Pitfalls in performing genome-wide association studies on ratio traits

Zachary R McCaw,[1,*] Rounak Dey,[1] Hari Somineni,[1] David Amar,[1] Sumit Mukherjee,[1]

Kaitlin Sandor,[1] insitro Research Team,[1] Theofanis Karaletsos,[1] Daphne Koller,[1]

George Davey Smith,[2] Daniel MacArthur,[3,4] Colm O'Dushlaine,[1] Thomas W Soare[1,*]

Abstract

Genome-wide association studies (GWAS) are often performed on ratios composed of a numerator trait divided by a denominator trait. Examples include body mass index (BMI) and the waist-to-hip ratio, among many others. Explicitly or implicitly, the goal of forming the ratio is typically to adjust the numerator for the denominator. While forming ratios may be clinically expedient, there are several important issues with performing GWAS on ratios. Forming a ratio does not "adjust" for the denominator in the sense of holding it constant, and it is unclear whether associations with ratios are attributable to the numerator, the denominator, or both. Here we demonstrate that associations arising in ratio GWAS can be entirely denominator-driven, implying that at least some associations uncovered by ratio GWAS may be due solely to a putative adjustment variable. In a survey of 10 exemplar ratios, we find that the ratio model disagrees with the adjusted model (performing GWAS on the numerator while conditioning on the denominator) at around 1/3 of loci. Using BMI as an example, we show that variants detected by only the ratio model are more strongly associated with the denominator (height), while variants detected by only the adjusted model are more strongly associated with the numerator (weight). Although the adjusted model provides effect sizes with a clearer interpretation, it is susceptible to collider bias. We propose and validate a simple method of correcting for the genetic collider bias via leave-one-chromosome-out polygenic scoring.

Keywords: Genome-wide association study; Polygenic scores; Proportion; Ratio; Body mass index; Collider bias

[1] Insitro, South San Francisco, CA, USA.

[2] MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK.

[3] Centre for Population Genomics, Garvan Institute of Medical Research, Sydney, New South Wales, AU.

[4] Centre for Population Genomics, Murdoch Children's Research Institute, Melbourne, Victoria, AU.
* Correspondence: ZRM (zmccaw@insitro.com), TWS (tsoare@insitro.com).

Introduction

Ratio traits, those composed of a numerator trait divided by a denominator trait, are widely used in clinical practice, where they provide convenient scalar indicators of health and disease [1]. Examples include body mass index (BMI) in obesity [2]; forced expiratory volume (FEV1) to forced vital capacity (FVC) in chronic obstructive pulmonary disease (COPD) [3] and asthma [4]; phenylalanine to tyrosine concentrations in phenylalanine hydroxylase deficiency [5]; left ventricular ejection fraction (LVEF) in heart failure [6] and dilated cardiomyopathy [7]; serum aspartate (AST) to alanine aminotransferase (ALT) in hepatic cirrhosis [8]; and vertical cup-to-disc ratio (VCDR) in glaucoma [9]. As biomarkers of health, ratio traits have been used as targets for many genome-wide association studies (GWAS) since the first applications of the method (e.g. BMI [10], serum metabolite ratios [11], WHR [12]), a practice that continues to the present (e.g. adipose tissue volumes [13], AST-ALT ratio [14], blood cell proportions [15], BMI [16], LVEF [17], metabolite ratios [18, 19], protein ratios [20], seated to standing height [21, 22], skeletal proportions [23], urinary albumin-creatinine ratio (UACR) [24], VCDR [25, 26], WHR [27]). In such ratio trait GWAS, the explicit or implicit goal is generally to uncover the genetic underpinnings of the numerator while controlling for the denominator.

However, the use of ratio variables in regression models has several well-known statistical issues [28]. Importantly, performing GWAS on a ratio does not actually test for a direct effect of genotype on the numerator while holding the denominator constant. Moreover, the genetic effect estimated by the ratio model is difficult to interpret because the association may be due to the numerator alone, the denominator alone, or both the numerator and denominator. An alternative to the ratio model is to condition on the denominator as a covariate, which we refer to as the "adjusted model." This model correctly estimates the direct effect but can introduce collider bias [29, 30].

Here we demonstrate several pitfalls that can arise when performing GWAS of ratio traits. As a running example, we consider BMI, the ratio of weight in kilograms to height^2 in

meters^2. Through simulations and analyses of real data in which the numerator (i.e. weight) is permuted, we demonstrate that associations with ratio traits can be entirely denominator-driven, and that the probability of detecting such associations increases with denominator heritability. In a survey of 10 exemplar ratio traits, we find that the ratio and adjusted models lead to discrepant conclusions at roughly 1/3 of loci, and that associations detected by only the ratio model are enriched for denominator-driven signals. Intuitively, variants identified by only the ratio model tend to be more strongly associated with height (the denominator), while variants identified by only the adjusted model tend to be more strongly associated with weight (the numerator). We scrutinize the practice of correlating effect sizes from two different models (e.g. ratio and adjusted) fit to the same data set, and show that doing so tends to overstate the agreement of the models. Finally, we address the issue of heritable covariate bias and propose a method of correcting for collider bias that arises from common genetic causes.

Results

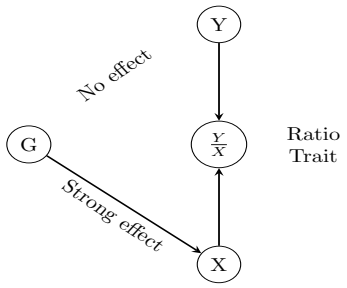
Ratio model associations can be entirely denominator-driven

Figure 1(a) presents a simple schematic of how a genetic variant can influence a ratio via an effect on the denominator only. To empirically demonstrate that associations obtained from ratio GWAS can be entirely denominator-driven, we conducted analyses of a null phenotype, permuted body weight, within the the UK Biobank (UKB; $N = 356$K) [31, 32]. Permutation keeps the mean and standard deviation (SD) of weight unchanged but abolishes any association with genotype. **Figure 1(c)** compares Manhattan plots for GWAS of weight (unpermuted), height, permuted weight adjusted for height and height^2 (the "adjusted model"), and the ratio of permuted weight to height^2 (the "ratio model"). As desired, the adjusted model detects no genome-wide significant (GWS) loci ($P \leq 5 \times 10^{-8}$). By contrast, the ratio model detects 142 independent GWS loci, and exhibits clear similarity to the marginal Manhattan plots for height. Of these 142 GWS loci, 140 (98.6%) were GWS for height, and the remaining 2 were suggestively significant ($P \leq 1 \times 10^{-6}$). Because the numerator was

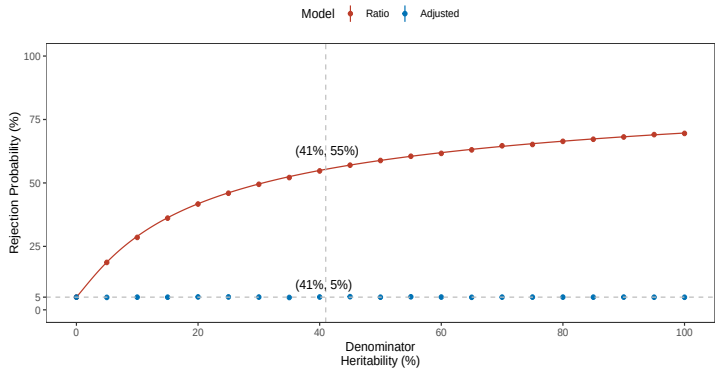
permuted, and consequently had no association with genotype, the signals detected by the ratio model in this experiment must be entirely attributable to the denominator.

Propensity to detect denominator-driven associations increases with denominator heritability

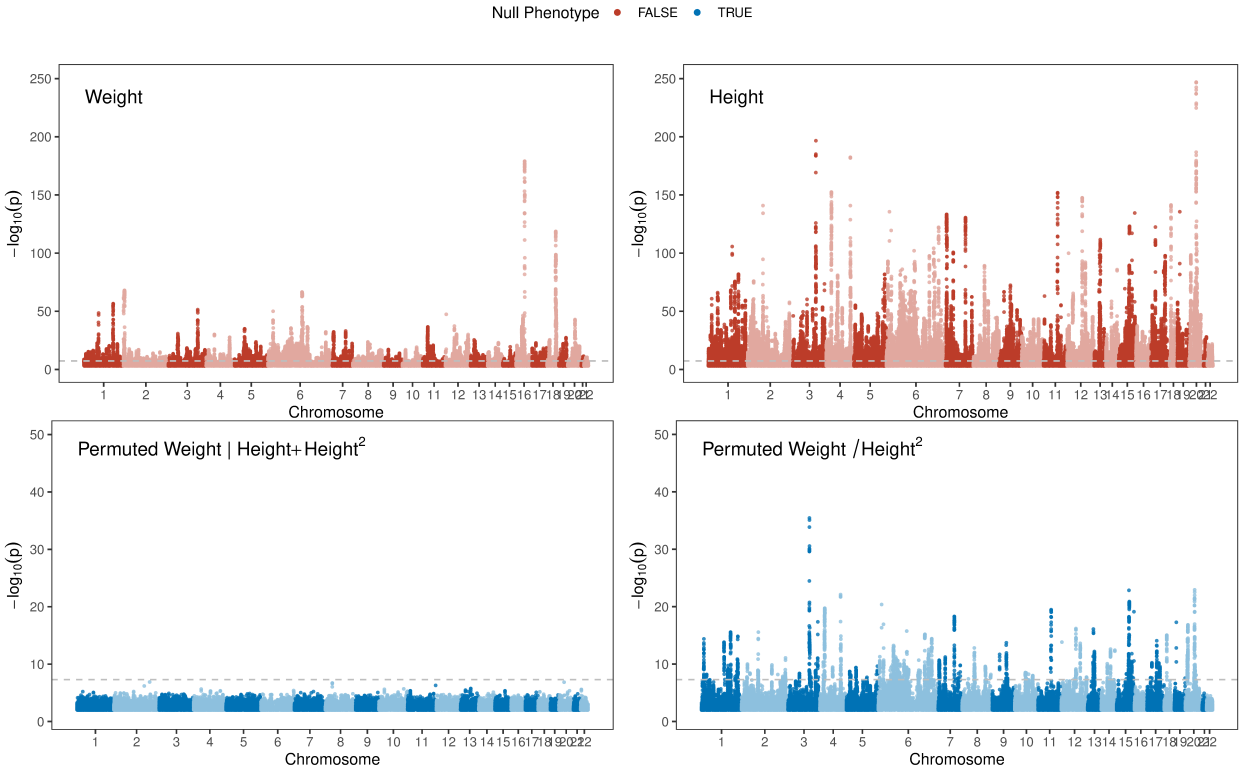
To study the effect of numerator heritability on the probability of observing an association with a ratio phenotype, a random sample of 10K subjects was drawn from the UKB. The numerator was permuted body weight. To enable control of the heritability, the denominator was simulated from an infinitesimal model in such a way that the mean and SD matched the observed values for height^2 in UKB (Supplemental Methods). **Figure 1(b)** tracks the rejection probability for the ratio and adjusted models. The probability that the adjusted model detects an association remains flat at the type I error rate of 5%, which is expected given that the numerator is unassociated with genotype. By contrast, the probability that the ratio model detects an association grows in concert with the heritability of the denominator (representing height^2). The solid red line is a theoretical calculation of the expected rejection probability as a function of denominator h^2 (Supplemental Methods). At the empirical heritability for height, estimated by linkage-disequilibrium (LD) score regression [33] in the full cohort, the ratio model detected an association with 55% probability. This experiment demonstrates that the risk of detecting denominator-driven associations increases with the denominator heritability. The interpretation of associations detected by the ratio model is further complicated by the fact that the number of association detected depends on the mean of the numerator, even when the numerator is null (**Figure S1**).



(a) **Variants can affect a ratio via the denominator only.**



(b) **Probability of detecting an association increasing with denominator heritability.**



(c) **Ratio GWAS of permuted weight over height^2 detects loci for height.**

Figure 1: **Simulated and empirical evidence that associations detected in ratio GWAS can be entirely denominator-driven.** (a) Example data generating process in which a variant G affects a ratio phenotype Y/X by influencing the denominator X only. (b) Simulation study of rejection probability as a function of denominator heritability. The numerator was permuted weight, and the denominator was simulated to control the heritability while matching the mean and variance to the observed values for height^2. (c) Manhattan plots for GWAS conducted in the UK Biobank ($N = 356$K).

Ratio and adjusted models can reach discrepant conclusions

We surveyed the extent to which ratio traits are used in the literature by querying the NHGRI-EBI GWAS Catalog [34] for trait names including: concentration, fraction, index, percent(age), percentile, proportion, rate, and ratio; as well as specific traits known to be ratios (e.g. BMI, WHR). We manually reviewed the 667 traits identified for evidence of having a heritable denominator, excluding, for example, per-volume concentrations or per-time rates. Using this approach, we ascertained that 362 traits, representing 3.2% of all traits and 7.8% of all reported associations in the GWAS Catalog, involved ratios with a heritable denominator (Supplementary Data).

To understand the extent to which the ratio and adjusted models differ in real data, we conducted two GWAS within the UKB on each of 10 phenotypes analyzed as ratios: one with the ratio as the outcome and the other with the numerator as the outcome, adjusting for the denominator (the "adjusted" analysis). The numbers of genome-wide significant (GWS) loci identified by the ratio and adjusted models are presented in **Table 1**. The union of loci detected by the ratio and adjusted models was partitioned into 3 sets (ratio-only, adjusted-only, or both) based on whether or not the lead variant from each ratio locus was within 250kb of the lead variant for an adjusted locus, and vice versa. On average the ratio model detected 386 loci while the adjusted model detected 292 loci. That the ratio model tends to identify more loci is expected given that a variant may associate with the ratio either through the numerator or the denominator. On average, 67.6% of all loci detected were detected by both models, 22.8% were unique to the ratio model, and 9.6% were unique to the adjusted model. Thus, although the ratio and adjusted models are generally in agreement, for 32.6% of loci the two analyses may reach differing conclusions. We also assessed the percentage of signals from the ratio and adjusted models that were within 250kb of the lead variant for the denominator trait, but not within 250kb of the lead variant for the numerator trait ("% denominator-driven"). On average, 48% of ratio-only signals were shared with the denominator trait, while only 8% of adjusted-only signals were shared with the denominator trait. Thus loci detected by only the ratio model are enriched for denominator-driven associations.

Traits			GWS Loci			
Numerator	Denominator	N	Ratio	Adjusted	Ratio-only (% Denominator-driven)	Adjusted-only (% Denominator-driven)
Albumin (blood)	Creatinine (blood)	311,122	944	556	727 (71%)	314 (0%)
Albumin (urine)	Creatinine (urine)	105,501	6	4	6 (0%)	4 (0%)
Weight	Height^2	355,627	844	830	89 (34%)	70 (4%)
FEV_1	FVC	325,220	538	474	86 (13%)	49 (2%)
Stroke volume	End diastolic volume	29,171	6	0	6 (0%)	0 (.)
Phenylalanine	Tyrosine	200,159	251	174	54 (56%)	10 (0%)
Platelets	Lymphocytes	52,133	2	0	2 (0%)	0 (.)
Seated height	Standing height	355,373	440	115	330 (45%)	39 (8%)
Trunk fat mass	Whole body fat mass	349,533	359	240	132 (0%)	46 (0%)
Waist circumference	Hip circumference	356,112	465	529	116 (1%)	115 (37%)
	Average	243,995	385.5	292.2	154.8 (48%)	64.7 (8%)

Table 1: **Survey of discrepancies between the ratio and adjusted models across traits analyzed as ratios.** Number of independent genome-wide significant loci detected by the ratio and adjusted models, categorized by whether the signal was detected by the ratio model only or the adjusted model only. Loci genome-wide significant for the denominator, but not the numerator, were considered "denominator-driven." Loci were obtained by clumping full-genome summary statistics at $R^2 \leq 0.10$ and $P \leq 5 \times 10^{-8}$.

Effect size correlation overstates concordance of ratio and adjusted models

Several studies of ratio phenotypes have performed sensitivity analyses in which the effect sizes from a ratio model are compared with those from an unadjusted model [10, 12, 17, 13] or an adjusted model [23]. Comparisons among effect sizes from different models fit to the same data should be interpreted with caution. As demonstrated analytically in the Supplemental Methods, such estimates are inherently correlated due to having been obtained from the same subjects. To enable an unbiased comparison of effect sizes across models, we split the full UKB cohort ($N = 356$K) into two equally sized subsets, A and B ($N = 178$K each). Subset B serves as the validation cohort for associations discovered in subset A and vice versa. Importantly, effect sizes estimated from models fit in A can fairly be correlated with effect sizes estimated in B.

Figure S3 presents the effect size correlations between BMI and a comparator GWAS where the effect sizes were either estimated in the same or in different (split) samples.

The comparators include marginal GWAS of height and weight, as well as GWAS of weight adjusted for height (and height^2). Although correlation remains, the magnitude of correlation declines appreciably when the effect sizes are measured in independent samples. For example, in the case of weight adjusted for height, the correlation with BMI drops from 97.8% to 81.1%.

Ratio associations are enriched for denominator signal

Comparing the ratio and adjusted models enables variants to be partitioned into 3 classes: variants detected by the ratio model only ("ratio-only"), variants detected by the adjusted model only ("adjusted-only"), and variants detected by both ("both"). To understand the biological significance of these three groups of variants, in **Figure S4** we explore the enrichment of signal from marginal associations, calculated in the full UKB cohort, for the numerator (weight) and denominator (height) traits at these variants. Intuitively, the mean χ^2 of height was larger at ratio-only loci (standard T-test, $T = 3.21$; $p = 0.0015$), as compared with adjusted-only loci, whereas the mean χ^2 for weight was larger at adjusted-only loci ($T = 3.19$; $p = 0.0017$), as compared to ratio-only loci.

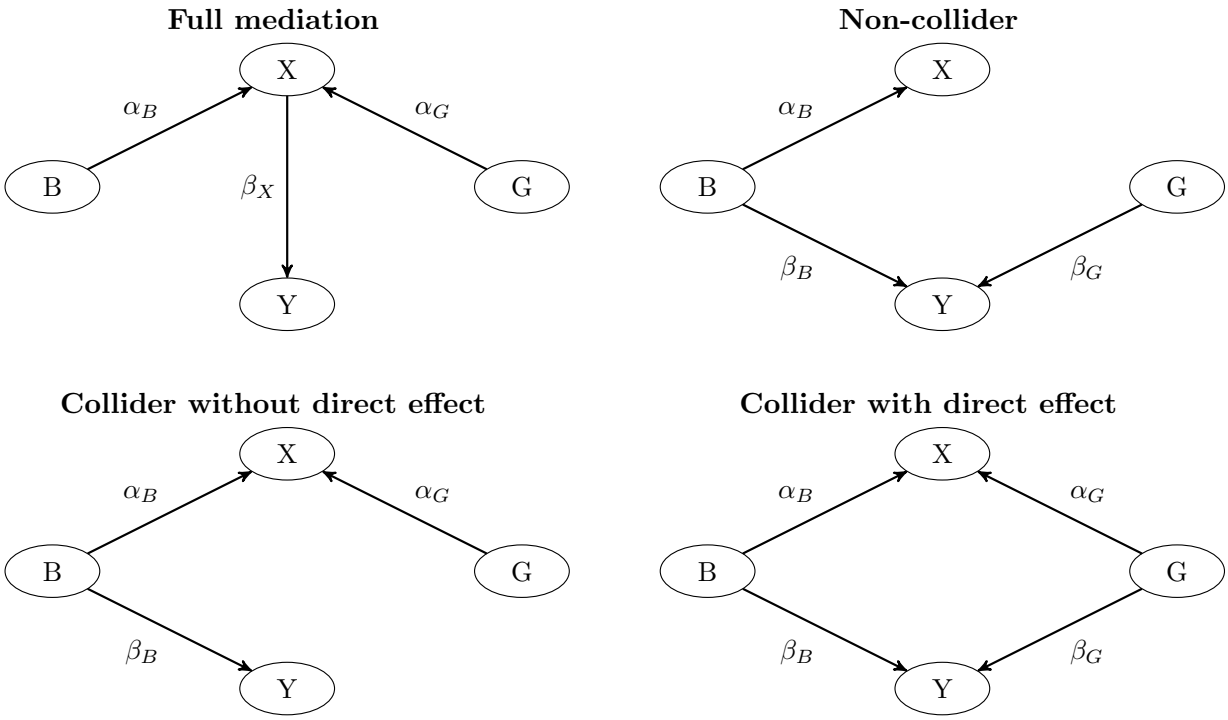
We also compared the power of the ratio and adjusted models at GWS loci for obesity-related traits compiled from external sources (excluding the UKB cohort where possible): body fat percentage [35], chronically-elevated alanine aminotransferase (cALT) [36], coronary artery disease (CAD) [37], hip circumference [38], low-density lipoprotein (LDL) cholesterol [39]), and type 2 diabetes (T2D) [40]. The adjusted model had larger χ^2 statistics at variants defined by several obesity-relevant traits, including body fat percentage (paired T-test, $T = 2.42$; $p = 0.033$), hip circumference ($T = 8.4$; $p = 4.3 \times 10^{-13}$), and LDL cholesterol ($T = 2.67$; $p = 0.0074$, **Table S4**).

Leave-one-chromosome-out polygenic scoring can partially correct for heritable covariate bias

The denominators of the ratios considered in **Table 1** are themselves heritable (**Table S3**). Important work by [29, 30] underscores that adjusting for a heritable covariate can introduce

collider bias. Consider the data generating processes represented in **Figure 3(a)**. Here Y is the phenotype, G is genotype, X is a covariate, and B represents background common causes of X and Y. Supposing B has an effect on both X ($\alpha_B \neq 0$) and Y ($\beta_B \neq 0$), if G has an effect on X ($\alpha_G \neq 0$), then conditioning on X opens a backdoor path [41] between G and Y, namely $Y \leftarrow B \rightarrow \boxed{\text{X}} \leftarrow G$. The presence of this backdoor path biases estimation of β_G, the genetic effect of interest. In cases where G has no effect on X ($\alpha_G = 0$), X is no longer a collider and the bias vanishes. Supposing $\alpha_G \neq 0$, another way to remove collider bias would be to condition on B. When B is genetic background, this can be achieved by performing conditional GWAS for Y adjusting for all variants that affect X. In practice, adjusting for all variants affecting X may be impractical. Instead, we pursue the strategy of conditioning on a polygenic score (PGS) for X, effectively blocking the path from $B \rightarrow X$.

(a) **Data generating processes**



(b) **Bias simulations**

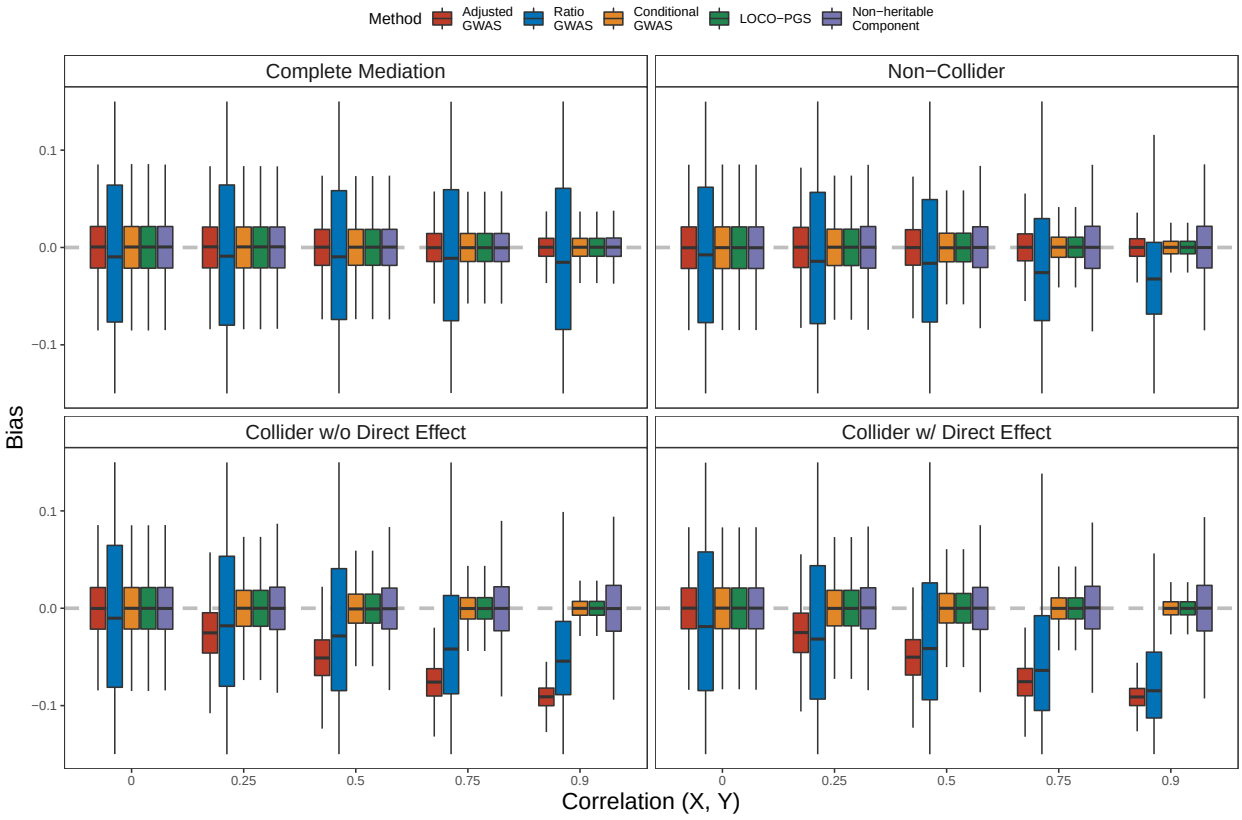


Figure 3: **Leave-one-chromosome-out polygenic scoring corrects for collider bias due to the genetic background.** (a) Data generating processes considered in the simulation study. In each case, B is genetic background, G is genotype, X is a heritable covariate, and Y is the phenotype of interest. (b) Bias in estimating β_G as a function of the correlation between X and Y due to B. The sample size is 10^3 and each box represents 10^4 simulations replicates. Adjusted and ratio are susceptible to bias; conditional is unbiased and theoretically optimal; LOCO-PGS and non-heritable component are the proposed bias-correction methods.

Drawing on ideas from BOLT-LMM [42] and Regenie [43], we propose constructing a leave-one-chromosome-out (LOCO) PGS for the heritable covariate X, then performing GWAS adjusting for both X and its LOCO-PGS. A related strategy is to regress the LOCO-PGS out of the heritable covariate then adjust for the residual "non-heritable" component (in practice, the residual component may retain some heritability if the PGS does not capture all genetic contributions to the heritable covariate [44]). In either case, the PGS serves to correct for collider bias by breaking the association between the genetic background and the heritable covariate. In the Supplemental Methods, we demonstrate analytically that excluding variants in LD with the focal variant is necessary (i.e. adjusting for a global rather than a LOCO PGS would introduce bias), and that LOCO-PGS adjustment successfully removes collider bias due to the genetic background. **Figure 3(b)** empirically assess the bias of adjusted, ratio, and conditional GWAS alongside the proposed methods of correcting for genetic background under each data generating process in **Figure 3(a)**. Adjusted GWAS is unbiased when the heritable covariate X fully mediates the effects of genotype G and genetic background B on the phenotype, or when genotype has no effect on X, but incurs bias when X is a collider. Conditional GWAS is the optimal strategy of controlling collider bias, but requires conditioning on B. Both LOCO-PGS and non-heritable covariate adjustment remove bias due to the genetic background, but LOCO-PGS is more efficient (i.e. produces smaller standard errors). Ratio GWAS is often biased, and inefficient, because the association model is misspecified. Specifically, genotype affects the numerator and/or denominator individually, not as a ratio.

Figure 4 compares validation effect sizes from several models with discovery effect size from the adjusted model using real data from the UKB. Correlation of the adjusted effect sizes with those for height is minimal, while correlation with effect sizes for weight and BMI is high. The similarity of the effect sizes from the adjusted, LOCO-PGS, and non-heritable component models suggests either that most variants detected by the adjusted model are not subject to substantial collider bias, or that the collider bias is due primarily to factors other than genetic background, for which the LOCO-PGS cannot account.

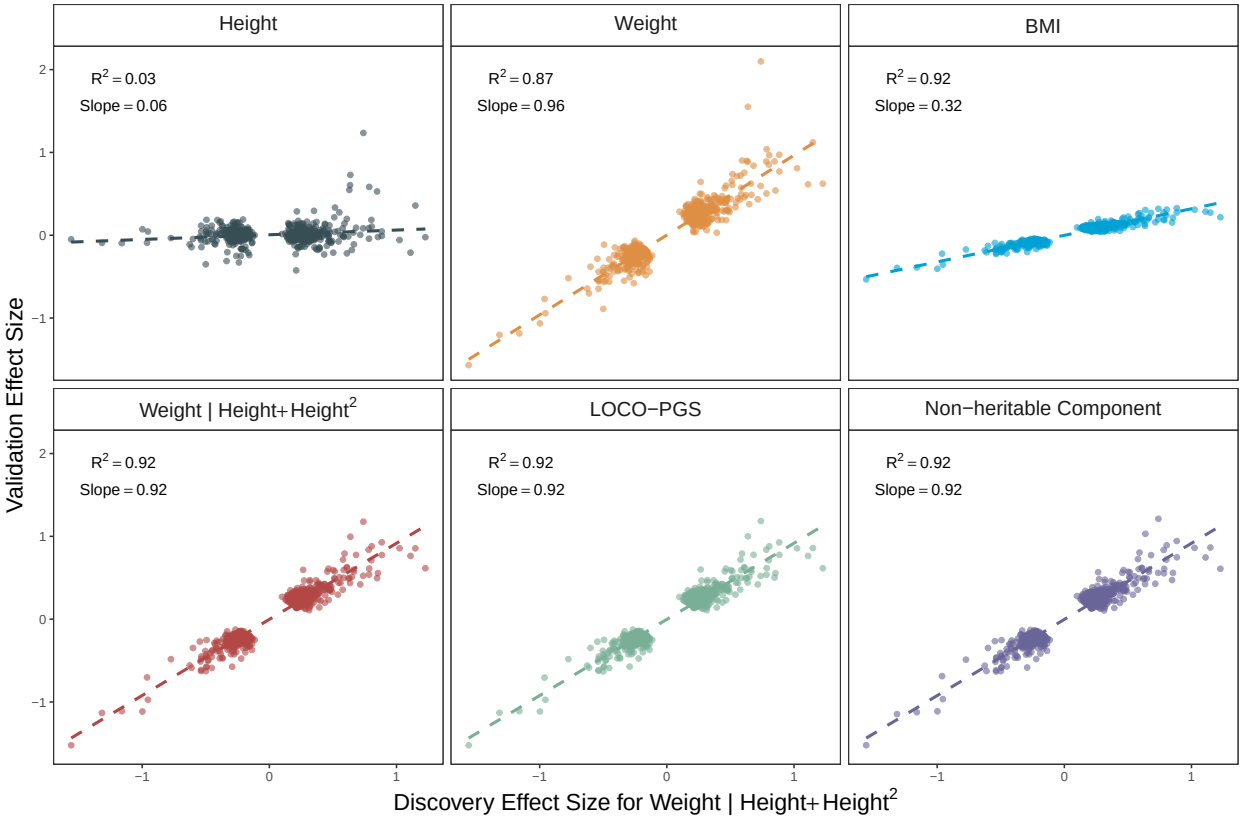


Figure 4: **Empirically, effect sizes from the adjusted and LOCO-PGS analyses are similar.** Validation effect sizes from several GWAS are compared with the discovery effect size from the adjusted model, and independent ($R^2 \leq 0.1$) genome-wide significant ($P \leq 5 \times 10^{-8}$) loci for the latter. The discovery and validation effect sizes were estimated in two independent subsets of the UK Biobank ($N = 178$K each).

Discussion

Ratio traits are useful heuristics in clinical practice, but introduce statistical challenges when used in regression models [28]. GWAS on ratio traits can uncover associations entirely driven by the denominator, and it is not possible to tell from the summary statistics for a ratio analysis alone which associations are likely denominator-driven. While many associations detected in ratio GWAS do in fact affect the numerator, it is worth scrutinizing the results of ratio GWAS to understand how those associations arose. We recommend comparing the effect sizes from ratio GWAS with those from unadjusted (marginal) GWAS of the numerator and denominator, and from the adjusted model. Variants not strongly associated with the denominator that have similar effect sizes in the unadjusted and adjusted analyses are likely numerator-driven. Conversely, variants strongly associated with the denominator that display markedly different effect sizes in the adjusted and unadjusted analyses may be denominator-driven.

Pleiotropic effects on the numerators and denominators of ratio traits appear widespread. In the survey of discrepancies between the ratio and adjusted models, two-thirds of loci were detected by both models (**Table S1**), and the genetic correlation between the numerator and denominator traits was often substantial (**Table S2**). Thus, not all of the associations identified by ratio GWAS are erroneous. Instead, the conclusion is that at least some associations identified by ratio GWAS may be due to the denominator. If the goal of a ratio GWAS is to detect variants associated with the numerator while accounting for the denominator, then an adjusted analysis may be more appropriate. If the goal is to detect pleiotropic variants, i.e. those affecting both the numerator and the denominator, then a multivariate GWAS may outperform either taking the intersection of individual GWAS or performing ratio GWAS [45, 46, 47]. A more recent approach to disentangling pleiotropic from trait-specific effects is genomic structural equation modeling [48].

Among the traits considered in the survey of discrepancies between the ratio and adjusted models, waist-to-hip ratio (WHR) is an outlier. Whereas for other traits, the ratio model

detected more associations (as expected), for WHR the adjusted model detected more associations. Moreover, for WHR only, more of the adjusted-only associations than the ratio-only associations were putatively denominator-driven. Possible explanations for this include pleiotropy between waist and hip circumference (using our summary statistics, the genetic correlation was estimated at 0.88 by LDSC; **Table S2**), and the potential for strong environmental effects on both waist and hip circumference (e.g. diet). While unraveling the distinctive behavior of WHR is not the focus of the present work, this does provide an exemplar trait where the adjusted model is likely inappropriate and a multivariate analysis would be preferred.

Recognizing that the purpose of BMI is to provide an index of adiposity that is independent of height, but that the ratio of weight to height^2 is not always optimal, Stergiakouli *et al* considered performing GWAS on BMI[x] = weight/height^x, where the exponent x was selected in an age-dependent manner to minimize phenotypic correlation between BMI[x] and height [49]. Although this is a step in the right direction, this strategy is unlikely to fully resolve the issues with performing GWAS on ratios. Selecting x to minimize the phenotypic correlation of BMI[x] with height does not guarantee that the final correlation will be exactly zero, and even if zero correlation could be achieved, this does not imply that BMI[x] and height will be free of higher-order dependencies. Absent the ability to select $x \neq 0$ such that BMI[x] and height are completely independent, it remains possible for variants to associate with BMI[x] via height.

When the denominator is heritable, adjusting for the denominator as a covariate can introduce collider bias [29, 30]. It is worth noting that adjusting for a heritable covariate does not automatically result in collider bias. As shown here and discussed previously [29], there are causal architectures that do not incur collider bias, such as when the effect of background is fully mediated through the heritable covariate. For settings where collider bias is expected, we proposed additionally adjusting for a LOCO-PGS for the heritable covariate. To further control for effects of genetic background on numerator and denominator, this approach could be extended by including a random effect with covariance

proportional to the genetic relatedness matrix [50, 42, 51]. When performing GWAS of weight adjusted for height, we found that addition of a LOCO-PGS had little effect on the estimated effect sizes, suggesting that if collider bias is present, it is primarily due to factors other than genetic background. A limitation of the LOCO-PGS approach is that it cannot adjust for collider bias introduced by non-genetic common causes, such as the environment. Methods for addressing collider bias due to potentially unobserved factors include mtCOJO [52], which employs Mendelian randomization, and SlopeHunter [53], which utilizes variants associated with the heritable covariate to estimate a correction factor subsequently applied to all variants. Another limitation is that LOCO-PGS does not, by default, account for collider bias due to variants on the same chromosome. While the bias introduced by variants on the same chromosome will often be minimal, in cases where it is substantial, this bias can be removed by combining LOCO-PGS with conditional GWAS. Specifically, the association model would adjust for the heritable covaraite, its LOCO-PGS, and any variants on the same chromosome with strong effects on the covariate.

With increasingly high-dimensional phenotypic data, such as that provided by metabolomics and proteomics panels, there is a temptation to conduct hypothesis-free association testing on the ratios of all possible pairs of variables [54, 55]. In the metabolic context, a rationale for considering ratios is that the ratio of product to substrate may capture information about binding affinity or reaction rate [11, 55, 18, 56]. While GWAS of ratios may be justified in some situations, we suggest that the scientific hypothesis be carefully elaborated before a ratio phenotype is selected. The fact that the denominator is likely related to the numerator is not sufficient justification for using a ratio phenotype as the relationship between numerator and denominator can be modeled in other ways. If potential collider bias precludes adjusting for the denominator as a covariate, then a multivariate analysis that separately estimates the effect of genotype on the numerator and denominator will typically provide greater interpretability. Notably, the hypotheses that genotype is associated with both the numerator and denominator, or with one or the other, can all be evaluated within a multivariate framework [57]. Performing GWAS on ratios is generally admissible when the denominator is a non-heritable normalizing factor, such as volume in the case of

concentration measurements (e.g. HbA1c) or time in the case of certain rates (e.g. 24-hour urinary albumin excretion rate). However, even here it should be considered whether treating the denominator as an offset or covariate is not more appropriate given the other covariates in the model [28].

In conclusion, GWAS of ratio traits can identify associations that are entirely denominator-driven, and when the rationale for forming the ratio was to adjust for the denominator, such associations may be considered false positives. We recommend reanalyzing ratio GWAS in order to discern whether the associations are attributable to the numerator, the denominator, or both. Downstream analyses based on summary statistics from ratio GWAS (e.g. predictions based on polygenic scores, causal effect estimates based on Mendelian randomization) may also require reconsideration.

Methods

GWAS catalog analysis

We downloaded all studies in the NHGRI-EBI GWAS Catalog [34] (accessed 07-APR-2023) and searched for the following terms in the "MAPPED_TRAIT" field of the studies table: concentration, fraction, index, percent, percentile, proportion, rate, ratio, percentage; BMI, chronic kidney disease, chronic obstructive pulmonary disease, CKD, COPD, and WHR. This returned a collection of 667 traits, which we manually reviewed for evidence of GWAS having been performed on a ratio outcome with a heritable denominator, examining the title, abstract, and study as necessary. The list of 362 ratio traits is provided as Supplementary Data.

GWAS

GWAS among unrelated White-British subjects from the UK Biobank [31, 32] were performed using `PLINK` (v1.9) [58]. We filtered to imputed genotypes with minor allele frequency $> 0.1\%$, INFO score > 0.8, and Hardy-Weinberg equilibrium $P > 1 \times 10^{-10}$. Samples were filtered to those used in the genetic PCA calculation [32]; self-reported "White-British", "White", or "Irish" ancestry; and without sex chromosome aneuploidy. For GWAS, samples with phenotype values in the bottom 0.01 or top 99.99 percentiles were removed. To facilitate direct comparison of effect sizes, all models included a standard set of covariates, namely age, sex, genotyping array, and the top 20 genetic principal components.

Clumping

Loci independent at $R^2 \leq 0.1$ and genome-wide significant (GWS) at $P \leq 5 \times 10^{-8}$ loci were identified by applying `PLINK`'s clumping functionality to genome-wide summary statistics with a window size of 250kb. Any reference to independent, GWS loci or clumping refers to clumping with these parameters, unless otherwise noted.

Effect size correlations

To obtain unbiased estimates of effect size correlation, the full UKB cohort ($N = 356$K) was randomly split into two independent subsets of size $N = 178$K, labeled A and B. Subset A was arbitrarily designated the discovery cohort and subset B the validation cohort. GWAS was performed in both cohorts, and correlations were estimated across cohorts, restricting to variants that were independent and GWS in the discovery cohort. As sensitivity analyses, we explored taking B as the discovery cohort and A as the validation cohort, as well as generating additional random splits. Results from these analyses were qualitatively similar to the results presented.

Variant group definition

The full set of loci detected by the ratio and adjusted models was partitioned into 3 sets (ratio-only, adjusted-only, or both) based on whether or not the lead variant from each ratio locus was within 250kb of the lead variant for an adjusted locus, and vice versa. Variants having a suggestive association with height ($p < 1 \times 10^{-6}$) were excluded from downstream analysis. Summary statistics from GWAS on obesity-related traits, excluding the UKB cohort where possible, were downloaded from GWAS Catalog, dbGaP, or directly from consortia websites. The summary statistics were lifted over to GRCh38 where needed, and independent, GWS loci were identified by clumping $R^2 \leq 0.5$.

Leave-one-chromosome-out polygenic scores

Leave-one-chromosome-out (LOCO) polygenic scoring was performed within the split-sample framework. First, GWAS for height was performed in each of subsets A and B. Using weights from the independent, GWS variants for height from subset B, a whole-genome polygenic score (PGS) S was calculated for each subject in subset A and vice versa for B. Next, for each chromosome k, a LOCO-PGS S_k was formed by removing from S the contributions of all variants on chromosome k. This resulted in 22 LOCO-PGSs per subject, one for each autosome (our GWAS were confined to autosomes). Finally, GWAS was performed,

separately for each chromosome, using the association model:

$$Y = G\beta_G + X\beta_X + S_k\beta_S + \varepsilon,$$

where G is genotype, X is a set of covariates including height, and S_k is the LOCO-PGS for the chromosome k on which G resides. For the non-heritable component analysis, height was regressed on the LOCO-PGS to obtain residuals $H_k^{\perp}$, then GWAS was performed, separately for each chromosome, according to the model:

$$Y = G\beta_G + X_{-H}\beta_X + H_k^{\perp}\beta_H + \varepsilon.$$

Here the covariate set X_{-H} now excludes height, and $H_k^{\perp}$ is formed using the S_k for the chromosome on which G resides.

Acknowledgements

The authors would like thank the participants of the UK Biobank, whose data were used with permission. This research was conducted using the UK Biobank Resource under approved Application Number 51766.

Author Contributions

ZRM, HS, SM, and TWS conceived of the project. RD performed theoretical derivations. ZRM, HS, DA, SM, and TWS performed analyses. KS reviewed the list of ratio traits from GWAS catalog. All authors provided scientific input. DK, GDS, DM, and CO provided early feedback and direction on the manuscript. ZRM and TWS wrote the first draft of the manuscript. All authors contributed to critical revision of the final manuscript.

insitro Research Team Banner and Contribution Statements

All contributors are listed in alphabetical order.

Downloading, preprocessing, and curation of UK Biobank data

- Francesco Paolo Casale,[1,2,3,4] Eilon Sharon,[1] Thomas W. Soare,[1] James Warren[1].

Software engineering

- **Statistical genetics pipeline**: Magdalena Borecka,[1] Francesco Paolo Casale,[1,2,3,4] Anna Merkoulovitch,[1] Colm O'Dushlaine,[1] Thomas W. Soare,[1] Paul Sud,[1] Baris Ungun[1].
- **redun development**: Robin Betz,[1] Edward Chee,[1] Patrick R. Conrad,[1] Kevin Ford,[1] Christoph Klein,[1] Donald Naegely,[1] Matthew Rasmussen[1].

Affiliations:

[1] insitro, South San Francisco, CA, USA.

[2] Institute of AI for Health, Helmholtz Munich, Neuherberg, Germany.

[3] Helmholtz Pioneer Campus, Helmholtz, Munich, Neuherberg, Germany.

[4] School of Computation, Information and Technology, Technical University of Munich, Munich, Germany.

Declaration of Interests

ZRM, RD, HS, DA, SM, KS, members of the insitro Research Team, TK, DK, CO, and TWS are current or former employees and shareholders of Insitro. GDS and DM are advisors to Insitro.

References

[1] JS Garrow and J Webster. Quetelet's index (w/h2) as a measure of fatness. *Int J Obes*, 9(2):147 – 153, 1985.

[2] AJ Orzano and JG Scott. Diagnosis and treatment of obesity in adults: An applied evidence-based review. *The Journal of the American Board of Family Medicine*, 17(5):359 – 369, 2004.

[3] J Vestbo, SS Hurd, AG Agusti, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: Gold executive summary. *Am J Respir Crit Care Med*, 187(4):347 – 365, 2013.

[4] Global Initiative for Asthma. Global strategy for asthma management and prevention.

[5] J Vockley, HC Andersson, KM Antshel, et al. Phenylalanine hydroxylase deficiency: diagnosis and management guideline. *Genetics in Medicine*, 16(2):188 – 200, 2014.

[6] Writing Committee Members, CW Yancy, M Jessup, et al. 2013 accf/aha guideline for the management of heart failure: a report of the american college of cardiology foundation/american heart association task force on practice guidelines. *Circulation*, 128(16):e240 – 327, 2013.

[7] EM McNally and L Mestroni. Dilated cardiomyopathy: genetic determinants and mechanisms. *Circ Res*, 121:731 – 748, 2017.

[8] ALB Williams and Hoofnagle JH. Ratio of serum aspartate to alanine aminotransferase in chronic hepatitis relationship to cirrhosis. *Gastroenterology*, 95(3):734 – 739, 1988.

[9] DF Garway-Heath, ST Ruben, A Viswanathan, and RA Hitchings. Vertical cup/disc ratio in relation to optic disc size: its value in the assessment of the glaucoma suspect. *British Journal of Ophthalmology*, 82(10):1118 – 1124, 1998.

[10] RJ Loos, CM Lindgren, S Li, et al. Common variants near mc4r are associated with fat mass, weight and risk of obesity. *Nat Genet*, 40(6):768 – 775, 2008.

[11] T Illig, C Gieger, G Zhai, et al. A genome-wide perspective of genetic variation in human metabolism. *Nat Genet*, 42:137 – 141, 2010.

[12] CM Lindgren, IM Heid, JC Randall, et al. Genome-wide association scan meta-analysis identifies three loci influencing adiposity and fat distribution. *PLOS Genetics*, 5(6):e1000508, 2009.

[13] S Agrawal, M Wang, MDR Klarqvist, et al. Inherited basis of visceral, abdominal subcutaneous and gluteofemoral fat depots. *Nat Commun*, 13:3771, 2022.

[14] N Sinnott-Armstrong, Y Tanigawa, D Amar, et al. Genetics of 35 blood and urine biomarkers in the uk biobank. *Nat Genet*, 53(2):185 – 194, 2021.

[15] DeWan AT Kachuri L, Jeon S et al. Genetic determinants of blood-cell traits influence susceptibility to childhood acute lymphoblastic leukemia. *Am J Hum Genet*, 108(10):1823 – 1835, 2021.

[16] L Yengo, J Sidorenko, KE Kemper, et al. Meta-analysis of genome-wide association studies for height and body mass index in 700000 individuals of european ancestry. *Hum Mol Genet*, 27(20):3641 – 3649, 2018.

[17] JP Pirruccello, A Bick, M Wang, et al. Analysis of cardiac magnetic resonance imaging in 36,000 individuals yields genetic insights into dilated cardiomyopathy. *Nat Commun*, 11(1):2254, 2020.

[18] Y Cheng, P Schlosser, J Hertel, et al. Rare genetic variants affecting urine metabolite levels link population variation to inborn errors of metabolism. *Nat Commun*, 12:964, 2021.

[19] Y Chen, T Lu, U Pettersson-Kymmer, et al. Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases. *Nat Genet*, 55(1):44 – 53, 2023.

[20] K Suhre. Genetic associations with ratios between protein levels detect new pqtls and reveal protein-protein interactions. *medRxiv*, 2023.

[21] Y Chan, RM Salem, YH Hsu, et al. Genome-wide analysis of body proportion classifies height-associated variants by mechanism of action and implicates genes important for skeletal development. *Am J Hum Genet*, 96(5):695 – 708, 2015.

[22] E Bartell, K Lin, K Tsuo, et al. Genetics of skeletal proportions in two different populations. *bioRxiv*, 2023.

[23] E Kun, EM Javan, O Smith, et al. The genetic architecture and evolution of the human skeletal form. *Science*, 381:eadf8009, 2023.

[24] A Teumer, Y Li, S Ghasemi, et al. Genome-wide association meta-analyses and fine-mapping elucidate pathways influencing albuminuria. *Nat Commun*, 10:4130, 2019.

[25] Alipanahi B, Hormozdiari F, B Behsaz, et al. Large-scale machine-learning-based phenotyping significantly improves genomic discovery for optic nerve head morphology. *Am J Hum Genet*, 108(7):1217 – 1230, 2021.

[26] X Han, P Gharahkhani, AR Hamel, et al. Large-scale multitrait genome-wide association analyses identify hundreds of glaucoma risk loci. *Nat Genet*, 55:1116 – 1125, 2023.

[27] GT Hansen, DR Sobreira, ZT Weber, et al. Genetics of sexually dimorphic adipose distribution in humans. *Nat Genet*, 55:461 – 470, 2023.

[28] RA Kronmal. Spurious correlation and the fallacy of the ratio standard revisited. *JRSSA*, 156(3):379 – 392, 1993.

[29] H Aschard, Bj Vilhjalmsson, AD Joshi, AL Price, and P Kraft. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am J Hum Genet*, 96(2):329 – 339, 2015.

[30] FR Day, PR Loh, RA Scott, KK Ong, and JR Perry. A robust example of collider bias in a genetic association study. *Am J Hum Genet*, 98(2):392 – 393, 2016.

[31] C Sudlow, J Gallacher, N Allen, 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*, 12(3):e1001779, 2015.

[32] C Bycroft, C Freeman, D Petkova, et al. The uk biobank resource with deep phenotyping and genomic data. *Nature*, 562:203 – 209, 2018.

[33] BK Bulik-Sullivan, PR Loh, HK Finucane, et al. Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, 47(3):291 – 295, 2015.

[34] E Sollis, A Mosaku, A Abid, et al. The nhgri-ebi gwas catalog: knowledgebase and deposition resource. *Nucleic Acids Res*, 51(D1):D977 – D985, 2023.

[35] Y Lu, FR Day, S Gustafsson, et al. New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nat Commun*, 7:10495, 2016.

[36] M Vujkovic, S Ramdas, KM Lorenz, et al. A multiancestry genome-wide association study of unexplained chronic alt elevation as a proxy for nonalcoholic fatty liver disease with histological and radiological validation. *Nat Genet*, 54:761 – 771, 2022.

[37] CP Nelson, A Goel, AS Butterworth, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet*, 49(9):1385 – 1391, 2017.

[38] D Shungin, TW Winkler, DC Croteau-Chonka, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature*, 518(7538):187 – 196, 2015.

[39] SE Graham, SL Clarke, KHH Wu, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature*, 600:675 – 679, 2021.

[40] A Mahajan, CN Spracklen, W Zhang, et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat Genet*, 54(5):560 – 572, 2022.

[41] J Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669 – 688, 1995.

[42] PR Loh, G Tucker, BK Bulik-Sullivan, et al. Efficient bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet*, 47(3):284 – 290, 2015.

[43] J Mbatchou, L Barnard, J Backman, et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet*, 53:1097 – 1103, 2008.

[44] K Silventoinen, H Lahtinen, GD Smith, TT Morris, and P Martikainen. Height, social position and coronary heart disease incidence: the contribution of genetic and environmental factors. *J Epidemiol Community Health*, 77(6):384 – 390, 2023.

[45] MA Ferreira and SM Purcell. A multivariate test of association. *Bioinformatics*, 25(1):132 – 133, 2009.

[46] PF O'Reilly, CJ Hoggart, Y Pomyen, et al. Multiphen: joint model of multiple phenotypes can increase discovery in gwas. *PLoS One*, 7(5):e34861, 2012.

[47] X Zhou and M Stephens. Efficient algorithms for multivariate linear mixed models in genome-wide association studies. *Nat Methods*, 11(4):407 – 409, 2014.

[48] AD Grotzinger, M Rhemtulla, R de Vlaming, et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat Hum Behav*, 3:513 – 525, 2019.

[49] E Stergiakouli, R Gaillard, JM Tavare, et al. Genome-wide association study of height-adjusted bmi in childhood identifies functional variant in adcy3. *Obesity*, 22(10):2252 – 2259, 2014.

[50] J Yang, NA Zaitlen, ME Goddard, PM Visscher, and AL Price. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet*, 46(2):100 – 106, 2014.

[51] H Chen, C Wang, MP Conomos, et al. Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *Am J Hum Genet*, 98(4):653 – 666, 2016.

[52] Z Zhu, Z Zheng, F Zhang, et al. Causal associations between risk factors and common diseases inferred from gwas summary data. *Nat Commun*, 9(1):224, 2018.

[53] O Mahmoud, F Dudbridge, GD Smith, and M Munafo. A robust method for collider bias correction in conditional genome-wide association studies. *Nat Commun*, 13(1):10.1038/s41467–022–28119–9, 2022.

[54] AK Petersen, J Krumsiek, B Wägele, et al. On the hypothesis-free testing of metabolite ratios in genome-wide and metabolome-wide association studies. *BMC Bioinformatics*, 13:120, 2012.

[55] K Suhre and C Gieger. Genetic variation in metabolic phenotypes: study designs and applications. *Nat Rev Genet*, 13:759 – 769, 2012.

[56] K Suhre, MI McCarthy, and JM Schwenk. Genetics meets proteomics: perspectives for large population-based studies. *Nat Rev Genet*, 22:19 – 37, 2021.

[57] G Seber. *The Linear Model and Hypothesis*. Springer, 2015.

[58] S Purcell, B Neale, K Todd-Brown, et al. Plink: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 81(3):559 – 575, 2007.

Pitfalls in performing genome-wide association studies on ratio traits

Contents

1 Supplemental Methods **2**

1.1 Null numerator simulation 2

1.2 Theoretical analysis of rejection probability versus denominator heritability . 3

1.3 Effect size correlations 6

1.4 Leave-one-chromosome-out polygenic scoring 7

1.4.1 Generative model 7

1.4.2 Adjusted analysis 8

1.4.3 Model with a polygenic score 10

1.4.4 Case 1: PGS constructed only using variants not in LD with G . . . 11

1.4.5 Case 2: global PGS constructed based on all variants 12

2 Supplemental Results **12**

2.1 Null numerator experiment 12

2.2 Survey of ratio traits 15

2.3 Effect size correlations 17

2.4 Biological relevance of ratio and adjusted models 18

2.5 Leave-one-chromosome-out polygenic scoring 20

1 Supplemental Methods

1.1 Null numerator simulation

Genotypes for $N = 10^4$ independent subjects were simulated at $J = 10^2$ variants in linkage equilibrium. Each variant G_j was standardized to have mean 0 and variance 1. A base denominator phenotype, with heritability h^2, was simulated from an infinitesimal model:

$$Y_{h^2} = \sum_{j=1}^{J} G_j \beta_j + \epsilon,$$

where the effect sizes β_j were simulated as independent random effects $\beta_j \sim N(0, h^2/J)$, and the residual was simulated as $\epsilon \sim N(0, 1 - h^2)$. In this way:

$$\mathbb{V}(Y_{h^2}) = \sum_{j=1}^{J} \mathbb{V}(G_j)\mathbb{V}(\beta_j) + \mathbb{V}(\epsilon) = \sum_{j=1}^{J} 1 \cdot \left(\frac{h^2}{J}\right) + (1 - h^2) = h^2.$$

Finally the base phenotype Y was shifted by the observed mean μ_{obs} and scaled by the observed standard deviation σ_{obs} of height^2:

$$H^*_{h^2} = \sigma_{\text{obs}} Y_{h^2} + \mu_{\text{obs}}.$$

Across simulations, the mean and SD of BMI were 27.8 and 6.5 km/m^2, which are comparable with the mean and SD of BMI in the full cohort (27.4 and 4.8 km/m^2 respectively).

The numerator phenotype was permuted weight W, with the mean and variance left at their observed values. For ratio GWAS, $W/H^*_{h^2}$ was regressed on genotype G_j, while for the adjusted model, W was regressed on G_j adjusting for $H^*_{h^2}$. The rejection probability at $\alpha = 0.05$ was calculated as:

$$\mathbb{P}(\text{rejection}) = \frac{1}{R} \sum_{r=1}^{R} \sum_{j=1}^{J} \mathbb{I}(p_j \leq 0.05),$$

where p_j is the p-value for association of G_j with the outcome and R is the number of

simulation replicates.

1.2 Theoretical analysis of rejection probability versus denominator heritability

Suppose we have the following generative models for the two traits $Y^{(1)}$ (numerator) and $Y^{(2)}$ (denominator), where $Y^{(1)}$ is a null trait, meaning that it has no association with genotype, and $Y^{(2)}$ is affected by p independent causal variants $Z_1, \ldots, Z_p$.

$$\begin{aligned} \text{(Numerator)} \quad & Y^{(1)} = \mu_1 1 + \epsilon_1, \quad \epsilon_1 \sim N(0, \sigma_1^2 I) \\ \text{(Denominator)} \quad & Y^{(2)} = \mu_2 1 + \sum_{j=1}^{p} Z_j \alpha_j + \epsilon_2, \\ & \alpha_j \overset{\text{i.i.d}}{\sim} N(0, N\tau^2/p), \quad \epsilon \sim N(0, \sigma^2 I), \end{aligned} \tag{1}$$

where I denotes the identity matrix (of dimension same as the sample-size N), μ_1, μ_2 are fixed scalars, ϵ_1, ϵ_2 are independent errors, and 1 is the vector of all ones. The genotypes are centered, i.e., $Z_i^\top 1 = 0$, and scaled, i.e., $Z_i^\top Z_i = 1$ for all $i = 1, \ldots, p$. The narrow-sense heritability is then defined as $h^2 = \tau^2/(\sigma^2 + \tau^2) = \tau^2/\sigma_2^2$ where $\sigma_2^2 = \sigma^2 + \tau^2$.

Now, without loss of generality, let us assume we are testing the first causal SNP $G = Z_1$ for association with the ratio trait $Y = Y^{(1)}/Y^{(2)}$ based on the linear model $Y = G\beta + \epsilon_Y$. The squared Wald test statistic for testing $H_0 : \beta = 0$ vs $H_1 : \beta \neq 0$ will be $S^2 = \hat{\beta}^2/\hat{\mathbb{V}}(\hat{\beta})$. Then, the following result provides the asymptotic distribution of S^2.

Result 1. *Assuming the number of causal SNPs (p) for the denominator trait, and sample-size (N) to be large, asymptotically* $S^2 \sim (1+\eta)\chi_1^2$*, where,*

$$\eta = \frac{\left(1 + \frac{3\sigma_2^2}{\mu_2^2}\right)^2 \gamma \sigma_2^2}{\mu_2^2 \left[\frac{\sigma_1^2}{\mu_1^2} + \frac{\sigma_2^2}{\mu_2^2}\right]} h^2, \quad \gamma = \lim_{N,p \to \infty} \frac{N}{p}.$$

This result implies that the test will have inflated chi-square statistics for variants that are causal to the denominator trait, and the asymptotic limit of the inflation factor is $(1 + \eta)$.

In addition, when the mean of the numerator trait $\mu_1 \to 0$, then $\eta \to 0$ and therefore the test will have nominal rejection probability when $\mu_1 = 0$, as shown in **Figures 1(b) and S1**.

Proof. First, we re-parametrize the generative model for $Y^{(2)}$ as following,

$$Y^{(2)} = \mu_2 1 + G\alpha + \epsilon_2, \qquad \alpha = \alpha_1, \qquad \epsilon_2 = \sum_{j=2}^{p} Z_j \alpha_j + \epsilon,$$

where $\mathbb{E}(\epsilon_2) = 0$ and $\mathbb{V}(\epsilon_2) = (\sigma^2 + \frac{p-1}{p}\tau^2)I \approx \sigma_2^2 I$. Then, using the second order Taylor series expansion around $Y_i^{(1)} = \mu_1$ and $Y_i^{(2)} = \mu_2 + G_i\alpha$ (with the index i denotes the element corresponding to the i-th subject), we have,

$$\begin{aligned}
\mu_{Y_i} = \mathbb{E}(Y_i) &\approx \frac{\mu_1}{\mu_2 + G_i\alpha}\left[1 + \frac{\sigma_2^2}{(\mu_2 + G_i\alpha)^2}\right] \\
\sigma_{Y_i}^2 = \mathbb{V}(Y_i) &\approx \frac{\mu_1^2}{(\mu_2 + G_i\alpha)^2}\left[\frac{\sigma_1^2}{\mu_1^2} + \frac{\sigma_2^2}{(\mu_2 + G_i\alpha)^2}\right].
\end{aligned}$$

Assuming $G_i\alpha = O_p(1/\sqrt{p})$ where p is large, we can further apply Taylor series expansion of μ_{Y_i} and $\sigma_{Y_i}^2$ around $G_i\alpha = 0$ and ignore terms that are $O_p(1/p)$,

$$\begin{aligned}
\mu_{Y_i} &\approx \frac{\mu_1}{\mu_2}\left[1 + \frac{\sigma_2^2}{\mu_2^2} - \frac{1}{\mu_2}\left(1 + \frac{3\sigma_2^2}{\mu_2^2}\right) G_i\alpha\right] \\
\sigma_{Y_i}^2 &\approx \frac{\mu_1^2}{\mu_2^2}\left[\frac{\sigma_1^2}{\mu_1^2} + \frac{\sigma_2^2}{\mu_2^2} - \frac{2}{\mu_2}\left(\frac{\sigma_1^2}{\mu_1^2} + \frac{2\sigma_2^2}{\mu_2^2}\right) G_i\alpha\right]
\end{aligned} \tag{2}$$

Since G is centered and scaled ($G^\top G = 1$), a linear regression of Y on G will yield the effect-size estimate and the estimate of its variance

$$\hat{\beta} = G^\top Y, \qquad \hat{\mathbb{V}}(\hat{\beta}) = \frac{1}{N-2} Y^\top (I - P_{[1:G]}) Y,$$

where P_A denotes the projection matrix of the matrix A, and $[1 : G]$ denotes the matrix that includes as columns the ones vector and the genotype vector G.

Let us denote Σ_Y to be a diagonal matrix with i-th diagonal element equal to $\sigma_{Y_i}^2$. We can

then show that,

$$\begin{aligned}
\hat{\mathbb{V}}(\hat{\beta}) \xrightarrow{p} \lim_{N\to\infty} E\left(\hat{\mathbb{V}}(\hat{\beta})\right) &= \lim_{N\to\infty} \frac{1}{N-2}\left[\mathrm{tr}((I-P_{[1:G]})\Sigma_Y) + \mu_Y^\top (I-P_{[1:G]})\mu_Y\right] \\
&= \lim_{N\to\infty} \frac{1}{N-2}\mathrm{tr}((I-P_{[1:G]})\Sigma_Y) \\
&= \lim_{N\to\infty} \frac{1}{N-2}\mathrm{tr}\left[(I-\frac{1}{N}11^\top - GG^\top)\Sigma_Y\right] \\
&= \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\sigma_{Y_i}^2 = \frac{\mu_1^2}{\mu_2^2}\left[\frac{\sigma_1^2}{\mu_1^2}+\frac{\sigma_2^2}{\mu_2^2}\right],
\end{aligned}$$

The second step follows from the fact that μ_Y can be written as $\mu_Y = a + bG$ for some constants a and b (2), and hence $(I - P_{[1:G]})\mu_Y = 0$. Now, using (1) and (2),

$$\begin{aligned}
\mathbb{E}(\hat{\beta}) &= \mathbb{E}\left(\mathbb{E}(\hat{\beta}|\alpha)\right) = \mathbb{E}\left[-\frac{\mu_1}{\mu_2^2}\left(1+\frac{3\sigma_2^2}{\mu_2^2}\right)\alpha\right] = 0, \\
\mathbb{V}(\hat{\beta}) &= \mathbb{E}\left(\mathbb{V}(\hat{\beta}|\alpha)\right) + \mathbb{V}\left(\mathbb{E}(\hat{\beta}|\alpha)\right) \\
&= \mathbb{E}\left[\sum_{i=1}^{N} G_i^2\sigma_{Y_i}^2\right] + \mathbb{V}\left[-\frac{\mu_1}{\mu_2^2}\left(1+\frac{3\sigma_2^2}{\mu_2^2}\right)\alpha\right] \\
&= \frac{\mu_1^2}{\mu_2^2}\left[\frac{\sigma_1^2}{\mu_1^2}+\frac{\sigma_2^2}{\mu_2^2}\right] - \mathbb{E}\left[\frac{2\mu_1^2}{\mu_2^3}\left(\frac{\sigma_1^2}{\mu_1^2}+\frac{2\sigma_2^2}{\mu_2^2}\right)\alpha\sum_{i]1}^{N} G_i^3\right] + \frac{\mu_1^2}{\mu_2^4}\left(1+\frac{3\sigma_2^2}{\mu_2^2}\right)^2\frac{N\tau^2}{p} \\
&= \frac{\mu_1^2}{\mu_2^2}\left[\frac{\sigma_1^2}{\mu_1^2}+\frac{\sigma_2^2}{\mu_2^2}\right]\left(1+\frac{\left(1+\frac{3\sigma_2^2}{\mu_2^2}\right)^2\frac{N\tau^2}{p}}{\mu_2^2\left[\frac{\sigma_1^2}{\mu_1^2}+\frac{\sigma_2^2}{\mu_2^2}\right]}\right).
\end{aligned}$$

Therefore, $|\hat{\mathbb{V}}(\hat{\beta}) - \mathbb{V}(\hat{\beta})/(1+\eta)| \xrightarrow{p} 0$. By replacing the $\hat{\mathbb{V}}(\hat{\beta})$ with its in-probability limit in the expression of the Z-score, we complete the proof,

$$S^2 = \frac{\hat{\beta}^2}{\hat{\mathbb{V}}(\hat{\beta})} \approx \frac{\hat{\beta}^2}{\mathbb{V}(\hat{\beta})}(1+\eta) \sim (1+\eta)\chi^2 \quad \text{asymptotically.}$$

□

1.3 Effect size correlations

Consider performing GWAS for two related phenotypes Y_1 and Y_2 via the following models:

$$Y_1 = G\alpha_G + X_1\alpha_X + \epsilon_1, \qquad Y_2 = G\beta_G + X_2\beta_X + \epsilon_2. \tag{3}$$

Here G is genotype at a particular variant, X_1 and X_2 are two sets of covariates, not necessarily including the same variables, and ϵ_1 and ϵ_2 are residuals. To simplify subsequent analysis, we can remove dependence on covariates by regressing X_k out of the association model for Y_k, $k \in \{1, 2\}$. Define the projection matrices:

$$Q_1 = I - X_1(X_1^T X_1)^{-1} X_1^T, \qquad Q_2 = I - X_2(X_2^T X_2)^{-1} X_2^T.$$

The genetic effect sizes (α_G, β_G) from (3) are consistently estimated via the models:

$$e_{Y_1} = e_{G_1}\alpha_G + \varepsilon_1, \qquad e_{Y_2} = e_{G_2}\beta_G + \varepsilon_2, \tag{4}$$

where $e_{Y_k} = Q_k Y_k$ and $e_{G_k} = Q_k G$. That is, e_{Y_k} is the residual after regressing Y_k on X_k and e_{G_k} is the residual after regressing G on X_k. Now, the ordinary least squares (OLS) estimates of α_G and β_G are:

$$\hat{\alpha}_G = (e_{G_1}^T e_{G_1})^{-1} e_{G_1}^T e_{Y_1}, \qquad \hat{\beta}_G = (e_{G_2}^T e_{G_2})^{-1} e_{G_2}^T e_{Y_2}. \tag{5}$$

Upon substitution of (4) into (5):

$$\hat{\alpha}_G = \alpha_G + (e_{G_1}^T e_{G_1})^{-1} e_{G_1}^T \varepsilon_1, \qquad \hat{\beta}_G = \beta_G + (e_{G_2}^T e_{G_2})^{-1} e_{G_2}^T \varepsilon_2$$

Taking the covariance of the estimated effect sizes:

$$\mathbb{C}(\hat{\alpha}_G, \hat{\beta}_G) = \mathbb{C}(\alpha_G, \beta_G) + (e_{G_1}^T e_{G_1})^{-1} e_{G_1}^T \mathbb{C}(\varepsilon_1, \varepsilon_2) e_{G_2} (e_{G_2}^T e_{G_2})^{-1}.$$

Observe that covariance between the estimated effect sizes has two sources: covariance of the true effect sizes $\mathbb{C}(\alpha_G, \beta_G)$ plus a term that depends on the covariance of residuals $\mathbb{C}(\varepsilon_1, \varepsilon_2)$.

When $\hat{\alpha}_G$ and $\hat{\beta}_G$ are estimated using independent subjects, then Y_1 and Y_2 are independent, as are the residuals ϵ_1 and ϵ_2 in (3), and by extension the residuals ε_1 and ε_2 in the reduced model (4). Thus, in the setting of independent data sets $\mathbb{C}(\hat{\alpha}_G, \hat{\beta}_G)$ provides an unbiased estimate of the true effect size covariance $\mathbb{C}(\alpha_G, \beta_G)$. However, when Y_1 and Y_2 are measured on the same subjects, then in general ϵ_1 and ϵ_2 are dependent, and $\mathbb{C}(\varepsilon_1, \varepsilon_2) \neq 0$. Thus, when the same data are utilized to estimate both $\hat{\alpha}_G$ and $\hat{\beta}_G$, $\mathbb{C}(\hat{\alpha}_G, \hat{\beta}_G)$ is biased as an estimate of $\mathbb{C}(\alpha_G, \beta_G)$ with bias depending on $\mathbb{C}(\varepsilon_1, \varepsilon_2)$, which in turn depends on the covariance between the initial phenotypes, Y_1 and Y_2.

1.4 Leave-one-chromosome-out polygenic scoring

In this section, we analytically show the following:

1. The adjusted model $Y \sim G + X$ is subject to bias when there exist background variants that affect both X and Y.

2. This bias can be removed by adjusting for a polygenic score (PGS) for X constructed from variants <u>not</u> in linkage disequilibrium (LD) with the variant of interest G.

3. Including G and variants in LD with G in the PGS S results in bias.

1.4.1 Generative model

Consider the following generative models for the numerator (Y) and the denominator (X) traits,

$$\begin{aligned} Y &= G\beta_{GY} + \sum Z_i \beta_{Yi} + \epsilon_Y, \\ X &= G\beta_{GX} + \sum Z_i \beta_{Xi} + \epsilon_X. \end{aligned} \tag{6}$$

Here, G is the variant of interest, Z_i-s are the other variants affecting the numerator and/or denominator, and $\epsilon_X \sim N(0, \sigma_X^2 I_N), \epsilon_Y \sim N(0, \sigma_Y^2 I_N)$ are independent errors. For notational simplicity, we assume G and Z_i-s are all centered and scaled such that $G^\top G = Z_i^\top Z_i = N$

for all i. We further assume that:

$$Z_i = G\alpha_i + \epsilon_i.$$

The parameters β_{Yi}, β_{Xi}, and α_i can assume the value zero. If $\alpha_i = 0$, the variant Z_i is not in LD with G, and $Z_i = \epsilon_i$. If $\beta_{Xi} = 0$, then the variant Z_i is not causal to X, and similarly for β_{Yi}.

1.4.2 Adjusted analysis

Suppose we fit the adjusted model $Y \sim G + X$:

$$Y = 1\mu + G\beta + X\gamma + \varepsilon, \tag{7}$$

which is a mis-specified model compared to the true generative model in (6). Let us denote $\tilde{X} = [1 : X]$ to be the covariate matrix X augmented to include the intercept, and $P_A = A(A^\top A)^{-1}A^\top$ to be the matrix for orthogonal projection onto A. Then, the effect-size estimate of G from (7) will be,

$$\begin{aligned}
\hat{\beta} &= \left(G^\top(I - P_{\tilde{X}})G\right)^{-1} G^\top(I - P_{\tilde{X}})Y \\
&= \left(G^\top(I - P_{\tilde{X}})G\right)^{-1} G^\top(I - P_{\tilde{X}}) \left(\mu_Y 1 + G\beta_{GY} + \sum Z_i\beta_{Yi} + \epsilon_Y\right) \\
&= \beta_{GY} + \sum \beta_{Yi} \left(G^\top(I - P_{\tilde{X}})G\right)^{-1} G^\top(I - P_{\tilde{X}})Z_i \\
&\qquad + \left(G^\top(I - P_{\tilde{X}})G\right)^{-1} G^\top(I - P_{\tilde{X}})\epsilon_Y.
\end{aligned}$$

Taking expectations,

$$\begin{aligned}
\mathbb{E}(\hat{\beta}) &= \beta_{GY} + \sum \beta_{Yi} E\left[\left(G^\top(I - P_{\tilde{X}})G\right)^{-1} G^\top(I - P_{\tilde{X}})Z_i\right] \\
&= \beta_{GY} + \sum \beta_{Yi} \mathbb{E}(\hat{\beta}_{\{G:Z_i \sim G+X\}})
\end{aligned}$$

Here, $\hat{\beta}_{\{G:Z_i \sim X+G\}} = \left(G^\top(I - P_{\tilde{X}})G\right)^{-1} G^\top(I - P_{\tilde{X}})Z_i$ is the estimated coefficient for G in the model $Z_i \sim X + G$. The following derivation closely follows Aschard *et al* [1]. First,

note that:

$$(I - P_{\tilde{X}})G = (I - P_{[1:X]})G = (I - P_1 - P_{(I-P_1)X})G = (I - P_{X^{(c)}})G,$$

where $X^{(c)} = (I - P_1)X$ is the centered denominator. Next, we have,

$$G^\top (I - P_{\tilde{X}})G = N - \frac{(G^\top X^{(c)})^2}{X^{(c)\top} X^{(c)}} = N\left(1 - \frac{\hat{\beta}_{GX}^2}{v_X}\right),$$

where $\hat{\beta}_{GX}$ is the coefficient estimate of G in the regression of X on G, and $v_X = N^{-1} X^{(c)\top} X^{(c)}$ the sample variance of X. Based on our generative models, $\hat{\beta}_{GX}$ is unbiased to β_{GX}. Next,

$$\begin{aligned} G^\top (I - P_{\tilde{X}})Z_i &= G^\top Z_i - \frac{(G^\top X^{(c)})(X^{(c)\top} Z_i)}{X^{(c)\top} X^{(c)}} \\ &= N\left(\hat{\alpha}_i - \frac{\hat{\beta}_{GX}\hat{\rho}_{X,Z_i}}{\sqrt{v_X}}\right), \end{aligned}$$

where ρ_{X,Z_i} is the sample correlation of X and Z_i. Therefore,

$$\hat{\beta}_{\{G:Z_i \sim G+X\}} = \frac{\hat{\alpha}_i - \frac{\hat{\beta}_{GX}\hat{\rho}_{X,Z_i}}{\sqrt{v_X}}}{1 - \frac{\hat{\beta}_{GX}^2}{v_X}} \approx \hat{\alpha}_i - \frac{\hat{\beta}_{GX}\hat{\rho}_{X,Z_i}}{\sqrt{v_X}},$$

assuming $\hat{\beta}_{GX}/v_X \ll 1$ as is typical for human traits (this assumption follows [1]). For large sample-sizes, we can replace the sample estimates to their corresponding limits in the population to obtain,

$$\begin{aligned} \mathbb{E}(\hat{\beta}_{\{G:Z_i \sim G+X\}}) &\approx \alpha_i - \frac{\beta_{GX}\rho_{X,Z_i}}{\sqrt{V_X}}, \\ V_X &= \mathbb{E}\left(\frac{1}{N} X^\top (I - P_1) X\right), \\ \rho_{X,Z_i} &= V_X^{-1/2} \mathbb{E}\left(\frac{1}{N} X^\top (I - P_1) Z_i\right), \end{aligned}$$

where ρ_{X,Z_i} is the marginal correlation between X and Z_i and V_X is the marginal variance of X marginalized over all Z_i-s, G, and ϵ_X. Therefore,

$$\mathbb{E}(\hat{\beta}) \approx \beta_{GY} + \sum \beta_{Yi}(\alpha_i - V_X^{-1/2}\rho_{X,Z_i}\beta_{GX}). \tag{8}$$

In equation (8), β_{GY} is the effect size of interest and the summation is a bias term. If $\beta_{Yi} = 0$, meaning the background variants do not affect the numerator trait, then the bias vanishes. However, when $\beta_{Yi} \neq 0$, bias is expected if Z_i is in LD with G, such that $\alpha_i \neq 0$, or if

i. Z_i has an effect on X, such that $\rho_{X,Z_i} \neq 0$, and

ii. G has an effect on X, such that $\beta_{GX} \neq 0$.

1.4.3 Model with a polygenic score

Lets S denote a PGS for the denominator trait X. Suppose we update the association model to $Y \sim G + X + S$:

$$Y = 1\mu + G\beta + X\gamma + S\delta + \varepsilon. \tag{9}$$

Then, the effect-size estimate of G from (9) will be:

$$\begin{aligned}
\hat{\beta} &= \left(G^\top(I - P_{[\tilde{X}:S]})G\right)^{-1} G^\top(I - P_{[\tilde{X}:S]})Y \\
&= \beta_{GY} + \sum \beta_{Yi} \left(G^\top(I - P_{[\tilde{X}:S]})G\right)^{-1} G^\top(I - P_{[\tilde{X}:S]})Z_i \\
&\qquad + \left(G^\top(I - P_{[\tilde{X}:S]})G\right)^{-1} G^\top(I - P_{[\tilde{X}:S]})\epsilon_Y.
\end{aligned}$$

Taking expectations,

$$\begin{aligned}
\mathbb{E}(\hat{\beta}) &= \beta_{GY} + \sum \beta_{Yi}\mathbb{E}\left[\left(G^\top(I - P_{[\tilde{X}:S]})G\right)^{-1} G^\top(I - P_{[\tilde{X}:S]})Z_i\right] \\
&= \beta_{GY} + \sum \beta_{Yi}\mathbb{E}(\hat{\beta}_{\{G:Z_i \sim G+X+S\}})
\end{aligned}$$

where $\hat{\beta}_{\{G:Z_i \sim G+X+S\}}$ is the coefficient on G from the model $Z_i \sim G+X+S$. In the following, we write:

$$\begin{aligned}(I - P_{[\tilde{X}:S]})G &= (I - P_1 - P_{(I-P_1)S} - P_{(I-P_{[1:S]})X})G \\ &= (I - P_{S^{(c)}} - P_{X^{(r)}})G,\end{aligned}$$

where $S^{(c)}$ is the centered PGS, and $X^{(r)}$ is the residual after least squares regression of X on S, $X \sim S$.

1.4.4 Case 1: PGS constructed only using variants not in LD with G

Suppose S is constructed only using variants that are not in LD with G. Here, $S^{(c)}$ and G are independent, thus $P_{S^{(c)}}G \xrightarrow{p} 0$. Therefore,

$$\begin{aligned}\hat{\beta}_{\{G:Z_i \sim G+X+S\}} &= \left(G^\top (I - P_{[\tilde{X}:S]})G\right)^{-1} G^\top (I - P_{[\tilde{X}:S]})Z_i \\ &\approx \left(G^\top (I - P_{X^{(r)}})G\right)^{-1} G^\top (I - P_{X^{(r)}})Z_i \\ &= \hat{\beta}_{\{G:Z_i \sim G+X^{(r)}\}}\end{aligned}$$

We can analyze the above expression as we did for the adjusted model, with the only difference being X is replaced by $X^{(r)}$:

$$\mathbb{E}(\hat{\beta}_{\{G:Z_i \sim G+X^{(r)}\}}) \approx \alpha_i - \frac{\beta_{GX^{(r)}} \rho_{X^{(r)},Z_i}}{\sqrt{V_X^{(r)}}}.$$

Now if the ith variant is not in LD with G, then $\alpha_i = 0$ and $\rho_{X^{(r)},Z_i} \xrightarrow{p} 0$ because $X^{(r)}$ is the residual of X having regressed out the effects of variants not in LD with G (including Z_i). Therefore, using a PGS S for X constructed only from variants not in LD with G can remove the heritable covariate bias arising from variants not in LD with G.

1.4.5 Case 2: global PGS constructed based on all variants

Suppose instead S is constructed from all variants, including G and variants in LD with G. In this case $S^{(c)}$ is no longer uncorrelated with G. Then,

$$\begin{aligned}\hat{\beta}_{\{G:Z_i\sim G+X+S\}} &= \left(G^\top(I-P_{[\tilde{X}:S]})G\right)^{-1} G^\top(I-P_{[\tilde{X}:S]})Z_i \\ &\approx \left(G^\top(I-P_{S^{(c)}}-P_{X^{(r)}})G\right)^{-1} G^\top(I-P_{S^{(c)}}-P_{X^{(r)}})Z_i.\end{aligned}$$

Following a similar derivation as before, we can obtain the individual bias contribution terms,

$$\mathbb{E}(\hat{\beta}_{\{G:Z_i\sim G+X+S\}}) \approx \alpha_i - V_X^{(r)-1/2}\beta_{GX^{(r)}}\rho_{X^{(r)},Z_i} - V_S^{(c)-1/2}\beta_{GS^{(c)}}\rho_{S^{(c)},Z_i}.$$

In this case, even when the i-th variant is not in LD with G, such that $\alpha_i = 0$ and $\rho_{X^{(r)},Z_i} \xrightarrow{p} 0$, there remains a bias term:

$$\mathbb{E}(\hat{\beta}_{\{G:Z_i\sim G+X+S\}}) \approx -V_S^{(c)-1/2}\beta_{GS^{(c)}}\rho_{S^{(c)},Z_i}.$$

$\beta_{GS^{(c)}}$ and $\rho_{S^{(c)},Z_i}$ both will be non-ignorable when G and Z_i are included in the PGS S.

2 Supplemental Results

2.1 Null numerator experiment

Weight was permuted to form a null phenotype W, then centered and scaled to have mean 0 and variance 1. The mean of W was shifted by adding a constant $\mu \in \{0, \ldots, 250\}$:

$$W_\mu = W + \mu.$$

W was analyzed either as a ratio, dividing by the observed value of height^2 to emulate BMI, or via an adjusted model, conditioning on height and height^2. In **Figure S1**, the number of independent genome-wide significant loci is reported as a function of null numerator's mean. Whereas no loci are discovered by the adjusted model, the number of loci detected

by the ratio model depends artifactually on the null numerator's mean. This result is not particular to weight; replacing W with independent random draws from the standard normal distribution led to nearly identical results. **Figure S2** shows that the Manhattan plots for observed height and for permuted height, shifted to mean 250, divided by height^2 are nearly identical, underscoring that the associations being detected by the ratio model are exactly those associated with the denominator trait.

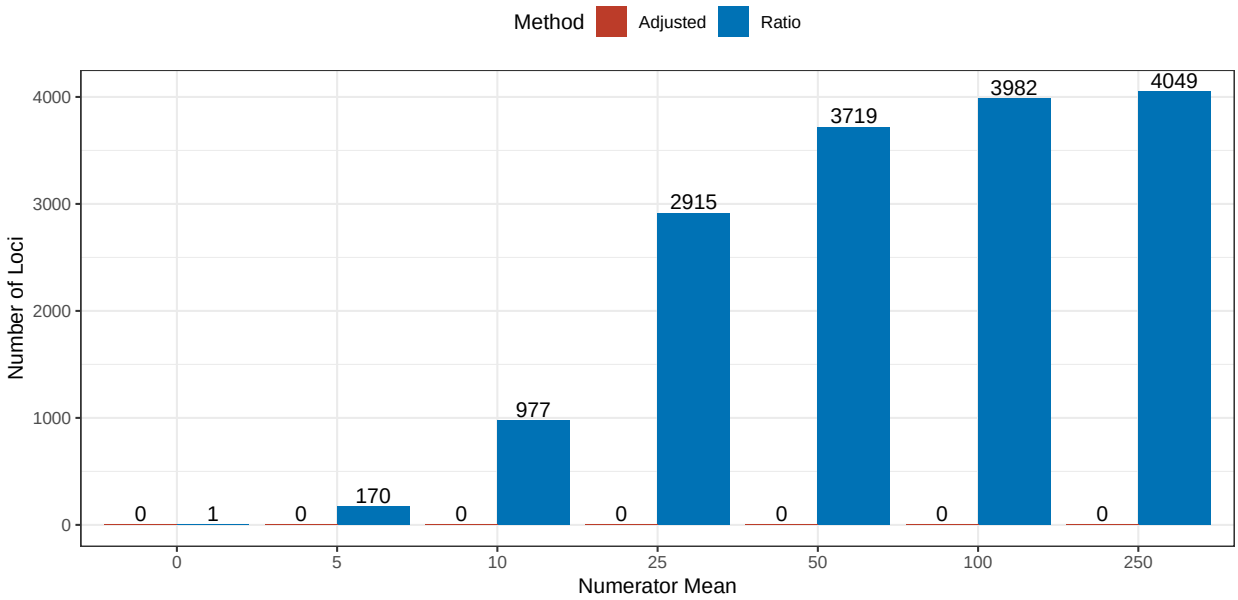
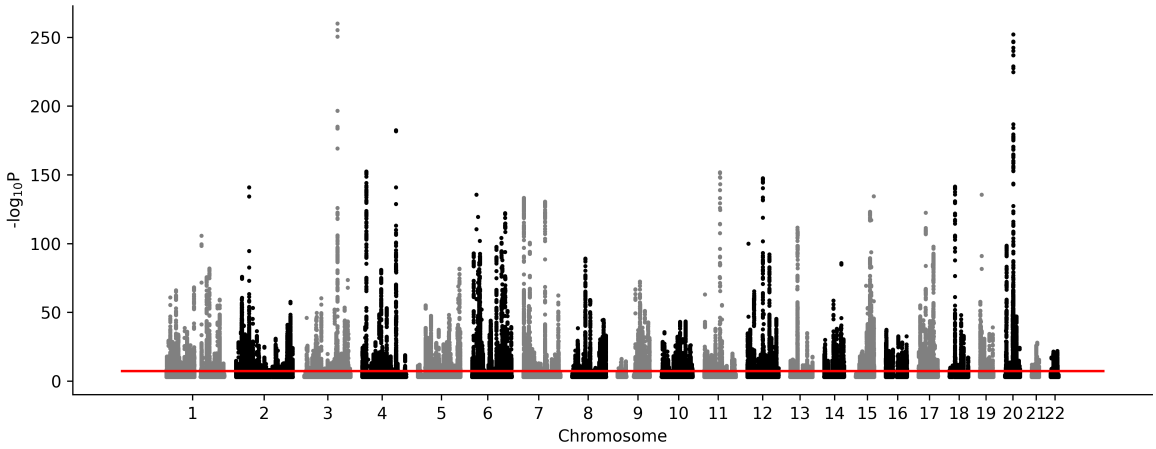
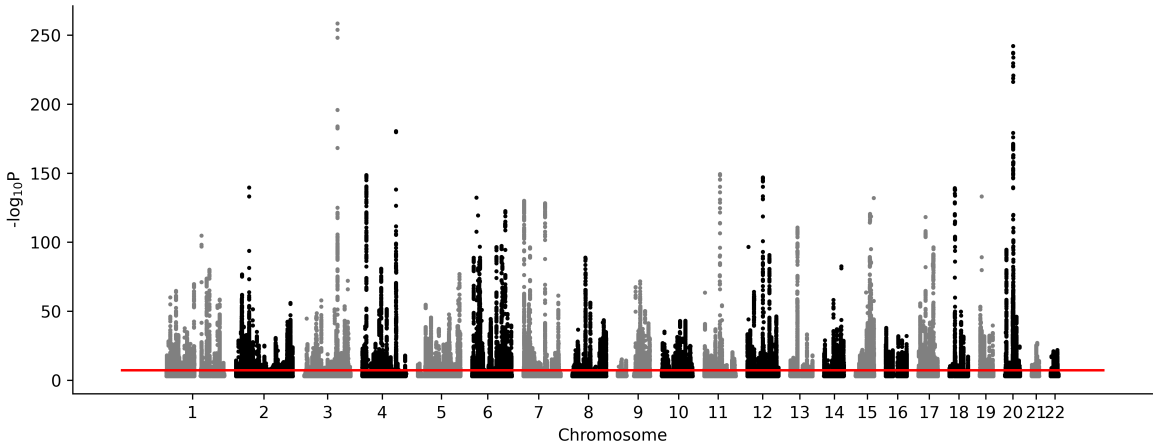


Figure S1: **Artifactual dependence of the number of loci discovered by the ratio model on the mean of the numerator.** A null phenotype was formed by permuting weight, scaling to unit variance, then shifting to mean μ. The null phenotype was either analyzed as a ratio, dividing by the observed value of height^2 (to emulate BMI), or via an adjusted model, conditioning on height and height^2. Shown are the number of independent ($R^2 \leq 0.1$) genome-wide significant ($P \leq 5 \times 10^{-8}$) loci.



(a) **Height**



(b) **(Permuted Weight + 250) / Height^2**

Figure S2: **Manhattan plots for height and for permuted weighted, shifted to have mean 250, divided by height^2.** Because weight is permuted, any associations in the lower plot are attributable to height. GWAS were performed among $N = 356$K independent subjects from the UK Biobank.

2.2 Survey of ratio traits

Traits			GWS Loci		
Numerator	Denominator	N	Ratio	Adjusted	Both (% of Total)
Albumin (blood)	Creatinine (blood)	311,122	944	556	459 (31%)
Albumin (urine)	Creatinine (urine)	105,501	6	4	0 (0%)
Weight	Height^2	355,627	844	830	1515 (91%)
FEV_1	FVC	325,220	538	474	877 (87%)
Stroke volume	End diastolic volume	29,171	6	0	0 (0%)
Phenylalanine	Tyrosine	200,159	251	174	361 (85%)
Platelets	Lymphocytes	52,133	2	0	0 (0%)
Seated height	Standing height	355,373	440	115	186 (34%)
Trunk fat mass	Whole body fat mass	349,533	359	240	421 (70%)
Waist circumference	Hip circumference	356,112	465	529	763 (77%)
	Average	243,995	385.5	292.2	458.2 (68%)

Table S1: **Overlap of loci detected by the ratio and adjusted models.** Numbers of independent genome-wide significant loci detected by the ratio and adjusted models, the count of signals detected by both models, and the percentage of the total number of signals detected by both models. Loci were obtained by clumping full-genome summary statistics at $R^2 \leq 0.10$ and $P \leq 5 \times 10^{-8}$. A locus detected by the ratio model was considered tagged by the adjusted model if within 250kb of an adjusted locus, and vice versa.

Numerator	Denominator	Genetic Rho (%)	SE (%)	P
Albumin (blood)	Creatinine (blood)	-9.80	2.80	0.00
Albumin (urine)	Creatinine (urine)	-23.50	14.90	0.12
Weight	Height	40.50	1.60	0.00
FEV1	FVC	92.20	0.40	0.00
Stroke volume	End diastolic volume	73.80	10.50	0.00
Phenylalanine	Tyrosine	49.90	6.50	0.00
Platelets	Lymphocytes	-34.30	31.60	0.28
Seated height	Standing height	90.80	0.60	0.00
Trunk fat mass	Whole body fat mass	99.30	0.00	0.00
Waist circumference	Hip circumference	88.50	0.80	0.00

Table S2: **Genetic correlations between the numerators and denominators of traits analyzed as ratios.** Genetic correlations were calculated from genome-wide summary statistics using LD score regression.

Numerator				Denominator			
Trait	Loci	h^2 (%)	SE (%)	Trait	Loci	h^2	SE (%)
Albumin (blood)	557	11.30	0.70	Creatinine (blood)	543	12.70	0.70
Albumin (urine)	4	0.80	0.50	Creatinine (urine)	10	6.00	0.60
Weight	1069	23.70	0.90	Height	4132	41.10	1.70
FEV_1	708	18.20	0.70	FVC	851	19.00	0.80
Stroke volume	18	13.00	2.20	End diastolic volume	68	7.00	1.80
Phenylalanine	79	3.60	0.70	Tyrosine	193	7.90	1.60
Platelets	2	0.70	1.10	Lymphocytes	109	19.20	1.70
Seated height	1330	18.60	0.90	Standing height	4120	41.10	1.70
Trunk fat mass	795	21.70	0.70	Whole body fat mass	789	21.80	0.70
Waist circumference	1212	19.50	0.70	Hip circumference	1742	19.50	0.70

Table S3: **Counts of independent genome-wide significant loci and estimated heritability for the numerators and denominators of traits analyzed as ratios.** Loci were obtained by clumping full-genome summary statistics at $R^2 \leq 0.10$ and $P \leq 5 \times 10^{-8}$. Heritability was estimated by LD score regression.

2.3 Effect size correlations

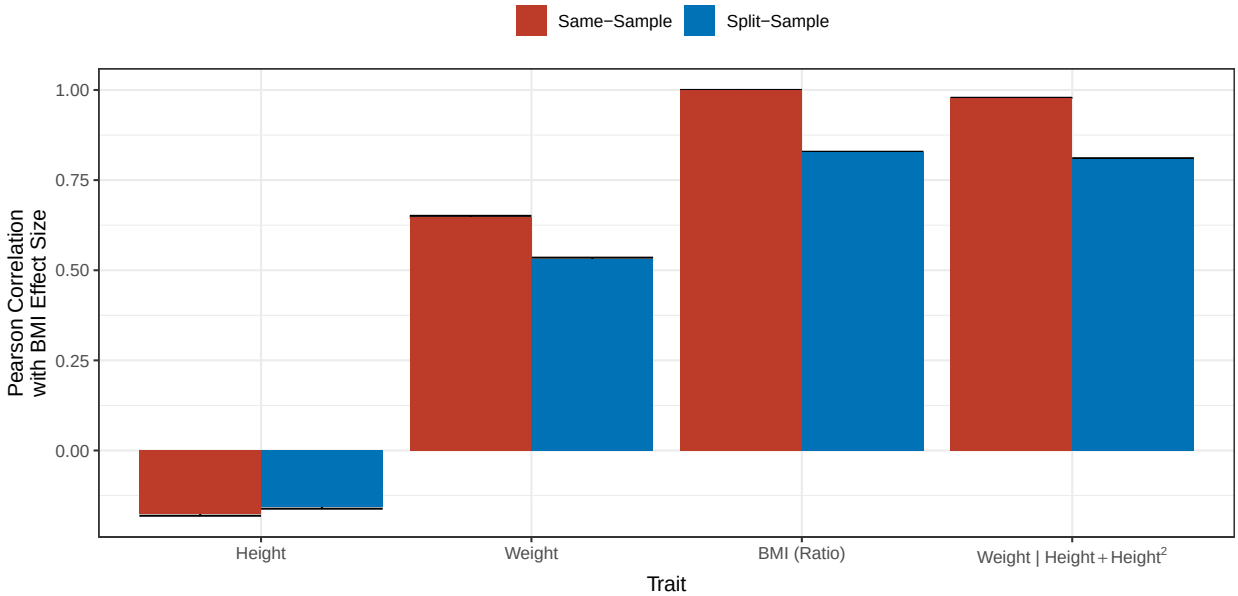


Figure S3: **Correlation of effect sizes with those for BMI, measured either in the same sample or in different samples.** Same-sample refers to correlating effect sizes from two GWAS fit to the same data set, namely the full UKB cohort ($N = 356$K). Split-sample refers to correlating effect sizes estimated in equally-sized independent subsets of the UKB ($N = 178$K each). Effect sizes are correlated across the set of variants significantly ($p \leq 5 \times 10^{-8}$) associated with at least 1 phenotype (BMI, height, weight, weight adjusted for height) in the full sample.

2.4 Biological relevance of ratio and adjusted models

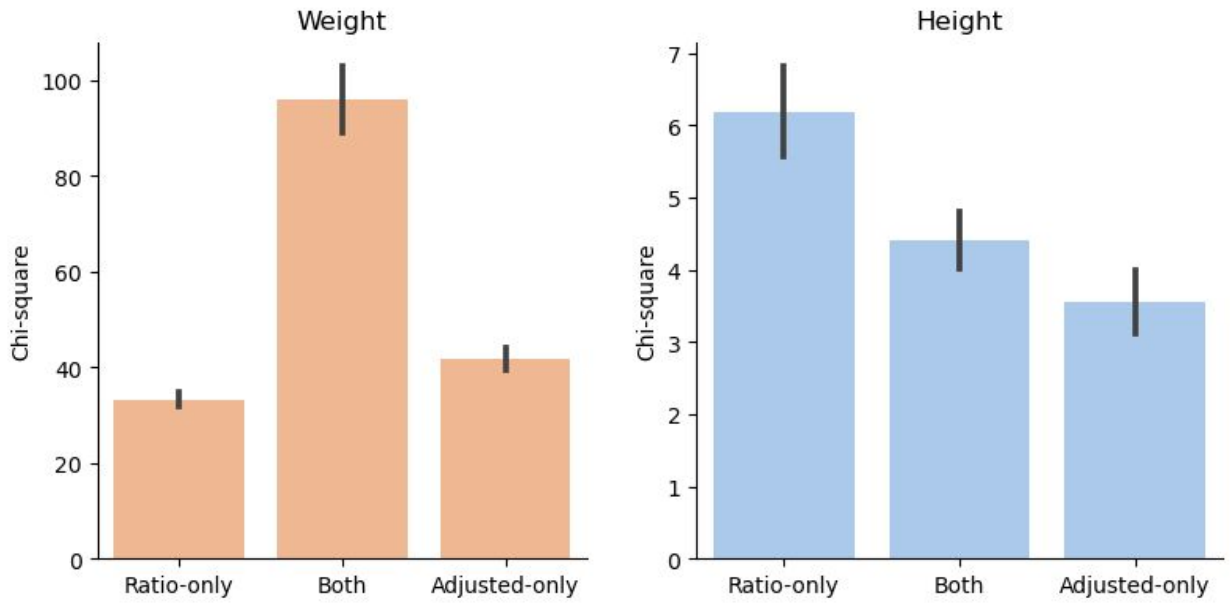


Figure S4: **Adjusted model loci are enriched for numerator signals and ratio model loci are enriched for denominator signals.** Average χ^2 statistics for weight (the numerator trait) and height (the denominator trait) partitioned by whether the locus was detected by the ratio model only (i.e. BMI), the adjusted model only, or both. Error bars are +/- 1 standard error.

Trait	N	h^2 (SE)	Loci	Adjusted-model average χ^2 (SE)	Ratio-model average χ^2 (SE)	Paired T	P
Body fat % [2]	100,716	0.07 (0.01)	13	213.95 (70.31)	204.59 (68.10)	2.4169	0.0325
Hip circumference [3]	224,459	0.14 (0.01)	96	95.82 (11.82)	88.59 (11.50)	8.401	4.28×10^{-13}
T2D [4]	898,130	0.05 (0.002)	921	13.46 (1.52)	13.41 (1.47)	0.4371	0.6622
CAD [5]	250,000	0.04 (0.003)	143	5.78 (0.92)	5.79 (0.95)	-0.0927	0.9263
cALT [6]	218,595	.	225	4.60 (0.60)	4.63 (0.59)	-0.2675	0.7894
LDL cholesterol [7]	700,000	0.09 (0.01)	4035	4.47 (0.14)	4.39 (0.14)	2.6782	0.0074

Table S4: **The adjusted model has more power at loci defined by obesity-related traits.** Loci were obtained by clumping full-genome summary statistics at $R^2 \leq 0.50$ and $P \leq 5 \times 10^{-8}$. Genome-wide summary statistics were not available for chronically-elevated ALT (cALT) [6]. T2D: Type 2 Diabetes, CAD: coronary artery disease.

2.5 Leave-one-chromosome-out polygenic scoring

Following Aschard *et al* [1], data were generated from the 4 data generating processes in **Figure 3(a)**. These cases can be described as:

a. **Complete mediation**, where the effects of genotype G and background B on the phenotype Y are entirely due to X.

b. **Non-collider**, where the background B influences both X and Y, and G has an effect on phenotype Y but not the covariate X.

c. **Collider without direct effect**, where B influences both X and Y, and G has an effect on the covariate X but not the phenotype Y.

d. **Collider with direct effect**, where B and G both influence X and Y.

For the simulation study, G represents genotype at the variant of interest and B represents the cumulative effect of variants **not** in LD with G (i.e. the genetic background). Coefficients were chosen such that the total variation in each variable is 1.0, and the correlation between between X and Y was $\rho \in \{0.00, 0.25, 0.50, 0.75, 0.90\}$. Wherever an edge from G is present, G explains 1% of the variation in the downstream variable (i.e. $h^2 = 1\%$). The simulation was repeated 10^4 times at each level of ρ.

Five estimators are compared:

i. **Adjusted GWAS** of Y on G adjusting for X. This estimator is expected to exhibit bias for data generating processes where X is a collider.

ii. **Ratio GWAS** of Y/X on G. This estimator is expected to regularly exhibit bias because the association model is misspecified. In practice, the fact that genotype G is rarely expected to act directly on the ratio, as opposed to the individual phenotypes, should be considered before performing a ratio GWAS.

iii. **Conditional GWAS** of Y on G adjusting for X and B. This estimator is unbiased and theoretically optimal but requires knowledge of B.

iv. **LOCO-PGS GWAS** of Y on G adjusting for X and a LOCO-PGS for X. The weights of the PGS were estimated by regressing Y on B within the same sample, then constructing a score S for each subject. Note that in the real data analysis, the PGS weights were estimated in an independent sample. Here, we estimate the PGS weights in the same sample to determine what effect, if any, this has on the performance of the estimator. S emulates a LOCO-PGS because B is not in LD with, and in fact is independent from, G.

v. **Non-heritable component GWAS** of Y on G adjusting for the residual after X is regressed on the LOCO-PGS from (iv).

References

[1] H Aschard, Bj Vilhjalmsson, AD Joshi, AL Price, and P Kraft. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am J Hum Genet*, 96(2):329 – 339, 2015.

[2] Y Lu, FR Day, S Gustafsson, et al. New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nat Commun*, 7:10495, 2016.

[3] D Shungin, TW Winkler, DC Croteau-Chonka, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature*, 518(7538):187 – 196, 2015.

[4] A Mahajan, CN Spracklen, W Zhang, et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat Genet*, 54(5):560 – 572, 2022.

[5] CP Nelson, A Goel, AS Butterworth, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet*, 49(9):1385 – 1391, 2017.

[6] M Vujkovic, S Ramdas, KM Lorenz, et al. A multiancestry genome-wide association study of unexplained chronic alt elevation as a proxy for nonalcoholic fatty liver disease with histological and radiological validation. *Nat Genet*, 54:761 – 771, 2022.

[7] SE Graham, SL Clarke, KHH Wu, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature*, 600:675 – 679, 2021.