ONLINE SUPPLEMENT

1. Path model equations for heritability

Equation S1 below is used to calculate the heritability of trait one, or objective BMI, in all of our bivariate models:

$$h^2_{Objective\ BMI} = \frac{a^2_{11}}{a^2_{11}+e^2_{11}} \quad \text{(Equation S1)}$$

Equation S2 below is used to calculate the heritability of trait 2, or perceived weight status, in all of our bivariate models:

$$h^2_{Perceived\ Weight\ Status} = \frac{a^2_{21}+a^2_{22}}{a^2_{21}+e^2_{21}+a^2_{22}+e^2_{22}} \quad \text{(Equation S2)}$$

2. Robustness check for calculating genetic variance unique to perceived weight status

Another way to determine the residual genetic variance unique to perceived weight status (and thus independent of objective BMI) is to residualize the perceived weight status phenotype out of the objective BMI phenotype and therefore determine the univariate heritability of those residual values. In Table S1 below, we use only monozygotic and same-sex dizygotic twins to perform this robustness check. As can be seen in this table, our estimates of this residual phenotype are very similar to those calculated in the full Cholesky model (Table 3).

Table S1
Wave 1-4 univariate twin AE parameter estimates for heritability of perceived weight status, independent of objective BMI. Fit statistics and 95% confidence intervals for heritability estimates included. $p > 0.05$ indicates ability to drop C parameter from model.

	RESIDUAL PERCEIVED WEIGHT STATUS (PERCEIVED WEIGHT STATUS ~ OBJECTIVE BMI)					
AE Model Wave	A (h^2)	E	95% h^2 CI Minimum	95% h^2 CI Maximum	-2LL	p <
Wave 1	0.23	0.77	0.12	0.33	1811.52	0.6
Wave 2	0.21	0.8	0.09	0.32	1547.4	0.34
Wave 3	0.28	0.72	0.14	0.4	1287.01	1
Wave 4	0.25	0.75	0.12	0.36	1452.55	1

3. Testing for qualitative and quantitative sex differences in univariate models

We first formally test for qualitative sex-limitation in our univariate models, or the possibility that different genetic influences matter for men and women. (Because these are univariate tests, no a_{21} or e_{21} paths exist in these models.) To perform this test, we fix the correlation between genetic factors for opposite sex DZ twins to be equal to those of same-sex DZ twins, and we then compare the χ^2 statistic of this fixed model to that of a model where we freely estimate this correlation. As shown by the "Model 1 to 2 Diff. p" column in Table S2 below, we find some evidence for qualitative sex-limitation across our waves and phenotypes. However, the results are mixed and on average, either insignificant or borderline.

We next test for quantitative sex-limitation in our univariate models, or the possibility that the magnitude of genetic influences differs across gender. This time, we further fix both the additive genetic and nonshared environmental pathways across males and females to be equal. We compare the χ^2 statistic of this model to the previous model, where we fixed only the opposite sex DZ twin genetic correlations. As shown by the "Model 2 to 3 Diff. p" column in Table S2 below, we again see mixed but relatively weak evidence that univariate quantitative sex differences exist.

Table S2
Tests of univariate qualitative and quantitative sex-limitation for objective BMI and perceived weight status, by wave.

	UNIVARIATE QUALITATIVE AND QUANTITATIVE SEX LIMITATION				
	Free χ^2 (Model 1)	Qualitative Fixed χ^2 (Model 2)	Quantitative Fixed χ^2 (Model 3)	Model 1 to 2 Diff. p	Model 2 to 3 Diff. p
	Objective BMI				
Wave 1	28.08	31.48	37.91	0.07	0.09
Wave 2	44.14	44.61	45.92	0.49	0.73
Wave 3	35.73	39.08	62.79	0.07	0.00
Wave 4	28.91	34.73	61.98	0.02	0.00
	Perceived Weight Status				
Wave 1	10.35	10.42	12.64	0.79	0.53
Wave 2	29.94	34.30	41.69	0.04	0.06
Wave 3	21.49	23.33	25.86	0.17	0.47
Wave 4	19.50	26.84	38.73	0.01	0.01

Freely estimated model (Model 1) has 18 degrees of freedom; fixed, qualitative model (Model 2) has 19; fixed, quantitative model (Model 3) has 22.

4. Robustness check for quantitative sex differences in the bivariate model

Because we found some mild evidence of qualitative sex differences in our phenotypes across waves, we want to be certain that these potential qualitative sex differences are not confounding our test of quantitative sex differences in our bivariate model. One way to be more certain that this is not the case is to drop our opposite sex pairs from our analysis and perform our bivariate test again with only MZ and *same-sex* DZ twin pairs. As reported in Table S3 below, when we perform this robustness check, we observe very similar patterns in this test as in the test with opposite sex DZ twins (Table 4), providing evidence that the quantitative sex differences we observe in our bivariate model are not confounded with potential qualitative sex differences.

Table S3
Test of quantitative sex-limitation for the bivariate model by wave. Includes only MZ and same-sex DZ twin pairs.

	BIVARIATE QUANTITATIVE SEX LIMITATION		
	Free parameters across gender χ^2 (Model 1)	Fixed parameters across gender χ^2 (Model 2)	Model Diff. p
Wave 1	49.54	66.45	0.03
Wave 2	43.19	51.96	0.36
Wave 3	58.44	99.95	0.00
Wave 4	48.87	89.65	0.00

Freely estimated model (Model 1) has 40 degrees of freedom; fixed model (Model 2) has 48.

5. Considerations of the twin model and molecular genetic data

We recognize that the twin modeling technique is limited by the assumptions implied by the model. For instance, a key assumption of these models is that MZ and DZ twins share similar environments. While this equal environments assumption has been criticized, the assumption has been directly tested and largely upheld (Kendler et al., 1993; Turkheimer, 2011). Further, twin models have limited power to estimate additive genetic and dominant genetic effects simultaneously, as well as epistatic (GxG) effects. Yet, even as more and more studies begin to turn toward molecular genetic data as it becomes more available in well-powered samples (i.e., many millions of individuals to estimate epistatic effects), we also note one important continued utility of twin- and family-based genetic designs. Unlike single nucleotide polymorphism (SNP)-based approaches, which can only explain variation in a phenotype related to the SNPs actually measured in a given sample, the broad-sense estimates from twin and family studies continue to define an "upper threshold" of additive genetic variation that may explain overall variation in a phenotype. Twin and family methods also allow for tests of hypotheses that are unavailable to molecular methods. For example, we found no evidence of shared environmental effects on either BMI or subjective weight perception, a finding only possible when multiple family members are assessed. Our study goals—and those of much social science research—align more with partitioning variance and identifying patterns of covariation that unfold over development, rather than identifying specific polymorphisms. As such, twin models, which maximize power for such investigations, remain a key and underused research tool in the social sciences.

We also note that, though our twin sample is relatively small, we show consistent results across our 4 waves of data, which bolsters the validity of our findings. In well-phenotyped datasets like Add Health, using genome-wide molecular methods (as opposed to twin- or family-

based methods) to analyze a phenotype such as subjective weight that is scarcely collected in survey datasets can be difficult. This is due to either a lack of molecular data or sample sizes that are not highly powered enough to perform genome-wide association studies. In contrast, twin and family methods maximize power by leveraging the contrast between genetically identical relatives and relatives that share approximately 50% of segregating genetic material. However, through consortium-driven efforts (e.g., Okbay et al., 2016), these sorts of datasets are becoming increasingly available and well-powered. Thus, while it goes beyond the scope of this paper to detail the specific genes and corresponding biological pathways responsible for variation in weight identity in the population, we nevertheless encourage future researchers to test specific pathways through which genes are differentially associated with different health-related phenotypes as a function of gender and biological sex. For instance, a number of researchers have used genome-wide methods to investigate genetic contributions to, and gene-environment interactions related to, BMI and other objective phenotypes of interest (Boardman et al., 2014; Domingue et al., 2016). Using some of these same methods, future researchers might use genome-wide complex trait analysis (GCTA, Yang et al., 2010; Yang et al., 2011) and Linkage Disequilibrium Score Regression (LDSR, Bulik-Sullivan et al., 2015) to measure the heritability of *self-perceived* weight by sex, and in so doing bypass the assumptions of the twin model by relying instead on measured genotypic data to obtain heritability estimates. Further, researchers might use a genome-wide association study (GWAS) to search for causal genetic variants related to self-perceived weight, or polygenic scores to search for genetic moderation across age, gender, and objective weight.