabolic correlates of body mass index. *Metabolomics : Official journal of the Metabolomic Society* **10**, 259–269 (2014).
- 6292 329. Wishart, D. S. *et al.* HMDB 4.0: The human metabolome database for 2018. *Nucleic
6293 Acids Research* **46**, D608–D617 (2018).
- 6294 330. Sampson, J. N. *et al.* Metabolomics in epidemiology: sources of variability in metabolite
6295 measurements and implications. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **22**, 631–640 (2013).

- 6296 331. Darst, B. F., Koscik, R. L., Hogan, K. J., Johnson, S. C. & Engelman, C. D. Longitudinal
6297 plasma metabolomics of aging and sex. *Aging* **11**, 1262–1282 (2019).
- 6298 332. Rosato, A. *et al.* From correlation to causation: analysis of metabolomics data using
6299 systems biology approaches. *Metabolomics : Official journal of the Metabolomic Society* **14**, 37 (2018).
- 6300 333. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal
6301 inference in epidemiological studies. *Human molecular genetics* **23**, R89–R98 (2014).
- 6302 334. Timpson, N. J. *et al.* C-reactive protein and its role in metabolic syndrome: mendelian
6303 randomisation study. *Lancet* **366**, 1954–1959 (2005).
- 6304 335. Yarmolinsky, J. *et al.* Circulating Selenium and Prostate Cancer Risk: A Mendelian
6305 Randomization Analysis. *Journal of the National Cancer Institute* **110**, 1035–1038 (2018).
- 6306 336. Davey Smith, G. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology
6307 contribute to understanding environmental determinants of disease?*. *International
Journal of Epidemiology* **32**, 1–22 (2003).
- 6308 337. Smith, G. D. & Ebrahim, S. Mendelian randomization: prospects, potentials, and
6309 limitations. *Int J Epidemiol* **33**, 30–42 (2004).
- 6310 338. Davies, N. M., Holmes, M. V. & Davey Smith, G. Reading Mendelian randomisation
6311 studies: a guide, glossary, and checklist for clinicians. *BMJ* **362**, k601 (2018).
- 6312 339. Burgess, S., Daniel, R. M., Butterworth, A. S., Thompson, S. G. & Consortium, E.-I.
6313 Network Mendelian randomization: using genetic variants as instrumental variables to
investigate mediation in causal pathways. *International journal of epidemiology* **44**,
484–495 (2015).
- 6314 340. Bowden, J. & Holmes, M. V. Meta-analysis and Mendelian randomization: A review.
6315 *Research Synthesis Methods* **0**, (2019).
- 6316 341. Lawlor, D. *et al.* A Mendelian Randomization dictionary: Useful definitions and de-
6317 scriptions for undertaking, understanding and interpreting Mendelian Randomization
studies. (2019). doi:[10.31219/osf.io/6yzs7](https://doi.org/10.31219/osf.io/6yzs7)

- 6318 342. Pierce, B. L. & Burgess, S. Efficient Design for Mendelian Randomization Studies:
6319 Subsample and 2-Sample Instrumental Variable Estimators. *American Journal of
Epidemiology* **178**, 1177–1184 (2013).
- 6320 343. Hartwig, F. P., Davies, N. M. & Davey Smith, G. Bias in Mendelian randomization due
6321 to assortative mating. *Genetic epidemiology* **42**, 608–620 (2018).
- 6322 344. Brumpton, B. *et al.* Within-family studies for Mendelian randomization: avoiding
6323 dynastic, assortative mating, and population stratification biases. *bioRxiv* 602516
(2019). doi:[10.1101/602516](https://doi.org/10.1101/602516)
- 6324 345. Sanderson, E., Davey Smith, G., Bowden, J. & Munafò, M. R. Mendelian randomisation
6325 analysis of the effect of educational attainment and cognitive ability on smoking
behaviour. *Nature Communications* **10**, 2949 (2019).
- 6326 346. Haworth, S. *et al.* Apparent latent structure within the UK Biobank sample has
6327 implications for epidemiological analysis. *Nature Communications* **10**, 333 (2019).
- 6328 347. Berg, J. J. *et al.* Reduced signal for polygenic adaptation of height in UK Biobank.
6329 *eLife* **8**, e39725 (2019).
- 6330 348. Smith, G. D. Mendelian Randomization for Strengthening Causal Inference in Obser-
6331 vational Studies: Application to Gene \times Environment Interactions. *Perspectives on
Psychological Science* **5**, 527–545 (2010).
- 6332 349. Geiler-Samerotte, K., Sartori, F. M. O. & Siegal, M. L. Decanalizing thinking on genetic
6333 canalization. *Seminars in cell & developmental biology* **88**, 54–66 (2019).
- 6334 350. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From
6335 Polygenic to Omnipgenic. *Cell* **169**, 1177–1186 (2017).
- 6336 351. Sohail, M. *et al.* Polygenic adaptation on height is overestimated due to uncorrected
6337 stratification in genome-wide association studies. *eLife* **8**, e39702 (2019).
- 6338 352. Carter, A. R. *et al.* Mendelian randomisation for mediation analysis: current methods
6339 and challenges for implementation. *bioRxiv* 835819 (2020). doi:[10.1101/835819](https://doi.org/10.1101/835819)
- 6340 353. Relton, C. L. & Davey Smith, G. Two-step epigenetic Mendelian randomization: a
6341 strategy for establishing the causal role of epigenetic processes in pathways to disease.
Int J Epidemiol **41**, 161–176 (2012).

- 6342 354. Sanderson, E., Davey Smith, G., Windmeijer, F. & Bowden, J. An examination of
6343 multivariable Mendelian randomization in the single-sample and two-sample summary
data settings. *International Journal of Epidemiology* **48**, 713–727 (2018).
- 6344 355. VanderWeele, T. J. Mediation Analysis: A Practitioner's Guide. *Annual Review of Public
6345 Health* **37**, 17–32 (2016).
- 6346 356. Anette, V. *et al.* Remnant Cholesterol, Low-Density Lipoprotein Cholesterol, and Blood
6347 Pressure as Mediators From Obesity to Ischemic Heart Disease. *Circulation Research*
116, 665–673 (2015).
- 6348 357. Xu, L., Borges, M. C., Hemani, G. & Lawlor, D. A. The role of glycaemic and lipid
6349 risk factors in mediating the effect of BMI on coronary heart disease: a two-step,
two-sample Mendelian randomisation study. *Diabetologia* **60**, 2210–2220 (2017).
- 6350 358. Marouli, E. *et al.* Mendelian randomisation analyses find pulmonary factors mediate
6351 the effect of height on coronary artery disease. *Communications Biology* **2**, 119
(2019).
- 6352 359. Carter, A. R. *et al.* Understanding the consequences of education inequality on
6353 cardiovascular disease: mendelian randomisation study. *BMJ* **365**, l1855 (2019).
- 6354 360. Davies, N. M. *et al.* Multivariable two-sample Mendelian randomization estimates of
6355 the effects of intelligence and education on health. *eLife* **8**, e43990 (2019).
- 6356 361. Richardson, T. G. *et al.* Apolipoprotein B underlies the causal relationship of
6357 circulating blood lipids with coronary heart disease. *medRxiv* 19004895 (2019).
doi:[10.1101/19004895](https://doi.org/10.1101/19004895)
- 6358 362. Johnson, K. E. *et al.* Assessing a causal relationship between circulating lipids
6359 and breast cancer risk: Mendelian randomization study. *bioRxiv* 794594 (2019).
doi:[10.1101/794594](https://doi.org/10.1101/794594)
- 6360 363. Santos Ferreira, D. L. *et al.* Association of pre-pregnancy body mass index with
6361 offspring metabolic profile: Analyses of 3 European prospective birth cohorts. *PLOS
Medicine* **14**, e1002376 (2017).
- 6362 364. Lau, C.-H. E. *et al.* Metabolic Signatures of Gestational Weight Gain and Postpartum
6363 Weight Loss in a Lifestyle Intervention Study of Overweight and Obese Women. **10**,
(2020).

- 6364 365. Brachem, C. *et al.* Associations of BMI and Body Fat with Urine Metabolome in
6365 Adolescents Are Sex-Specific: A Cross-Sectional Study. **10**, (2020).
- 6366 366. Stevens, V. L., Carter, B. D., McCullough, M. L., Campbell, P. T. & Wang, Y. Metabolomic
6367 Profiles Associated with BMI, Waist Circumference, and Diabetes and Inflammation
Biomarkers in Women. *Obesity (Silver Spring, Md.)* **28**, 187–196 (2020).
- 6368 367. O'Keeffe, L. M. *et al.* Sex-specific associations of adiposity with cardiometabolic traits:
6369 multi-life-stage cohort study with repeat metabolomics. *medRxiv* 2020.11.30.20240895
(2020). doi:[10.1101/2020.11.30.20240895](https://doi.org/10.1101/2020.11.30.20240895)
- 6370 368. Wulaningsih, W., Proitsi, P., Wong, A., Kuh, D. & Hardy, R. Metabolomic correlates
6371 of central adiposity and earlier-life body mass index. *Journal of lipid research* **60**,
1136–1143 (2019).
- 6372 369. Frigerio, G. *et al.* Plasma Metabolomic Profiling in 1391 Subjects with Overweight and
6373 Obesity from the SPHERE Study. **11**, (2021).
- 6374 370. Neeland, I. J. *et al.* Metabolomics Profiling of Visceral Adipose Tissue: Results From
6375 MESA and the NEO Study. *Journal of the American Heart Association* **8**, e010810
(2019).
- 6376 371. Bachlechner, U. *et al.* Associations of anthropometric markers with serum metabolites
6377 using a targeted metabolomics approach: results of the EPIC-potsdam study. *Nutrition
& Diabetes* **6**, e215–e215 (2016).
- 6378 372. Rangel-Huerta, O. D., Pastor-Villaescusa, B. & Gil, A. Are we close to defining a
6379 metabolomic signature of human obesity? A systematic review of metabolomics
studies. *Metabolomics* **15**, 93 (2019).
- 6380 373. Boone, S. *et al.* Metabolomics: a search for biomarkers of visceral fat and liver fat
6381 content. *Metabolomics* **15**, 139 (2019).
- 6382 374. Ahola-Olli, A. V. *et al.* Circulating metabolites and the risk of type 2 diabetes: a
6383 prospective study of 11,896 young adults from four Finnish cohorts. *Diabetologia* **62**,
2298–2309 (2019).
- 6384 375. Cavus, E. *et al.* Association of Circulating Metabolites With Risk of Coronary Heart
6385 Disease in a European Population: Results From the Biomarkers for Cardiovascular
Risk Assessment in Europe (BiomarCaRE) Consortium. *JAMA Cardiology* **4**, 1270–
1279 (2019).

- 6386 376. Bot, M. *et al.* Metabolomics Profile in Depression: A Pooled Analysis of 230 Metabolic
6387 Markers in 5283 Cases With Depression and 10,145 Controls. *Biological psychiatry*
6387 **87**, 409–418 (2020).
- 6388 377. Onuh, J. O. & Aliani, M. Metabolomics profiling in hypertension and blood pressure
6389 regulation: a review. *Clinical hypertension* **26**, 23 (2020).
- 6390 378. Gowda, G. A. N. *et al.* Metabolomics-based methods for early disease diagnostics.
6391 *Expert review of molecular diagnostics* **8**, 617–633 (2008).
- 6392 379. Beger, R. D. A Review of Applications of Metabolomics in Cancer. **3**, (2013).
6393
- 6394 380. Verma, M. & Banerjee, H. N. Metabolomic Approaches in Cancer Epidemiology. **3**,
6395 (2015).
- 6396 381. Cruickshank-Quinn, C. I. *et al.* Metabolomics and transcriptomics pathway approach
6397 reveals outcome-specific perturbations in COPD. *Scientific Reports* **8**, 17132 (2018).
- 6398 382. Santos Ferreira, D. L. *et al.* Associations between Blood Metabolic Profile at 7 Years
6399 Old and Eating Disorders in Adolescence: Findings from the Avon Longitudinal Study
of Parents and Children. *Metabolites* **9**, 191 (2019).
- 6400 383. Hartley, A., Santos Ferreira, D. L., Anderson, E. L. & Lawlor, D. A. Metabolic profiling
6401 of adolescent non-alcoholic fatty liver disease [version 2; peer review: 2 approved].
Wellcome Open Research **3**, (2019).
- 6402 384. Chi, N.-F. *et al.* Untargeted metabolomics predicts the functional outcome of ischemic
6403 stroke. *Journal of the Formosan Medical Association = Taiwan yi zhi* **120**, 234–241
(2021).
- 6404 385. Volani, C. *et al.* Metabolic Signature of Arrhythmogenic Cardiomyopathy. **11**, (2021).
6405
- 6406 386. Ho, J. E. *et al.* Metabolomic Profiles of Body Mass Index in the Framingham Heart
6407 Study Reveal Distinct Cardiometabolic Phenotypes. *PLoS one* **11**, e0148361 (2016).
- 6408 387. Newgard, C. B. *et al.* A branched-chain amino acid-related metabolic signature that
6409 differentiates obese and lean humans and contributes to insulin resistance. *Cell
metabolism* **9**, 311–326 (2009).

- 6410 388. Cheng, S. *et al.* Metabolite Profiling Identifies Pathways Associated With Metabolic
6411 Risk in Humans. *Circulation* **125**, 2222–2231 (2012).
- 6412 389. Fall, T. *et al.* The role of adiposity in cardiometabolic traits: a Mendelian randomization
6413 analysis. *PLoS Med* **10**, e1001474 (2013).
- 6414 390. Holmes, M. V. *et al.* Causal Effects of Body Mass Index on Cardiometabolic Traits
6415 and Events: A Mendelian Randomization Analysis. *The American Journal of Human
Genetics* **94**, 198–208 (2014).
- 6416 391. Liu, J. *et al.* A Mendelian Randomization Study of Metabolite Profiles, Fasting Glucose,
6417 and Type 2 Diabetes. *Diabetes* **66**, 2915 LP–2926 (2017).
- 6418 392. Bull, C. J. *et al.* Adiposity, metabolites, and colorectal cancer risk: Mendelian random-
6419 zation study. *BMC medicine* **18**, 396 (2020).
- 6420 393. Hsu, Y.-H. H. *et al.* Integrating untargeted metabolomics, genetically informed causal
6421 inference, and pathway enrichment to define the obesity metabolome. *International
Journal of Obesity* **44**, 1596–1606 (2020).
- 6422 394. Porcu, E. *et al.* Triangulating evidence from longitudinal and Mendelian randomization
6423 studies of metabolomic biomarkers for type 2 diabetes. *Scientific Reports* **11**, 6197
(2021).
- 6424 395. Burgess, S. & Harshfield, E. Mendelian randomization to assess causal effects of
6425 blood lipids on coronary heart disease: lessons from the past and applications to the
future. *Current opinion in endocrinology, diabetes, and obesity* **23**, 124–130 (2016).
- 6426 396. Lord, J. *et al.* Mendelian randomization identifies blood metabolites previously linked
6427 to midlife cognition as causal candidates in Alzheimer’s disease. *Proceedings of the
National Academy of Sciences* **118**, e2009808118 (2021).
- 6428 397. Moayyeri, A. *et al.* Metabolomic Pathways to Osteoporosis in Middle-Aged Women: A
6429 Genome-Metabolome-Wide Mendelian Randomization Study. *Journal of Bone and
Mineral Research* **33**, 643–650 (2018).
- 6430 398. Qian, L. *et al.* Genetically Determined Levels of Serum Metabolites and Risk of
6431 Neuroticism: A Mendelian Randomization Study. *International Journal of Neuropsychopharmacology* **24**, 32–39 (2021).

- 6432 399. Qin, Y. *et al.* Genome-wide association and Mendelian randomization analysis prioritizes bioactive metabolites with putative causal effects on common diseases. *medRxiv* 6433 2020.08.01.20166413 (2020). doi:10.1101/2020.08.01.20166413
- 6434 400. Deeks, J. J., Higgins, J. P. T., Altman, D. G. & Group, on behalf of the C. S. M. Analysing 6435 data and undertaking meta-analyses. 241–284 (2019). doi:<https://doi.org/10.1002/9781119536604.ch10>
- 6436 401. Bramer, W. M., Giustini, D., Jonge, G. B. de, Holland, L. & Bekhuis, T. De-duplication 6437 of database search results for systematic reviews in EndNote. *Journal of the Medical Library Association : JMLA* **104**, 240–243 (2016).
- 6438 402. Ouzzani, M., Hammady, H., Fedorowicz, Z. & Elmagarmid, A. Rayyan-a web and 6439 mobile app for systematic reviews. *Systematic reviews* **5**, 210 (2016).
- 6440 403. Pingault, J.-B., Cecil, C. A. M., Murray, J., Munafò, M. R. & Viding, E. Causal Inference 6441 in Psychopathology: A Systematic Review of Mendelian Randomisation Studies Aiming to Identify Environmental Risk Factors for Psychopathology. *Psychopathology Review* **a4**, 4–25 (2016).
- 6442 404. Storm, C. S., Kia, D. A., Almramhi, M. & Wood, N. W. Using Mendelian randomization 6443 to understand and develop treatments for neurodegenerative disease. *Brain Communications* **2**, (2020).
- 6444 405. Kuźma, E. *et al.* Which Risk Factors Causally Influence Dementia? A Systematic 6445 Review of Mendelian Randomization Studies. *Journal of Alzheimer's disease : JAD* **64**, 181–193 (2018).
- 6446 406. Mamluk, L. *et al.* Evidence of detrimental effects of prenatal alcohol exposure 6447 on offspring birthweight and neurodevelopment from a systematic review of quasi-experimental studies. *International Journal of Epidemiology* **49**, 1972–1995 (2020).
- 6448 407. Riaz, H. *et al.* Association Between Obesity and Cardiovascular Outcomes: A Systematic 6449 Review and Meta-analysis of Mendelian Randomization Studies. *JAMA Network Open* **1**, e183788–e183788 (2018).
- 6450 408. Li, X. *et al.* Serum uric acid levels and multiple health outcomes: umbrella review 6451 of evidence from observational studies, randomised controlled trials, and Mendelian randomisation studies. *BMJ* **357**, j2376 (2017).

- 6452 409. Zhang, X. *et al.* Non-genetic biomarkers and colorectal cancer risk: Umbrella review
6453 and evidence triangulation. *Cancer medicine* **9**, 4823–4835 (2020).
- 6454 410. Markozannes, G. *et al.* Global assessment of C-reactive protein and health-related
6455 outcomes: an umbrella review of evidence from observational studies and Mendelian
randomization studies. *European Journal of Epidemiology* **36**, 11–36 (2021).
- 6456 411. Köhler, C. A. *et al.* Mapping risk factors for depression across the lifespan: An umbrella
6457 review of evidence from meta-analyses and Mendelian randomization studies. *Journal
of Psychiatric Research* **103**, 189–207 (2018).
- 6458 412. Kim, M. S., Kim, W. J., Khera, A. V. & Won, H.-H. Association between adiposity
6459 and cardiovascular outcomes: an umbrella review and meta-analysis. *medRxiv*
2020.08.18.20176578 (2020). doi:[10.1101/2020.08.18.20176578](https://doi.org/10.1101/2020.08.18.20176578)
- 6460 413. Diemer, E. W., Labrecque, J. A., Neumann, A., Tiemeier, H. & Swanson, S. A.
6461 Mendelian randomisation approaches to the study of prenatal exposures: A systematic
review. *Paediatric and Perinatal Epidemiology* **35**, 130–142 (2021).
- 6462 414. Cheng, T. S., Day, F. R., Lakshman, R. & Ong, K. K. Association of puberty timing
6463 with type 2 diabetes: A systematic review and meta-analysis. *PLOS Medicine* **17**,
e1003017 (2020).
- 6464 415. Abbasi, A. *et al.* A Systematic Review of Biomarkers and Risk of Incident Type 2
6465 Diabetes: An Overview of Epidemiological, Prediction and Aetiological Research
Literature. *PLOS ONE* **11**, e0163721 (2016).
- 6466 416. Grau-Perez, M., Agha, G., Pang, Y., Bermudez, J. D. & Tellez-Plaza, M. Mendelian
6467 Randomization and the Environmental Epigenetics of Health: a Systematic Review.
Current Environmental Health Reports **6**, 38–51 (2019).
- 6468 417. Davey Smith, G. *et al.* STROBE-MR: Guidelines for strengthening the reporting of
6469 Mendelian randomization studies. *PeerJ Preprints* **7**, e27857v1 (2019).
- 6470 418. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in
6471 two-sample Mendelian randomization. *Genet Epidemiol* **40**, 597–608 (2016).
- 6472 419. Balduzzi, S., Rücker, G. & Schwarzer, G. How to perform a meta-analysis with R: a
6473 practical tutorial. *Evidence Based Mental Health* **22**, 153 LP–160 (2019).

- 6474 420. Higgins, J. P. T., Thompson, S. G. & Spiegelhalter, D. J. A re-evaluation of random-
6475 effects meta-analysis. *Journal of the Royal Statistical Society. Series A, (Statistics in
6476 Society)* **172**, 137–159 (2009).
- 6477 421. Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. A basic introduction
6478 to fixed-effect and random-effects models for meta-analysis. *Research synthesis
6479 methods* **1**, 97–111 (2010).
- 6480 422. Hartung, J. & Knapp, G. On tests of the overall treatment effect in meta-analysis with
6481 normally distributed responses. *Statistics in Medicine* **20**, 1771–1782 (2001).
- 6482 423. Hartung, J. & Knapp, G. A refined method for the meta-analysis of controlled clinical
6483 trials with binary outcome. *Statistics in Medicine* **20**, 3875–3889 (2001).
- 6484 424. IntHout, J., Ioannidis, J. P. A. & Borm, G. F. The Hartung-Knapp-Sidik-Jonkman method
6485 for random effects meta-analysis is straightforward and considerably outperforms the
6486 standard DerSimonian-Laird method. *BMC Medical Research Methodology* **14**, 25
6487 (2014).
- 6488 425. Langan, D. *et al.* A comparison of heterogeneity variance estimators in simulated
6489 random-effects meta-analyses. *Research Synthesis Methods* **10**, 83–98 (2019).
- 6490 426. RC, P. & MANDEL, J. CONSENSUS VALUES AND WEIGHTING FACTORS. (1982).
- 6491 427. Veroniki, A. A. *et al.* Methods to estimate the between-study variance and its uncer-
6492 tainty in meta-analysis. *Research synthesis methods* **7**, 55–79 (2016).
- 6493 428. E., D. C. *et al.* Causal Associations of Adiposity and Body Fat Distribution With
6494 Coronary Heart Disease, Stroke Subtypes, and Type 2 Diabetes Mellitus. *Circulation*
6495 **135**, 2373–2388 (2017).
- 6496 429. Gao, C. *et al.* Mendelian randomization study of adiposity-related traits and risk
6497 of breast, ovarian, prostate, lung and colorectal cancer. *International Journal of
Epidemiology* **45**, 896–908 (2016).
- 6498 430. Censin, J. C. *et al.* Causal relationships between obesity and the leading causes of
6499 death in women and men. *PLOS Genetics* **15**, e1008405 (2019).
- 6500 431. Gharahkhani, P. *et al.* Effect of increased body mass index on risk of diagnosis or
6501 death from cancer. *British Journal of Cancer* **120**, 565–570 (2019).

- 6498 432. Jarvis, D. *et al.* Mendelian randomisation analysis strongly implicates adiposity with
6499 risk of developing colorectal cancer. *British Journal of Cancer* **115**, 266–272 (2016).
- 6500 433. Wang, N. *et al.* Type 2 Diabetes and Adiposity Induce Different Lipid Profile Disorders:
6501 A Mendelian Randomization Analysis. *The Journal of Clinical Endocrinology & Metabolism* **103**, 2016–2025 (2018).
- 6502 434. Kar, S. P. *et al.* The association between weight at birth and breast cancer risk revisited
6503 using Mendelian randomisation. *European journal of epidemiology* **34**, 591–600
(2019).
- 6504 435. Shu, X. *et al.* Associations of obesity and circulating insulin and glucose with breast
6505 cancer risk: a Mendelian randomization analysis. *International Journal of Epidemiology*
48, 795–806 (2019).
- 6506 436. Østergaard, S. D. *et al.* Associations between Potentially Modifiable Risk Factors and
6507 Alzheimer Disease: A Mendelian Randomization Study. *PLOS Medicine* **12**, e1001841
(2015).
- 6508 437. Brower, M. A. *et al.* Bidirectional Mendelian randomization to explore the causal
6509 relationships between body mass index and polycystic ovary syndrome. *Human
Reproduction* **34**, 127–136 (2019).
- 6510 438. Day, F. *et al.* Large-scale genome-wide meta-analysis of polycystic ovary syndrome
6511 suggests shared genetic architecture for different diagnosis criteria. *PLOS Genetics*
14, e1007813 (2018).
- 6512 439. Kivimäki, M. *et al.* Lifetime body mass index and later atherosclerosis risk in young
6513 adults: examining causal links using Mendelian randomization in the Cardiovascular
Risk in Young Finns study. *European Heart Journal* **29**, 2552–2560 (2008).
- 6514 440. Derek, K., A., E. C., Pradeep, N., F., C. M. & Sekar, K. Genetic Analysis of Venous
6515 Thromboembolism in UK Biobank Identifies the ZFPM2 Locus and Implicates Obesity
as a Causal Risk Factor. *Circulation: Cardiovascular Genetics* **10**, e001643 (2017).
- 6516 441. Larsson, S. C. *et al.* Type 2 diabetes, glucose, insulin, BMI, and ischemic stroke
6517 subtypes. *Neurology* **89**, 454 LP–460 (2017).
- 6518 442. Larsson, S. C., Burgess, S. & Michaëlsson, K. Genetic association between adiposity
6519 and gout: a Mendelian randomization study. *Rheumatology* **57**, 2145–2148 (2018).

- 6520 443. Lindström, S. *et al.* Assessing the causal relationship between obesity and venous
6521 thromboembolism through a Mendelian Randomization study. *Human genetics* **136**,
897–902 (2017).
- 6522 444. Lv, W.-Q. *et al.* Genetically driven adiposity traits increase the risk of coronary artery
6523 disease independent of blood pressure, dyslipidaemia, glycaemic traits. *European
Journal of Human Genetics* **26**, 1547–1553 (2018).
- 6524 445. Lyall, D. M. *et al.* Association of Body Mass Index With Cardiometabolic Disease in
6525 the UK Biobank: A Mendelian Randomization Study. *JAMA Cardiology* **2**, 882–889
(2017).
- 6526 446. Nordestgaard, L. T., Tybjærg-Hansen, A., Nordestgaard, B. G. & Frikke-Schmidt, R.
6527 Body Mass Index and Risk of Alzheimer's Disease: A Mendelian Randomization Study
of 399,536 Individuals. *The Journal of Clinical Endocrinology & Metabolism* **102**,
2310–2320 (2017).
- 6528 447. Painter, J. N. *et al.* Genetic Risk Score Mendelian Randomization Shows that Obesity
6529 Measured as Body Mass Index, but not Waist:Hip Ratio, Is Causal for Endometrial
Cancer. *Cancer Epidemiology Biomarkers & Prevention* **25**, 1503 LP–1510
(2016).
- 6530 448. Palmer, T. M. *et al.* Instrumental Variable Estimation of Causal Risk Ratios and Causal
6531 Odds Ratios in Mendelian Randomization Analyses. *American Journal of Epidemiology*
173, 1392–1403 (2011).
- 6532 449. Richardson, T. G., Harrison, S., Hemani, G. & Davey Smith, G. An atlas of polygenic
6533 risk score associations to highlight putative causal relationships across the human
phenome. *eLife* **8**, e43657 (2019).
- 6534 450. Shapland, C. Y., Thompson, J. R. & Sheehan, N. A. A Bayesian approach to Mendelian
6535 randomisation with dependent instruments. *Statistics in Medicine* **38**, 985–1001
(2019).
- 6536 451. Skaaby, T. *et al.* Estimating the causal effect of body mass index on hay fever, asthma
6537 and lung function using Mendelian randomization. *Allergy* **73**, 153–164 (2018).
- 6538 452. Speed, M. S., Jefsen, O. H., Børglum, A. D., Speed, D. & Østergaard, S. D. Investigating
6539 the association between body fat and depression via Mendelian randomization.
Translational Psychiatry **9**, 184 (2019).

- 6540 453. Tyrrell, J. *et al.* Using genetics to understand the causal influence of higher BMI on
6541 depression. *International Journal of Epidemiology* **48**, 834–848 (2019).
- 6542 454. Broek, N. van den *et al.* Causal associations between body mass index and mental
6543 health: a Mendelian randomisation study. *Journal of Epidemiology and Community
Health* **72**, 708 LP–710 (2018).
- 6544 455. H., W. K. *et al.* Assessing the Causal Role of Body Mass Index on Cardiovascular
6545 Health in Young Adults. *Circulation* **138**, 2187–2201 (2018).
- 6546 456. Wang, T. *et al.* Causal Association of Overall Obesity and Abdominal Obesity with
6547 Type 2 Diabetes: A Mendelian Randomization Analysis. *Obesity* **26**, 934–942 (2018).
- 6548 457. Yarmolinsky, J. *et al.* Appraising the role of previously reported risk factors in ep-
6549 ithelial ovarian cancer risk: A Mendelian randomization analysis. *PLoS medicine* **16**,
e1002893 (2019).
- 6550 458. Hippel, P. T. von. The heterogeneity statistic I^2 can be biased in small meta-analyses.
6551 *BMC Medical Research Methodology* **15**, 35 (2015).
- 6552 459. Winkler, T. W. *et al.* A joint view on genetic variants for adiposity differentiates subtypes
6553 with distinct metabolic implications. *Nature communications* **9**, 1946 (2018).
- 6554 460. Todd, J. N. *et al.* Genetic Evidence for a Causal Role of Obesity in Diabetic Kidney
6555 Disease. *Diabetes* **64**, 4238–4246 (2015).
- 6556 461. Emdin, C. A. *et al.* Genetic Association of Waist-to-Hip Ratio With Cardiometabolic
6557 Traits, Type 2 Diabetes, and Coronary Heart Disease. *JAMA* **317**, 626–634 (2017).
- 6558 462. Millard, L. A. C. *et al.* MR-PheWAS: hypothesis prioritization among potential causal ef-
6559 fects of body mass index on many outcomes, using Mendelian randomization. *Scientific
Reports* **5**, 16645 (2015).
- 6560 463. Mumby, H. S. *et al.* Mendelian Randomisation Study of Childhood BMI and Early
6561 Menarche. *Journal of obesity* **2011**, 180729 (2011).
- 6562 464. Fan, H.-Y. *et al.* Birthweight, time-varying adiposity growth and early menarche in
6563 girls: A Mendelian randomisation and mediation analysis. *Obesity research & clinical
practice* **12**, 445–451 (2018).

- 6564 465. Budu-Aggrey, A. *et al.* Evidence of a causal relationship between body mass index and
6565 psoriasis: A mendelian randomization study. *PLOS Medicine* **16**, e1002739 (2019).
- 6566 466. Jackson, D. & Turner, R. Power analysis for random-effects meta-analysis. *Research
6567 Synthesis Methods* **8**, 290–302 (2017).
- 6568 467. O'Mara, T. A. *et al.* Identification of nine new susceptibility loci for endometrial cancer.
6569 *Nature Communications* **9**, 3166 (2018).
- 6570 468. Taylor, K. *et al.* Differences in pregnancy metabolic profiles and their determinants
6571 between white European and south Asian women: Findings from the born in Bradford
Cohort. *Metabolites* **9**, (2019).
- 6572 469. Bell, J. A. *et al.* Early Metabolic Features of Genetic Liability to Type 2 Diabetes:
6573 Cohort Study With Repeated Metabolomics Across Early Life. *Diabetes Care* **43**, 1537
LP–1545 (2020).
- 6574 470. Tufte, E. R. The visual display of quantitative information. (1983). doi:[10.2307/530384](https://doi.org/10.2307/530384)
6575
- 6576 471. May-Wilson, S. *et al.* Pro-inflammatory fatty acid profile and colorectal cancer risk:
6577 A Mendelian randomisation analysis. *European journal of cancer (Oxford, England : 1990)* **84**, 228–238 (2017).
- 6578 472. Soares, A. L. G. *et al.* Sex and area differences in the association between adiposity
6579 and lipid profile in Malawi. *BMJ Global Health* **4**, e001542 (2019).
- 6580 473. Hartley, A. *et al.* Metabolomics analysis in adults with high bone mass identifies a
6581 relationship between bone resorption and circulating citrate which replicates in the
general population. *Clinical Endocrinology* **92**, 29–37 (2020).
- 6582 474. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics.
6583 *Genome research* **19**, 1639–1645 (2009).
- 6584 475. Saben, J. *et al.* A comprehensive analysis of the human placenta transcriptome.
6585 *Placenta* **35**, 125–131 (2014).
- 6586 476. Kettunen, J. *et al.* Genome-wide association study identifies multiple loci influencing
6587 human serum metabolite levels. *Nature Genetics* **44**, 269–276 (2012).

- 6588 477. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize implements and enhances
6589 circular visualization in R . *Bioinformatics* **30**, 2811–2812 (2014).
- 6590 478. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations
6591 in multidimensional genomic data. *Bioinformatics (Oxford, England)* **32**, 2847–2849
(2016).
- 6592 479. R Core Team. R: A Language and Environment for Statistical Computing. *R Foundation
6593 for Statistical Computing, Vienna, Austria* <https://www.R-project.org> (2019).
- 6594 480. Chang, W., Cheng, J., Allaire, J. J., Xie, Y. & McPherson, J. shiny: Web Application
6595 Framework for R. (2019).
- 6596 481. Bibbins-Domingo, K., Coxson, P., Pletcher, M. J., Lightwood, J. & Goldman, L. Adolescent
6597 overweight and future adult coronary heart disease. *The New England journal of
medicine* **357**, 2371–2379 (2007).
- 6598 482. Hannon, T. S., Rao, G. & Arslanian, S. A. Childhood Obesity and Type 2 Diabetes
6599 Mellitus. *Pediatrics* **116**, 473 LP–480 (2005).
- 6600 483. Baer, H. J. *et al.* Body fatness during childhood and adolescence and incidence of
6601 breast cancer in premenopausal women: a prospective cohort study. *Breast cancer
research : BCR* **7**, R314–25 (2005).
- 6602 484. Swerdlow, A. J. *et al.* Risk factors for breast cancer at young ages in twins: an
6603 international population-based study. *Journal of the National Cancer Institute* **94**,
1238–1246 (2002).
- 6604 485. Biro, F. M. & Wien, M. Childhood obesity and adult morbidities. *The American Journal
6605 of Clinical Nutrition* **91**, 1499S–1505S (2010).
- 6606 486. Fraser, A. *et al.* Cohort Profile: the Avon Longitudinal Study of Parents and Children:
6607 ALSPAC mothers cohort. *International journal of epidemiology* **42**, 97–110 (2013).
- 6608 487. Boyd, A. *et al.* Cohort Profile: the 'children of the 90s'—the index offspring of the Avon
6609 Longitudinal Study of Parents and Children. *International journal of epidemiology* **42**,
111–127 (2013).
- 6610 488. Northstone, K. *et al.* The Avon Longitudinal Study of Parents and Children (ALSPAC):
6611 an update on the enrolled sample of index children in 2019 [version 1; peer review: 2
approved]. *Wellcome Open Research* **4**, (2019).

- 6612 489. Altmaier, E. *et al.* Bioinformatics analysis of targeted metabolomics—uncovering old
6613 and new tales of diabetic mice under medication. *Endocrinology* **149**, 3478–3489
(2008).
- 6614 490. Gieger, C. *et al.* Genetics meets metabolomics: a genome-wide association study of
6615 metabolite profiles in human serum. *PLoS Genet* **4**, e1000282 (2008).
- 6616 491. Suhre, K. *et al.* Metabolic Footprint of Diabetes: A Multiplatform Metabolomics Study
6617 in an Epidemiological Setting. *PLOS ONE* **5**, e13953 (2010).
- 6618 492. Soininen, P. *et al.* High-throughput serum NMR metabonomics for cost-effective holistic
6619 studies on systemic metabolism. *Analyst* (2009). doi:[10.1039/b910205a](https://doi.org/10.1039/b910205a)
- 6620 493. Inouye, M. *et al.* Metabonomic, transcriptomic, and genomic variation of a population
6621 cohort. *Molecular Systems Biology* (2010). doi:[10.1038/msb.2010.93](https://doi.org/10.1038/msb.2010.93)
- 6622 494. Soininen, P., Kangas, A. J., Würtz, P., Suna, T. & Ala-Korpela, M. Quantitative serum nu-
6623 clear magnetic resonance metabolomics in cardiovascular epidemiology and genetics.
Circulation: Cardiovascular Genetics **8**, 192–206 (2015).
- 6624 495. Ong, K. K. *et al.* Insulin sensitivity and secretion in normal children related to size at
6625 birth, postnatal growth, and plasma insulin-like growth factor-I levels. *Diabetologia* **47**,
1064–1070 (2004).
- 6626 496. Hughes, D. A. *et al.* metaboprep: an R package for pre-analysis data description and
6627 processing. *bioRxiv* 2021.07.07.451488 (2021). doi:[10.1101/2021.07.07.451488](https://doi.org/10.1101/2021.07.07.451488)
- 6628 497. Yu, Z. *et al.* Human serum metabolic profiles are age dependent. *Aging cell* **11**,
6629 960–967 (2012).
- 6630 498. Brennan, L. & Gibbons, H. Sex matters: a focus on the impact of biological sex on
6631 metabolomic profiles and dietary interventions. *Proceedings of the Nutrition Society*
79, 205–209 (2020).
- 6632 499. Zajacova, A. & Lawrence, E. M. The Relationship Between Education and Health:
6633 Reducing Disparities Through a Contextual Approach. *Annual Review of Public Health*
39, 273–289 (2018).
- 6634 500. Bonevski, B., Regan, T., Paul, C., Baker, A. L. & Bisquera, A. Associations between
6635 alcohol, smoking, socioeconomic status and comorbidities: Evidence from the 45 and
Up Study. *Drug and Alcohol Review* **33**, 169–176 (2014).

- 6636 501. Afshin, A. *et al.* Health effects of dietary risks in 195 countries, 1990–2017: a
6637 systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **393**,
1958–1972 (2019).
- 6638 502. O'Donoghue, G. *et al.* Socio-economic determinants of physical activity across the life
6639 course: A "DEterminants of Diet and Physical ACTivity" (DEDIPAC) umbrella literature
review. *PLoS one* **13**, e0190737–e0190737 (2018).
- 6640 503. Anderson, E. L. *et al.* Estimating Trajectories of Energy Intake Through Childhood and
6641 Adolescence Using Linear-Spline Multilevel Models. *Epidemiology* **24**, (2013).
- 6642 504. Singh, A. S., Mulder, C., Twisk, J. W. R., Van Mechelen, W. & Chinapaw, M. J. M.
6643 Tracking of childhood overweight into adulthood: a systematic review of the literature.
Obesity Reviews **9**, 474–488 (2008).
- 6644 505. Buscot, M.-J. *et al.* BMI Trajectories Associated With Resolution of Elevated Youth BMI
6645 and Incident Adult Obesity. *Pediatrics* **141**, e20172003 (2018).
- 6646 506. Haufe, S. *et al.* Branched-chain and aromatic amino acids, insulin resistance and liver
6647 specific ectopic fat storage in overweight to obese subjects. *Nutrition, metabolism, and
cardiovascular diseases : NMCD* **26**, 637–642 (2016).
- 6648 507. Fattuoni, C. *et al.* Preliminary metabolomics analysis of placenta in maternal obesity.
6649 *Placenta* **61**, 89–95 (2018).
- 6650 508. Zheng, Y. *et al.* Weight-loss diets and 2-y changes in circulating amino acids in 2
6651 randomized intervention trials. *The American journal of clinical nutrition* **103**, 505–511
(2016).
- 6652 509. Yu, H.-T. *et al.* Untargeted metabolomics approach (UPLC-Q-TOF-MS) explores the
6653 biomarkers of serum and urine in overweight/obese young men. *Asia Pacific journal of
clinical nutrition* **27**, 1067–1076 (2018).
- 6654 510. Xie, G. *et al.* The metabolite profiles of the obese population are gender-dependent.
6655 *Journal of proteome research* **13**, 4062–4073 (2014).
- 6656 511. Kim, J. Y. *et al.* Metabolic profiling of plasma in overweight/obese and lean men using
6657 ultra performance liquid chromatography and Q-TOF mass spectrometry (UPLC-Q-TOF
MS). *Journal of proteome research* **9**, 4368–4375 (2010).

- 6658 512. Feingold, K. R. Introduction to Lipids and Lipoproteins. in (eds. Feingold, K. R. et al.)
6659 (2000).
- 6660 513. März, W. *et al.* HDL cholesterol: reappraisal of its clinical relevance. *Clinical research*
6661 *in cardiology : official journal of the German Cardiac Society* **106**, 663–675 (2017).
- 6662 514. Moore, S. C. *et al.* A Metabolomics Analysis of Body Mass Index and Postmenopausal
6663 Breast Cancer Risk. *Journal of the National Cancer Institute* **110**, 588–597 (2018).
- 6664 515. Perng, W. *et al.* Metabolite Profiles of the Relationship between Body Mass Index
6665 (BMI) Milestones and Metabolic Risk during Early Adolescence. *Metabolites* **10**, 316
(2020).
- 6666 516. Rauschert, S. *et al.* Sex differences in the association of phospholipids with compo-
6667 nents of the metabolic syndrome in young adults. *Biology of Sex Differences* **8**, 10
(2017).
- 6668 517. Mody, S. K. & Han, M. Obesity and contraception. *Clinical obstetrics and gynecology*
6669 **57**, 501–507 (2014).
- 6670 518. Lawlor, D. A., Tilling, K. & Davey Smith, G. Triangulation in aetiological epidemiology.
6671 *International Journal of Epidemiology* **45**, 1866–1886 (2016).
- 6672 519. Bell, J. A. *et al.* Dominant role of abdominal adiposity in circulating lipoprotein, lipid,
6673 and metabolite levels in UK Biobank: Mendelian randomization study. *medRxiv*
2021.05.29.21258044 (2021). doi:[10.1101/2021.05.29.21258044](https://doi.org/10.1101/2021.05.29.21258044)
- 6674 520. Moore, C. *et al.* The INTERVAL trial to determine whether intervals between blood
6675 donations can be safely and acceptably decreased to optimise blood supply: study
protocol for a randomised controlled trial. *Trials* **15**, 363 (2014).
- 6676 521. McCaw, Z. R., Lane, J. M., Saxena, R., Redline, S. & Lin, X. Operating characteristics of
6677 the rank-based inverse normal transformation for quantitative trait analysis in genome-
wide association studies. *Biometrics* **76**, 1262–1272 (2020).
- 6678 522. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across
6679 the human phenome. *eLife* **7**, e34408 (2018).
- 6680 523. Lawson, D. J. *et al.* Is population structure in the genetic biobank era irrelevant, a
6681 challenge, or an opportunity? *Human genetics* **139**, 23–41 (2020).

- 6682 524. Sarmanova, A., Morris, T. & Lawson, D. J. Population stratification in GWAS meta-analysis should be standardized to the best available reference datasets. *bioRxiv* 2020.09.03.281568 (2020). doi:[10.1101/2020.09.03.281568](https://doi.org/10.1101/2020.09.03.281568)
- 6683
- 6684 525. Sanderson, E., Richardson, T. G., Hemani, G. & Davey Smith, G. The use of negative control outcomes in Mendelian randomization to detect potential population stratification. *International Journal of Epidemiology* (2021). doi:[10.1093/ije/dyaa288](https://doi.org/10.1093/ije/dyaa288)
- 6685
- 6686 526. Dixon, S. C. *et al.* Adult body mass index and risk of ovarian cancer by subtype: a Mendelian randomization study. *International journal of epidemiology* **45**, 884–895 (2016).
- 6687
- 6688 527. Nead, K. T. *et al.* Evidence of a Causal Association Between Insulinemia and Endometrial Cancer: A Mendelian Randomization Analysis. *JNCI: Journal of the National Cancer Institute* **107**, (2015).
- 6689
- 6690 528. Hägg, S. *et al.* Adiposity as a cause of cardiovascular disease: a Mendelian randomization study. *International Journal of Epidemiology* **44**, 578–586 (2015).
- 6691
- 6692 529. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
- 6693
- 6694 530. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nature Genetics* **50**, 906–908 (2018).
- 6695
- 6696 531. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
- 6697
- 6698 532. Pulit, S. L., With, S. A. J. de & Bakker, P. I. W. de. Resetting the bar: Statistical significance in whole-genome sequencing-based association studies of global populations. *Genetic epidemiology* **41**, 145–151 (2017).
- 6699
- 6700 533. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature genetics* **47**, 284–290 (2015).
- 6701
- 6702 534. Bowden, J. *et al.* A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Statistics in medicine* **36**, 1783–1802 (2017).
- 6703
- 6704 535. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International journal of epidemiology* **44**, 512–525 (2015).
- 6705

- 6706 536. Burgess, S. *et al.* Dissecting Causal Pathways Using Mendelian Randomization with
6707 Summarized Genetic Data: Application to Age at Menarche and Risk of Breast Cancer.
Genetics **207**, 481–487 (2017).
- 6708 537. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data
6709 Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol*
(2017). doi:[10.1093/ije/dyx102](https://doi.org/10.1093/ije/dyx102)
- 6710 538. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between
6711 imprecisely measured traits using GWAS summary data. *PLOS Genetics* **13**, e1007081
(2017).
- 6712 539. Yaghoobkar, H. *et al.* Genetic Evidence for a Link Between Favorable Adiposity and
6713 Lower Risk of Type 2 Diabetes, Hypertension, and Heart Disease. *Diabetes* **65**,
2448–2460 (2016).
- 6714 540. Ritchie, S. A. *et al.* Reduced levels of hydroxylated, polyunsaturated ultra long-chain
6715 fatty acids in the serum of colorectal cancer patients: implications for early screening
and detection. *BMC medicine* **8**, 13 (2010).
- 6716 541. Ni, Y., Xie, G. & Jia, W. Metabonomics of human colorectal cancer: new approaches for
6717 early diagnosis and biomarker discovery. *Journal of proteome research* **13**, 3857–3870
(2014).
- 6718 542. Lin, Y. *et al.* NMR-based fecal metabolomics fingerprinting as predictors of earlier
6719 diagnosis in patients with colorectal cancer. *Oncotarget* **7**, 29454–29464 (2016).
- 6720 543. Brown, D. G. *et al.* Metabolomics and metabolic pathway networks from human
6721 colorectal cancers, adjacent mucosa, and stool. *Cancer & metabolism* **4**, 11 (2016).
- 6722 544. Bahado-Singh, R. O. *et al.* Metabolomics and first-trimester prediction of early-onset
6723 preeclampsia. *The journal of maternal-fetal & neonatal medicine : the official journal of
the European Association of Perinatal Medicine, the Federation of Asia and Oceania
Perinatal Societies, the International Society of Perinatal Obstetricians* **25**, 1840–1847
(2012).
- 6724 545. Bahado-Singh, R. O. *et al.* Metabolomic prediction of fetal congenital heart defect in
6725 the first trimester. *American journal of obstetrics and gynecology* **211**, 240.e1–240.e14
(2014).

- 6726 546. Le Gall, G. *et al.* Metabolomics of fecal extracts detects altered metabolic activity of
6727 gut microbiota in ulcerative colitis and irritable bowel syndrome. *Journal of proteome
research* **10**, 4208–4218 (2011).
- 6728 547. Hong, Y.-S. *et al.* Metabonomic understanding of probiotic effects in humans with
6729 irritable bowel syndrome. *Journal of clinical gastroenterology* **45**, 415–425 (2011).
- 6730 548. Bjerrum, J. T. *et al.* Metabonomics of human fecal extracts characterize ulcerative
6731 colitis, Crohn's disease and healthy individuals. *Metabolomics : Official journal of the
Metabolomic Society* **11**, 122–133 (2015).
- 6732 549. Ikeda, A. *et al.* Serum metabolomics as a novel diagnostic approach for gastrointestinal
6733 cancer. *Biomedical Chromatography* **26**, 548–558 (2012).
- 6734 550. Zhang, L. *et al.* Distinguishing pancreatic cancer from chronic pancreatitis and healthy
6735 individuals by ¹H nuclear magnetic resonance-based metabonomic profiles. *Clinical
Biochemistry* **45**, 1064–1069 (2012).
- 6736 551. Goedert, J. J. *et al.* Fecal metabolomics: assay performance and association with
6737 colorectal cancer. *Carcinogenesis* **35**, 2089–2096 (2014).
- 6738 552. Dossus, L. *et al.* Prospective analysis of circulating metabolites and endometrial
6739 cancer risk. *Gynecologic Oncology* **162**, 475–481 (2021).
- 6740 553. Venugopal, S. K., Anoruo, M. & Jialal, I. *Biochemistry, Low Density Lipoprotein.* in
6741 (2021).
- 6742 554. Kosmas, C. E. *et al.* High-density lipoprotein (HDL) functionality and its relevance to
6743 atherosclerotic cardiovascular disease. *Drugs in context* **7**, 212525 (2018).
- 6744 555. Ference, B. A. *et al.* Low-density lipoproteins cause atherosclerotic cardiovascular
6745 disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus
statement from the European Atherosclerosis Society Consensus Panel. *European
heart journal* **38**, 2459–2472 (2017).
- 6746 556. White, J. *et al.* Association of Lipid Fractions With Risks for Coronary Artery Disease
6747 and Diabetes. *JAMA Cardiology* **1**, 692–699 (2016).
- 6748 557. Holmes, M. V. & Smith, G. D. Dyslipidaemia: Revealing the effect of CETP inhibition in
6749 cardiovascular disease. *Nature reviews. Cardiology* **14**, 635–636 (2017).

- 6750 558. Rohatgi, A. *et al.* HDL Cholesterol Efflux Capacity and Incident Cardiovascular Events.
6751 *New England Journal of Medicine* **371**, 2383–2393 (2014).
- 6752 559. Kuusisto, S. *et al.* Direct Estimation of HDL-Mediated Cholesterol Efflux Capacity from
6753 Serum. *Clinical Chemistry* **65**, 1042–1050 (2019).
- 6754 560. Pirro, M. *et al.* High density lipoprotein cholesterol and cancer: Marker or causative?
6755 *Progress in Lipid Research* **71**, 54–69 (2018).
- 6756 561. Ala-Korpela, M., Kuusisto, S. & Holmes, M. V. Commentary: Big data bring big
6757 controversies: HDL cholesterol and mortality. *International Journal of Epidemiology*
50, 913–915 (2021).
- 6758 562. Richardson, T. G. *et al.* Evaluating the relationship between circulating lipoprotein lipids
6759 and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian
randomisation analysis. *PLOS Medicine* **17**, e1003062 (2020).
- 6760 563. Ferrari, R. *et al.* Genetic architecture of sporadic frontotemporal dementia and overlap
6761 with Alzheimer's and Parkinson's diseases. *Journal of neurology, neurosurgery, and
psychiatry* **88**, 152–164 (2017).
- 6762 564. Kliemann, N. *et al.* Metabolic signatures of greater body size and their associations with
6763 risk of colorectal and endometrial cancers in the European Prospective Investigation
into Cancer and Nutrition. *BMC Medicine* **19**, 101 (2021).
- 6764 565. Wild, C., Weiderpass, E. & Stewart, B. *World Cancer Report: Cancer Research for
6765 Cancer Prevention*. 613 (2020).
- 6766 566. Critchley, H. O. D., Maybin, J. A., Armstrong, G. M. & Williams, A. R. W. Physiology
6767 of the Endometrium and Regulation of Menstruation. *Physiological Reviews* **100**,
1149–1179 (2020).
- 6768 567. Amant, F. *et al.* Endometrial cancer. *The Lancet* **366**, 491–505 (2005).
6769
- 6770 568. Lacey, J. V. J. *et al.* Insulin-like growth factors, insulin-like growth factor-binding proteins,
and endometrial cancer in postmenopausal women: results from a U.S. case-control
study. *Cancer epidemiology, biomarkers & prevention : a publication of the American
Association for Cancer Research, cosponsored by the American Society of Preventive
Oncology* **13**, 607–612 (2004).
6771

- 6772 569. Setiawan, V. W. *et al.* Type I and II endometrial cancers: have they different risk
6773 factors? *Journal of clinical oncology : official journal of the American Society of Clinical
Oncology* **31**, 2607–2618 (2013).
- 6774 570. Gaudet, M. M. *et al.* Analysis of Serum Metabolic Profiles in Women with Endometrial
6775 Cancer and Controls in a Population-Based Case-Control Study. *The Journal of
Clinical Endocrinology & Metabolism* **97**, 3216–3223 (2012).
- 6776 571. Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-
6777 analyses. *Nature protocols* **9**, 1192–1212 (2014).
- 6778 572. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes
6779 of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine* **12**,
e1001779 (2015).
- 6780 573. Julkunen, H., Cichońska, A., Slagboom, P. E., Würtz, P. & Initiative, N. H. U. K. B.
6781 Metabolic biomarker profiling for identification of susceptibility to severe pneumonia
and COVID-19 in the general population. *eLife* **10**, e63033 (2021).
- 6782 574. Würtz, P. *et al.* Quantitative serum nuclear magnetic resonance metabolomics in large-
6783 scale epidemiology: a primer on-omic technologies. *American journal of epidemiology*
186, 1084–1096 (2017).
- 6784 575. Mitchell, R. *et al.* MRC IEU UK Biobank GWAS pipeline version 2. (2019).
6785 doi:[10.5523/bris.pnoat8cxo0u52p6ynfaekeig](https://doi.org/10.5523/bris.pnoat8cxo0u52p6ynfaekeig)
- 6786 576. Sanderson, E., Davey Smith, G., Windmeijer, F. & Bowden, J. An examination of
6787 multivariable Mendelian randomization in the single-sample and two-sample summary
data settings. *International Journal of Epidemiology* **48**, 713–727 (2019).
- 6788 577. Sanderson, E., Spiller, W. & Bowden, J. Testing and correcting for weak and pleiotropic
6789 instruments in two-sample multivariable Mendelian randomization. *Statistics in
Medicine* **n/a**, (2021).
- 6790 578. Burgess, S., Thompson, S. G. & Collaboration, C. R. P. C. H. D. G. Avoiding bias
6791 from weak instruments in Mendelian randomization studies. *International Journal of
Epidemiology* **40**, 755–764 (2011).
- 6792 579. Burgess, S. & Thompson, S. G. Multivariable Mendelian randomization: the use of
6793 pleiotropic genetic variants to estimate causal effects. *American journal of epidemiology* **181**, 251–260 (2015).

- 6794 580. Lindemann, K., Vatten, L. J., Ellstrøm-Engh, M. & Eskild, A. Serum lipids and endome-
6795 trial cancer risk: Results from the HUNT-II study. *International Journal of Cancer* **124**,
2938–2941 (2009).
- 6796 581. Ulmer, H. *et al.* Serum triglyceride concentrations and cancer risk in a large cohort
6797 study in Austria. *British Journal of Cancer* **101**, 1202–1206 (2009).
- 6798 582. Trabert, B. *et al.* Metabolic Syndrome and Risk of Endometrial Cancer in the United
6799 States: A Study in the SEER–Medicare Linked Database. *Cancer Epidemiology
Biomarkers & Prevention* **24**, 261 LP–267 (2015).
- 6800 583. Luo, Y.-Z. *et al.* Pretreatment triglycerides-to-high density lipoprotein cholesterol ratio
6801 in postmenopausal women with endometrial cancer. *The Kaohsiung Journal of Medical
Sciences* **35**, 303–309 (2019).
- 6802 584. Cust, A. E. *et al.* Metabolic syndrome, plasma lipid, lipoprotein and glucose levels, and
6803 endometrial cancer risk in the European Prospective Investigation into Cancer and
Nutrition (EPIC). *Endocrine-related cancer* **14**, 755–767 (2007).
- 6804 585. National Center for Biotechnology Information. PubChem Compound Summary for
6805 CID 11148, Trimyristin. (2021).
- 6806 586. Corbin, L. J. *et al.* Metabolic characterisation of disturbances in the APOC3/triglyceride-
6807 rich lipoprotein pathway through sample-based recall by genotype. *Metabolomics* **16**,
69 (2020).
- 6808 587. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg
6809 effect: the metabolic requirements of cell proliferation. *Science (New York, N.Y.)* **324**,
1029–1033 (2009).
- 6810 588. DeBerardinis, R. J. & Thompson, C. B. Cellular metabolism and disease: what do
6811 metabolic outliers teach us? *Cell* **148**, 1132–1144 (2012).
- 6812 589. Zhu, J. & Thompson, C. B. Metabolic regulation of cell growth and proliferation. *Nature
6813 Reviews Molecular Cell Biology* **20**, 436–450 (2019).
- 6814 590. Pavlova, N. N. & Thompson, C. B. The Emerging Hallmarks of Cancer Metabolism.
6815 *Cell metabolism* **23**, 27–47 (2016).
- 6816 591. Raffone, A. *et al.* Metabolomics in endometrial cancer diagnosis: A systematic review.
6817 *Acta Obstetricia et Gynecologica Scandinavica* **99**, 1135–1146 (2020).

- 6818 592. Lieu, E. L., Nguyen, T., Rhyne, S. & Kim, J. Amino acids in cancer. *Experimental &*
6819 *Molecular Medicine* **52**, 15–30 (2020).
- 6820 593. Green, C. R. *et al.* Branched-chain amino acid catabolism fuels adipocyte differentiation
6821 and lipogenesis. *Nature chemical biology* **12**, 15–21 (2016).
- 6822 594. Gates, E. J., Hirschfield, L., Matthews, R. P. & Yap, O. W. S. Body mass index as a
6823 prognostic factor in endometrioid adenocarcinoma of the endometrium. *Journal of the
National Medical Association* **98**, 1814–1822 (2006).
- 6824 595. Bjørge, T., Engeland, A., Tretli, S. & Weiderpass, E. Body size in relation to cancer of
6825 the uterine corpus in 1 million Norwegian women. *International Journal of Cancer* **120**,
378–383 (2007).
- 6826 596. Crosbie, E. J., Zwahlen, M., Kitchener, H. C., Egger, M. & Renehan, A. G. Body
6827 Mass Index, Hormone Replacement Therapy, and Endometrial Cancer Risk: A Meta-
Analysis. *Cancer Epidemiology Biomarkers & Prevention* **19**, 3119 LP–3130
(2010).
- 6828 597. Yang, H. P. *et al.* Endometrial cancer risk factors by 2 main histologic subtypes: the
6829 NIH-AARP Diet and Health Study. *American journal of epidemiology* **177**, 142–151
(2013).
- 6830 598. Aune, D. *et al.* Anthropometric factors and endometrial cancer risk: a systematic review
6831 and dose-response meta-analysis of prospective studies. *Annals of oncology : official
journal of the European Society for Medical Oncology* **26**, 1635–1648 (2015).
- 6832 599. Gao, Y. *et al.* Body Mass Index is Negatively Associated with Endometrial Cancer
6833 Stage, Regardless of Subtype and Menopausal Status. *Journal of Cancer* **9**, 4756–
4761 (2018).
- 6834 600. Sulc, J. *et al.* Composite trait Mendelian Randomization reveals distinct metabolic and
6835 lifestyle consequences of differences in body shape. *medRxiv* 2020.09.03.20187567
(2021). doi:[10.1101/2020.09.03.20187567](https://doi.org/10.1101/2020.09.03.20187567)
- 6836 601. Simpson, E. R. Sources of estrogen and their importance. *The Journal of Steroid
6837 Biochemistry and Molecular Biology* **86**, 225–230 (2003).
- 6838 602. Alseekh, S. *et al.* Mass spectrometry-based metabolomics: a guide for annotation,
6839 quantification and best reporting practices. *Nature Methods* **18**, 747–756 (2021).

- 6840 603. Audet-Delage, Y., Villeneuve, L., Grégoire, J., Plante, M. & Guillemette, C. Identification
6841 of Metabolomic Biomarkers for Endometrial Cancer and Its Recurrence after Surgery
in Postmenopausal Women. **9**, 87 (2018).
- 6842 604. Njoku, K. *et al.* Metabolomic Biomarkers for the Detection of Obesity-Driven Endome-
6843 trial Cancer. **13**, (2021).
- 6844 605. Cheng, Y. *et al.* Rare genetic variants affecting urine metabolite levels link population
6845 variation to inborn errors of metabolism. *Nature Communications* **12**, 964 (2021).
- 6846 606. Nounu, A. *et al.* A Combined Proteomics and Mendelian Randomization Approach
6847 to Investigate the Effects of Aspirin-Targeted Proteins on Colorectal Cancer. *Cancer
Epidemiology Biomarkers & Prevention* **30**, 564 LP–575 (2021).
- 6848 607. Safaei, M., Sundararajan, E. A., Driss, M., Boulila, W. & Shapi'i, A. A systematic
6849 literature review on obesity: Understanding the causes & consequences of obesity and
reviewing various machine learning approaches used to predict obesity. *Computers in
Biology and Medicine* **136**, 104754 (2021).
- 6850 608. Djalalinia, S., Qorbani, M., Peykari, N. & Kelishadi, R. Health impacts of Obesity.
6851 *Pakistan journal of medical sciences* **31**, 239–242 (2015).
- 6852 609. Afshin, A. *et al.* Health Effects of Overweight and Obesity in 195 Countries over 25
6853 Years. *The New England journal of medicine* **377**, 13–27 (2017).
- 6854 610. Guh, D. P. *et al.* The incidence of co-morbidities related to obesity and overweight: A
6855 systematic review and meta-analysis. *BMC Public Health* **9**, 88 (2009).
- 6856 611. Hruby, A. *et al.* Determinants and Consequences of Obesity. *American journal of
6857 public health* **106**, 1656–1662 (2016).
- 6858 612. Must, A. *et al.* The disease burden associated with overweight and obesity. *JAMA* **282**,
6859 1523–1529 (1999).
- 6860 613. Fontaine, K. R. & Barofsky, I. Obesity and health-related quality of life. *Obesity Reviews*
6861 **2**, 173–182 (2001).
- 6862 614. Wade, K. H. & Davey Smith, G. Adiposity and Cardiometabolic Outcomes: What Can
6863 Meta-analyses of Mendelian Randomization Studies Contribute? *JAMA Network Open*
1, e183778–e183778 (2018).

- 6864 615. Thelwall, M. *et al.* Is useful research data usually shared? An investigation of genome-
6865 wide association study summary statistics. *PLOS ONE* **15**, e0229578 (2020).
- 6866 616. Elsworth, B. *et al.* The MRC IEU OpenGWAS data infrastructure. *bioRxiv*
6867 2020.08.10.244293 (2020). doi:[10.1101/2020.08.10.244293](https://doi.org/10.1101/2020.08.10.244293)
- 6868 617. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide associa-
6869 tion studies, targeted arrays and summary statistics 2019. *Nucleic acids research* **47**,
D1005–D1012 (2019).
- 6870 618. Burgess, S., Scott, R. A., Timpson, N. J., Davey Smith, G. & Thompson, S. G. Using
6871 published data in Mendelian randomization: a blueprint for efficient identification of
causal risk factors. *European journal of epidemiology* **30**, 543–552 (2015).
- 6872 619. Heid, I. M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio
6873 and reveals sexual dimorphism in the genetic basis of fat distribution. *Nature genetics*
42, 949–960 (2010).
- 6874 620. Karlsson, T. *et al.* Contribution of genetics to visceral adiposity and its relation to
6875 cardiovascular and metabolic disease. *Nature medicine* **25**, 1390–1395 (2019).
- 6876 621. Temprosa, M. *et al.* COMETS Analytics: An online tool for analyzing and meta-
6877 analyzing metabolomics data in large research consortia. *American journal of epidemi-
ology* (2021). doi:[10.1093/aje/kwab120](https://doi.org/10.1093/aje/kwab120)
- 6878 622. Wishart, D. S. *et al.* HMDB: the Human Metabolome Database. *Nucleic acids research*
6879 **35**, D521–D526 (2007).
- 6880 623. Sadreev, I. I. *et al.* Navigating sample overlap, winner's curse and weak instru-
6881 ment bias in Mendelian randomization studies using the UK Biobank. *medRxiv*
2021.06.28.21259622 (2021). doi:[10.1101/2021.06.28.21259622](https://doi.org/10.1101/2021.06.28.21259622)
- 6882 624. Yengo, L., Yang, J. & Visscher, P. M. Expectation of the intercept from bivariate LD
6883 score regression in the presence of population stratification. *bioRxiv* 310565 (2018).
doi:[10.1101/310565](https://doi.org/10.1101/310565)
- 6884 625. Davies, N. M. *et al.* Within family Mendelian randomization studies. *Human molecular*
6885 *genetics* **28**, R170–R179 (2019).

- 6886 626. Dixon, P., Davey Smith, G., Hinke, S. von, Davies, N. M. & Hollingworth, W. Estimating Marginal Healthcare Costs Using Genetic Variants as Instrumental Variables: Mendelian Randomization in Economic Evaluation. *PharmacoEconomics* **34**, 1075–1086 (2016).
- 6887
- 6888 627. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* **38**, 904–909 (2006).
- 6889
- 6890 628. Silventoinen, K., Kaprio, J., Lahelma, E., Viken, R. J. & Rose, R. J. Assortative mating by body height and BMI: Finnish twins and their spouses. *American journal of human biology : the official journal of the Human Biology Council* **15**, 620–627 (2003).
- 6891
- 6892 629. Gibson, D. A., Collins, F., Cousins, F. L., Esnal Zufiaurre, A. & Saunders, P. T. K. The impact of 27-hydroxycholesterol on endometrial cancer proliferation. *Endocrine-Related Cancer* **25**, 381–391 (2018).
- 6893
- 6894 630. Elsworth, B. *et al.* MELODI: Mining Enriched Literature Objects to Derive Intermediates. *International Journal of Epidemiology* **47**, 369–379 (2018).
- 6895
- 6896 631. Stergiakouli, E. *et al.* Genome-wide association study of height-adjusted BMI in childhood identifies functional variant in ADCY3. *Obesity* n/a–n/a (2014). doi:[10.1002/oby.20840](https://doi.org/10.1002/oby.20840)
- 6897
- 6898 632. Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R. & Kievit, R. A. Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome open research* **4**, 63 (2019).
- 6899