A data-driven approach for studying the role of body mass in multiple diseases: a phenome-wide registry-based case-control study in the UK Biobank

A data-driven approach for studying the role of body mass in multiple diseases: a phenome-wide registry-based case-control study in the UK Biobank

Elina Hyppönen, Anwar Mulugeta, Ang Zhou, Vimaleswaran Karani Santhanakrishnan

Summary

Background Mendelian randomisation allows for the testing of causal effects in situations where clinical trials are challenging to do. In this hypothesis-free, data-driven phenome-wide association study (PheWAS), we sought to assess possible associations of high body-mass index (BMI) with multiple disease outcomes.

Methods For this registry-based case-control PheWAS, we used genome-wide data available from the UK Biobank to construct a genetic risk score of 76 variants related to BMI. Eligible UK Biobank participants were aged 37–73 years during recruitment, were white British, were unrelated to each other, and had available genetic information. Disease outcomes from these participants were mapped to a phenotype code (phecode). Participants with a phecode of interest were recoded as cases, whereas participants without a phecode of interest or any codes under a parent phecode were classified as controls. We did a PheWAS to analyse possible associations between the BMI genetic risk score and a range of disease outcomes. Disease associations passing stringent correction for multiple testing (Bonferroni corrected threshold $p<5{\cdot}4\times10^{-5}$, false discovery rate corrected $p<0{\cdot}0074$) were assessed for causal association with use of inverse-variance weighted mendelian randomisation. We did sensitivity analyses to assess pleiotropy and stability of estimation with use of weighted median, weighted mode, Egger regression, and mendelian randomisation pleiotropy residual sum and outlier methods.

Findings Our study population comprised 337 536 UK Biobank participants, and analyses were done for 925 unique phecodes from 17 different disease categories. After Bonferroni correction, PheWAS identified that BMI genetic risk score was associated with hospital-diagnosed obesity and 58 other outcomes; 30 distinct disease associations were supported by the mendelian randomisation analyses. 30 distinct disease associations were supported by the mendelian randomisation analyses. In inverse-variance weighted mendelian randomisation, genetically determined BMI was associated with endocrine disorders (odds ratio per one SD or 4·1 kg/m² higher BMI 2·72, 95% CI 2·33–3·29 for type 2 diabetes; 2·11, 1·62–2·76 for type 1 diabetes; and 1·46, 1·25–1·70 for hypothyroidism), circulatory diseases (1·96, 1·53–2·51 for phlebitis and thrombophlebitis; 1·89, 1·39–2·57 for cardiomegaly; 1·68, 1·35–2·09 for congestive heart failure; 1·55, 1·37–1·76 for hypertension; 1·31, 1·13–1·52 for ischaemic heart disease; and 1·25, 1·14–1·37 for cardiac dysrhythmias), and inflammatory or dermatological conditions (2·00, 1·72–2·23 for superficial cellulitis and abscess; 3·37, 2·17–5·25 for chronic ulcers of leg and foot; 4·99, 2·54–9·82 for gangrene; and 2·24, 1·53–3·28 for atopy). Mendelian randomisation analyses provided further support for a causal effect of BMI on renal failure, osteoarthrosis, neurological (insomnia and peripheral nerve disorders) and respiratory diseases (asthma and chronic bronchitis), structural problems (hernias and knee derangement), and chemotherapy treatment. Mendelian randomisation with Egger regression produced consistently wider CIs compared with those of other methods. 26 of 72 distinct diseases detected under false discovery rate correction produced consistent estimates across at least four mendelian randomisation methods, and consistent evidence across all five approaches was obtained for 14 diseases.

Interpretation Our data-driven approach identified a range of diseases as possibly affected by high BMI. This population-level screening approximated the accumulated consequences of high BMI, whereas the true effects might be more complex and vary by life stage. Our results highlight the importance of obesity prevention and effective management of obesity-related comorbidities.

Funding National Health and Medical Research Council of Australia.

Lancet Digital Health 2019; 1: e116–26

See **Comment** page e101

Australian Centre for Precision Health, University of South Australia Cancer Research Institute, Adelaide, SA, Australia (Prof E Hyppönen PhD, A Mulugeta MSc, A Zhou PhD); **South Australian Health and Medical Research Institute, Adelaide, SA, Australia** (Prof E Hyppönen); **Department of Pharmacology, School of Medicine, College of Health Science, Addis Ababa University, Addis Ababa, Ethiopia** (A Mulugeta); **and Hugh Sinclair Unit of Human Nutrition, University of Reading, Reading, UK** (V K Santhanakrishnan PhD)

Correspondence to:
Prof Elina Hyppönen, Australian Centre for Precision Health, University of South Australia Cancer Research Institute, Adelaide, SA 5001, Australia.
elina.hypponen@unisa.edu.au

Introduction

The global obesity epidemic is a major public health concern. Each year, obesity leads to 2·8 million preventable deaths, making it the second most important modifiable risk factor after smoking.[1] Obesity-related comorbidities lead to a notable loss in disease-free years of life,[2] incurring a major economic burden to health services. A global, rising trend is evident in health-care costs related to obesity and, in the USA alone, the annual costs have been estimated to exceed US$342 billion.[3]

Research in context

Evidence before this study

We searched the MEDLINE database using a combination of search terms that included alternative spellings and synonyms: "body mass index OR BMI OR obes*" AND "phenome-wide OR phenomewide OR phewas OR Mendelian randomization OR Mendelian randomisation". We included studies published up to April 11, 2019, and the review was restricted to studies with information on disease outcomes and published in English. The search identified 343 papers, 95 of which were considered relevant on the basis of the title and were reviewed for further information. We identified 80 mendelian randomisation studies. We obtained convincing evidence of a probable causal effect of high BMI for type 2 diabetes and cardiovascular diseases, which were the most commonly studied outcomes. Evidence supported an association between high BMI and a lower risk of breast cancer (but reduced survival), whereas positive associations were suggested for lung cancer (squamous cell and small cell cancer, but not adenocarcinoma), meningioma, and colorectal, oesophageal, renal, pancreatic, ovarian, and endometrial cancer; no evidence of association was seen for prostate cancer, glioma, or multiple myeloma. Mendelian randomisation studies also supported adverse effects on asthma, psoriasis, osteoarthritis, depression, gout, multiple sclerosis, and type 1 diabetes and a possible protection in patients with Parkinson's disease, but no effect on Alzheimer's disease. To date, three phenome-wide association studies (PheWAS) have been published, one in children and two in adults. The only PheWAS focusing on disease outcomes that we found used *FTO* variants (n=24 198) and had evidence supporting the adverse associations of high BMI with type 2 diabetes, sleep apnoea, and non-alcoholic fatty liver disease, whereas an inverse association was seen of high BMI with fibrocystic breast disease. A PheWAS published in 2019, using PHESANT software, tested for associations of high BMI with all continuous, categorical, and binary outcomes in the UK Biobank, proposing adverse effects on type 2 diabetes, cardiovascular diseases, and multiple other diseases and traits, and suggesting protection from various psychosocial traits.

Added value of this study

Our data-driven analyses provided some evidence for a possible causal effect of high BMI in 30 distinct diseases under Bonferroni correction (72 under false discovery rate correction) including circulatory, endocrine and metabolic, digestive, neurological, dermatological, musculoskeletal, respiratory, and genitourinary disorders, in addition to overall cancer risk, injuries and poisonings (postoperative shock and internal knee derangement), and other symptoms such as gangrene. All associations reflected adverse effects of high BMI, except for a suggested protection from inguinal hernias, cystitis, and breast cancer.

Implications of all the available evidence

Data-driven approaches are of increasing value in informing public health. Large-scale genetic studies provide consistent evidence of the widespread implications of a high BMI, highlighting the importance of obesity prevention and careful management of related comorbidities. Our population-level screening implicated an extensive range of diseases as potentially affected by high BMI. True causal effects of obesity are likely to differ by life stage and might be more complex than what can be captured by this type of genetic design.

Globally, nearly 2 billion adults are overweight or obese.[4] Obesity is associated with increased risks of major chronic conditions including diabetes, cardiovascular diseases, and cancer.[5] However, the estimation of causal effects is challenging, due to the scarcity of efficient, evidence-based, well defined, and applicable interventions to prevent obesity.[6] In the absence of robust randomised controlled trials, mendelian randomisation provides an alternative approach for assessing evidence on causality, because it is less affected by reverse causation and confounding, which can lead to bias in other types of observational studies.[7] Body-mass index (BMI) is a simple measure to assess weight for height, commonly used to define overweight and obesity.[4] Genome-wide association studies (GWAS) have led to major advances in the understanding of the genetic architecture of obesity through the discovery of BMI-associated loci,[8] providing opportunities to use these variants in mendelian randomisation studies to disentangle the causal health effects of obesity. However, although excess body fat and obesity are likely to affect a broad range of disease outcomes, the mendelian randomisation studies done to date have been largely hypothesis driven, typically focusing on assessing the causal link with one disease or outcome at a time. By contrast, a hypothesis-free approach covering a broader range of disease outcomes can help to approximate the total burden of comorbidity and, because it is not constrained by pre-existing expectations of disease-outcome associations, this approach might also lead to novel insights on disease risk factors.

In this data-driven phenome-wide association study (PheWAS), we used registry-based information from individuals from the UK Biobank to explore the range of comorbidities associated with high BMI. By contrast with GWAS, which test several genetic loci for their association with one phenotype, our PheWAS used a genetic risk score to approximate high BMI when testing for its associations with multiple disease outcomes. The diseases included in our PheWAS cover all major causes of death and morbidity in this population and, for all associations passing a stringent multiple testing correction, we used several complementary mendelian randomisation approaches to examine evidence supporting a causal role for high BMI.

Methods

Study design and participants

We did a registry-based case-control PheWAS using the UK Biobank, which comprised more than 500 000 participants aged 37–73 years during recruitment, done between March 13, 2006, and Oct 1, 2010.[9] The analyses in this study were restricted to white British individuals (confirmed by self-report and genetic data) who were unrelated (no first-degree, second-degree, or third-degree relatives) and for whom genetic information was available.

We identified disease outcomes on the basis of hospital episode statistics and causes of death from mortality registrations up to March 1, 2016. We extracted International Classification of Diseases (ICD; ninth and tenth editions) codes from the hospital admissions and mortality registrations and mapped them to a phenotype code (phecode) using a previously published method.[10] Compared with ICD coding, phecodes have been shown to provide a grouping of disease codes that are closely aligned with diseases commonly mentioned in clinical practice and genomic studies.[10] We recoded participants with a phecode of interest as cases, whereas participants without a parent phecode or any phecodes under a parent phecode were classified as controls. We excluded all phecodes with fewer than 200 cases.[11]

This research has been done with the UK Biobank, under application 20175. The UK Biobank study was approved by the National Information Governance Board for Health and Social Care and North West Multicentre Research Ethics Committee (11/NW/0382). Participants provided electronic consent to use their anonymised data and samples for health-related research, to be recontacted for further substudies, and for the UK Biobank to access their health-related records.[9]

Genetic risk score

Using genome-wide data available from the UK Biobank,[12] we constructed a genetic risk score consisting of 76 BMI-related genetic variants identified by GWAS in populations of European ancestry (appendix pp 3–7).[8] Weighted genetic risk scores were calculated as the sum of the number of risk increasing alleles weighted by the coefficients from variant-exposure associations taken from the discovery GWAS.[8] The weighted genetic risk scores were rescaled to express the associations per BMI-increasing allele.

Statistical analysis

The main analyses involved two stages: phenome-wide association and mendelian randomisation analyses. Power calculations for each disease outcome were done with a previously established method (appendix p 5).[13] Among the 925 diseases tested for causal association with BMI, our study was powered to detect at least a 20% increase in risk per each SD increase for 108 diseases and at least a 50% increase for 467 diseases (α=5%, power 80%, r^2=2·7%;[8] appendix pp 17–26).

We did a PheWAS to obtain evidence for an association between the BMI genetic risk score and a range of diseases. We used the R package titled phewas to run a logistic regression of each disease outcome against the genetic risk score, adjusting for age, sex, assessment centre, type of genotyping array, and top 15 principal components. Comparisons with models adjusting for 40 principal components and birth location are shown in the appendix (p 27). We examined associations between BMI genetic risk score and confounders (appendix p 8). Because of suggested non-linear association in smokers,[14] we repeated the PheWAS, restricting the sample to participants who never smoked (appendix pp 4, 28). We corrected for multiple testing using Bonferroni correction ($p<5\cdot4\times10^{-5}$; based on α=0·05 divided by the number of outcomes tested) and false discovery rate (p<0·0074).

For all outcomes that passed multiple testing correction in the PheWAS, we did two-sample mendelian randomisation analyses. These analyses used estimates for associations between genetic risk score and BMI from the original GWAS,[8] while the associations between genetic risk score and phecode were tested in the UK Biobank. We did these main analyses using inverse-variance weighted mendelian randomisation (IVWMR), which produces reliable causal estimates in the absence of directional pleiotropy.[15] For sensitivity analyses to assess the stability of estimation, we used several complementary mendelian randomisation methods that operate in differing ways and rely on different assumptions. Although the stability of estimation across methods improves our confidence on the association, no single method exists that outperforms all others across scenarios.[16] The weighted median mendelian randomisation method gives consistent estimates when variants contributing 50% or more of the total weight are valid instruments.[15] The weighted mode method is flexible regarding any variants that violate pleiotropy assumption, assuming the largest weights are contributed by valid instruments.[17] An adaptation of the Egger regression for mendelian randomisation analyses (MR-Egger) allows for the testing of directional pleiotropy, but it produces conservative estimates and will be biased if the "instrument strength independent of direct effect"[16] assumption is violated—for example, by a pleiotropic association with a confounder or residual confounding.

See Online for appendix

The overall estimation in the mendelian randomisation pleiotropy residual sum and outlier (MR-PRESSO)[18] method produces results similar to IVWMR, but the method does three informative tests: a global test to detect the presence of pleiotropy, an outlier test to detect potentially pleiotropic outlier variants for each genetically determined exposure–outcome association, and a distortion test to identify changes in the causal estimates of mendelian randomisation after exclusion of the pleiotropic outlier variant (or variants).[18] MR-PRESSO was re-run with the exclusion of detected outliers. In the presence of pleiotropy (indicated by

	Number of participants	Participants with BMI ≥30 kg/m²	p value	Number of comorbidities	p value
Total	336 442	81 163 (24·1%)	..	3 (1–6)	..
Sex	..	..	$p<1·5\times10^{-47}$	..	$p<1·4\times10^{-8}$
Women	181 268 (53·7%)	41 763 (23·1%)	..	3 (1–6)	..
Men	156 268 (46·3%)	39 400 (25·3%)	..	3 (0–6)	..
Age (years)	..	..	$p<1·6\times10^{-38}$	..	$p<1·0\times10^{-300}$
39·0–44·9	31 721 (9·4%)	6573 (20·8%)	..	1 (0–4)	..
45·0–49·9	42 136 (12·5%)	9328 (22·2%)	..	2 (0–4)	..
50·0–54·9	50 249 (14·9%)	12 460 (24·9%)	..	2 (0–5)	..
55·0–59·9	61 042 (18·1%)	15 404 (25·3%)	..	2 (0–6)	..
60·0–64·5	85 447 (25·3%)	21 197 (24·9%)	..	3 (1–7)	..
65·0–73·0	66 941 (19·8%)	16 201 (24·3%)	..	5 (2–9)	..
History of comorbidity	..	..	$p<1·0\times10^{-300}$	..	$p<1·0\times10^{-300}$
No history	82 337 (24·4%)	15 079 (18·3%)	..	0 (0–0)	..
1	44 564 (13·2%)	8279 (18·6%)	..	1 (1–1)	..
2 to 3	68 845 (20·4%)	14 542 (21·2%)	..	2 (2–3)	..
4 to 5	45 582 (13·5%)	11 247 (24·7%)	..	4 (4–5)	..
6 or more	96 208 (28·5%)	32 016 (33·5%)	..	9 (7–13)	..
General health	..	..	$p<1·0\times10^{-300}$	..	$p<1·0\times10^{-300}$
Excellent	56 541 (16·8%)	5658 (10·0%)	..	1 (0–3)	..
Good	197 195 (58·4%)	41 835 (21·2%)	..	2 (0–5)	..
Fair	68 634 (20·3%)	26 479 (38·8%)	..	5 (2–10)	..
Poor	13 984 (4·1%)	6762 (49·2%)	..	10 (5–18)	..
Not known	1182 (0·4%)	429 (36·3%)	..	6 (2–12)	..

Data are n (%) or median (IQR). p values from logistic regression from models adjusted for age, sex and assessment centre. 1·32% of variance in participants with body-mass index (BMI) 30 kg/m² or higher and 0·52% of variance in the number of comorbidities were explained by the genetic risk score of 76 BMI-related variants from models further adjusted for genotyping array and 15 principal components.

Table: **Distribution of obesity and number of comorbidities across population characteristics**

MR-Egger $p_{intercept}$ <0·050) we re-ran the analyses excluding all variants with greater estimates for variant–disease association than those for variant–obesity associations.[19] Finally, for all associations between BMI genetic risk score and phenotype that passed the multiple testing correction, we searched MR-Base[20] for related variant–outcome association estimates from analyses that did not include the UK Biobank. All statistical analyses were done with R package (version 3.3.1).

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. AM, AZ, and EH had full access to all the data in the study. The corresponding author had final responsibility for the decision to submit for publication.

Results

Of the more than 500 000 participants in the UK Biobank, our study population comprised 337 536 participants, 181 268 (53·7%) of whom were women (table). The prevalence of obesity (defined as BMI ≥30 kg/m²) was higher in men than in women and more common in older participants than in younger ones (table). Analyses were done for 925 unique phecodes from 17 different disease categories (appendix p 9).

BMI genetic risk score explained 1·32% of the variation in obesity, and 1·89% of the variation in BMI in participants. The strongest association in PheWAS was between the risk score and hospital-diagnosed obesity ($p=2·3\times10^{-115}$), whereas associations with 58 other phecodes passed Bonferroni correction for multiple testing ($p<5·4\times10^{-5}$; appendix pp 10–11). BMI genetic risk score showed the strongest signals for type 2 diabetes ($p=9·6\times10^{-71}$), hypertension ($p=3·7\times10^{-47}$), and osteoarthrosis ($p=2·5\times10^{-29}$; figure 1), but evidence for an association was also seen for multiple other diseases, covering the spectrum of disease categories (figure 1; appendix pp 10–11). Associations between BMI genetic risk score with outcomes including breast cancer, depression, gout, and psoriasis passed false discovery rate correction (p<0·0074 for all), but did not pass the more stringent Bonferroni threshold (appendix pp 10–11).

After removing overlapping phecodes, the two-sample mendelian randomisation analysis provided evidence for associations between BMI and 30 distinct conditions that passed the initial PheWAS screening under the Bonferroni correction (figure 2; appendix pp 10–11). The IVWMR method supported the role of high BMI in endocrine diseases, circulatory diseases, and inflammatory or dermatological conditions (figure 2). We found genetic evidence for BMI-related increases in the risk of peripheral nerve disorders, osteoarthrosis, renal failure, bronchitis, asthma, insomnia, and internal derangement of the knee. The only evidence for a potential protective role of higher BMI was observed for inguinal hernia, whereas the risks of umbilical hernia and cholelithiasis were increased. Association results were similar with the IVWMR, weighed median, weighed mode, and MR-PRESSO approaches, whereas evidence for an association was less strong when we used the MR-Egger method (figure 2, appendix pp 29–33). 11 of 30 associations were significant across all mendelian randomisation analysis approaches (type 2 diabetes, type 1 diabetes, hypothyroidism, superficial cellulitis and abscess, chronic ulcer of leg or foot, inguinal hernia, cholelithiasis, osteoarthrosis, insomnia, other peripheral nerve disorders, and hereditary retinal dystrophies). Another nine associations (hypertension, ischaemic heart disease, phlebitis and thrombophlebitis of lower extremities, renal failure, umbilical hernia, other arthropathies, gangrene, chemotherapy, and non-rheumatic aortic valve disorders) were significant with four of the five methods. The associations for four outcomes (congestive heart failure, obstructive chronic bronchitis, internal knee derangement, and oedema) were significant with IVWMR, weighed median mendelian randomisation, and MR-PRESSO, whereas the remaining associations were supported only by IVWMR and MR-PRESSO (figure 2; appendix pp 10–11). MR-Egger pleiotropy test suggested the

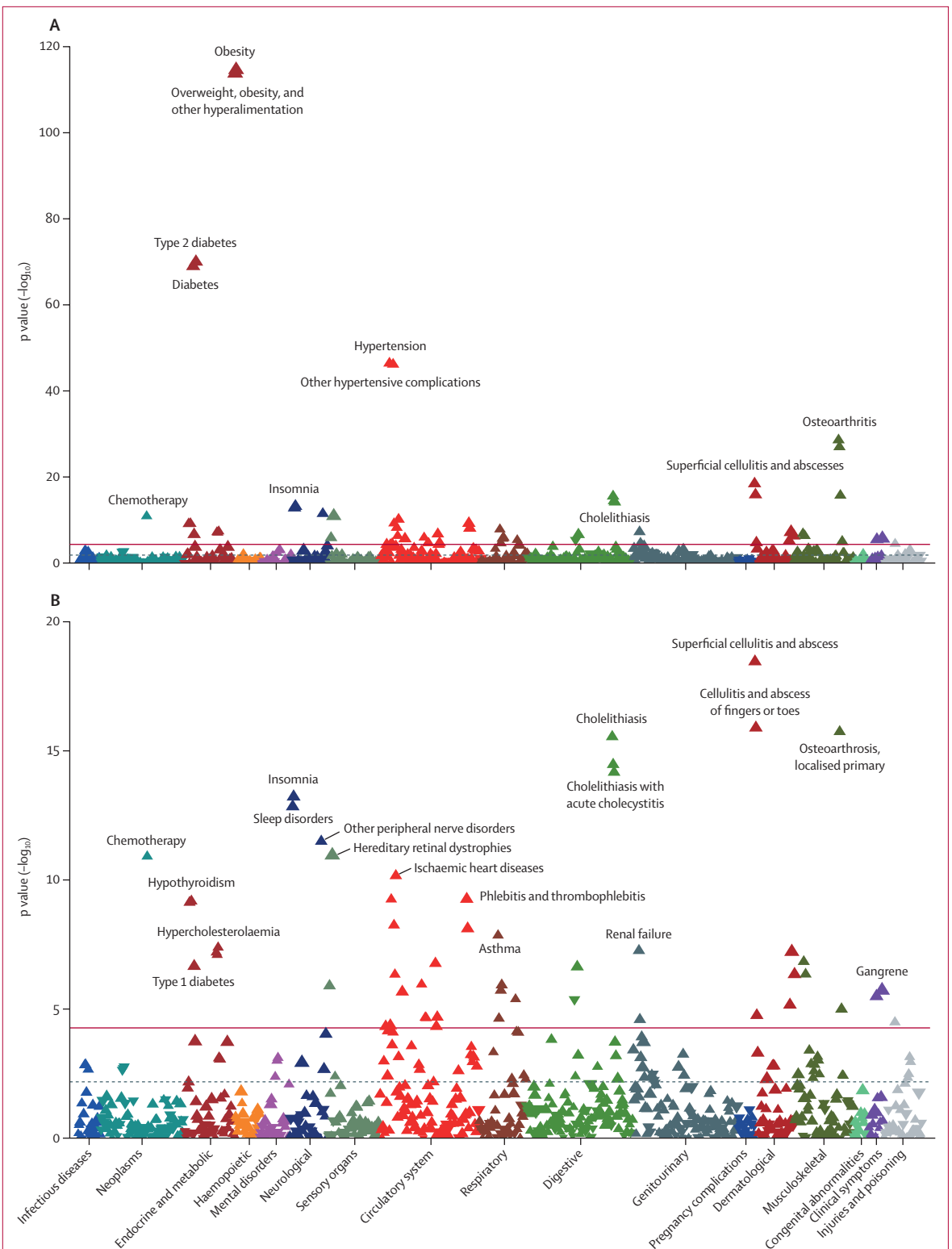
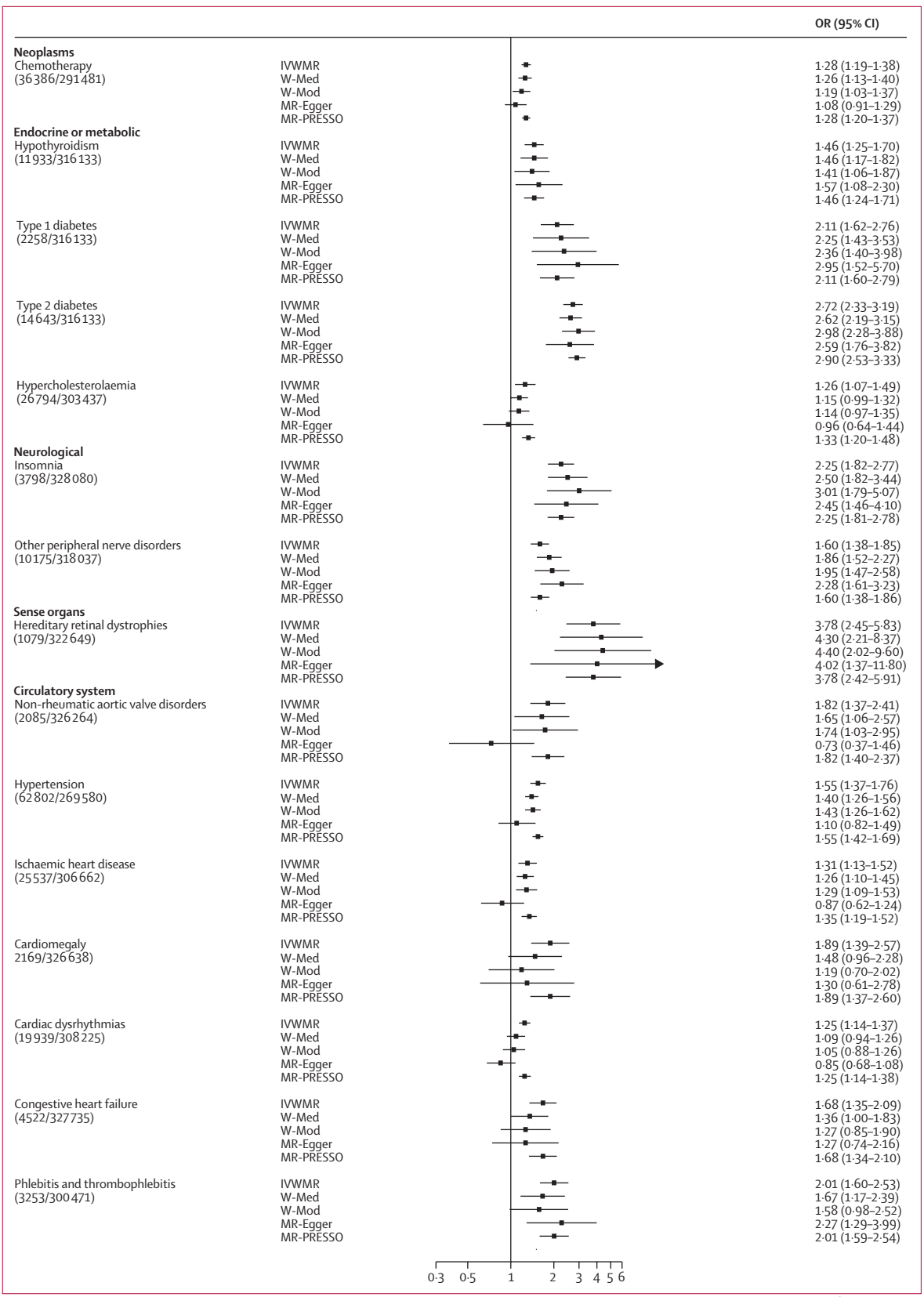


Figure 1: **Manhattan plot showing the phenome-wide association between BMI genetic risk score and disease outcomes**
Bonferroni corrected threshold (p<0·0001) is represented by the red line and false discovery rate-corrected threshold (p<0·0074) by the dashed line. The direction of the arrow reflects whether variants increasing body-mass index (BMI) were associated with increased (up arrow) or decreased (down arrow) odds of disease. (A) Global view with all disease outcomes. (B) Zoomed view, showing signals within the range of $p>1\times10^{-20}$ to $p<5{\cdot}4\times10^{-5}$.



(Figure 2 continued on next page)

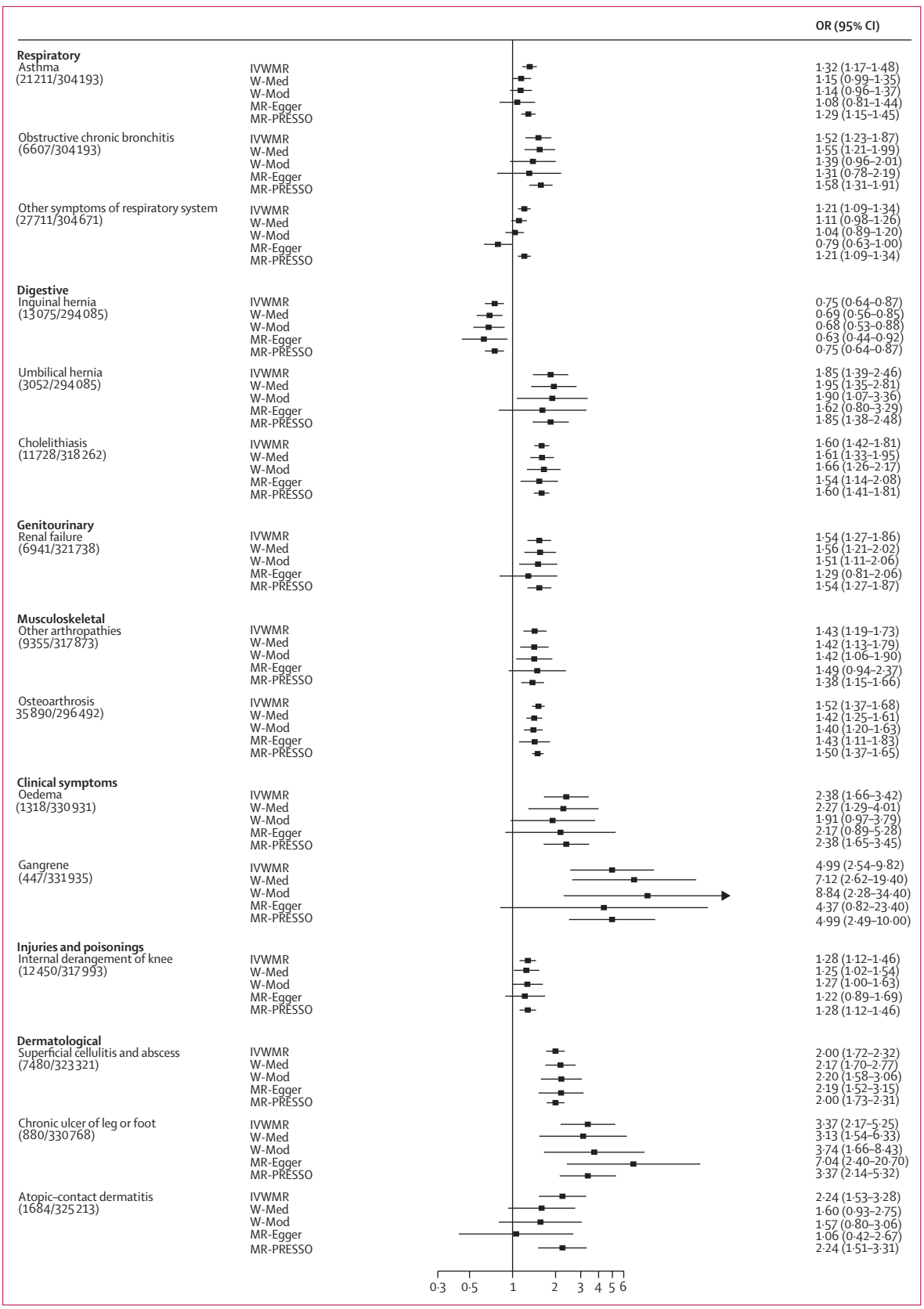


***Figure 2*: Mendelian randomisation analyses on the top 30 distinct BMI–disease associations**
Analyses were done with inverse-variance weighted mendelian randomisation (IVWMR), weighted median mendelian randomisation (W-Med), weighted mode mendelian randomisation (W-Mod), adapted Egger regression for mendelian randomisation (MR-Egger), and mendelian randomisation pleiotropy residual sum and outlier (MR-PRESSO) method. Estimates are odds ratios (OR; 95% CI) per SD (4·1 kg/m^2) higher body-mass index (BMI).

presence of directional pleiotropy for seven of the 30 BMI–disease associations (appendix pp 12–13, 34–63), including for hypertension, ischaemic heart diseases, non-rheumatic aortic valve disorders, cardiac dysrhythmias, other symptoms of respiratory system, other peripheral nerve disorders, and chemotherapy (a proxy phecode indicating all cancer types). Of the 30 diseases associated with BMI, MR-PRESSO identified outlying variants for type 2 diabetes, hypercholesterolaemia, hypertension, ischaemic heart diseases, inguinal hernia, osteoarthrosis, and chronic bronchitis (appendix pp 12–13). However, the exclusion of these potentially pleiotropic outlier variants did not notably affect the effect estimates (for all, $p_{distortion}>0\cdot25$; appendix pp 10–11).

There were 42 additional distinct disease associations that were detected under false discovery rate correction (appendix pp 10–11, 29–33). The analyses on outcomes under false discovery rate correction provided consistent support across mendelian randomisation methods for BMI association with nephritis, cystitis, and inflammatory neuropathy, whereas associations with autonomic nervous system disorders, respiratory failure, and ulcer of the oesophagus were supported by all but MR-Egger (appendix pp 34–63).

We ran individual PheWAS analyses for all 76 variants included in the BMI genetic risk score to identify pleiotropic variants with strong individual associations with disease risk (appendix pp 14–15, 64–79). To reduce the risk of bias due to pleiotropy, we re-ran the mendelian randomisation analysis excluding the 17 pleiotropic single nucleotide polymorphisms identified in the individual variant PheWAS. In these mendelian randomisation analyses using 59 variants, we observed no evidence for pleiotropy in the BMI associations with hypertension and chemotherapy (appendix pp 10–11). Evidence for pleiotropy remained for ischaemic heart diseases, non-rheumatic aortic valve disorders, cardiac dysrhythmias, other peripheral nerve disorders, and other symptoms of respiratory system, after excluding the pleiotropic variants.

We used the available consortia-based studies identified from MR-Base[20] or elsewhere to replicate the association between BMI and multiple disease outcomes. In IVWMR, BMI was associated with type 2 diabetes ($p=1\cdot2\times10^{-14}$), ischaemic heart disease ($p=6\cdot1\times10^{-9}$), knee and hip osteoarthritis ($p=6\cdot4\times10^{-4}$), asthma ($p=0\cdot046$), gout ($p=1\cdot9\times10^{-6}$), and breast cancer ($p=1\cdot7\times10^{-7}$; appendix p 16). We also observed associations for relevant continuous traits including high-density lipoprotein (HDL) cholesterol ($p=1\cdot4\times10^{-6}$, related to hypercholesterolaemia) and renal function-related measures such as microalbuminuria ($p=9\cdot7\times10^{-4}$) and urinary albumin-to-creatinine ratio ($p<0\cdot025$). MR-Egger and weighted median mendelian randomisation methods supported these associations ($p<0\cdot010$ for all comparisons), except for HDL-cholesterol ($p=0\cdot089$).

Discussion

Our data-driven, mendelian randomisation PheWAS done in 337 536 participants from the UK Biobank suggests wide-ranging implications for excess bodyweight and obesity on health. Our analyses, using multiple mendelian randomisation methods, supported the well known effects of high BMI on cardiovascular diseases and showed that, in addition to ischaemic heart disease and hypertension, the downstream consequences of high BMI include increases in the risks of cardiomegaly, dysrhythmias, and heart failure. We also showed evidence supporting not only the well known association between high BMI and type 2 diabetes, but also the less well established association with type 1 diabetes, and we provided novel evidence for the role of high BMI as a risk factor for—rather than a consequence of[21]—hypothyroidism. Furthermore, although clinical evidence supports associations between obesity and many outcomes identified in our analyses, our study also provides causal evidence regarding inflammatory conditions (eg, cellulitis and bronchitis), retinal dystrophies, hypothyroidism, chronic ulcers of leg and foot, gangrene, and structural problems such as hernias and knee derangement.

The burden of excess weight on cardiovascular health is well known,[5] and the range of diseases that are adversely affected by weight is extensive. Our hypothesis-free approach picked up adverse associations of genetically determined high BMI with ischaemic heart disease (eg, angina pectoris and atherosclerosis), cardiomegaly, dysrhythmias (atrial fibrillation), heart failure, hypertension, phlebitis and thrombophlebitis, and hypercholesterolaemia. High BMI was associated with both type 2 and type 1 diabetes, as well as with peripheral nerve disorders, chronic leg or foot ulcers, and gangrene, which can all occur as diabetic complications. Our reported association of high BMI with renal failure is relevant, supporting findings from an earlier mendelian randomisation study[22] that showed a link between obesity and end-stage renal disease in type 1 diabetes, highlighting the importance of managing obesity as a way of reducing the risk of complications.

Obesity is a strong predictor for the incidence and progression of musculoskeletal diseases, including osteoarthritis, a leading cause of disability worldwide.[23] Evidence suggests that the risk of osteoarthritis increases with long exposure to high BMI, leading obese individuals to have more severe joint degeneration in the knees compared with that of non-obese people.[24] A causative role for high BMI on osteoarthritis is supported by an earlier mendelian randomisation study[25] and, in our analyses, the association identified with use of the UK Biobank was further supported by look-up data from the MR-Base.[13] We also observed evidence for an effect of high BMI on structural problems, including knee derangement and hernias. Obesity is a risk factor for umbilical hernias,[26] and this adverse association was supported by our

analyses. By contrast, our findings and those of previous observational studies[27] suggest a protective effect of high BMI on the risk of inguinal hernia. The development of hernias, and the location where this occurs in a susceptible individual, depends on the strength of the supporting tissues and might be affected by the presence and location of excess fat, differentially increasing internal pressure on the tissue, according to location.

Obesity is a well known risk factor for cancer, and our primary PheWAS picked up the association of high BMI with chemotherapy treatment, used as a proxy indicator for any type of cancer, while little evidence for association was seen with specific types of cancer. However, one of the rare protective associations of high BMI was with breast cancer, which was detected in our PheWAS under the false discovery rate and was supported by independent replication.[28] This proposed protective association is in line with studies[29] on the lower risk of premenopausal breast cancer in obese women compared with that of non-obese women, reported for oestrogen receptor-positive breast cancer. More work is required to establish causality and underlying mechanisms, because the association between obesity and breast cancer is likely to be complex, and might depend upon the timing of obesity, oestrogen receptor positivity, and menopausal status.

We observed an association between high BMI and insomnia, whereas a previous PheWAS[30] using *FTO* variants to approximate BMI (n=24198) found an association of high BMI with sleep apnoea, another sleep disorder linked to insomnia. Obesity can affect aspects of mental health and, in our study, we observed an association between high BMI and depression, which passed false discovery rate correction. We have previously reported[19] evidence of a causal association between obesity and depression in women, but not in men. Indeed, because of societal and other differences in the perception of body image, some mental health consequences of obesity can differ between genders. However, we did not do sex-stratified analyses, because this would have markedly reduced the statistical power and increased the burden of multiple testing.

An important strength of our study was the large sample size and the availability of comprehensive information on hospitalisations and mortality registrations. Although another PheWAS[31] using mendelian randomisation, published in 2019, also examined outcome associations related to BMI in the UK Biobank, this study used a different approach, implementing the analyses with use of the PHESANT open-source phenome scan tool. Both the PHESANT study[31] and our study were able to find expected associations of BMI with outcomes such as diabetes and cardiovascular diseases, whereas the PHESANT study also highlighted influences of BMI on factors such as self-reported psychosocial traits, which were not covered by our study. By contrast, our registry-based case-control study identified associations with a broader range of diseases and included extensive analyses to assess the robustness of causal evidence for all outcomes.

There are several limitations with the type of discovery to causal inference approach implemented in our study. Although the usefulness of the mendelian randomisation approach in the context of BMI is well established, valid inferences on causal support for an association can only be made conditional on three core assumptions: the genetic instrument needs to be associated with the exposure, there should be no joint causal influence affecting both instrument and outcome, and the instrument should not affect the outcome through any mechanism other than exposure.[7] Despite our extensive modelling strategy, we could not fully exclude bias due to pleiotropic effects. Although pleiotropy does not discount a causal effect of BMI on these outcomes, it suggests a presence of positive or negative bias in the estimated effect sizes, with some individual variants reflecting a disproportionate effect than that expected on the basis of their primary association with BMI. Pleiotropy might also introduce reverse causality, where the variants included in the BMI genetic risk score primarily influence other disease outcomes, and the apparent effect on high BMI arises through disease-inflicted changes on BMI. The GWAS[8] that formed the basis of our instrument identification was based on an adult population, with the discovered variants reflecting genetic influences on BMI that were accumulated over the life course. This has relevance for the broader interpretation of our findings, because estimates derived with use of a genetic risk score (in our study or others), might not adequately reflect the effects of BMI on disease risk for all life stages.

Another methodological limitation was residual confounding, which can arise, for example, through population stratification and could be aggravated by the use of a genetic risk score in the discovery stage.[32] Our study proceeded from PheWAS discovery to mendelian randomisation analyses, where associations that are the most affected by residual confounding are, by default, the ones most likely to be taken forward to causality testing. We were reassured to find that the strongest associations of high BMI were with obesity and well established conditions such as type 2 diabetes and hypertension and that our findings were not sensitive to the strategy to control for population stratification. However, future studies implementing an approach such as ours should consider using prespecified positive or negative controls based on a priori statements of expected effects, because this would allow the establishment of a reference point for the assessment of influences by unmeasured residual confounding.

Our study focused on linear increases of BMI and we might not have fully captured the effects of high BMI, reflecting more severe forms of obesity. A mendelian randomisation study,[14] published in 2019, reported genetic evidence for a J-shaped association between BMI and mortality risk, with evidence for increased mortality

risk in underweight individuals who smoke regularly. Because a high BMI was associated with increases in disease risk in our study (with very few exceptions), a failure to account for the adverse effects of underweight would have led to an underestimation of the BMI–disease risk associations. Sensitivity analyses in the subsample of participants who never smoked did not strengthen our associations; however, the sample size was halved by this restriction, and stratification can lead to collider bias.[33] Collider bias might also arise because of selection in the UK Biobank, which only had a 5% participation rate; however, the effect of collider bias caused by selection on mendelian randomisation studies has been suggested to be less than that of other biases.[33] We used the MR-Base[20] with the aim of further supporting our findings in datasets with different confounding structures and determinants of selection. However, for most outcomes, independent data were not available. Another limitation of our study was the inability to provide conclusive evidence for an absence of an association. Although a simulation study has suggested 200 reported cases as a threshold for reasonable power in PheWAS,[11] this estimation is based on the ability to detect associations of single nucleotide polymorphisms and disease risk, and our study was not formally powered to detect many of the associations in the mendelian randomisation framework.[13] Inclusion of outcomes with a low case threshold also led to our conservative approach to multiple testing, which can help to reduce the likelihood of type I errors.

In conclusion, although our mendelian randomisation PheWAS provides a promising data-driven approach to disease risk factor discovery, challenges remain with its implementation. Our findings implicate an extensive range of diseases as being affected by high BMI. However, the true causal effects of excess weight on disease risk are likely to vary by life stage and be more complex than indicated by our population-level screening.

Contributors

EH conceptualised and led the study and advised on analyses. VKS made the first draft of the manuscript. EH and VKS wrote the paper and searched literature. AM and AZ conducted, described, and advised on the statistical analyses; managed data; and did the figures. AM, AZ, and EH vouch for the data and analysis. All authors were involved in revising the manuscript, in data interpretation, and reviewed the manuscript before submission.

Declaration of interests

EH reports grants from the National Health and Medical Research Council Australia, during the conduct of the study. All other authors declare no competing interests.

Data sharing

All data is available through the UK Biobank repository on application.

Acknowledgments

This work was funded by the National Health and Medical Research Council Australia (GNT1123603 and GNT1157281). The authors thank the participants of UK Biobank who answered surveys and made this work possible. This work was conducted under UK Biobank application 20175.

References

1 Xu L, Lam TH. Stage of obesity epidemic model: learning from tobacco control and advocacy for a framework convention on obesity control. *J Diabetes* 2018; **10:** 564–71.

2 Nyberg ST, Batty GD, Pentti J, et al. Obesity and loss of disease-free years owing to major non-communicable diseases: a multicohort study. *Lancet Public Health* 2018; **3:** e490–97.

3 Biener A, Cawley J, Meyerhoefer C. The high and rising costs of obesity to the US health care system. *J Gen Intern Med* 2017; **32**(suppl 1): 6–8.

4 WHO. Obesity and overweight. 2018. http://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight. (accessed April 17, 2019).

5 Dixon JB. The effect of obesity on health outcomes. *Mol Cell Endocrinol* 2010; **316:** 104–08.

6 Chiolero A. Why causality, and not prediction, should guide obesity prevention policy. *Lancet Public Health* 2018; **3:** e461–62.

7 Davies NM, Holmes MV, Davey Smith G. Reading mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* 2018; **362:** k601.

8 Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015; **518:** 197–206.

9 Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015; **12:** e1001779.

10 Denny JC, Bastarache L, Ritchie MD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nature Biotech* 2013; **31:** 1102–10.

11 Verma A, Bradford Y, Dudek S, et al. A simulation study investigating power estimates in phenome-wide association studies. *BMC Bioinformatics* 2018; **19:** 120.

12 Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018; **562:** 203–09.

13 Burgess S. Sample size and power calculations in mendelian randomization with a single instrumental variable and a binary outcome. *Int J Epidemiol* 2014; **43:** 922–29.

14 Sun YQ, Burgess S, Staley JR, et al. Body mass index and all cause mortality in HUNT and UK Biobank studies: linear and non-linear mendelian randomisation analyses. *BMJ* 2019; **364:** l1042.

15 Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016; **40:** 304–14.

16 Burgess S, Foley CN, Allara E, Staley JR, Howson JMM. A robust and efficient method for mendelian randomization with hundreds of genetic variants: unravelling mechanisms linking HDL-cholesterol and coronary heart disease. *bioRxiv* 2019; published online March 4. DOI:10.1101/566851 (preprint).

17 Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* 2017; **46:** 1985–98.

18 Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nat Genet* 2018; **50:** 693–98.

19 Tyrrell J, Mulugeta A, Wood AR, et al. Using genetics to understand the causal influence of higher BMI on depression. *Int J Epidemiol* 2018; published online Nov 13. DOI:10.1093/ije/dyy223.

20 Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018; **7:** e34408.

21 Chaker L, Bianco AC, Jonklaas J, Peeters RP. Hypothyroidism. *Lancet* 2017; **390:** 1550–62.

22 Todd JN, Dahlstrom EH, Salem RM, et al. Genetic evidence for a causal role of obesity in diabetic kidney disease. *Diabetes* 2015; **64:** 4238–46.

23 Kulkarni K, Karssiens T, Kumar V, Pandit H. Obesity and osteoarthritis. *Maturitas* 2016; **89:** 22–28.

24 Wills AK, Black S, Cooper R, et al. Life course body mass index and risk of knee osteoarthritis at the age of 53 years: evidence from the 1946 British birth cohort study. *Ann Rheum Dis* 2012; **71:** 655–60.

25 Panoutsopoulou K, Metrustry S, Doherty SA, et al. The effect of FTO variation on increased osteoarthritis risk is mediated through body mass index: a mendelian randomisation study. *Ann Rheum Dis* 2014; **73:** 2082–86.

26 Lau B, Kim H, Haigh PI, Tejirian T. Obesity increases the odds of acquiring and incarcerating noninguinal abdominal wall hernias. *Am Surg* 2012; **78:** 1118–21.

27 Ruhl CE, Everhart JE. Risk factors for inguinal hernia among adults in the US population. *Am J Epidemiol* 2007; **165:** 1154–61.

28 Michailidou K, Beesley J, Lindstrom S, et al. Genome-wide association analysis of more than 120 000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* 2015; **47:** 373–80.

29 Premenopausal Breast Cancer Collaborative G, Schoemaker MJ, Nichols HB, et al. Association of body mass index and age with subsequent breast cancer risk in premenopausal women. *JAMA Oncol* 2018; **4:** e181771.

30 Cronin RM, Field JR, Bradford Y, et al. Phenome-wide association studies demonstrating pleiotropy of genetic variants within FTO with and without adjustment for body mass index. *Front Genet* 2014; **5:** 250.

31 Millard LAC, Davies NM, Tilling K, Gaunt TR, Davey Smith G. Searching for the causal effects of body mass index in over 300 000 participants in UK Biobank, using mendelian randomization. *PLoS Genet* 2019; **15:** e1007951.

32 Burgess S, Thompson SG. Use of allele scores as instrumental variables for mendelian randomization. *Int J Epidemiol* 2013; **42:** 1134–44.

33 Gkatzionis A, Burgess S. Contextualizing selection bias in mendelian randomization: how bad is it likely to be? *Int J Epidemiol* 2018; published online Oct 15. DOI:10.1093/ije/dyy202.