Systematic review and meta-analyses: What has the application of Mendelian randomization told us about the causal effect of adiposity on health outcomes?

Matthew A Lee[*1,2,3], Charlie Hatcher[1,2], Luke A McGuinness[1,2,**], Nancy McBride[1,2,**], Thomas Battram[1,2,**], Wenxin Wan[1,2,**], Si Fang[1,2,**], Kaitlin H Wade[1,2], Laura J Corbin[1,2], Nicholas J Timpson[1,2]

[1] Medical Research Council Integrative Epidemiology Unit at University of Bristol, Bristol, BS8 2BN, UK

[2] Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, BS8 2BN, UK

[3] International Agency for Research on Cancer, World Health Organisation, Lyon, 69008, France

[*] Corresponding author: Matthew A Lee – leem@iarc.who.int

[**] Contributed equally

Abstract

Mendelian randomization (MR) is increasingly used for generating estimates of the causal impact of exposures on outcomes. Evidence suggests a causal role of excess adipose tissue (adiposity) on many health outcomes. However, this body of work has not been systematically appraised.

We systematically reviewed and meta-analysed results from MR studies investigating the association between adiposity and health outcomes prior to the SARS-CoV-2/COVID-19 pandemic (PROSPERO: CRD42018096684). We searched Medline, EMBASE, and bioRxiv up to February 2019 and obtained data on 2,214 MR analyses from 173 included articles. 29 meta-analyses were conducted using data from 34 articles (including 66 MR analyses) and results not able to be meta-analysed were narratively synthesised.

Body mass index (BMI) was the predominant exposure used and was primarily associated with an increase in investigated outcomes; the largest effect in the meta-analyses was observed for the association between sex-combined BMI and female-specific polycystic ovary syndrome (estimates reflect odds ratios (OR) per standard deviation change in each adiposity measure): OR = 2.55; 95% confidence interval (CI) = 1.22–5.33. Only colorectal cancer (sex-combined) was investigated with two exposures in the meta-analysis: BMI (sex-combined; OR = 1.18; 95% CI = 1.01–1.37) and waist-hip ratio (sex-combined; WHR; OR = 1.48; 95% CI = 1.08–2.03). Broadly, results were consistent across the meta-analyses and narrative synthesis.

Consistent with many observational studies, this work highlights the impact of adiposity across a broad spectrum of health outcomes, enabling targeted follow-up analyses. However, missing and incomplete data mean results should be interpreted with caution.

Keywords

Systematic review, epidemiology, Mendelian randomization, adiposity, obesity

Introduction

Observational epidemiological studies have indicated that adiposity is strongly associated with all-cause and cause-specific mortality[1,2] as well as numerous health outcomes[3]. This includes many common diseases, such as cardiovascular disease (CVD)[4] and many cancers[5], as well as commonly accepted risk factors for diseases such as high blood pressure[6]. Mendelian randomization (MR) studies can be used alongside conventional observational studies to strengthen evidence for causality within an association (or indeed provide evidence against an association)[7], and there has been a steady increase in their publication since being widely reported on in 2003[8]. There is now a large body of evidence from MR studies for a causal effect of adiposity on many outcomes, including many cancers[9,10].

Systematic reviews enable a global overview of the literature and provide avenues for hypothesis generation. In combination with meta-analyses, systematic reviews can be used as a method for improved causal inference as pooled estimates can be more precise than estimates from individual studies[11]. As the MR literature has not been systematically appraised with respect to the association between adiposity and health outcomes, we set out to systematically review MR studies investigating adiposity as an exposure and provide pooled estimates where appropriate. During the recent severe acute respiratory syndrome coronavirus 2/coronavirus disease 2019 (SARS-CoV-2/COVID-19) pandemic, there was an explosion of work focused on body mass index (BMI) related traits and outcomes/intermediates or infection impact. This is extremely important, but is complicated by both the parameterisation of infection as a target and the

nature of exhaustive genetic instruments for adiposity. This work has been brought together to recount the body of work undertaken immediately before this event and hence presents a pre-pandemic overview of the literature. Further work is now, of course, needed to distil the post-pandemic literature; however, that is not within the remit of this review.

Here, a hypothesis-free systematic review and meta-analyses are presented alongside a narrative synthesis of 173 articles reporting 2,214 MR analyses. This work was pre-published on PROSPERO (Extended Data File 1), is accompanied by Extended Data (10.5281/zenodo.7377442), and a GitHub repository (10.5281/zenodo.7377406) and data browser where all data, scripts, results, and figures are available. A narrative synthesis of non-meta-analysed studies is given in Extended Data 6.

Methods

Data sources and search strategy

EMBASE and MEDLINE were searched from inception (EMBASE = 1974; MEDLINE = 1946) until 18 February, 2019 using detailed search strategies including free text and controlled vocabulary terms, and used synonyms for both adiposity and MR terms (Extended Data File 2 and on GitHub). The pre-print service, bioRxiv, was also searched from inception (November 2013) until 18 February, 2019. Due to the limited search functionality and inability to include Boolean operators ('AND', 'OR', 'NOT') in bioRxiv searches, four free-text terms in four independent searches were used: 'Mendelian randomization', 'Mendelian randomisation', 'causal inference', and 'causal analysis'.

Study selection

Articles returned through the searches of EMBASE and MEDLINE were imported into EndNote (version X8.2; Clarivate Analytics), and de-duplication was performed using pagination identifiers[12]. Articles returned from bioRxiv were imported into Mendeley and de-duplication performed using the Mendeley de-duplication function. Titles and abstracts of all remaining articles were screened by two independent reviewers (MAL and LJM) using Rayyan[13], with discrepancies resolved through discussion. Articles that met the pre-defined inclusion criteria (see below) were combined and, in instances where the bioRxiv study had been published and this was identified in either the EMBASE or MEDLINE search, the bioRxiv version of the study was excluded. The full texts of all studies that met inclusion criteria were then screened by the two reviewers.

For title and abstract screening and for full-text screening, articles must have met the following pre-defined inclusion criteria: be written in English; be available in full text (or in the case of conference abstracts, the authors must be contactable to obtain the relevant data); be published in a peer-reviewed journal or bioRxiv; use MR methodology to investigate the causal effect of adiposity on any outcome. Adiposity was considered to be any measure which aimed to assess the amount of adipose tissue an individual possessed. If a study focused on adiposity alongside other exposures, the effect of each adiposity measure was reported separately if available. If it was not available, the joint effect was reported. Articles in which an MR approach was used but not explicitly called 'Mendelian randomization' were included. More specifically, any study in which genetic variants were used as instrumental variables (IVs) or the direct association between a genetic variant and outcome was employed was eligible (as described previously[8]), provided it met the other inclusion criteria.

Data extraction

In the first instance, data extraction was performed by eight reviewers (MAL, CH, LJM, NM, TB, WW, SF, and KHW), with articles split evenly between them, using a data extraction form (Extended data 3) designed using a pre-publication version of the STROBE-MR guidelines[14] in order to obtain all relevant data from each study. Once all articles had been reviewed, two reviewers (MAL and CH) extracted data from all articles they did not review in the first instance. The same two reviewers then checked all extracted data for

discrepancies, which were resolved through a third review of individual articles and subsequent discussion. In some cases, articles included in the data extraction contained more than one relevant MR analysis. As such, the words "study" and "studies" refer to the MR analyses within an article. The following data were extracted from each of the studies from all contributing articles: exposure(s), outcome(s), study design and sample characteristics, genetic variant IV selection, MR methodology, sensitivity analysis, and causal estimates. Where relevant data was not reported by the article, "Not discussed" was entered into the data extraction form.

Once data extraction was completed, three columns were added to summarise the type of outcome being studied: column 1 ("outcome") was used as a general categorisation of all outcomes across articles (*e.g.*, the outcome "oestrogen receptor negative (ER-) breast cancer" would have the value "breast cancer"); column 2 ("outcome info") reported the outcome-specific information that distinguished outcomes within categories defined in column 1 (*e.g.*, column 2 would contain the value "ER- breast cancer" for the same breast cancer example); and column 3 ("outcome group") categorised outcomes more generally than values defined in column 1 (*e.g.*, the breast cancer example would be categorised as "cancer"). Outcome categories were assigned based on prior biological knowledge and aimed to collapse the large number of outcomes. Where there were too few outcomes to make a category, they were grouped into an "other" category.

Quality assessment

There is currently no risk of bias tool to assess the quality of MR analyses. Here, the tool used by Mamluk et al.(2020)[15] was adapted and used for quality assessment of studies included in the meta-analyses. The quality of each study (MR analysis) within an article was assessed on a three-point scale (low = 3, medium = 2, high = 1; Extended Data 5) across 12 questions, including the five used by Mamluk et al., (2020)[15]. Additional questions which aimed to assess instrument selection, sample overlap, sensitivity analyses, descriptive data, data availability (data missingness), and statistical parameters were included based on a pre-publication version of the STROBE-MR guidelines. Quality assessment was not used as a prerequisite for inclusion or exclusion in the meta-analyses. Rather, it was used to supplement the meta-analyses and aid interpretation, with studies grouped into three rankings based on their quality assessment score: low (total score 12-19), medium (total score 20-27) or high quality (total score 28-35).

Meta-analysis

Studies were included for meta-analysis if they met a series of rules that ensured the exposure and outcome were consistent across studies. To be meta-analysed, study methods had to be compatible, for example, the same MR method(s) and units of measurement. As sample overlap can induce bias in MR studies[16], no population overlap between the different studies that provided data for an outcome being meta-analysed across multiple MR studies or between the different studies that provided the exposure and outcome data were permitted within a meta-analysis (**Error! Reference source not found.**). Where there

was sample overlap between studies, the study with the larger sample size was retained. Studies using the same population samples for the exposure data were included as the risk of bias is low[16].

In a fixed-effects meta-analysis, the assumption is that all effect estimates estimate the same effect. In MR analyses, we assume that studies using the same exposure and outcome will be estimating the same effect, but that the exposure and outcome is subtly different among different populations given instrumentation and measurement error. We therefore consider these to be related effects[17,18]. In an inverse variance weighted fixed-effects model, a weighted average is calculated as:

$$weighted\ average\ = \frac{\sum y_i(1/SE_i^2)}{\sum(1/SE_i^2)}$$

Where, y_i is the causal effect estimates in the i^{th} MR study, SE_i is the standard error of that estimate, and the summation (Σ) is across all studies. In a random-effects model, SE_i is adjusted to incorporate heterogeneity among study effects (τ^2). In this, a random-effects model will weight smaller studies more than a fixed-effects model would, as they provide more information on the distribution of effects as opposed to more information on the overall effect. This does not mean that random-effects models account for heterogeneity; random- and fixed-effects models will give identical results when there is no heterogeneity.

Following this and considerations in the Cochrane handbook, an inverse variance weighted random-effects model using estimates and standard errors was performed using the meta[19] package in R and the function metagen. Where standard errors and effect estimates were

not available for a study (*e.g.*, confidence intervals (CIs) and odds ratios were available), these were back-calculated manually. For both binary and continuous outcomes, the Hartung and Knapp method to adjust CIs to reflect uncertainty in the estimation of between-study heterogeneity[20,21], which is recommended for random-effects models[22,23], was used where ≥ 5 studies were included in the meta-analysis[22]. Between-study variance was estimated for all meta-analyses using the Paule-Mandel estimator[24], for which simulation studies have shown good performance compared to other estimators[25].

Forest plots were used to visualise results. For binary outcomes, the relevant summary method was used for odds ratios, risk ratios, hazard ratios, among others. For continuous outcomes, the mean difference was used for the underlying summary method. When presenting results, "increase" and "positive" refer to, for example, a higher BMI or an increase in the risk of type 2 diabetes; "decrease" and "negative" refer to, for example, a lower BMI or a decreased risk of type 2 diabetes.

Narrative synthesis

A narrative synthesis of all studies not included in the meta-analyses was performed in order to gain a global picture of reported causal effects. The narrative synthesis summarised the reported directions of effect estimates across outcome categories, including a summary of the evidence for selected exposures and outcomes. The outcome categories were used to guide the synthesis. Given the non-independence of studies and the focus on summarising directions of effect estimates, the synthesis should be interpreted as an overview and not

as definitive evidence for a causal effect. For a complete picture, or to look at specific exposure-outcome pairs, data extracted from all included studies are available from Extended Data 3 and can be browsed online.

Results

Literature search and data extraction

A total of 173 articles met the pre-defined inclusion criteria after full text screening (**Error! Reference source not found.**; PDFs for each article available on GitHub) – articles from bioRxiv included in data extraction were replaced with their published version if available. Of the 23 included bioRxiv articles, 18 were published once data extraction began and these published versions were included instead of the bioRxiv article. One bioRxiv article was excluded as the published version did not include the MR analysis. The remaining four bioRxiv articles were included. Most of the 173 articles were published in the past five years (**Error! Reference source not found.**). Data were extracted for 2,214 studies performed across the 173 articles (*i.e.*, many articles conducted multiple MR analyses) and one-sample MR was the predominant analysis performed (**Error! Reference source not found.**). This included 30 exposures and 659 outcomes. The majority of studies (68%) used BMI as the exposure (

Table 1). The largest proportion of outcomes were grouped into the metabolic (18%) and cancer categories (16%) (Table 2). The "other" category included 118 methylation outcomes, 68 mortality outcomes, and a handful of the following outcomes: age related macular degeneration, cataract, disease count, hernia, sleep, and physical activity.

Table 1 Number and frequency of exposures used across all 2,214 Mendelian randomization analyses

Exposure	N	%
BMI	1509	68.16
WHR adjusted for BMI	156	7.05
WHR	112	5.06

Birth weight	102	4.61
WC	50	2.26
BF	45	2.03
Fat mass	37	1.67
BMI increasing and WHR decreasing	20	0.90
BMI increasing and WHR increasing	20	0.90
Fat free mass	15	0.68
Obesity	15	0.68
WC adjusted for BMI	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
HC adjusted for BMI	9	0.41
Favourable adiposity	7	0.32
Overweight	7	0.32
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; WC = waist circumference; HC = hip circumference; BF = body fat percentage.

230

231

232

233 *Table 2 Number and frequency of outcomes within each outcome category across all 2,214 Mendelian randomization*
234 *analyses*

Outcome 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

235

236 ***Meta-analysis and quality assessment***

237 In total, 66 studies from 34 articles were included in 29 meta-analyses – studies
238 investigating the effect of adjusted variables (*i.e.*, WHRadjBMI) in two-sample settings
239 were excluded given recent evidence of biased estimates when using adjusted traits in MR
240 studies[26]. Most of the 2,214 studies were excluded due to a lack of meta-analysable data
241 (*e.g.*, only one MR analysis looked at a given exposure-outcome pair). The a15verage
242 quality assessment score across the 66 studies was 24 (standard deviation (SD) = 2.8; **Error!**
243 **Reference source not found.**). Only the study of the association between BMI and
244 haemorrhagic stroke by Dale *et al.*, (2017)[27] was ranked as high quality. All low scoring
245 studies showed consistent directions of effect with the other studies with which they were

meta-analysed. Quality assessment scores for each study are presented alongside the meta-analysis results (**Error! Reference source not found.** and **Error! Reference source not found.**). The majority of studies included in the meta-analyses used sex-combined data for the exposure and outcome. As such, we consider meta-analysis results to be the sex-combined effect of the exposure on the outcome. The exception is for the sex-specific outcomes endometrial, ovarian, and prostate cancer and polycystic ovary syndrome which used sex-specific outcome data. For these four outcomes only, we consider the effect on the outcome to be sex-specific.

All results are given per SD unit increase. For all binary outcomes, results are given as an odds ratio (OR) and reflect the OR of the outcome per SD unit increase in the exposure. For continuous outcomes, results are given as the mean difference (MD) and reflect an average unit change in the outcome per SD unit increase in the exposure. The term "effect estimate" is used throughout.

Of the 20 binary (**Error! Reference source not found.**) and 9 continuous (**Error! Reference source not found.**) outcomes, 5 meta-analyses had negative effect estimates: birthweight on ER-breast cancer and colon cancer, and BMI on high-density lipoprotein cholesterol (HDL-C; analysed with SD and mmol/L units) and low-density lipoprotein cholesterol (LDL-C; mmol/L). 14 of the remaining tests had positive effect estimates with CIs that did not span the null. The remaining 10 tests had positive effect estimates with CIs that spanned the null. There was little difference between effect estimates from studies contributing to

individual meta-analyses that had a low-quality assessment score and studies with a medium or high-quality assessment score. One outcome was investigated using more than one exposure, colorectal cancer with BMI and WHR. There was evidence for an increasing effect of both measures on colorectal cancer: WHR (OR = 1.48; 95% CI = 1.08–2.03); BMI (OR = 1.18; 95% CI = 1.01–1.37.

BMI was the predominant exposure and was found to be associated with an increase in the risk of all cancers tested (colorectal, endometrial, lung, ovarian, and prostate), CIs crossed the null only for prostate cancer (OR = 1.08; 95% CI = 0.91–1.28). There was weak evidence for an association between BMI and ischemic and haemorrhagic stroke, hypertension, arthritis, and Alzheimer's disease, with effect estimates close to the null and CIs spanning the null.

There was evidence of heterogeneity within the included studies, 8 of 20 binary outcomes and 5 of 9 continuous outcomes had heterogeneity statistics with p-values ≤ 0.05. However, given no meta-analysis met the requirements for heterogeneity statistics (≥ 5 studies)[28] these results should be interpreted with caution.

Narrative synthesis

A total of 2,144 studies were not included in the meta-analyses. A complete summary for each outcome category is available as Extended Data 6. All extracted data are available from Extended Data 3 and can be browsed online. Briefly, of the 2,144 studies, 1,343 reported a

positive direction of effect and 597 reported a negative direction of effect. The remaining 204 studies either did not report an effect estimate or the effect estimate was null. The largest number of studies and articles investigated the association between adiposity and metabolic or cancer outcomes which are summarised here. In this synthesis we discuss directions of effect across all studies and do not account for sex in this regard.

For the metabolic category, 380 studies were reported across 51 articles. 89 studies reported a positive effect estimate and 266 studies reported a negative effect estimate, the remaining studies did not report an effect estimate. For example, there was weak evidence for an increasing effect of BMI on cholesterol, but strong evidence for an increasing effect of WHRadjBMI on cholesterol. Evidence was strongest for outcomes analysed by multiple studies and articles. For example, there was strong evidence for an increasing effect of BMI, birth weight, childhood BMI, WHR, WHRadjBMI, and WC on diabetes (type 1, type 2, and all).

For the cancer category, 332 studies were reported across 39 articles. Overall, 189 studies reported a positive effect estimate and 137 studies reported a negative effect estimate; the remaining studies reported an effect estimate equal to the null: most studies reported CIs which spanned the null. A total of 31 cancer outcomes were investigated across the 332 studies, with breast cancer the most common, followed by lung, ovarian, and colorectal cancers. Negative effect estimates were found for cervical (with BMI and WHRadjBMI), clear cell (with BMI), and gastric (with BMI) cancers. Positive effect estimates were found

for Barrett's esophagus (with BMI), colon (with BMI), esophageal (with BMI), lymphoid (with BMI), meningioma (with BMI, WC, and BF), rectal (with BMI), renal (with BMI, WHR, and BF), skin (including melanoma; with BMI), stomach and esophageal (with BMI), and low malignant potential tumours (with BMI). Positive and negative effect estimates were found for the remaining cancer outcomes, including breast, colorectal, endometrial, glioma, kidney, lung, multiple myeloma, ovarian, pancreatic, prostate, testicular, and upper aerodigestive cancers. Broadly, results suggest adiposity increases overall cancer risk and risk of mortality. However, this risk is modulated by cancer type and subtype.

Discussion

Here, 173 articles and 2,214 MR analyses were reviewed. Meta-analyses and a narrative synthesis of these studies provide an overview of the causal landscape of adiposity. Broadly, evidence points to an increasing effect of adiposity on a wide array of outcomes, including many cancers as well as cardiovascular traits, and type-2 diabetes. It was not possible to summarise the effect of adiposity on each outcome in the narrative synthesis. Instead, extracted data from all 2,214 studies are available as Extended Data 5 and via a data browser. Broadly, results from the meta-analyses were consistent with the narrative synthesis. However, there was variability within outcomes.

There were some inconsistencies between evidence from the meta-analyses and narrative synthesis. For example, there was evidence for an increasing effect of adiposity on endometrial and colorectal cancer in the meta-analysis, but within the narrative synthesis, there were studies that reported evidence of an increasing, protective, and null effect of adiposity on both cancers. This is expected to some degree since in meta-analyses the sample size is considered, and studies are weighted by this. In contrast, in the narrative synthesis, only the direction of effect was used to summarise the effect of adiposity. Additionally, studies included in the meta-analyses were non-overlapping, whereas the narrative synthesis will have included numerous studies of the same exposure-outcome pair with overlapping samples. As a result, effects from the same population are likely repeated in the narrative synthesis, which may have biased the summation of the overall effect of adiposity.

342

343 Many of the consistent effects observed across the meta-analyses and narrative synthesis
344 are supported by observational studies, including increased risk of CVD[4] and hypertension[6].
345 However, there are some inconsistencies with the observational literature, notably for the
346 effect of adiposity on haemorrhagic stroke, where evidence for an effect of adiposity was
347 weak in meta-analysis but is strong in observational analyses[29]. There was also evidence in
348 the narrative synthesis for an effect of adiposity on a broad number of metabolites which
349 is also found in the observational literature[6,30]. However, there was weak evidence in the
350 meta-analyses for a decreasing effect of BMI on HDL-C and an increasing effect on LDL-C
351 (*e.g.*, the estimate with SD units was positive and had less heterogeneity across the studies
352 meta-analysed), which is repeatedly found in observational studies[6,30].

353

354 A particular consideration from this work is the shallowness of the identified exposure-
355 outcome pairs. That is, many outcomes have been assessed, but these have predominantly
356 been assessed with BMI as the exposure. Although there is some replication of the results
357 of the association between BMI and various outcomes, they are concentrated on more
358 heavily studied diseases such as cancer and CVD. An additional component of this
359 observation is the use of meta-analyses and biobanks, whereby the same exposure-outcome
360 association has been assessed using ever larger samples, which include the same
361 populations. This poses a potential problem for future work, whereby large studies using
362 meta-GWAS and biobanks, due to their size, are able to capture population structure[31]. If
363 not controlled within GWASs and MR analyses, this population structure may bias MR

analyses and meta-analyses of MR results due to the introduction of genetic confounding and violation of the second MR assumption.

Data extraction was based on the STROBE-MR guidelines, which includes information on interpretability and reproducibility. It was not possible to extract all data from many of the 2,214 studies included in the review. Although some of this data related to reproducibility guidelines (*e.g.*, software used) a large proportion was related to interpretability (*e.g.*, SNPs used). This also included data on sex, which was routinely missing or difficult to extract from both the MR studies and original GWAS publications from which the MR studies obtained exposure and/or outcome data. This limited the scope of the narrative synthesis to an overall summary of the direction of effect estimates and did not allow for sex-specific summaries. As the STROBE-MR guidelines have now been published[14], it is expected that the reporting quality of studies will improve. The omission of methodological detail is unlikely to affect the results of an analysis but does impact on reproducibility and the reuse of results in meta-analyses such as those presented here.

Most studies employed similar instrumentation approaches, using a p-value threshold of 5×10^{-8} and a linkage disequilibrium R2 threshold of 0.0001 (the default for the TwoSampleMR R package) to identify independent instruments. This has the advantage that many studies will likely have used the same SNPs for the same exposure. Similarly, most studies used the same methodologies; however, there was little investigation of non-linear effects.

386

387 *Strengths and limitations*

388 The majority of the 29 meta-analyses included just two MR analyses; this was primarily a
389 result of overlapping outcome samples across studies which would ultimately bias results
390 towards the confounded observational estimate[16]. This overlap suggests replication within
391 the literature but also the use of meta-GWAS to obtain ever larger populations for MR
392 analyses. The limited number of analyses included in each meta-analysis (*i.e.*, < 5 studies)
393 prevents meaningful interpretation of heterogeneity statistics[28] and prevented the
394 assessment of publication bias.

395

396 Given the incomplete and often poor reporting of MR analyses, results here should be
397 interpreted cautiously. Studies were excluded from meta-analysis if there was overlap
398 between the outcome data between studies or between the exposure data and outcome data
399 between studies. However, it is possible that this was not completely accurate given that
400 not all studies reported the cohorts used in their analyses. Additional limitations of MR
401 analyses, including homogeneity and monotonicity, may be especially important in meta-
402 analysis results given effects among different populations may not be homogeneous (*i.e.*,
403 the effect of the IV or exposure is not the same for all populations) or monotonic (*i.e.*, the
404 effect of the IV on the exposure is differential among populations).

405

406 *Conclusions*

407 Adiposity is shown to exert a predominantly increasing effect on numerous outcomes
408 including many cancers, cardiovascular outcomes, and metabolic traits. Results here are
409 broadly consistent with the observational literature and provide corroborative evidence for
410 associations with several traits. However, these results are not definitive and should instead
411 be used as a guide for future investigations aiming to triangulate evidence of association[7].
412 There is a need to update this work, especially considering the large body of work
413 conducted during the SARS-CoV-2/COVID-19 pandemic, and it is hoped this will become
414 easier as the quality of studies improves with the adoption of the STROBE-MR guidelines.
415

Data availability

Underlying data

All data, scripts, results, and figures are available on GitHub (10.5281/zenodo.7377406). All data obtained from the data extraction process can be accessed via Extended Data 3 and can be searchable online.

Extended data

All Extended Data, including the preregistration document and PRISMA checklists, are available from Zenodo: 10.5281/zenodo.7377442. Extended data includes:

1. PROSPERO preregistration document
2. Search strategy
3. Data extraction manual, data extraction form with raw data, and formatted extracted data
4. Formatted results from meta-analyses
5. Quality assessment tool and results
6. Narrative synthesis of all non-meta-analysed studies
7. PRISMA checklists
8. Letter from editor of IJE and response to reviewer comments
9. PRISMA flowchart

Author contributions

MAL: conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, supervision, validation, visualization, writing (original draft preparation), writing (review & editing)

CH: data curation, investigation, validation, writing (review & editing)

LJM: data curation, methodology, writing (review & editing)

NM: data curation, writing (review & editing)

TB: data curation, writing (review & editing)

WW: data curation, writing (review & editing)

SF: data curation, writing (review & editing)

KHW: resources, data curation, supervision, writing (review & editing)

LJC: resources, supervision, writing (review & editing)

NJT: resources, funding acquisition, supervision, writing (review & editing)

463 Competing interests

464 No competing interests were disclosed

465

466 Funding

467 This work was supported by the Wellcome Trust through a Wellcome Trust Investigator
468 award to NJT (202802/Z/16/Z). MAL is funded by a Medical Research Council GW4
469 studentship (grant number: MR/R502340/1). NJT is a Wellcome Trust Investigator
470 (202802/Z/16/Z), is the PI of the Avon Longitudinal Study of Parents and Children (MRC
471 & WT 217065/Z/19/Z), is supported by the University of Bristol NIHR Biomedical
472 Research Centre (BRC-1215-2001), the MRC Integrative Epidemiology Unit
473 (MC_UU_00011/1) and works within the CRUK Integrative Cancer Epidemiology
474 Programme (C18281/A29019). L.J.C. is supported by N.J.T.'s Wellcome Investigator
475 Award (202802/Z/16/Z). All authors work in the MRC Integrative Epidemiology Unit at
476 the University of Bristol, which is supported by the Medical Research Council (grant
477 numbers: MC_UU_00011/1-7) and the University of Bristol. LAM is supported by an
478 NIHR Doctoral Research Fellowship (DRF-2018-11-ST2-048). The views expressed in this
479 article are those of the authors and do not necessarily represent those of the NHS, the
480 NIHR, or the Department of Health and Social Care.

481

482 This research was funded in whole, or in part, by the Wellcome Trust [202802/Z/16/Z,
483 217065/Z/19/Z]. For the purpose of Open Access, the author has applied a CC BY public
484 copyright licence to any Author Accepted Manuscript version arising from this submission.

Acknowledgements

This work is based on Chapter 2 of the PhD thesis of the first author Matthew A Lee, available from the University of Bristol.

References

1. Flegal, K. M., Kit, B. K., Orpana, H. & Graubard, B. I. Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. *JAMA* **309**, 71–82 (2013).
2. Rost, S. *et al.et al.* New indexes of body fat distribution and sex-specific risk of total and cause-specific mortality: a prospective cohort study. *BMC Public Health* **18**, 427 (2018).
3. Cornier, M.-A. *et al.et al.* Assessing Adiposity. *Circulation* **124**, 1996–2019 (2011).
4. Koliaki, C., Liatis, S. & Kokkinos, A. Obesity and cardiovascular disease: revisiting an old relationship. *Metabolism* **92**, 98–107 (2019).
5. Bhaskaran, K. *et al.et al.* Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults. *Lancet Lond. Engl.* **384**, 755–765 (2014).
6. Würtz, P. *et al.et al.* Metabolic Signatures of Adiposity in Young Adults: Mendelian Randomization Analysis and Effects of Weight Change. *PLOS Med.* **11**, e1001765 (2014).
7. Lawlor, D. A., Tilling, K. & Davey Smith, G. Triangulation in aetiological epidemiology. *Int. J. Epidemiol.* **45**, 1866–1886 (2016).
8. Davey Smith, G. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*. *Int. J. Epidemiol.* **32**, 1–22 (2003).
9. Lauby-Secretan, B. *et al.et al.* Body Fatness and Cancer — Viewpoint of the IARC Working Group. *N. Engl. J. Med.* **375**, 794–798 (2016).

10. Brennan, P. *et al.et al.* Obesity and cancer: Mendelian randomization approach utilizing the FTO genotype. *Int. J. Epidemiol.* **38**, 971–975 (2009).
11. Deeks, J. J., Higgins, J. P. T., Altman, D. G. & Group, on behalf of the C. S. M. Analysing data and undertaking meta-analyses. *Cochrane Handbook for Systematic Reviews of Interventions* 241–284 Preprint at https://doi.org/10.1002/9781119536604.ch10 (2019).
12. Bramer, W. M., Giustini, D., de Jonge, G. B., Holland, L. & Bekhuis, T. De-duplication of database search results for systematic reviews in EndNote. *J. Med. Libr. Assoc. JMLA* **104**, 240–243 (2016).
13. Ouzzani, M., Hammady, H., Fedorowicz, Z. & Elmagarmid, A. Rayyan-a web and mobile app for systematic reviews. *Syst. Rev.* **5**, 210 (2016).
14. Skrivankova, V. W. *et al.et al.* Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ* **375**, n2233 (2021).
15. Mamluk, L. *et al.et al.* Evidence of detrimental effects of prenatal alcohol exposure on offspring birthweight and neurodevelopment from a systematic review of quasi-experimental studies. *Int. J. Epidemiol.* **49**, 1972–1995 (2020).
16. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol* **40**, 597–608 (2016).
17. Higgins, J. P. T., Thompson, S. G. & Spiegelhalter, D. J. A re-evaluation of random-effects meta-analysis. *J. R. Stat. Soc. Ser. A Stat. Soc.* **172**, 137–159 (2009).

18. Borenstein, M., Hedges, L. V, Higgins, J. P. T. & Rothstein, H. R. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res. Synth. Methods* **1**, 97–111 (2010).

19. Balduzzi, S., Rücker, G. & Schwarzer, G. How to perform a meta-analysis with R: a practical tutorial. *Evid. Based Ment. Health* **22**, 153 LP – 160 (2019).

20. Hartung, J. & Knapp, G. On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Stat. Med.* **20**, 1771–1782 (2001).

21. Hartung, J. & Knapp, G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Stat. Med.* **20**, 3875–3889 (2001).

22. IntHout, J., Ioannidis, J. P. A. & Borm, G. F. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med. Res. Methodol.* **14**, 25 (2014).

23. Langan, D. *et al.et al.* A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Res. Synth. Methods* **10**, 83–98 (2019).

24. RC, P. & MANDEL, J. CONSENSUS VALUES AND WEIGHTING FACTORS. (1982).

25. Veroniki, A. A. *et al.et al.* Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res. Synth. Methods* **7**, 55–79 (2016).

26. Hartwig, F. P., Tilling, K., Davey Smith, G., Lawlor, D. A. & Borges, M. C. Bias in two-sample Mendelian randomization when using heritable covariable-adjusted summary associations. *Int. J. Epidemiol.* **50**, 1639–1650 (2021).

552 27. E., D. C. *et al.et al.* Causal Associations of Adiposity and Body Fat Distribution With
553 Coronary Heart Disease, Stroke Subtypes, and Type 2 Diabetes Mellitus. *Circulation*
554 **135**, 2373–2388 (2017).
555 28. von Hippel, P. T. The heterogeneity statistic I2 can be biased in small meta-analyses.
556 *BMC Med. Res. Methodol.* **15**, 35 (2015).
557 29. N., K. W., E., I. S., Carla, S., F., M. R. & L., F. K. Obesity. *Stroke* **44**, 278–286 (2013).
558 30. Cirulli, E. T. *et al.et al.* Profound Perturbation of the Metabolome in Obesity Is
559 Associated with Health Risk. *Cell Metab* **29**, 488-500 e2 (2019).
560 31. Tan, V. Y. & Timpson, N. J. The UK Biobank: A Shining Example of Genome-Wide
561 Association Study Science with the Power to Detect the Murky Complications of Real-
562 World Epidemiology. *Annu. Rev. Genomics Hum. Genet.* (2022) doi:10.1146/annurev-
563 genom-121321-093606.
564
565

Figures and tables

*Figure 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 with regards to sample overlap. * = MR analyses had to use the same exposure and the same outcome to be compatible, e.g., for the exposure, body mass index (BMI) could not be meta-analysed with any other exposure that was not BMI. This also applies to outcomes, e.g., the outcome oestrogen receptor negative (ER-) breast cancer could not be meta-analysed with breast cancer, it could only be meta-analysed with ER- breast cancer.*

Figure 2 PRISMA flowchart. N gives the number of articles at each stage. MR = Mendelian randomization.

Figure 3 Distribution of publication year and average exposure and outcome sample sizes across included studies up to the search date of 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 for binary outcomes.

Figure 4 Distribution of study design across 173 included articles. The Y axis gives the MR study design and the X axis gives the number of studies for that study design. 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 MR meta-analysis indicates that the meta-analysis included MR analyses that were both one- and two-sample designs. Generalized summary data-based MR allows for, and models, correlated SNPs within the instrument. Factorial MR is analogous to a factorial randomized controlled trial, whereby individuals are grouped using genetic scores (generally in a 2 x 2 approach). An MR-PheWAS is the investigation of a single trait on many, potentially hundreds, of outcomes. Direct G-O refers to an MR analysis which used instruments from a single locus, e.g., the FTO locus.

Figure 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.

Figure 6 Meta-analysis: effect estimates and 95% confidence intervals for binary outcomes. Forest plot shows effect estimates and 95% confidence intervals (CIs) from a meta-analysis of 22 different exposure-outcome pairs. Mendelian randomization analyses included based on criteria in ***Error! Reference source not found.****. P-values are given for the heterogeneity statistics. QA = quality assessment score; OR = odds ratio. Available on GitHub. Forest plots of individual meta-analyses are also available on GitHub.*

Figure 7 Meta-analysis: effect estimates and 95% confidence intervals for continuous outcomes. Forest plot shows effect estimates and 95% confidence intervals (CIs) from a meta-analysis of 9 different exposure-outcome pairs. Mendelian randomization analyses included based on criteria in ***Error! Reference source not found.****. P-values are given for the heterogeneity statistics. QA = quality assessment score; OR = odds ratio. Available on GitHub. Forest plots of individual meta-analyses are also available on GitHub.*

Table 3 Number and frequency of exposures used across all 2,214 Mendelian randomization analyses

Table 4 Number and frequency of outcomes within each outcome category across all 2,214 Mendelian randomization analyses