

Interactions between Polygenic Scores and Environments: Methodological and Conceptual Challenges

Benjamin W. Domingue,^a Sam Trejo,^b Emma Armstrong-Carter,^a Elliot M. Tucker-Drob^c

a) Stanford University; b) University of Wisconsin–Madison; c) University of Texas at Austin

Abstract: Interest in the study of gene–environment interaction has recently grown due to the sudden availability of molecular genetic data—in particular, polygenic scores—in many long-running longitudinal studies. Identifying and estimating statistical interactions comes with several analytic and inferential challenges; these challenges are heightened when used to integrate observational genomic and social science data. We articulate some of these key challenges, provide new perspectives on the study of gene–environment interactions, and end by offering some practical guidance for conducting research in this area. Given the sudden availability of well-powered polygenic scores, we anticipate a substantial increase in research testing for interaction between such scores and environments. The issues we discuss, if not properly addressed, may impact the enduring scientific value of gene–environment interaction studies.

Keywords: polygenic score; gene–environment interaction

Citation: Domingue, Benjamin W., Sam Trejo, Emma Armstrong-Carter, and Elliot M. Tucker-Drob. 2020. "Interactions between Polygenic Scores and Environments: Methodological and Conceptual Challenges." *Sociological Science* 7: 465-486.

Received: June 5, 2020

Accepted: August 24, 2020

Published: September 21, 2020

Editor(s): Olav Sorenson

DOI: 10.15195/v7.a19

Copyright: © 2020 The Author(s). This open-access article has been published under a Creative Commons Attribution License, which allows unrestricted use, distribution and reproduction, in any form, as long as the original author and source have been credited. 

OVER the past decade, the world has witnessed a massive expansion of our ability to conduct biological inquiry into human behavior (Visscher et al. 2017). Genome-wide association studies (GWAS) (Pearson and Manolio 2008) have established that a broad array of behavioral traits (e.g., mental well-being, cognitive function, tobacco use, and risk-taking) and biomedical traits (e.g., height, body mass index, cholesterol, and cardiovascular disease) are highly polygenic (Boyle, Li, and Pritchard 2017; Chabris et al. 2015). Thus, population variation in these traits is attributable to many genetic variants, each individually exhibiting a relatively small effect. This has led many researchers to forego the study of specific genetic variants in favor of genome-wide composite measures (Dudbridge 2013). These composite measures, known as polygenic scores (PGSs), summarize the cumulative effects of many variants across the genome and aim to index an individual's genetic liability for a given trait. PGSs constructed from large GWAS are robustly predictive of a sizable proportion of variance in consequential outcomes, such as educational attainment and lifespan (Cesarini and Visscher 2017; Lee et al. 2018; Sugrue and Desikan 2019).¹ In fact, many PGSs are predictive of important biobehavioral and social science outcomes that were not the target of the original GWAS. Although PGS are neither pure (they may capture, e.g., correlated nongenetic factors [Morris et al. 2020]) nor universal (i.e., they may not generalize to environmental contexts not captured in the original GWAS from which they were constructed [Mostafavi et al. 2020]) measures, they have still sparked substantial interest. Many have argued that PGSs may advance our understanding of the behavioral and biomedical sciences

(Belsky and Harden 2019; Conley and Fletcher 2017; Dudbridge 2016; Harden and Koellinger 2020). Sociologists, in particular, have begun to offer frameworks for thinking about how the discipline may benefit from such work (Freese 2018; Mills and Tropf 2020).

The increasing adoption of genetic approaches in social and behavioral science research has not diminished interest in the environment. Indeed, how social and environmental factors combine and interact with biological factors to produce individual differences is a question at the forefront of many research agendas in the social and behavioral sciences. Researchers have long posited that genetic effects likely vary as a function of environment (Feldman and Lewontin 1975).² For example, in the twin study literature, there has been substantial interest in whether decompositions of observed variation in a phenotype into genetic and environmental components differ by socioeconomic context or age (Purcell 2002). Other research designs have tested interactions between measured genotypes (i.e., individual genetic variants) and environmental features. Although such an approach has intuitive appeal, it has proven technically challenging to implement (Duncan and Keller 2011). Some of the challenges of this research agenda may be attributable to unrealistically large expectations for effect sizes of individual variants and thus circumvented through use of PGSs. Yet, even when gene–environment interaction (G×E) results are robust and replicable, the interpretational and practical implications of such research can be unclear.

Polygenic scores are rapidly becoming widely available. Data sets such as the Health and Retirement Study (HRS; Ware et al. 2017), Add Health (Braudt and Harris 2018), and the Wisconsin Longitudinal Study (Okbay, Benjamin, and Visscher 2018) are posting preconstructed scores for use by researchers, and catalogs of polygenic scores are being made available (Lambert et al. 2020). This novel data resource may offer new and more robust avenues for exploration of G×E. However, challenges remain. Given the emergence of this new tool, we aim to provide timely guidance on how to conduct high-quality G×E research using PGSs. In this article, we have two main objectives. First, we outline several concerns associated with performing G×E research that future work may benefit from considering. Second, we offer some guidelines for designing, implementing, and interpreting high-quality G×E research using PGSs.

The Standard G×E Model

We consider some outcome, Y , to be a function of an individual's genotype, G , and some (potentially continuously varying) environmental exposure, E . We generically describe this data-generating model as

$$E(Y | \dots) = f(G, E). \quad (1)$$

Equation (1) accommodates both complex interplay between genotype and environment as well as outcomes that are not normally distributed (e.g., Y may have a Bernoulli distribution). We supplement this simple model with a few crucial assumptions. We assume that we have reasonable proxies available for G and E and

some identifiable approximation to $f()$. We comment on each of these assumptions below.

With respect to G , we assume that we can characterize genetic influence on the trait as a PGS,

$$\text{PGS}_i = \sum \beta_j (N \text{ Alleles})_{ij},$$

that is, a sum wherein the number (N) of alleles (0, 1, or 2) that an individual i has for each single nucleotide polymorphism (SNP) j is weighted by the effect, β_j , identified via GWAS. We note a few assumptions implicit in the above. We are focusing on traits that have a genetic architecture appropriately characterized by effects that are *additive with respect to one another* (although they may be nonadditive in terms of their potential to interact with environmental contexts) and dispersed over many loci. We view the assumption of additivity as an acceptable simplification given both the success of additive GWAS and the relative lack of strong empirical support for dominance or epistasis (i.e., gene–gene interaction) models (Polderman et al. 2015).

The assumption that genetic effects spread over many loci is not especially restrictive. Empirical work has indicated that many traits of interest in population health—body mass index, cardiovascular disease—are highly polygenic. Not all traits necessarily have this characteristic; consider, for example, monogenic diseases such as cystic fibrosis. However, as the sample size of the GWAS used to generate a PGS increases, weights (i.e., β_j) for SNPs that are not relevant to the phenotype of interest will go toward zero; thus, a summative approach can still potentially be used in such cases. Moreover, much of our discussion still applies when using genetic predictors constructed from a smaller number of variants or even using a single variant allele count (e.g., Boardman et al. 2012; Rosenquist et al. 2015).

We also note that GWAS results (i.e., β_j) are themselves potentially a function of both trait-specific biology and contextual features of the data used to derive them: for example, the social and policy landscape governing behavior of participants in the GWAS, selection issues associated with being a part of GWAS sample, et cetera. (Mostafavi et al. 2020; Pirastu et al. 2020). PGSs index the genetic propensity within the environmental context and demographic characteristics of participants in the original discovery GWAS on which the PGS is constructed.³ An interaction between PGS and environment may then indicate that the influence of genetic factors on the outcome is larger in some environments than others, that the sample in one environment is more similar to the sample from the discovery GWAS than in others, or some mixture of the two. This ambiguity regarding interpretation is important to keep in mind when findings from polygenic score research are interpreted. However, we focus the current article on inferential and statistical issues pertaining to the samples in which the PGSs are constructed and analyzed (i.e., we do not focus on the potential mismatch between that sample and the GWAS discovery sample).

With respect to E , we assume that researchers use specific measures of the environment, which we denote ENV. At present, research typically focuses on variation in measured environments that have relatively large main effects on Y . We consider this topic in detail later. In general, we emphasize that there are numerous

challenges associated with identification of the appropriate ENV in GxE research (Boardman, Daw, and Freese 2013). The identification of appropriate ENV measures merits additional scrutiny in future work. Following selection of a candidate ENV, more questions follow. Are we measuring the environmental characteristic at the appropriate level (e.g., household vs. neighborhood vs. community)? Are we measuring a salient exposure given the respondents' ages? Can we measure the environmental exposure of interest with high fidelity? Are the exposures and contexts of interest correlated with other, unmeasured, environmental or genetic variables that are themselves the driving forces in the identified GxE?

Finally, we assume that the unknown function $f()$ is well approximated by a relatively simple model. In particular, many GxE studies aim to shed light on Equation (1) using regression models of the form

$$E(Y|\dots) = b_0 + b_1 \cdot \text{PGS} + b_2 \cdot \text{ENV} + b_3 \cdot \text{PGS} \cdot \text{ENV} + \text{covariates}. \quad (2)$$

The aim is to have Equation (2) elucidate key properties of the (unknown) data-generating process, even if Equation (2) is only a rough approximation of Equation (1). There are several concerns that apply to such regression models. We review two important issues that have been the subject of previous scrutiny below before then considering several novel issues of specific relevance when conducting GxE studies in the next section.

First, environmental exposures are typically partly endogenous (Jaffee and Price 2007), creating complex patterns of correlations between genes, focal environments, and other relevant exposures that lead to inferential challenges for the identification of GxE. We do not provide an in-depth treatment of this issue here as it has been discussed in depth elsewhere (Briley et al. 2019; Dudbridge and Fletcher 2014; Fletcher and Conley 2013). The question of endogeneity is, of course, closely related to the question of whether the measured environment that statistically moderates PGS effects has a causal effect. This is of course a crucial question; whether the effect is causal has direct implications for whether direct manipulation of that environment will produce changes in the genotype–phenotype association. Second, misspecification bias is a generic problem that introduces additional complexities in the case of interaction research. For example, care must be taken to distinguish between models containing interactions between two variables versus those with no interactions but nonlinear (e.g., quadratic) terms in one or both of the two variables (Lubinski and Humphreys 1990; MacCallum and Mar 1995). In particular, GxE research must also attend to the issue raised by Keller (2014) focusing on the covariates included in Equation (2). When covariates are included in Equation (2), specification error may result if additional interaction terms between the covariates and both E and the PGS are not included. This is because the main effects of the covariates are insufficient controls in the case where there is covariation between both the covariate and genotype or the covariate and the environment. Fortunately, there is a straightforward solution. Researchers simply need include the full suite of interaction terms between the PGSs and the covariates when estimating Equation (2).

Study Design Issues in GxE Research

The Environmental Exposure

The problem. A great deal of research in the social sciences focuses exclusively on the effects of environments. For example, there is substantial interest in the effects of poverty, reflected primarily in the home environment of a young child, on the developing brain and related cognitive functioning (Duncan and Magnuson 2012; Johnson, Riis, and Noble 2016). GxE research has tended to emphasize environmental variables, like poverty, for which large main effects have been well documented (Barr et al. 2018; Gould et al. 2018; Musci et al. 2019). However, the environmental features having large main effects need not also be the features that lead to nuanced GxE effects. GxE research may benefit from additional attention to the theorized nature of the candidate environmental variables deployed in GxE research.

To better frame our argument, we consider two stylized patterns of GxE interaction. We emphasize here that these two patterns are not an exhaustive taxonomy of GxE interaction. Rather, they serve as illustrations of the considerations that we encourage. First, consider GxE interactions in which the environment functions as a “dimmer” on genetic effects. Dimmers, as in switches responsible for dimming or brightening lights, may magnify or constrict genetic effects on an outcome without changing their sign. Investigating dimmer-type GxE may be of high substantive interest in many contexts. For instance, it is of strong practical and theoretical importance to determine whether an educational policy with a robust positive average effect for the population disproportionately benefits children at highest genetic risk or those at lowest genetic risk or has uniform effect across the spectrum of genotypes.

However, as we discuss at greater length in the subsection on coarsened outcome variables, it is also important to be vigilant about the potential for GxE to arise as an artifact of more general effects on the distribution of the observed outcome itself. For instance, suppose the educational policy of interest is associated with an appreciable increase in both the mean and variation of math achievement in the student population. It is then possible that the intervention has increased the effect of the PGS on math achievement (i.e., a positive b_3 estimate in Eq. [2]) simply as a byproduct of more general increases in variation in math achievement. Because conventional ordinary least squares methods are blind to this type of heteroscedasticity, the concomitant increase in non-PGS variance may go overlooked.

Second, consider GxE interactions in which the environment functions as an image-inverting “lens” on genetic effects. An environment acts as such a lens when the direction of the effect of the PGS differs across the range of that environment. We refer to these environments as lenses based on the optical notion of a lens; in particular, certain glass lenses invert the orientation of objects.⁴ When considering lenses, the relative effect of a given genotype may be positive for a “low” level of the relevant environmental exposure and negative for “high” levels of the exposure, or vice versa. This has led to the hypothesis that what qualifies as a high- or low-risk genotype may depend upon the environmental context (Belsky and Pluess 2009;

Ellis et al. 2011; Obradović and Boyce 2009). Note that an environment may function as a lens even when it has a limited main effect.

Researchers frequently conceptualize environments to operate as lenses as a theoretical motivation for doing GxE research, and yet, in practice, many of the environmental measures typically used in GxE studies may be conceptually closer to the dimmer category. Moreover, the selection of PGS effects for examining lens-type GxE may be particularly challenging given that we construct PGSs from GWASs that only include main effects of SNPs (although this is perhaps changing in ways we expand upon below). This limitation may act as a strict limiting factor when it comes to identifying GxE with polygenic scores. A related issue is that, if the environmental context of the participants in the GWAS sample used to construct the PGS is similar to that in the test sample used to estimate GxE, then it is unlikely to include SNPs that demonstrate lens-type patterns, as the main effects of these SNPs will be close to zero.

We can also understand the difference between a dimmer and a lens in terms of their effect on the rank ordering of outcomes. All else being equal, a dimmer is order-preserving; that is, it preserves the order of the genotypes at different levels of the environment. Variation in the dimmer serves to vary the distance, in the outcome metric, between different levels of the PGS but never changes the rank orderings of the levels. In contrast, a lens reverses the order of genotypes; a PGS that predicts an outcome near the top of the distribution at one level of environmental exposure will predict an outcome near the bottom at another level of environmental exposure. Our dimmer/lens typology is similar in many respects to the ordinal/disordinal typology previously suggested (Widaman et al. 2012) but may be a useful conceptual distinction as GxE research becomes more common in the social sciences.

Recommendations. Conceptually, researchers will benefit from being attentive as to the form of GxE they expect; for example, is the candidate environment expected to operate as a lens, as a dimmer, or according to some more complex functional form? In our experience, GxE researchers will tend to observe that environments with large main effects on a phenotype act as dimmers. Such environments will moderate the magnitude of the effect of the polygenic score on the outcome without changing its sign. Although these observations may be of value, they need to be distinguished from the more dramatic patterns of sign reversal of PGS effects in different environments that have received a great deal of conceptual attention. In studies seeking to identify lens-type patterns (Troth et al. 2018), both the genetic and the environmental components are of crucial importance for testing hypotheses in which the environmental context determines whether a given genotype is risky or advantageous.

Analytically, we offer several suggestions that might be of interest in future work. We first emphasize the potential for analyses that take advantage of environmental variation without identifying a specific environmental feature of interest. In situations wherein individuals cluster into some unit, researchers may first want to consider the level of empirical support for GxE based on relatively omnibus measures of the environment. For example, one might test for variation observed in the relationship between phenotype and polygenic score across environmental

units (e.g., schools or census tracts); see Trejo et al. (2018) for one such example. Such analyses are informative in that they offer preliminary guidance on whether specific features of the environments deserve additional scrutiny as possible G×E targets.

Yet another approach that researchers might want to consider involves analyses of heritability and genetic correlation as captured by genomic techniques (Grotzinger et al. 2019; Yang et al. 2011). Polygenic scores collapse information from across the genome into a composite designed to predict a specific outcome in a novel data set. Because PGSs are constructed using a large number of GWAS regression weights that themselves are estimated with error, PGS prediction is biased downward in novel samples (and biased upward in the original GWAS samples). In contrast, genomic heritability and genetic correlation estimates are constructed using methods related to mixed effects modeling and are unbiased by measurement error. Such analyses can be used to, for example, study changing patterns of heritability (Tropf et al. 2017) across environments. Although analyses of heritability and genetic correlation do not provide scores for individual participants (because they estimate random effects to represent population variation, rather than individual estimates [de Vlaming et al. 2017]), they can still offer information about the way that genotypes are related to phenotypes.

We also note the increase in methodologies focused on identifying genetic variants that are associated with the amount of variation in the outcome rather than strictly the level (Conley et al. 2018; Wang et al. 2019; Yang et al. 2012; Young, Wauthier, and Donnelly 2018). Such approaches are generating data that may be useful in future G×E work. A natural question to ask of the genetic variants identified in such studies is whether environments interact with such variants to further modulate variation in the outcome. Although such approaches will presumably also involve novel methodological challenges, they are an exciting new resource that could be used to study gene–environment interplay.

Coarsened Outcome Variables

The problem. Characteristics of the distribution of Y may have crucial implications for conducting G×E studies. When Y is a discrete outcome coarsened from an underlying continuous variable, researchers encounter an opportunity to misinterpret affirmative findings of G×E. For simplicity of exposition, we focus on the simplest case where Y is dichotomous (though the phenomenon extends to coarsened variables that take more than two values). Suppose a dichotomous outcome Y is a coarsened version of some continuously varying latent indicator Y^* (so $Y = 1$ if $Y^* > \lambda$ for some scalar λ and 0 otherwise). For example, Y might be obesity or college completion (in which case Y^* would be body mass index or years of schooling, respectively). Suppose we estimate Equation (2) with ordinary least squares using Y^* instead of Y and yield a nonzero and statistically significant b_3 . How should we interpret such a finding? One possibility is that a finding of G×E suggests differences in the slope of association between G and Y^* . This, we argue, is what researchers generally have in mind when conducting studies testing for G×E. However, a second possibility is that a purely environmental shock may

shift the intercept of the association line between G and Y^* , thus resulting in a GxE finding (i.e., a nonzero and statistically significant b_3) with different interpretation.

We illustrate the basic problem in Figure 1. When we examine relationships between PGS and outcome in the context of the continuously measured version— Y^* in Figure 1A—we observe a constant linear association with genotype across two environments. However, when we observe a dichotomized version of the outcome— Y in Figure 1B—we have a relationship that is more challenging to interpret. In particular, Figure 1B suggests GxE when a linear probability model is used (i.e., the dotted curves are not parallel). In contrast, when a logistic regression model is used, we obtain unbiased estimates of GxE (i.e., b_3) but they may suffer from low power (and large confidence intervals) due to low variability in the dichotomized outcome at some regions of the environmental measure. This problem may be even more severe when gene–environment correlation results in a large shift in the distributions of the PGS along the range of the environmental measure.

Findings such as those in Figure 1B are worth noting, and they may be highly relevant in cases where the continuous Y^* is of less interest than the dichotomized Y (e.g., college completion may well matter more than years of schooling) or when Y^* is latent. However, we also need not confuse matters by misunderstanding the nature of the associations in question. If findings are driven by differences in intercepts and relatively consistent slopes, as in Figure 1A, this is important information to report. We expect that GxE research will benefit from distinguishing between these two possibilities; see also our discussion of this issue in an empirical context elsewhere (Trejo et al. 2018).⁵

Recommendations. When research uses coarsened outcome variables due to substantive interest in the coarsened outcome themselves (e.g., credentials, obesity indicators), sensitivity analyses that probe the issue considered here based on the underlying (noncoarsened) variable are essential. Such analyses will help to better contextualize findings from coarsened variables. In analysis of binary outcomes for which no underlying continuous variable is available (i.e., case-control status), utilization of multiple methods, such as both logistic and linear probability models, may be used to probe for sensitivity of the results to the functional form of the model. This will be especially important when the environment is itself nontrivially correlated with the outcome under study.

Although we do not focus here on coarsened outcomes that are nonbinary (e.g., ordered categorical, nominal, or censored/truncated outcomes), we note that many of the concerns raised here would be of relevance in those cases as well. At a minimum, sensitivity analyses probing the persistence of findings across a range of model specifications may be valuable. For example, in an analysis of the highest math course taken by high schools students (Harden et al. 2020), a variety of models—cumulative link, adjacent-category logit, locally estimated scatterplot smoothing (LOESS) based on dichotomizations—were used in an attempt to interrogate potential differences in course as a function of genotype when stratified by school socioeconomic status. GxE analysis in the context of such coarsened outcomes is likely to be challenging; future work describing methodological best practices in this domain would be welcome.

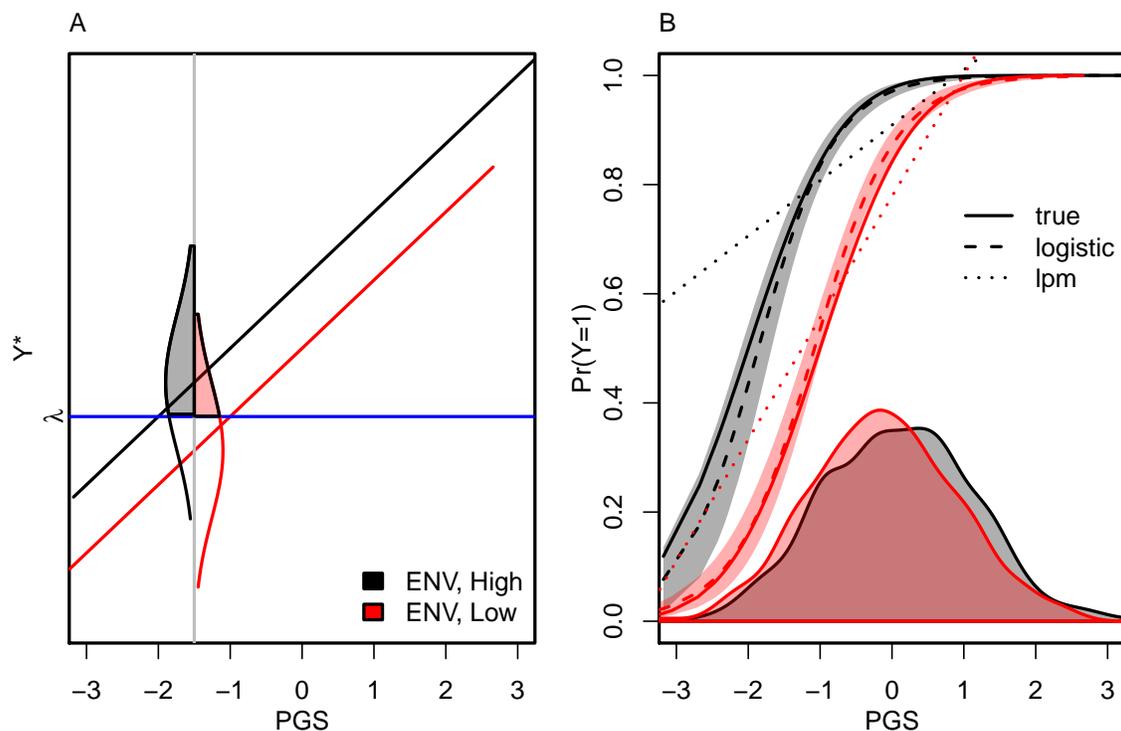


Figure 1: The challenge of studying GxE when using dichotomous outcomes. (A) True association between PGS and continuously varying outcome Y^* . Densities show distributions above horizontal blue line for those in high and low environments. (B) True associations and those estimated using either a logistic regression model or a linear probability model when Y^* has been dichotomized prior to analysis ($Y = 1$ when $Y^* > \lambda$). The linear probability model (lpm) produces the misimpression of GxE (nonparallel regression lines). The logistic regression model does not suffer from this bias but may still suffer from large standard errors and low power when we observe low variability in the dichotomous Y variable in one of the environments.

Measurement Error

The problem. Measurement error acts both to bias associated parameter estimates toward zero (Hutcheon, Chiolero, and Hanley 2010) and to distort power calculations. In the specific context of GxE studies, there are several concerns. Measurement error exists in both the operationalized PGS and ENV variables of Equation (2). Measurement error in G , which results from imprecise estimates of the GWAS betas used to construct the PGS, has received some attention (Conley et al. 2016; DiPrete, Burik, and Koellinger 2018; Tucker-Drob 2017). However, less attention has been paid to measurement error in E . Homoscedastic measurement error in E has implications for power (matters may be further complicated in the presence of nonhomoscedastic measurement error, but we focus on the simpler case here).

Figure 2 is a simple illustration of this via a simulation study.⁶ We assume that we measure both the PGS and the target environmental variable with error. We focus on variation in the reliability of the environmental measure (the x axis) and

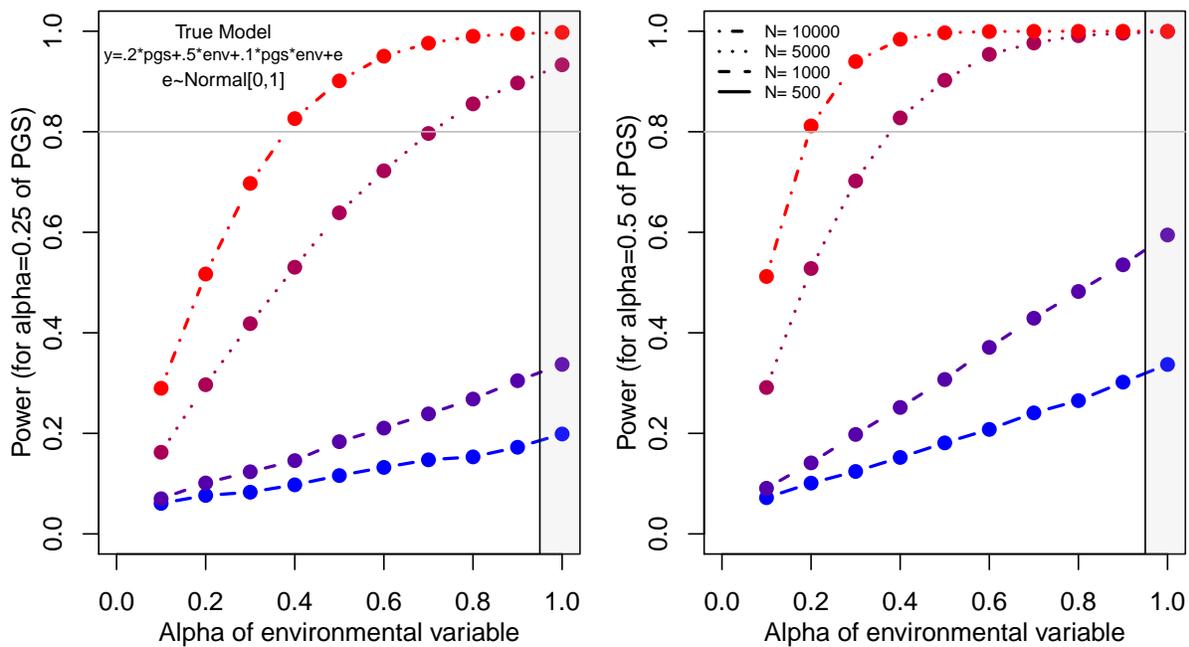


Figure 2: Reduction in power as a function of measurement error in both PGS and ENV. Left and right panels focus on relatively low ($\alpha = 0.25$) and high ($\alpha = 0.5$) reliability polygenic scores. Data-generating equation is shown in left-hand panel.

choose two levels of reliability (which we index as α) of 0.25 (on left) and 0.5 (on right) for the PGS; we view these reliabilities as representative of relatively weak and relatively strong polygenic scores given existing GWAS. The main takeaway is that ignoring measurement error with respect to the environment inevitably leads to inflated power calculations.

Let us first focus our attention on a PGS with relatively high reliability by current PGS standards ($\alpha = 0.5$) in the case where we have 1,000 respondents. We first assume that there is no error in our environmental measure (region emphasized in gray rectangle). In such a case, power is below standard levels of acceptability (power = 0.8). As the reliability of our environmental measure declines, however, power becomes increasingly poor. Even when the environment is measured with decent reliability ($\alpha = 0.7$), power is greatly reduced (power = 0.4). In the case where the PGS is of lower reliability, power is even worse (power = 0.2 for an environmental measure of reliability $\alpha = 0.7$). When the PGS is measured with substantial error ($\alpha = 0.25$), even relatively large samples (when considering population-based studies) of $N = 10,000$ will suffer from power limitations when the environment is also measured with substantial error. These calculations are based upon a toy model that might not be relevant in all cases, but given that interaction studies are power-hungry even without considering measurement error (McClelland and Judd 1993), our model emphasizes the need to carefully consider whether one has reasonable power before conducting GxEn studies.

Recommendations. We recommend that power analyses be the norm (and not the exception) in GxE research. Traditional power analyses are used to inform key design features, such as the sample size, prior to the implementation of a study. In contrast, power analyses of the type considered here offer information about the power of a study design given existing data (e.g., the sample sizes available from large longitudinal studies such as the HRS and Add Health) and key assumptions about the relevant parameters. In particular, power analyses specifically designed to offer insights into the level of power available given measurement error in both the polygenic score and the environment would be valuable additions when planning analyses of data that are already available. As Figure 2 illustrates, a failure to consider measurement error can lead to inflated estimates of power. Even for samples of several thousand, GxE analyses will be weakly powered absent highly penetrant genetic predictors or environments measured with little noise. Such power analyses are not cure-alls; rather, they hopefully help researchers to better understand the limitations that they face—specifically, the likelihood of observing false positives—in a given context.

Sample Selection Processes and Internal and External Validity

The problem. Selection processes complicate inference in observational settings in a number of ways, and studies of GxE are no exception. An often-underappreciated point is that sample selection issues threaten both external and internal validity. We discuss several (potentially overlapping) types of selection that are particularly relevant for GxE research. These sample selection processes limit the population to which GxE findings can be generalized and may lead to spurious results via collider bias (Elwert and Winship 2014). Notably, sample selection may pose a threat both in the discovery GWAS used to identify the betas needed to construct a PGS and in the prediction sample in which the PGS is actually constructed and used to estimate GxE.

We begin with mortality selection. Such selection occurs when a nonrandom subset is lost to mortality and therefore not observed. In studies of older respondents (e.g., the HRS), mortality selection tends to make the resulting sample “healthier, wealthier, and wiser” (Zajacova and Burgard 2013). Mortality selection is especially relevant to GxE research because genotyping is a relatively recent technology; participants in longstanding studies needed to survive long enough to make it into the genotyped subsample. Indeed, there is evidence to suggest that GxE findings may be sensitive to the presence of mortality selection (Domingue et al. 2017). When studying health-related traits, especially in older populations, we need to consider mortality selection’s role in shaping findings (Oliynyk 2019). In scenarios wherein mortality can be readily modeled with existing data, one possible analytic solution is to use inverse probability weighting (van der Wal and Geskus 2011) to correct for the role of mortality selection. A related issue is that individuals with certain genetic profiles—for example, those with high genetic liabilities for schizophrenia—may be underrepresented in various data sources (Martin et al. 2016; Meisner, Kundu, and Chatterjee 2019; Pirastu et al. 2020; Taylor et al. 2018). Such selection can also lead to issues of both bias and generalizability in subsequent studies.

A second issue is that demographic factors play a role in who gets included in genetic studies. This, in turn, has implications with respect to the populations to which results using genetic subsamples may generalize. Of particular note is the massive overrepresentation of European-descent individuals in both GWAS (Mills and Rahal 2019) and PGS (Duncan et al. 2019) studies. This problem is due to several factors, including both the overrepresentation of European-descent individuals in genetic studies and the fact that differences in linkage disequilibrium across groups leads to the GWAS findings performing better in the (predominantly European) samples from which they are derived. Efforts (Mills and Rahal 2020) are underway to monitor (with the hope of then remedying) this problem. In the meantime, researchers have noted that adoption of polygenic scores in precision medicine may exacerbate preexisting health disparities (Martin et al. 2019). A focus on homogeneous samples may lead to issues in GxE if it either severely constricts the relevant artificial variance or even potentially undermines the theoretical motivation suggesting a particular research question (which may necessitate a more diverse sample). In any event, equity concerns need to be in the foreground of genetics research; GxE is no exception.

These selection problems offer both internal and external validity threats to GxE studies that are important to consider carefully. An additional concern is that nonrandom selection into the analytic samples used in empirical studies may lead to reduced environmental variation further challenging attempts to make accurate inferences regarding GxE. As an illustration, we consider two key adolescent environments—the socioeconomic circumstances of home (Belsky et al. 2018) and the disadvantage of one's residential neighborhood (Belsky et al. 2019), both from Wave I of Add Health (Harris et al. 2019)—that may be of interest. As a function of the way the analytic sample becomes a selected portion of the full sample, we observe a decrease in environmental variance. These decreases will lead to even further reductions in our power to detect GxE effects; in particular, power analyses motivated by environmental variation observed in the full sample are likely to overstate true power given that empirical work will then take place with reduced environmental variation. Beyond power concerns, such selection can lead to a reduction in density in certain regions of the distribution of the measured environment that will increase the challenge of identifying the relevant functional form in that region.

Recommendations. Issues concerning selection require careful attention. Figure 3 suggests that they may have implications that need to be accounted for in other aspects of study design (i.e., are power analyses based on the appropriate quantities?). We further suggest two ways that research may approach these issues. First, the selection issues discussed here have implications for generalizability. Some forms of this problem are obvious. Given, for example, the problems of analysis in ancestrally heterogeneous samples and the subsequent work on samples of relatively limited genetic diversity, it would be imprudent to interpret GxE findings from such a study as applying in the broader population containing a fuller spectrum of genetic diversity. But it may also be the case that selection introduces other factors that limit generalizability. For example, long-lived smokers may be quite different

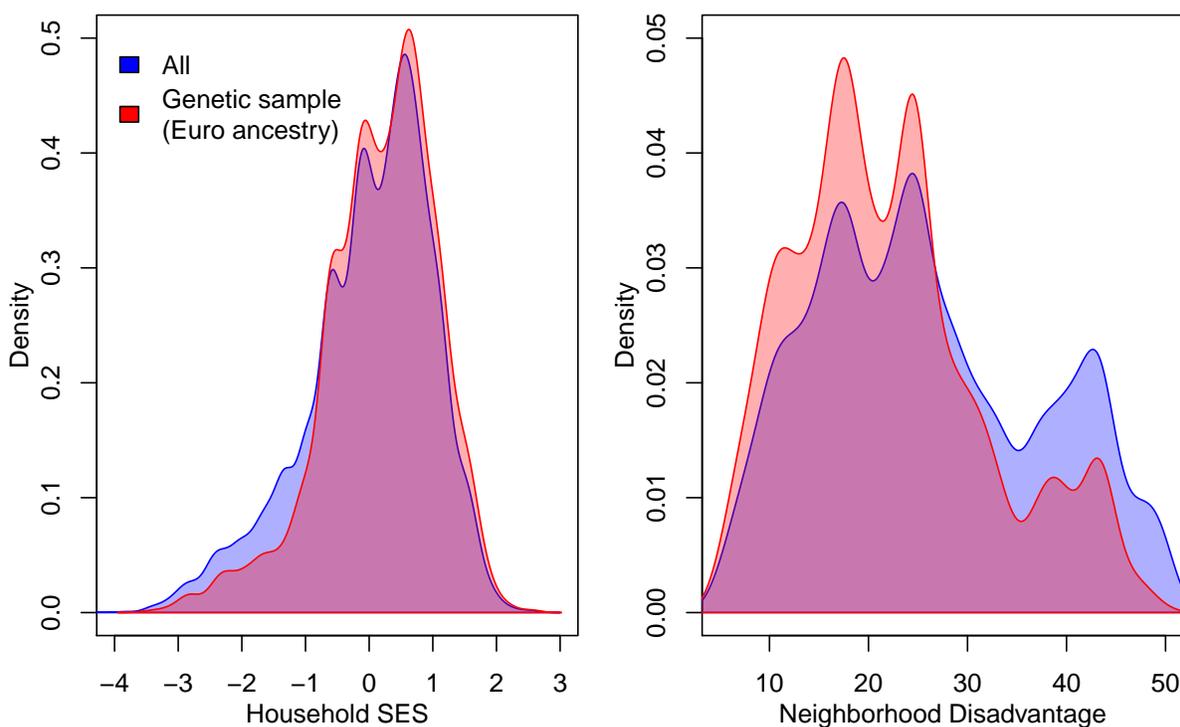


Figure 3: Distributions of two key environmental variables (household socioeconomic status [SES] and neighborhood disadvantage) taken from Wave I of Add Health (Harris et al. 2019). Note the reduction in variation of the distribution for the analytic sample (in red) versus that of the full sample (in blue). Reductions in the standard deviation are 11 percent for SES and 14 percent for neighborhood disadvantage; in variance terms, the reductions are 20 percent and 26 percent, respectively.

from the general population (Levine and Crimmins 2014); inference based on such samples may be misleading.

Second, on the analytic side, attempts to model the relevant selection processes may lead to direct insights into the degree of generalizability of patterns. For example, researchers may examine how results change when using formal techniques that correct for selection (e.g., inverse probability weighting [Cole and Hernán 2008]). Even less comprehensive analyses of selection processes may lead to insights about the nature of the analytic sample and offer guides to generalization that researchers can communicate alongside the relevant empirical results.

Conclusion

GxE characterizes both the environmental contingency of genetically linked processes and the genetic contingency of environmentally linked processes. In our view, GxE studies involving human behavior and polygenic scores may offer valuable insights but are also at risk of repeating many of the mistakes made by previous eras of research (e.g., the candidate gene era). Our goal has been to emphasize the need

for careful thinking about the rationale and methods underlying investigations of GxE.

In particular, we highlighted four issues—selection of the relevant environment, analysis of coarsened outcomes, the role of measurement error, and issues of sample selection—that deserve additional scrutiny in future research. We also attempted to offer recommendations for beginning to address each problem. We readily acknowledge that ours are a relatively modest set of recommendations that will not fully resolve the vast range of analytic and inferential challenges associated with GxE research with PGSs.

An overarching goal of research examining the combined genetic and environmental contributions to human behavior is to help construct *useful* models of human behavior. In our view, *useful* models avoid unnecessary complexity when accounting for messy data. At its best, GxE research can help inform the construction of such models by parsimoniously showcasing complexities from empirical reality that need to be accounted for. For instance, GxE research can help reveal important heterogeneity in developmental processes, treatment responses, and policy effects. To be informative, however, we must exercise care. Otherwise, GxE research threatens to introduce confusion into the already challenging landscape of social and behavioral science research.

Notes

- 1 Exactly how predictive a PGS is of a given trait depends on both on the trait's heritability and the sample size of the GWAS used to derive the effect size estimates; see Figure 2 of Harden and Koellinger (2020).
- 2 We note that one could alternatively discuss environmental effects differing as a function of genetics; we utilize the original formulation in this article but note that the latter may occasionally be the more germane.
- 3 In practice, polygenic scores may contain information on correlated nongenetic factors (e.g., population stratification and dynastic effects like genetic nurture) in addition to true genetic risk (Morris et al. 2020).
- 4 Specifically, convex lenses have such image-inverting properties. Here we use “lens” as shorthand for convex lens but note that concave lenses do not have this property.
- 5 For a similar observation in a different context, see https://twitter.com/Joni_Coleman/status/1220332653599186946?s=20.
- 6 Code available here: <https://gist.github.com/ben-domingue/6f14e3c4532ecb62df5f6e0c44c60411>.

References

- Barr, Peter B., Judy Silberg, Danielle M. Dick, and Hermine H. Maes. 2018. “Childhood Socioeconomic Status and Longitudinal Patterns of Alcohol Problems: Variation across Etiological Pathways in Genetic Risk.” *Social Science & Medicine* 209:51–58. <https://doi.org/10.1016/j.socscimed.2018.05.027>.
- Belsky, Daniel W., Avshalom Caspi, Louise Arseneault, David L. Corcoran, Benjamin W. Domingue, Kathleen Mullan Harris, Renate M. Houts, Jonathan S. Mill, Terrie E. Moffitt,

- Joseph Prinz, Karen Sugden, Jasmin Wertz, Benjamin Williams, and Candice L. Odgers. 2019. "Genetics and the Geography of Health, Behaviour and Attainment." *Nature Human Behaviour* 3:576–86. <https://doi.org/10.1038/s41562-019-0562-1>.
- Belsky, Daniel W., Benjamin W. Domingue, Robbee Wedow, Louise Arseneault, Jason D. Boardman, Avshalom Caspi, Dalton C. Conley, Jason M. Fletcher, Jeremy Freese, Pamela Herd, Terrie E. Moffitt, Richie Poulton, Kamil Sicinski, Jasmin Wertz, and Kathleen Mullan Harris. 2018. "Genetic Analysis of Social-Class Mobility in Five Longitudinal Studies." *Proceedings of the National Academy of Sciences* 115(31):E7275–84. <https://doi.org/10.1073/pnas.1801238115>.
- Belsky, Daniel W., and K. Paige Harden. 2019. "Phenotypic Annotation: Using Polygenic Scores to Translate Discoveries from Genome-Wide Association Studies from the Top Down." *Current Directions in Psychological Science* 28(1):82–90. <https://doi.org/10.1177/0963721418807729>.
- Belsky, Jay, and Michael Pluess. 2009. "Beyond Diathesis Stress: Differential Susceptibility to Environmental Influences." *Psychological Bulletin* 135(6):885–908. <https://doi.org/10.1037/a0017376>.
- Boardman, Jason D., Lisa L. Barnes, Robert S. Wilson, Denis A. Evans, and Carlos F. Mendes de Leon. 2012. "Social Disorder, APOE-E4 Genotype, and Change in Cognitive Function among Older Adults Living in Chicago." *Social Science & Medicine* 74(10):1584–90. <https://doi.org/10.1016/j.socscimed.2012.02.012>.
- Boardman, Jason D., Jonathan Daw, and Jeremy Freese. 2013. "Defining the Environment in Gene–Environment Research: Lessons from Social Epidemiology." *American Journal of Public Health* 103(S1):S64–72. <https://doi.org/10.2105/AJPH.2013.301355>.
- Boyle, Evan A., Yang I. Li, and Jonathan K. Pritchard. 2017. "An Expanded View of Complex Traits: From Polygenic to Omnigenic." *Cell* 169(7):1177–86. <https://doi.org/10.1016/j.cell.2017.05.038>.
- Braudt, David B., and Kathleen Mullan Harris. 2018. "Polygenic Scores (PGSs) in the National Longitudinal Study of Adolescent to Adult Health (Add Health)–Release 1." Carolina Population Center, University of North Carolina at Chapel Hill. <https://doi.org/10.17615/C6M372>.
- Briley, Daniel A., Jonathan Livengood, Jaime Derringer, Elliot M. Tucker-Drob, R. Chris Fraley, and Brent W. Roberts. 2019. "Interpreting Behavior Genetic Models: Seven Developmental Processes to Understand." *Behavior Genetics* 49(2):196–210. <https://doi.org/10.1007/s10519-018-9939-6>.
- Cesarini, David, and Peter M. Visscher. 2017. "Genetics and Educational Attainment." *NPJ Science of Learning* 2(1):4. <https://doi.org/10.1038/s41539-017-0005-6>.
- Chabris, Christopher F., James J. Lee, David Cesarini, Daniel J. Benjamin, and David I. Laibson. 2015. "The Fourth Law of Behavior Genetics." *Current Directions in Psychological Science* 24(4):304–12. <https://doi.org/10.1177/0963721415580430>.
- Cole, Stephen R., and Miguel A. Hernán. 2008. "Constructing Inverse Probability Weights for Marginal Structural Models." *American Journal of Epidemiology* 168(6):656–64. <https://doi.org/10.1093/aje/kwn164>.
- Conley, Dalton, and Jason Fletcher. 2017. *The Genome Factor: What the Social Genomics Revolution Reveals about Ourselves, Our History, and the Future*. Princeton: Princeton University Press. <https://doi.org/10.1515/9781400883240>.
- Conley, Dalton, Rebecca Johnson, Ben Domingue, Christopher Dawes, Jason Boardman, and Mark Siegal. 2018. "A Sibling Method for Identifying VQTLs." *PloS One* 13(4):e0194541. <https://doi.org/10.1371/journal.pone.0194541>.

- Conley, Dalton, Thomas M. Laidley, Jason D. Boardman, and Benjamin W. Domingue. 2016. "Changing Polygenic Penetrance on Phenotypes in the 20th Century among Adults in the US Population." *Scientific Reports* 6:30348. <https://doi.org/10.1038/srep30348>.
- de Vlaming, Ronald, Magnus Johannesson, Patrik K. E. Magnusson, M. Arfan Ikram, and Peter M. Visscher. 2017. "Equivalence of LD-Score Regression and Individual-Level-Data Methods." *bioRxiv*. Preprint, submitted October 31. <https://www.biorxiv.org/content/10.1101/211821v1>.
- DiPrete, Thomas A., Casper A. P. Burik, and Philipp D. Koellinger. 2018. "Genetic Instrumental Variable Regression: Explaining Socioeconomic and Health Outcomes in Nonexperimental Data." *Proceedings of the National Academy of Sciences* 115(22):E4970–79. <https://doi.org/10.1073/pnas.1707388115>.
- Domingue, Benjamin W., Daniel W. Belsky, Amal Harrati, Dalton Conley, David Weir, and Jason Boardman. 2017. "Mortality Selection in a Genetic Sample and Implications for Association Studies." *International Journal of Epidemiology* 46(4):1285–94. <https://doi.org/10.1093/ije/dyx041>.
- Dudbridge, Frank. 2013. "Power and Predictive Accuracy of Polygenic Risk Scores." *PLoS Genetics* 9(3):e1003348. <https://doi.org/10.1371/journal.pgen.1003348>.
- Dudbridge, Frank. 2016. "Polygenic Epidemiology." *Genetic Epidemiology* 40(4):268–72. <https://doi.org/10.1002/gepi.21966>.
- Dudbridge, Frank, and Olivia Fletcher. 2014. "Gene-Environment Dependence Creates Spurious Gene-Environment Interaction." *American Journal of Human Genetics* 95(3):301–7. <https://doi.org/10.1016/j.ajhg.2014.07.014>.
- Duncan, Greg J., and Katherine Magnuson. 2012. "Socioeconomic Status and Cognitive Functioning: Moving from Correlation to Causation." *Wiley Interdisciplinary Reviews: Cognitive Science* 3(3):377–86. <https://doi.org/10.1002/wcs.1176>.
- Duncan, L., H. Shen, B. Gelaye, J. Meijssen, K. Ressler, M. Feldman, R. Peterson, and B. Domingue. 2019. "Analysis of Polygenic Risk Score Usage and Performance in Diverse Human Populations." *Nature Communications* 10(1):3328. <https://doi.org/10.1038/s41467-019-11112-0>.
- Duncan, Laramie E., and Matthew C. Keller. 2011. "A Critical Review of the First 10 Years of Candidate Gene-by-Environment Interaction Research in Psychiatry." *American Journal of Psychiatry* 168(10):1041–49. <https://doi.org/10.1176/appi.ajp.2011.11020191>.
- Ellis, Bruce J., W. Thomas Boyce, Jay Belsky, Marian J. Bakermans-Kranenburg, and Marinus H. Van Ijzendoorn. 2011. "Differential Susceptibility to the Environment: An Evolutionary–Neurodevelopmental Theory." *Development and Psychopathology* 23(01):7–28. <https://doi.org/10.1017/S0954579410000611>.
- Elwert, Felix, and Christopher Winship. 2014. "Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable." *Annual Review of Sociology* 40:31–53. <https://doi.org/10.1146/annurev-soc-071913-043455>.
- Feldman, M. W., and R. C. Lewontin. 1975. "The Heritability Hang-Up." *Science* 190(4220):1163–68. <https://doi.org/10.1126/science.1198102>.
- Fletcher, Jason M., and Dalton Conley. 2013. "The Challenge of Causal Inference in Gene–Environment Interaction Research: Leveraging Research Designs from the Social Sciences." *American Journal of Public Health* 103(S1):S42–45. <https://doi.org/10.2105/AJPH.2013.301290>.
- Freese, Jeremy. 2018. "The Arrival of Social Science Genomics." *Contemporary Sociology* 47(5):524–36. <https://doi.org/10.1177/0094306118792214a>.

- Gould, Karen L., William L. Coventry, Richard K. Olson, and Brian Byrne. 2018. "Gene-Environment Interactions in ADHD: The Roles of SES and Chaos." *Journal of Abnormal Child Psychology* 46(2):251–263. <https://doi.org/10.1007/s10802-017-0268-7>.
- Grotzinger, Andrew D., Mijke Rhemtulla, Ronald de Vlaming, Stuart J. Ritchie, Travis T. Mallard, W. David Hill, Hill F. Ip, Riccardo E. Marioni, Andrew M. McIntosh, Ian J. Deary, Philipp D. Koellinger, K. Paige Harden, Michel G. Nivard, and Elliot M. Tucker-Drob. 2019. "Genomic Structural Equation Modelling Provides Insights into the Multivariate Genetic Architecture of Complex Traits." *Nature Human Behaviour* 3(5):513–25. <https://doi.org/10.1038/s41562-019-0566-x>.
- Harden, K. Paige, Benjamin W. Domingue, Daniel W. Belsky, Jason D. Boardman, Robert Crosnoe, Margherita Malanchini, Michel Nivard, Elliot M. Tucker-Drob, and Kathleen Mullan Harris. 2020. "Genetic Associations with Mathematics Tracking and Persistence in Secondary School." *NPJ Science of Learning* 5(1):1–8. <https://doi.org/10.1038/s41539-020-0060-2>.
- Harden, K. Paige, and Philipp D. Koellinger. 2020. "Using Genetics for Social Science." *Nature Human Behaviour* 4:567. <https://doi.org/10.1038/s41562-020-0862-5>.
- Harris, Kathleen Mullan, Carolyn Tucker Halpern, Eric A. Whitsel, Jon M. Hussey, Ley A. Killea-Jones, Joyce Tabor, and Sarah C. Dean. 2019. "Cohort Profile: The National Longitudinal Study of Adolescent to Adult Health (Add Health)." *International Journal of Epidemiology* 48(5):1415. <https://doi.org/10.1093/ije/dyz115>.
- Hutcheon, Jennifer A., Arnaud Chiolero, and James A. Hanley. 2010. "Random Measurement Error and Regression Dilution Bias." *BMJ* 340:c2289. <https://doi.org/10.1136/bmj.c2289>.
- Jaffee, Sara R., and Thomas S. Price. 2007. "Gene-Environment Correlations: A Review of the Evidence and Implications for Prevention of Mental Illness." *Molecular Psychiatry* 12(5):432–42. <https://doi.org/10.1038/sj.mp.4001950>.
- Johnson, Sara B., Jenna L. Riis, and Kimberly G. Noble. 2016. "State of the Art Review: Poverty and the Developing Brain." *Pediatrics* 137(4):e20153075. <https://doi.org/10.1542/peds.2015-3075>.
- Keller, Matthew C. 2014. "Gene \times Environment Interaction Studies Have Not Properly Controlled for Potential Confounders: The Problem and the (Simple) Solution." *Biological Psychiatry* 75(1):18–24. <https://doi.org/10.1016/j.biopsych.2013.09.006>.
- Lambert, Samuel A., Laurent Gil, Simon Jupp, Scott C. Ritchie, Yu Xu, Annalisa Buniello, Gad Abraham, Michael Chapman, Helen Parkinson, John Danesh, Jacqueline A. MacArthur, and Michael Inouye. 2020. "The Polygenic Score Catalog: An Open Database for Reproducibility and Systematic Evaluation." *medRxiv*. Preprint, submitted May 23. <https://www.medrxiv.org/content/10.1101/2020.05.20.20108217v1>.
- Lee, James J., Robbee Wedow, Aysu Okbay, Edward Kong, Omeed Maghziyan, Meghan Zacher, Tuan Anh Nguyen-Viet, Peter Bowers, Julia Sidorenko, Richard Karlsson Linnér, Mark Alan Fontana, Tushar Kundu, Chanwook Lee, Hui Li, Ruoxi Li, Rebecca Royer, Pascal N. Timshel, Raymond K. Walters, Emily A. Willoughby, Loïc Yengo, Maris Alver, Yanchun Bao, David W. Clark, Felix R. Day, Nicholas A. Furlotte, Peter K. Joshi, Kathryn E. Kemper, Aaron Kleinman, Claudia Langenberg, Reedik Mägi, Joey W. Trampush, Shefali Setia Verma, Yang Wu, Max Lam, Jing Hua Zhao, Zhili Zheng, Jason D. Boardman, Harry Campbell, Jeremy Freese, Kathleen Mullan Harris, Caroline Hayward, Pamela Herd, Meena Kumari, Todd Lencz, Jian'an Luan, Anil K. Malhotra, Andres Metspalu, Lili Milani, Ken K. Ong, John R. B. Perry, David J. Porteous, Marylyn D. Ritchie, Melissa C. Smart, Blair H. Smith, Joyce Y. Tung, Nicholas J. Wareham, James F. Wilson, Jonathan

- P. Beauchamp, Dalton C. Conley, Tõnu Esko, Steven F. Lehrer, Patrik K. E. Magnusson, Sven Oskarsson, Tune H. Pers, Matthew R. Robinson, Kevin Thom, Chelsea Watson, Christopher F. Chabris, Michelle N. Meyer, David I. Laibson, Jian Yang, Magnus Johannesson, Philipp D. Koellinger, Patrick Turley, Peter M. Visscher, Daniel J. Benjamin, and David Cesarini. 2018. "Gene Discovery and Polygenic Prediction from a Genome-Wide Association Study of Educational Attainment in 1.1 Million Individuals." *Nature Genetics* 50(8):1112–21. <https://doi.org/10.1038/s41588-018-0147-3>.
- Levine, Morgan, and Eileen Crimmins. 2014. "Not All Smokers Die Young: A Model for Hidden Heterogeneity within the Human Population." *PLoS One* 9(2):e87403. <https://doi.org/10.1371/journal.pone.0087403>.
- Lubinski, David, and Lloyd G. Humphreys. 1990. "Assessing Spurious 'Moderator Effects': Illustrated Substantively with the Hypothesized ('Synergistic') Relation between Spatial and Mathematical Ability." *Psychological Bulletin* 107(3):385–93. <https://doi.org/10.1037/0033-2909.107.3.385>.
- MacCallum, Robert C., and Corinne M. Mar. 1995. "Distinguishing between Moderator and Quadratic Effects in Multiple Regression." *Psychological Bulletin* 118(3):405–21. <https://doi.org/10.1037/0033-2909.118.3.405>.
- Martin, Alicia R., Masahiro Kanai, Yoichiro Kamatani, Yukinori Okada, Benjamin M. Neale, and Mark J. Daly. 2019. "Clinical Use of Current Polygenic Risk Scores May Exacerbate Health Disparities." *Nature Genetics* 51(4):584–91. <https://doi.org/10.1038/s41588-019-0379-x>.
- Martin, Joanna, Kate Tilling, Leon Hubbard, Evie Stergiakouli, Anita Thapar, George Davey Smith, Michael C. O'Donovan, and Stanley Zammit. 2016. "Association of Genetic Risk for Schizophrenia with Nonparticipation over Time in a Population-Based Cohort Study." *American Journal of Epidemiology* 183(12):1149–58. <https://doi.org/10.1093/aje/kww009>.
- McClelland, Gary H., and Charles M. Judd. 1993. "Statistical Difficulties of Detecting Interactions and Moderator Effects." *Psychological Bulletin* 114(2):376–90. <https://doi.org/10.1037/0033-2909.114.2.376>.
- Meisner, Allison, Prosenjit Kundu, and Nilanjan Chatterjee. 2019. "Case-Only Analysis of Gene-Environment Interactions Using Polygenic Risk Scores." *American Journal of Epidemiology* 188(11):2013–20. <https://doi.org/10.1093/aje/kwz175>.
- Mills, Melinda C., and Charles Rahal. 2019. "A Scientometric Review of Genome-Wide Association Studies." *Communications Biology* 2(1):9. <https://doi.org/10.1038/s42003-018-0261-x>.
- Mills, Melinda C., and Charles Rahal. 2020. "The GWAS Diversity Monitor Tracks Diversity by Disease in Real Time." *Nature Genetics* 52(3):242–43. <https://doi.org/10.1038/s41588-020-0580-y>.
- Mills, Melinda C., and Felix C. Tropf. 2020. "Sociology, Genetics, and the Coming of Age of Sociogenomics." *Annual Review of Sociology* 46:553–81. <https://doi.org/10.1146/annurev-soc-121919-054756>.
- Morris, Tim T., Neil M. Davies, Gibran Hemani, and George Davey Smith. 2020. "Population Phenomena Inflate Genetic Associations of Complex Social Traits." *Science Advances* 6(16):eaay0328. <https://doi.org/10.1126/sciadv.aay0328>.
- Mostafavi, Hakhamanesh, Arbel Harpak, Ipsita Agarwal, Dalton Conley, Jonathan K. Pritchard, and Molly Przeworski. 2020. "Variable Prediction Accuracy of Polygenic Scores within an Ancestry Group." *eLife* 9:e48376. <https://doi.org/10.7554/eLife.48376>.

- Musci, Rashelle J., Amie F. Bettencourt, Danielle Sisto, Brion Maher, Katherine Masyn, and Nicholas S. Ialongo. 2019. "Violence Exposure in an Urban City: A GxE Interaction with Aggressive and Impulsive Behaviors." *Journal of Child Psychology and Psychiatry* 60(1):72–81. <https://doi.org/10.1111/jcpp.12966>.
- Obradović, Jelena, and W. Thomas Boyce. 2009. "Individual Differences in Behavioral, Physiological, and Genetic Sensitivities to Contexts: Implications for Development and Adaptation." *Developmental Neuroscience* 31(4):300–8. <https://doi.org/10.1159/000216541>.
- Okbay, Aysu, Daniel Benjamin, and Peter Visscher. 2018. "Documentation." [Construction of Wisconsin Longitudinal Study Polygenic Scores.] University of Wisconsin—Madison. [https://www.ssc.wisc.edu/wlsresearch/documentation/GWAS/Lee_et_al_\(2018\)_PGS_WLS.pdf](https://www.ssc.wisc.edu/wlsresearch/documentation/GWAS/Lee_et_al_(2018)_PGS_WLS.pdf).
- Oliynyk, Roman Teo. 2019. "Age-Related Late-Onset Disease Heritability Patterns and Implications for Genome-Wide Association Studies." *PeerJ* 7:e7168. <https://doi.org/10.7717/peerj.7168>.
- Pearson, Thomas A., and Teri A. Manolio. 2008. "How to Interpret a Genome-Wide Association Study." *JAMA* 299(11):1335–44. <https://doi.org/10.1001/jama.299.11.1335>.
- Pirastu, Nicola, Mattia Cordioli, Priyanka Nandakumar, Gianmarco Mignogna, Abdel Abdellaoui, Ben Hollis, Masahiro Kanai, Veera Manikandan Rajagopal, Pietro Della Briotta Parolo, Nikolas Baya, Caitlin Carey, Juha Karjalainen, Thomas D. Als, Matthijs D. Van der Zee, Felix R. Day, Ken K. Ong, Finngen Study, Me Research Team, Consortium iPSYCH, Takayuki Morisaki, Eco de Geus, Rino Bellocco, Yukinori Okada, Anders D. Børglum, Peter Joshi, Adam Auton, David Hings, Benjamin M. Neale, Raymond K. Walters, Michel G. Nivard, John R. B. Perry, and Andrea Ganna. 2020. "Genetic Analyses Identify Widespread Sex-Differential Participation Bias." *bioRxiv*. Preprint, submitted March 23. <https://www.biorxiv.org/content/10.1101/2020.03.22.001453v1>.
- Polderman, Tinca, Beben Benyamin, Christiaan A. De Leeuw, Patrick F. Sullivan, Arjen Van Bochoven, Peter M. Visscher, and Danielle Posthuma. 2015. "Meta-Analysis of the Heritability of Human Traits Based on Fifty Years of Twin Studies." *Nature Genetics* 47(7):702–9. <https://doi.org/10.1038/ng.3285>.
- Purcell, Shaun. 2002. "Variance Components Models for Gene–Environment Interaction in Twin Analysis." *Twin Research and Human Genetics* 5(6):554–71. <https://doi.org/10.1375/136905202762342026>.
- Rosenquist, James Niels, Steven F. Lehrer, A. James O'Malley, Alan M. Zaslavsky, Jordan W. Smoller, and Nicholas A. Christakis. 2015. "Cohort of Birth Modifies the Association between FTO Genotype and BMI." *Proceedings of the National Academy of Sciences* 112(2):354–59. <https://doi.org/10.1073/pnas.1411893111>.
- Sugrue, Leo P., and Rahul S. Desikan. 2019. "What Are Polygenic Scores and Why Are They Important?" *JAMA* 321(18):1820–21. <https://doi.org/10.1001/jama.2019.3893>.
- Taylor, Amy E., Hannah J. Jones, Hannah Sallis, Jack Euesden, Evie Stergiakouli, Neil M. Davies, Stanley Zammit, Debbie A. Lawlor, Marcus R. Munafò, George Davey Smith, and Kate Tilling. 2018. "Exploring the Association of Genetic Factors with Participation in the Avon Longitudinal Study of Parents and Children." *International Journal of Epidemiology* 47(4):1207–16. <https://doi.org/10.1093/ije/dyy060>.
- Trejo, Sam, Daniel Belsky, Jason Boardman, Jeremy Freese, Kathleen Harris, Pam Herd, Kamil Sicinski, and Benjamin Domingue. 2018. "Schools as Moderators of Genetic Associations with Life Course Attainments: Evidence from the WLS and Add Health." *Sociological Science* 5:513–40. <https://doi.org/10.15195/v5.a22>.

- Tropf, Felix C., S. Hong Lee, Renske M. Verweij, Gert Stulp, Peter J. van der Most, Ronald de Vlaming, Andrew Bakshi, Daniel A. Briley, Charles Rahal, Robert Hellpap, Anastasia N. Iliadou, Tõnu Esko, Andres Metspalu, Sarah E. Medland, Nicholas G. Martin, Nicola Barban, Harold Snieder, Matthew R. Robinson, and Melinda C. Mills. 2017. "Hidden Heritability Due to Heterogeneity across Seven Populations." *Nature Human Behaviour* 1:757–65. <https://doi.org/10.1038/s41562-017-0195-1>.
- Troth, Ashley, Joshua R. Puzey, Rebecca S. Kim, John H. Willis, and John K. Kelly. 2018. "Selective Trade-Offs Maintain Alleles Underpinning Complex Trait Variation in Plants." *Science* 361(6401):475–78.
- Tucker-Drob, Elliot M. 2017. "Measurement Error Correction of Genome-Wide Polygenic Scores in Prediction Samples." *bioRxiv*. Preprint, submitted July 19. <https://www.biorxiv.org/content/10.1101/165472v1>.
- van der Wal, Willem M., and Ronald B. Geskus. 2011. "IPW: An R Package for Inverse Probability Weighting." *Journal of Statistical Software* 43(13):1–23. <https://doi.org/10.18637/jss.v043.i13>.
- Visscher, Peter M., Naomi R. Wray, Qian Zhang, Pamela Sklar, Mark I. McCarthy, Matthew A. Brown, and Jian Yang. 2017. "10 Years of GWAS Discovery: Biology, Function, and Translation." *American Journal of Human Genetics* 101(1):5–22. <https://doi.org/10.1016/j.ajhg.2017.06.005>.
- Wang, Huanwei, Futao Zhang, Jian Zeng, Yang Wu, Kathryn E. Kemper, Angli Xue, Min Zhang, Joseph E. Powell, Michael E. Goddard, Naomi R. Wray, Peter M. Visscher, Allan F. McRae, and Jian Yang. 2019. "Genotype-by-Environment Interactions Inferred from Genetic Effects on Phenotypic Variability in the UK Biobank." *Science Advances* 5(8):eaaw3538. <https://doi.org/10.1126/sciadv.aaw3538>.
- Ware, Erin B., Lauren L. Schmitz, Jessica D. Faul, Arianna Gard, Colter Mitchell, Jennifer A. Smith, Wei Zhao, David Weir, and Sharon L. R. Kardina. 2017. "Heterogeneity in Polygenic Scores for Common Human Traits." *bioRxiv*. Preprint, submitted February 5. <https://www.biorxiv.org/content/10.1101/106062v1>.
- Widaman, Keith F., Jonathan L. Helm, Laura Castro-Schilo, Michael Pluess, Michael C. Stallings, and Jay Belsky. 2012. "Distinguishing Ordinal and Disordinal Interactions." *Psychological Methods* 17(4):615–22. <https://doi.org/10.1037/a0030003>.
- Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher. 2011. "GCTA: A Tool for Genome-Wide Complex Trait Analysis." *American Journal of Human Genetics* 88(1):76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>.
- Yang, Jian, Ruth J. F. Loos, Joseph E. Powell, Sarah E. Medland, Elizabeth K. Speliotes, Daniel I. Chasman, Lynda M. Rose, Gudmar Thorleifsson, Valgerdur Steinthorsdottir, Reedik Magi, Lindsay Waite, Albert Vernon Smith, Laura M. Yerges-Armstrong, Keri L. Monda, David Hadley, Anubha Mahajan, Guo Li, Karen Kapur, Veronique Vitart, Jennifer E. Huffman, Sophie R. Wang, Cameron Palmer, Tonu Esko, Krista Fischer, Jing Hua Zhao, Ayse Demirkan, Aaron Isaacs, Mary F. Feitosa, Jian'an Luan, Nancy L. Heard-Costa, Charles White, Anne U. Jackson, Michael Preuss, Andreas Ziegler, Joel Eriksson, Zoltan Kutalik, Francesca Frau, Ilja M. Nolte, Jana V. Van Vliet-Ostaptchouk, Jouke-Jan Hottenga, Kevin B. Jacobs, Niek Verweij, Anuj Goel, Carolina Medina-Gomez, Karol Estrada, Jennifer Lynn Bragg-Gresham, Serena Sanna, Carlo Sidore, Jonathan Tyrer, Alexander Teumer, Inga Prokopenko, Massimo Mangino, Cecilia M. Lindgren, Themistocles L. Assimes, Alan R. Shuldiner, Jennie Hui, John P. Beilby, Wendy L. McArdle, Per Hall, Talin Haritunians, Lina Zgaga, Ivana Kolcic, Ozren Polasek, Tatijana Zemunik, Ben A. Oostra, M. Juhani Juntila, Henrik Gronberg, Stefan Schreiber, Annette Peters, Andrew A. Hicks, Jonathan Stephens, Nicola S. Foad, Jaana Laitinen, Anneli Pouta, Marika Kaakinen, Gonneke

- Willemsen, Jacqueline M. Vink, Sarah H. Wild, Gerjan Navis, Folkert W. Asselbergs, Georg Homuth, Ulrich John, Carlos Iribarren, Tamara Harris, Lenore Launer, Vilmundur Gudnason, Jeffrey R. O'Connell, Eric Boerwinkle, Gemma Cadby, Lyle J. Palmer, Alan L. James, Arthur W. Musk, Erik Ingelsson, Bruce M. Psaty, Jacques S. Beckmann, Gerard Waeber, Peter Vollenweider, Caroline Hayward, Alan F. Wright, Igor Rudan, Leif C. Groop, Andres Metspalu, Kay Tee Khaw, Cornelia M. van Duijn, Ingrid B. Borecki, Michael A. Province, Nicholas J. Wareham, Jean-Claude Tardif, Heikki V. Huikuri, L. Adrienne Cupples, Larry D. Atwood, Caroline S. Fox, Michael Boehnke, Francis S. Collins, Karen L. Mohlke, Jeanette Erdmann, Heribert Schunkert, Christian Hengstenberg, Klaus Stark, Mattias Lorentzon, Claes Ohlsson, Daniele Cusi, Jan A. Staessen, Melanie M. Van der Klauw, Peter P. Pramstaller, Sekar Kathiresan, Jennifer D. Jolley, Samuli Ripatti, Marjo-Riitta Jarvelin, Eco J. C. de Geus, Dorret I. Boomsma, Brenda Penninx, James F. Wilson, Harry Campbell, Stephen J. Chanock, Pim van der Harst, Anders Hamsten, Hugh Watkins, Albert Hofman, Jacqueline C. Witteman, M. Carola Zillikens, Andre G. Uitterlinden, Fernando Rivadeneira, M. Carola Zillikens, Lambertus A. Kiemeny, Sita H. Vermeulen, Goncalo R. Abecasis, David Schlessinger, Sabine Schipf, Michael Stumvoll, Anke Tonjes, Tim D. Spector, Kari E. North, Guillaume Lettre, Mark I. McCarthy, Sonja I. Berndt, Andrew C. Heath, Pamela A. F. Madden, Dale R. Nyholt, Grant W. Montgomery, Nicholas G. Martin, Barbara McKnight, David P. Strachan, William G. Hill, Harold Snieder, Paul M. Ridker, Unnur Thorsteinsdottir, Kari Stefansson, Timothy M. Frayling, Joel N. Hirschhorn, Michael E. Goddard, and Peter M. Visscher. 2012. "FTO Genotype Is Associated with Phenotypic Variability of Body Mass Index." *Nature* 490(7419):267–72. <https://doi.org/10.1038/nature11401>.
- Young, Alexander I., Fabian L. Wauthier, and Peter Donnelly. 2018. "Identifying Loci Affecting Trait Variability and Detecting Interactions in Genome-Wide Association Studies." *Nature Genetics* 50(11):1608–14. <https://doi.org/10.1038/s41588-018-0225-6>.
- Zajacova, Anna, and Sarah A. Burgard. 2013. "Healthier, Wealthier, and Wiser: A Demonstration of Compositional Changes in Aging Cohorts Due to Selective Mortality." *Population Research and Policy Review* 32(3):311–24. <https://doi.org/10.1007/s11113-013-9273-x>.

Acknowledgments: This work has been supported by the Russell Sage Foundation and the Ford Foundation (grant 96-17-04). S.T. was supported by the National Science Foundation (grant DGE-1656518) and the Institute of Education Sciences (grant R305B140009). E.M.T.-D. was supported by the National Institutes of Health (grants R01AG054628, R01MH120219, and R01HD083613) and by the Jacobs Foundation. Any opinions expressed are those of the authors alone and should not be construed as representing the opinions of any foundation. The authors would like to thank Jason Boardman and Jason Fletcher for comments on an early draft of this article.

Benjamin W. Domingue: Graduate School of Education, Stanford University.
E-mail: bdomingu@stanford.edu.

Sam Trejo: La Follette School of Public Affairs & Department of Sociology, University of Wisconsin–Madison. E-mail: sam.trejo@wisc.edu.

Emma Armstrong-Carter: Graduate School of Education, Stanford University.
E-mail: emmaac@stanford.edu.

Elliot M. Tucker-Drob: Department of Psychology and Population Research Center, University of Texas at Austin. E-mail: tuckerdrob@utexas.edu.