RESEARCH ARTICLE

AMERICAN JOURNAL OF medical genetics PART B
Neuropsychiatric Genetics

Revisiting Mendelian Randomization Studies of the Effect of Body Mass Index on Depression

Stefan Walter,[1,2]* Laura D. Kubzansky,[1] Karestan C. Koenen,[3] Liming Liang,[4,5] Eric J Tchetgen Tchetgen,[4,5] Marilyn C. Cornelis,[5] Shun-Chiao Chang,[1] Eric Rimm,[4,5] Ichiro Kawachi,[1] and M. Maria Glymour[1,2]

[1]Department of Social and Behavioral Sciences, Harvard School of Public Health, Boston, Massachusetts

[2]Department of Epidemiology & Biostatistics, University of California, San Fransisco, California

[3]Mailman School of Public Health, Colombia University, New York

[4]Biostatics, Harvard School of Public Health, Boston, Massachusetts

[5]Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts

Manuscript Received: 9 June 2014; Manuscript Accepted: 25 November 2014

Mendelian Randomization studies, which use genetic instrumental variables (IVs) as quasi-experiments to estimate causal effects, report inconsistent findings regarding effects of body mass index (BMI) on mental health. We used genetic IV to estimate effects of BMI on depression and evaluated validity of a commonly used IV. Female Nurse's Health Study participants (n = 6989, average age 56.4, [Standard Deviation 6.91] years at first depression assessment) self-reported BMI, which was averaged across eight reports prior to depression assessment (mean = 24.96, SD 4.50). Genetic instruments included fat mass and obesity-associated protein (FTO) alleles, melanocortin receptor 4 (MC4R) alleles, and polygenic risk scores based on 32 established polymorphisms for BMI. Depression was assessed using multiple symptom measures, scaled to the Geriatric Depression Scale 15, averaged across up to 7 biennial waves. We used over-identification tests to assess the validity of genetic IVs. In conventional estimates, each additional BMI point predicted 0.024 (95% Confidence Interval (CI): 0.020–0.029) higher average depression scores. Genetic IV estimates were not significant when based on FTO (beta: 0.064, CI: -0.014, 0.142), MC4R (beta: 0.005, CI: -0.146, 0.156), polygenic score excluding FTO (beta = -0.003, 95%-CI -0.051, 0.045), or mechanism-specific scores. The over-identification test comparing IV estimates based on FTO to estimates based on the polygenic score excluding FTO rejected equality of estimated effects ($P = 0.014$). Results provide no evidence against a null effect of BMI on depression and call into question validity of FTO as an instrument for BMI in Mendelian Randomization studies. © 2015 Wiley Periodicals, Inc.

Key words: obesity; adiposity; genetic; FTO; causation; instrumental variable analysis

How to Cite this Article:
Walter S,Kubzansky LD, Koenen KC, Liang L, Tchetgen Tchetgen EJ, Cornelis MC, Chang S-C, Rimm E, Kawachi I, Glymour MM. 2015. Revisiting Mendelian Randomization Studies of the Effect of Body Mass Index on Depression.
Am J Med Genet Part B 168B:108–115.

INTRODUCTION

Depression is a major cause of morbidity representing enormous social and economic burdens.[Brundtland 2001] The biological mechanisms underlying the etiology of depression remain uncertain, however, making it difficult to identify modifiable risk factors amenable to preventive interventions. The relationship between obesity and depression has long been controversial: many studies indicate that overweight and obesity are correlated with adverse mental health consequences while others report higher BMI predicts lower risk of depression. [Luppino et al., 2010] Conventional correlational studies may be biased by unobserved common causes of obesity and mental health or by reverse causation from psychological distress to higher body weight.

Recently, analyses based on genetic instrumental variables (IV), popularly termed "Mendelian Randomization" studies, have been proposed as an approach to avoid the confounding problems that beset observational studies in this area. [Davey et al., 2003] In genetic IV studies, genotypes that influence obesity are treated as natural experiments to estimate the mental health consequences of

Conflict of interest: None.
Grant sponsor: NIH/NIMH; Grant number: MH092707–01.
*Correspondence to:
Stefan Walter, 401 Park Drive, 02215 Boston, MA.
E-mail: swalter@hsph.harvard.edu
Article first published online in Wiley Online Library (wileyonlinelibrary.com)
DOI 10.1002/ajmg.b.32286

BMI. Because the genotype is determined at conception, reverse causation and unobserved confounding are less plausible in genetic IV studies. With a set of clearly described assumptions, such as those illustrated in the causal diagram in Figure 1, genotypic information can be used to estimate the effect of BMI on depression. Three genetic IV studies of the effect of obesity on mental health have been published to date, focusing on generalized measures of psychological distress and depression. [Jokela et al., 2012; Kivimäki et al., 2011; Lawlor et al., 2011] These studies had various limitations including small and non-representative samples [Jokela et al., 2012], non-specific or single-item measures of depression or psychological distress [Lawlor et al., 2011], and genetic IVs based on only one or two polymorphisms. [Kivimäki et al., 2011] Perhaps as a result of these limitations, findings have been inconsistent, with two studies suggesting obesity caused distress [Jokela et al., 2012; Kivimäki et al., 2011], while the other study suggested obesity promoted mental health. [Lawlor et al., 2011]

To try to resolve some of these inconsistencies, we examine the effect of BMI on depression, using a more comprehensive measure of depression than has previously been considered, and a genetic IV based on a large set of relevant, externally validated polymorphisms. We also introduce formal assessments of the validity of the genetic polymorphisms most commonly used in previous genetic IV analyses.

Taking advantage of findings from previously published meta-analyses of genetic determinants of BMI in over 250,000 participants [Speliotes et al., 2010] to define a genetic IV with optimum statistical power, we conducted genetic IV analyses to estimate the effects of BMI on depressive symptoms. We explored the validity of the genetic IV approach for this research question by using over-identification tests, which compare estimates from alternative genetic IVs, and by evaluating whether genotypes associated with BMI are directly associated with depression independently of the pathway involving BMI. Finally, we examined the possibility that the effects of BMI on depression may differ depending on the biological mechanism by which the genes influence BMI.

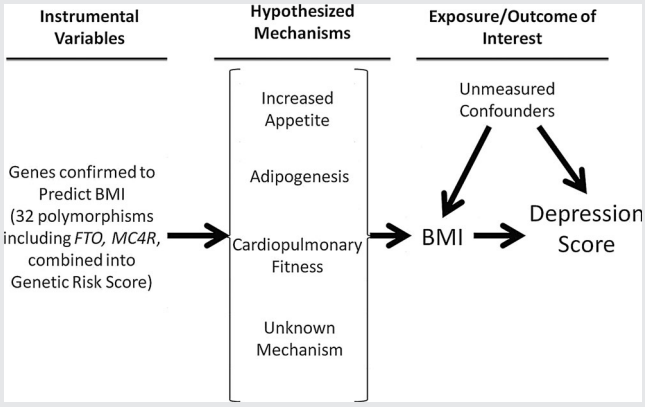

FIG. 1. Causal diagram representing the assumptions for genetic IV analyses to estimate the effect of BMI on depression. The causal diagram follows the rules for directed acyclic graphs (DAG) and describes the assumed causal structure motivating the IV analyses. Key assumptions represented in this diagram are: 1) the genotype is associated with BMI; 2) the genetic instrumental variables do not influence depression except via BMI; and 3) there are no common causes of genotype and depression. Under these assumptions, the genotype can be used as an instrumental variable to estimate the effect of BMI on depression, even when there are unmeasured confounders of BMI and depression.

METHODS

Population

All data were drawn from 4 nested case-control GWAS within the Nurses' Health Study (NHS). The NHS was established in 1976 when 121,700 female registered nurses aged 30–55 years and residing in 11 large U.S. states completed a mailed questionnaire on medical history and lifestyle characteristics.[Colditz and Hankinson 2005] Blood was collected from 32,826 participants between 1989 and 1990. DNA was extracted from white blood cells using the QIAmp^{TM} (QIAGEN Inc., Chatsworth, CA) blood protocol and all samples were processed in the same laboratory. Genome-wide scans were obtained from 4 independent GWAS nested within the cohort initially designed to examine type 2 diabetes (T2D), coronary heart disease (CHD), breast cancer (BrCa) and kidney stone (KS) disease. [Cornelis et al., 2011] After quality control (QC) and considering available information on depression, a total of 6989 genetically defined White participants were available from NHS (NHS T2D = 3084, NHS CHD = 1135, NHS BrCa = 2280, NHS KS = 490).

Ethics Statement

The NHS and this project were approved by the Human Subjects Committee of Brigham and Women's Hospital, Boston, MA. All participants in this study provided written informed consent.

Genetic Instrument

First, we constructed instruments imposing an additive genetic risk model. We selected two single genes previously used in genetic IV analyses of the effect of obesity on depression: FTO [Kivimäki et al., 2011] (rs1558902, A) and MC4R [Kivimäki et al., 2011; Lawlor et al., 2011] (rs571312, A). The IV for each polymorphism was the sum of the number of BMI increasing alleles multiplied by the estimated effect of the BMI increasing allele on BMI, with the estimated effect based on the beta-estimates from the largest current GWAS meta-analysis. [Speliotes et al., 2010] Both SNPs were highly significant when using an additive genetic model (rs1558902–$P = 4.8 \times 10^{-120}$, rs571312–p $= 6.4 \times 10^{-43}$) in the joint analysis of discovery and replication cohorts. [Speliotes et al., 2010] In addition, we separately used data on 32 single nucleotide polymorphisms (SNPs) (including the polymorphisms in FTO and MC4R) confirmed as genome-wide significant predictors of BMI to construct a Genetic Risk Score (GRS). [Speliotes et al., 2010] Information on all 32 SNPs was extracted from the imputed genotype data of all 4 sub-studies. We calculated the GRS for each individual *i* in our study sample as the sum of risk alleles, with each SNP weighted by

the beta estimate as published in Speliotes et al. (Supplementary Information Table S1, Equation 1). [Purcell et al., 2009; Speliotes et al., 2010]

$$GRS_i = \Sigma_{k=1}^{loci} \beta_k \cdot allelecount_{i,k} \quad (1)$$

Based on prior evidence suggesting a direct and protective effect of rs9939609 A, an intronic SNP located in the FTO gene in high linkage disequilibrium ($r^2 = 0.9$) rs1558902 A used in this study, on depression, we also constructed a GRS excluding FTO (GRS*ex*FTO). [Samaan et al., 2013].

Exact genotyping, quality control, and imputation protocols varied slightly by sample set (Tables S2 and S3). Principal components analyses were conducted to obtain eigenvectors to control for sub-threshold population stratification; based on these eigenvectors, we excluded self-reported white individuals who had substantial similarity to non-European reference samples. [Price et al., 2006] Each study imputed up to 2.5 million autosomal SNPs with NCBI build 36 of Phase II HapMap CEU data (release 22) as the reference panel using MACH.

We performed a search in PubMed and Google Scholar databases for scholarly articles identifying the functions of the 32 Speliotes SNPs; the NCBI Gene database was used as additional reference if published journal articles regarding SNP functions were not available. After SNP functions were identified, SNPs were categorized by key words and overall concept into the following 3 domains of functioning through which they are thought to influence BMI: Appetite, Adipogenesis, Cardio-pulmonary function. SNPs that could not be categorized into a clearly identifiable domain were combined into a separate domain referred to as Other/Unknown.

Phenotype

Height and weight were self-reported by all participants. Height was reported at the baseline questionnaire. BMI was calculated as weight (kg) divided by the square of height (m^2) and summarized as the mean of all measurements prior to the first assessment of depression in 1992 (8 reports of BMI between 1976 and 1990). In prior work with the NHS, self-reported weights, validated among 184 NHS participants, were highly correlated with measured weights ($r = 0.96$; mean difference [self-reported − measured weight] = − 1.5 kg). [Rimm et al., 1990] Because of the separate-sample IV design (see below), self-reported BMI was only used to confirm the validity of the instrument, it was not used to estimate the effect size of BMI on the outcome.

Outcome

In the NHS, a variety of assessments of depression or depressive symptoms were collected in successive questionnaire cycles from 1992 to 2008 (Table S4), ranging from standard symptom measures to reports of antidepressant use. To combine information on depressive symptoms across multiple sources of information and over 14 years of follow-up, we derived a standardized composite depression score for each questionnaire cycle as described previously. [Chang et al., 2014] Briefly, standardized composite depression scores were obtained by scaling all depression measures at each wave to the Geriatric Depression Scale 15 (GDS-15), a well-validated depression symptom screening tool in the elderly [Sharp and Lipsky 2002; Sheikh and JI, 1986], providing the most recent information available about depression in this sample. We then used these scores to derive a final 14-year long-term average depression measure as the average depression score across all available questionnaire cycles up to 2006 (up to 7 waves), to capture more accurately the long-term experience of the phenotype.

Analyses

We use Chi-Squared tests and Analysis of Variance for bivariate descriptive analyses to evaluate the association of the GRS with sociodemographic and cardiovascular conditions. We also confirmed that each of the genetic instruments predicted BMI in the NHS sample, using linear regression. All linear regression models were estimated with inverse probability weights to reconstruct the original cohort from the nested case-control genetic samples to circumvent the bias of selective sampling in this population. Additionally, all analyses were adjusted for age, age^2, and the top 3 eigenvectors to increase precision. We refer to the "conventional observational effect estimate" of BMI on depression as the linear regression coefficient from models adjusted for age and age-squared. Based on prior research, we anticipated that these estimates would be positively biased due to unmeasured confounding and reverse causation, each of which is likely to spuriously inflate the association between high BMI and risk of depression.

The IV analysis was executed using separate-sample IV analyses, which is a modification of the conventional two-stage least squares regression IV analysis and avoids weak-instruments bias. [Angrist and Krueger 1994, 1995] We use the 1st stage estimates from the previous meta-analysis of the genetic determinants of BMI as described above. In the 2nd stage of the separate-sample IV, we regressed the average depression score on the genetic IV, using a linear model. Under the assumptions for a valid IV (as in Fig. 1), the coefficients in this model are interpretable as the effect of BMI on average depression, with the caveat that this is specifically the effect of BMI changes induced by the genetic polymorphisms. We use the Wu-Hausman statistic to test the null that the IV estimate is consistent and less biased compared to the OLS estimate.

For inference, sandwich estimators were used to adjust for heteroskedasticity and inverse probability weights to correctly reweight the case-control populations to the Nurses' Health Study source population.[Monsees et al., 2009] All analyses were adjusted for age, age^2, and the first three genetic eigenvectors to control for population substructure. [Price et al., 2006] Results were based on two sided tests and $P < 0.05$ was considered statistically significant.

After deriving IV effect estimates, we used two approaches to evaluate the IV assumptions. First, we used the mechanism-specific categories of genetic risk scores to conduct over-identification tests assessing that the IV estimates from each of the genetic risk scores were identical to each other—as they should be because they instrument the same phenotype, i.e., BMI. Second, we assessed whether the association between the genetic IVs and depression was less than or equal to zero when additionally adjusted for BMI,

consistent with the idea that BMI would mediate the effect of the genes on depression. Although collider bias is a concern for this type of mediation models, it has been shown that in linear models and a broad range of non-linear models, collider bias in the context of a valid IV should flip the sign of the association between the instrument and the outcome making these "overadjusted models" a useful tool for testing IV assumptions. [Glymour et al., 2012]

RESULTS

The study population was on average 56.4 (Standard Deviation SD 6.91) years of age at the first depression assessment in 1992 (Table I). The mean BMI averaged across the years prior to the assessment of depression (1976–1990) was 24.96 (SD 4.50). The average depression score over the course of follow-up (a maximum time window of 14 years and 7 assessments) was 1.81 (SD 0.64).

Genetic IVs as Predictors of BMI and Other Characteristics

The GRS used in this analysis was strongly associated with BMI (Table II). Figure 2 shows the F statistic and the explained variance of BMI due to the GRS. The SNPs on the x axis are sorted by increasing *P*-value as reported in Speliotes et al. and the illustrated summary statistic corresponds to the association of the cumulative genetic risk score with self-reported BMI in our population, in models adjusted for eigenvectors to account for population substructure. For all but one SNP (rs987237 in TFAP2B) the directionality of the association was identical in our study population as reported in Speliotes et al. (Supplementary Table S1). The GRS explained about 2.4% and FTO 0.6% of the variation (partial R^2) in BMI when adjusted for age, age^2, and eigenvectors, consistent with prior reports.

Sociodemographic characteristics such as age ($P = 0.75$), education ($P = 0.33$), and marital status ($P = 0.76$) were not significantly associated with the GRS. Self-reported prevalence of cancer ($P = 0.01$), type 2 diabetes ($P < 0.001$) across all waves, and smoking status in 1992 ($P = 0.03$) were significantly associated with the GRS of BMI in this population, as expected given the strong link between BMI and each of these outcomes. Self-reported coronary heart disease was associated with the GRS at the $\alpha = 10\%$ level ($P = 0.09$).

TABLE I. NHS Participants With Genotype and Depressive Score Information

Variable	Mean (SD) /%
n	6989
Age at Blood Draw	56.4 (6.91)
Body Mass Index*	24.96 (4.50)
Average Depression Score	1.81 (0.64)
GRS**	3.99 (0.53)
GRSexFTO	3.67 (0.46)
FTO (A)	41.7%
MC4R (A)	23.7%
Marital Status at Blood Draw	Missing (3.32%)
Married	80.97%
Divorced/Separated	6.82%
Widowed	8.89%
Education	
RN	94.39%
BA/MA/PhD	5.61%
Smoking	Missing (2.64%)
Never	42.21%
Past	37.39%
Current	17.76%
Cancer	9.7%
Type 2 Diabetes	15.98%
Coronary Heart Disease	8.21%

*Body Mass Index (kg/m^2) was assessed as mean of all measurements available prior to the assessment of depression.
**Weighted sum of the number of alleles (0,1, or 2) with its published effect estimate for all 32 SNPs that were significantly associated with obesity published by Speliotes et al.

Conventional Observational Association Between BMI and Depression

Self-reported BMI significantly predicted depression, adjusted for age, age^2, and genetic eigenvectors. Each one unit increase in BMI was associated with 0.024 (95%-CI: 0.020, 0.029) higher long-term depression (partial $R^2 = 2.9\%$).

IV Estimates of the Effect of BMI on Depression

The three alternative genetic IV estimates of the effect of BMI on depression were inconsistent (Fig. 3). When using FTO as a genetic IV (rs1558902), a 1-unit increase in BMI was non-significantly associated with an increase in depression of 0.064 (95%-CI -0.014, 0.142) points, about 3 times the effect size estimated in the conventional observational analysis regressing depression on measured BMI. The IV effect estimate based on MC4R (rs571312) was small and non-significant (beta = 0.005; 95%-CI -0.146, 0.156). The IV estimate based on the polygenic GRS was more precisely estimated but also non-significant (beta = 0.013; 95%-CI: -0.028, 0.055) and smaller than the observational effect estimate. The polygenic GRS that excluded FTO still explained 1.8%(partial R^2) of BMI; the IV effect estimate of this GRS*ex*FTO was slightly less than zero (beta = -0.003, 95%-CI -0.051, 0.045).

In addition, the GRS was split into components according to likely biologic mechanism linking each SNP to BMI. None of the mechanism-specific gene scores ("Appetite", "Adiposity", "Cardio-Pulmonary Factors", "Other/Unknown") was significantly associated with depression (Fig. 3).

The Wu–Hausman Test did not reject the null hypothesis that the IV effect estimates were equal to the conventional estimate for any of the IV models (Supplementary Table 5).

Evaluating Genetic IV Model Assumptions

When the IV model that used FTO as the instrument was adjusted for self-reported BMI, the IV coefficient between the instrument and depression was attenuated by 47% to 0.033 (95%-CI: -0.044, 0.11). The estimated association between

TABLE II. First Stage Results of the Association between the Genetic Instrument and Body Mass Index*,**

Instrument	beta	95%-CI	*P*-value	R^{2***}
Single Locus Genetic IVs				
FTO	1.286	0.769, 1.803	<0.001	0.006
MC4R	0.856	−0.140, 1.852	0.092	0.001
Polygenic Risk Score IVs				
GRS	1.303	1.032, 1.575	<0.001	0.024
GRS*ex*FTO	1.306	0.994, 1.619	<0.001	0.018
Mechanism Specific IVs				
Adipogenesis	1.465	0.499,2.431	<0.003	0.003
Appetite	1.305	0.985, 1.625	<0.001	0.015
Cardio-pulmonary	1.374	0.361,2.387	<0.008	0.002
Other/unknown	1.361	0.790,1.931	<0.001	0.005

*Body Mass Index was assessed as mean of all measurements available prior to the assessment of anxiety. NHS: 1976–1990 (8 measurements).
**Linear regression model additionally adjusted for age, age^2, and the top 3 eigenvectors to increase precision.
***Partial R2.

GRSexFTO and depression was -0.003, and this estimate changed to -0.037 (95%-CI: -0.084, 0.011) upon adjustment for self-reported BMI. The associations between "Appetite", "Adiposity", "Cardio-Pulmonary Factors" genetic IVs were also attenuated towards the null by adjustment for BMI (Supplementary Table S5). The association between "Other/Unknown" gene score and depression, however, was strengthened ($P = 0.073$) and remained inversely associated with depression (-0.076, 95%-CI: -0.160, 0.007) after adjustment for BMI. The over-identification test comparing the four mechanism-specific genetic risk scores did not reject ($P = 0.16$) equality of the estimated effects. The test for over-identification comparing IV estimates based on FTO to IV estimates based on the GRSexFTO rejected the null hypothesis of equality of estimated effects ($P = 0.014$). This result suggests that either FTO or the GRSexFTO is not a valid IV for the effect of BMI on depression. When using all 32 weighted SNPs separately in the model, the over-identification test was not rejected ($P = 0.06$).

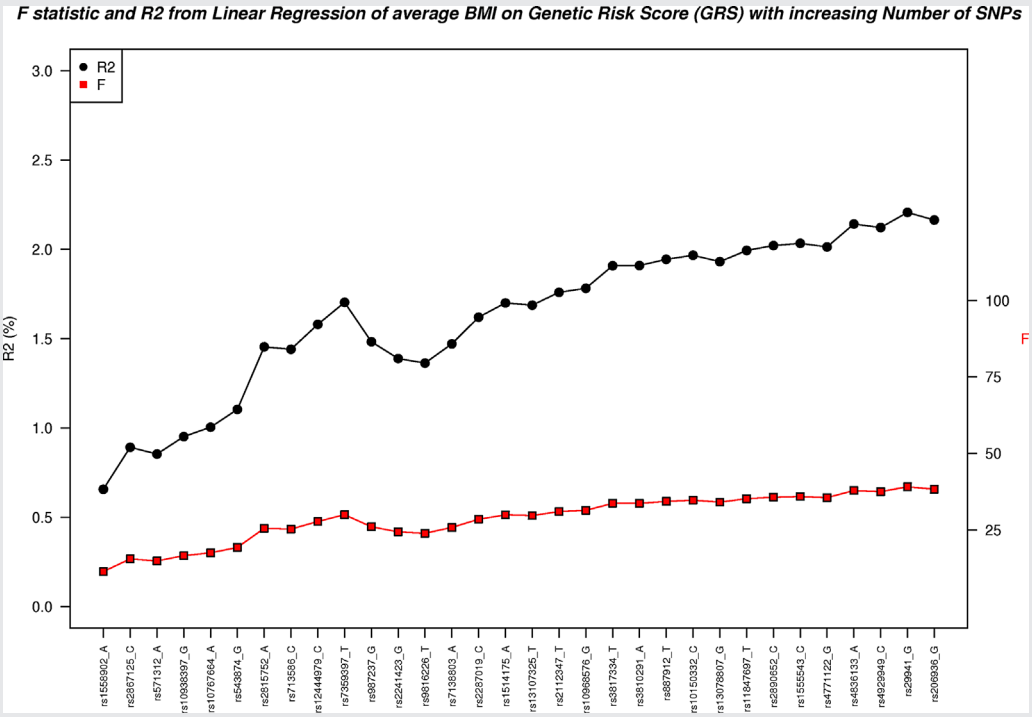

FIG. 2. Explained variation of body mass index by increasing number of single nucleotide polymorphism combined in the genetic risk score.

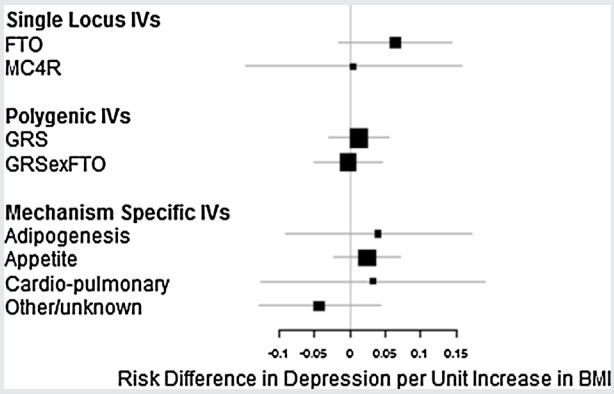

FIG. 3. Genetic IV estimates of the effect of BMI on Depression.

DISCUSSION

In a large sample of women, we found little evidence that genetically induced differences in BMI were associated with average depression as measured over a period of 14 years. The FTO polymorphism, associated with higher BMI, was non-significantly associated with higher depression, but this association was not fully attenuated by adjustment for BMI (as would be expected if FTO were a valid IV). This is evidence against collider bias given that the presence of collider bias should generally flip the sign of the associations in linear models; the modest attenuation of the FTO-depression association suggests the possibility of a violation of the exclusion restriction. Further, the IV effect estimate based on FTO was much larger than the conventional observational estimate, contrary to expectations that the conventional effect estimate is positively biased. The over-identification test comparing FTO to GRS*ex*FTO was rejected, implying that at least one of these is a biased IV. In total, this evidence suggests that FTO is not a valid instrument to estimate the causal effects of BMI on depression. None of the other IVs provided evidence for a causal effect of BMI on depression. Evidence from the "unknown mechanism" IV is most consistent with the hypothesis of strong positive confounding of the BMI-depression association. This mechanism-specific IV effect estimate was in fact below zero and lower than the conventional observational effect estimate. Our results suggest three conclusions: FTO alleles could directly affect depression via mechanisms independent of BMI; BMI itself has no direct effect on depression; or there are confounders—which may be genetic or environmental—that increase both BMI and depression.But the confidence intervals for all IV estimates were wide so it is not possible unequivocally confirm that there is no effect of BMI on depression.

Comparison With Other Studies

The results from this study build on prior findings from Kivimaki et al. in the Whitehall II study that reported FTO-induced BMI differences to be associated with psychological distress in men. [Kivimäki et al., 2011] Our empirical results are also consistent with some findings reported in a recent IV study applying a risk score based approach and using the same set of SNPs to investigate the association between BMI and depressive symptoms in 1731 participants of the Young Finns study. Findings from the Young Finns study were sensitive to the analytic approach used, but in the analyses that mirrored our approach (based on using external weights to construct the polygenic score, reported in supplementary information Table S2), they report no statistically significant IV effect estimate. Thus, these findings are similar to ours in that no convincing evidence against the null hypothesis to suggest that BMI influences depression was found. There are several additional differences between the two studies. For example, the Young Finn study contains men and women, while the NHS cohort contains only women. Differences due to sex specific effect of the SNPs on BMI and BMI on symptoms of depression cannot be ruled out. In addition, we also present analytic results that imply that FTO is a highly biased IV and further suggest a cautious interpretation of the results in the Young Finns and other studies particularly when FTO is included in the polygenic instrument. [Jokela et al., 2012]. The possibility of FTO violating the exclusion restriction is further supported by the recent finding of Samaan et al. who report a significant and protective association of rs9939609 A, an intronic SNP located in the FTO gene, and depression.[Samaan et al., 2013] Given that the direction of the FTO-depression association reported in the Samaan et al. paper is opposite to ours, additional caution is required when using FTO related variants either singly or in genetic risk scores in the genetic IV analysis of BMI and depression. It is important to note that this does not necessarily translate to studies with a different outcome. Richmond et al. did not find evidence of IV effect heterogeneity when analyzing the effect of BMI on physical activity. [Richmond et al., 2014]

Our study stands in largest contrast to findings from Lawlor et al., which suggested higher adiposity reduces mental distress, when instrumenting BMI and waist-to-hip ratio by FTO and MC4R. There is no obvious explanation for the discrepancy between our results and those reported in Lawlor et al. Plausible possibilities include differences in the outcome: the Lawlor study used 3 single-item questions on anxious state of mind and the use of anti-depressants. [Lawlor et al., 2011] Alternatively, there may be environmental differences in the population in the Lawlor study—Copenhagen residents aged 20+ in 2003—which modified the genotype-depression associations. This sample was primarily comprised of more recent birth cohorts than NHS and averaged higher BMI (26.2 in the Copenhagen sample versus 24.96 in NHS). In addition, these prior studies did not adjust for residual population stratification that might influence the study results in either direction.

Limitations

Our study population, although reweighted to represent the original study sample of 32,826 female nurses who provided a blood sample in 1990, is not representative of the general population. Our polygenic IV was not associated with age, education, or marital status providing evidence that the GRS can be considered a valid instrument plausibly mimicking a

randomized experiment. Type 2 diabetes and cancer were significantly associated with the GRS supporting the validity of the GRS, given that increases in BMI are expected to substantially increase risk of both type 2 diabetes and cancer, particularly in women. Similar to Jokela et al [Jokela et al., 2012] and Thorgeirsson et al [Thorgeirsson et al., 2013], we also observed a significant association of the GRS with smoking. This may reflect the effects of genetically induced increases in BMI on smoking, if smoking is a weight control strategy among individuals prone to overweight [Meyers et al., 1997] or if a common biological or genetic basis exists between obesity and nicotine addiction. [Thorgeirsson et al., 2013; Volkow et al., 2008]

Sample size is a limiting factor for Genetic IV studies. Even though the polygenic score used in this study is strongly predictive of BMI, it only accounts for a modest percent of the variation in the phenotype. This is reflected in the fact that the confidence interval for the IV effect estimate is consistent with both a small protective as well as a small hazardous effect. The benefit of the IV design is that these confidence bounds are unbiased under the IV assumptions. A larger sample size is required to further narrow the confidence intervals around the IV effect estimate.

Another limitation of the genetic IV analysis is that we assume linear effects of BMI on depressive symptoms. It might well be that the effects are different at the extremes of the BMI distribution, or there may even be thresholds for effects, such as obesity. Such non-linear effects would not invalidate our hypothesis test, because the GRS is associated with increases throughout the distribution of BMI (i.e., higher risk of being obese). However, if the effect of BMI on depression follows a completely non-linear model, the effect estimate from our MR analysis is not informative.

CONCLUSIONS

In conclusion, we present IV effect estimates and sensitivity analyses that suggest that FTO alleles possibly have direct effects on depression, not mediated by BMI. This implies prior evidence from IV estimates based on FTO should be reinterpreted. Our results with the much more powerful GRS found no convincing evidence against the null hypothesis to suggest that BMI influences depression. There is a need for more IV analyses excluding FTO from the set of instruments. Meta-analyzed effect estimates are likely to be most informative, given the limited statistical power in most Mendelian Randomization studies. Ideally, IV estimates based on genetic risk *scores excluding* FTO would provide inputs for the meta-analyses. Finally, we note that the mechanism-specific IV approach presented here, if linked with additional studies to improve statistical power, has the potential to elucidate heterogeneous effects of different types of adiposity. Such heterogeneity is biologically plausible and often assumed to occur in population samples, but otherwise difficult to evaluate in conventional observational studies with only simple measures of adiposity, such as BMI.

ACKNOWLEDGMENTS

We would like to thank the Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School for the effort in initiating and maintaining the NHS (P01 CA87969) and cohorts with the support of the aforementioned infrastructure grants. The authors also gratefully acknowledge funding from NIH/NIMH (MH092707-01). The funding institutions had no influence on design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript. Stefan Walter had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

REFERENCES

Angrist JD, Krueger AB. 1994. Split Sample Instrumental Variables. National Bureau of Economic Research Technical Paper 150.

Angrist JD, Krueger AB. 1995. Split-Sample Instrumental Variables Estimates of the Return to Schooling. J Bus Econ Stat 13 (2):225–235.

Brundtland GH. 2001. From the World Health Organization. Mental health: New understanding, new hope. JAMA 286(19):2391.

Colditz GA, Hankinson SE. 2005. The Nurses' Health Study: Lifestyle and health among women. Nat Rev Cancer 5(5):388–396.

Cornelis MC, Monda KL, Yu K, Paynter N, Azzato EM, Bennett SN, Berndt SI, Boerwinkle E, Chanock S, Chatterjee N, Couper D, Curhan G, Heiss G, Hu FB, Hunter DJ, Jacobs K, Jensen MK, Kraft P, Landi MT, Nettleton JA, Purdue MP, Rajaraman P, Rimm EB, Rose LM, Rothman N, Silverman D, Stolzenberg-Solomon R, Subar A, Yeager M, Chasman DI, van Dam RM, Caporaso NE. 2011. Genome-wide meta-analysis identifies regions on 7p21 (AHR) and 15q24 (CYP1A2) as determinants of habitual caffeine consumption. PLoS Genet 7(4):e1002033.

Chang S-C, Glymour MM, Walter S, Liang L, Koenen KC, Tchetgen EJ, Cornelis MC, Kawachi I, Rimm E, Kubzansky LD. 2014. Genome-wide polygenic scoring for a 14-year long-term average depression phenotype. Brain Behav 4(2):298–311.

Davey Smith, Ebrahim G. 2003. 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease?. Int J Epidemiol 32(1):1–22.

Glymour MM, Tchetgen EJ, Robins JM. 2012. Credible Mendelian randomization studies: Approaches for evaluating the instrumental variable assumptions. Am J Epidemiol 175(4):332–339.

Jokela M, Elovainio M, Keltikangas-Jarvinen L, Batty GD, Hintsanen M, Seppala I, Kahonen M, Viikari JS, Raitakari OT, Lehtimaki T, Kivimaki M. 2012. Body mass index and depressive symptoms: instrumental-variables regression with genetic risk score. Genes Brain Behav 11 (8): 942–948.

Kivimäki M, Jokela M, Hamer M, Geddes J, Ebmeier K, Kumari M, Singh-Manoux A, Hingorani A, Batty GD. 2011. Examining overweight and obesity as risk factors for common mental disorders using fat mass and obesity-sssociated (FTO) genotype-instrumented analysis. Am J Epidemiol 173(4):421–429.

Lawlor DA, Harbord RM, Tybjaerg Hansen, Palmer A, Zacho TM, Benn J, Timpson M, Davey NJ, Smith G, Nordestgaard BG. 2011. Using genetic loci to understand the relationship between adiposity and psychological distress: A mendelian randomization study in the copenhagen general population study of 53 221 adults. J Intern Med 269(5):525–537.

Luppino FS, de Wit LM, Bouvy PF, Stijnen T, Cuijpers P, Penninx BW, Zitman FG. 2010. Overweight, obesity, and depression: A systematic review and meta-analysis of longitudinal studies. Arch Gen Psychiatry 67(3):220–229.

Meyers AW, Klesges RC, Winders SE, Ward KD, Peterson BA, Eck LH. 1997. Are weight concerns predictive of smoking cessation? A prospective analysis. J Consult Clin Psychol 65(3):448–452.

Monsees GM, Tamimi RM, Kraft P. 2009. Genome-wide association scans for secondary traits using case-control samples. Genet Epidemiol 33(8): 717–728.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38(8):904–909.

Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P, Ruderfer DM, McQuillin A, Morris DW. 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460(7256):748–752.

Richmond RC, Davey Smith, Ness G, den Hoed AR, McMahon M, Timpson G. 2014. Assessing causality in the association between child adiposity and physical activity levels: a Mendelian randomization analysis. PLoS Med 11(3):e1001618.

Rimm EB, Stampfer MJ, Colditz GA, Chute CG, Litin LB, Willett WC. 1990. Validity of self-reported waist and hip circumferences in men and women. Epidemiology 1(6):466–473.

Samaan Z, Anand S, Zhang X, Desai D, Rivera M, Pare G, Thabane L, Xie C, Gerstein H, Engert JC, Craig I, Cohen-Woods S, Mohan V, Diaz R, Wang X, Liu L, Corre T, Preisig M, Kutalik Z, Bergmann S, Vollenweider P, Waeber G, Yusuf S, Meyre D. 2013. The protective effect of the obesity-associated rs9939609 A variant in fat mass- and obesity-associated gene on depression. Mol Psychiatry 18, 1281–1286.

Sharp LK, Lipsky MS. 2002. Screening for depression across the lifespan: A review of measures for use in primary care settings. Am Fam Physician 66(6):1001–1008.

Sheikh JI YJ. 1986. Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. In: Brink TL e, editor . Clinical Gerontology: A Guide to Assessment and Intervention. New York: The Haworth Press. p 165–173.

Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, Allen HL, Lindgren CM, Luan J, Mägi R. 2010. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 42(11):937–948.

Thorgeirsson TE, Gudbjartsson DF, Sulem P, Besenbacher S, Styrkarsdottir U, Thorleifsson G, Walters GB, Furberg H, Sullivan PF, Marchini J, McCarthy MI, Steinthorsdottir V, Thorsteinsdottir U, Stefansson K. 2013. A common biological basis of obesity and nicotine addiction. Transl Psychiatry 3:e308.

Volkow ND, Wang GJ, Fowler JS, Telang F. 2008. Overlapping neuronal circuits in addiction and obesity: Evidence of systems pathology. Philos Trans R Soc Lond B Biol Sci 363(1507):3191–3200.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article.