Author(s): Pehkonen, Jaakko; Viinikainen, Jutta; Böckerman, Petri; Lehtimäki, Terho; Pitkänen, Niina; Raitakari, Olli

Title: The challenges of GxE research : A rejoinder

Year: 2017

Version:

Please cite the original version:

Pehkonen, J., Viinikainen, J., Böckerman, P., Lehtimäki, T., Pitkänen, N., & Raitakari, O. (2017). The challenges of GxE research : A rejoinder. Social Science and Medicine, 188, 204-205. https://doi.org/10.1016/j.socscimed.2017.07.010

Accepted Manuscript

The challenges of GxE research: A rejoinder

Jaakko Pehkonen, Jutta Viinikainen, Petri Böckerman, Terho Lehtimäki, Niina Pitkänen, Olli Raitakari

PII: S0277-9536(17)30443-4

DOI: 10.1016/j.socscimed.2017.07.010

Reference: SSM 11320

To appear in: *Social Science & Medicine*

Received Date: 13 July 2017

Revised Date: 0277-9536 0277-9536

Accepted Date: 18 July 2017

Please cite this article as: Pehkonen, J., Viinikainen, J., Böckerman, P., Lehtimäki, T., Pitkänen, N., Raitakari, O., The challenges of GxE research: A rejoinder, *Social Science & Medicine* (2017), doi: 10.1016/j.socscimed.2017.07.010.

The Challenges of GxE Research: A Rejoinder

1. General remarks

XX (2017) correctly emphasizes three important points related to empirical research on genetic–environmental (GxE) interactions in human populations. First, the commentary stresses the importance of modeling genotypes as a moderator of social influences on human behavior. This research strategy, which directly incorporates genetic and biological information into data on various social behaviors, augments prior approaches based on twin and adoption designs. As noted in the commentary, large datasets containing genotyped individuals, such as WLS, HRS and Add Health, and subsequently incorporated into social cross-sectional or longitudinal surveys, have provided new insights for social scientists.

Second, the commentary emphasizes that research on GxE interactions may be of substantive social and economic importance. In particular, GxE analyses may identify how exogenous interventions modify genetic propensities. Thus, the existence of heterogeneous treatment effects (responses to stimuli in the environment) and their proper identification can provide policy-relevant information about the marginal returns of public health investments and eventually improve the targeting of health interventions.

Third, the commentary illustrates the key challenges of such empirical endeavors, highlighting two major concerns. The first concern relates to the measurement of genetic contributions—whether the distributions of candidate genes used in empirical analyses are randomly assigned across environments and whether the summary indicator, i.e., the mean level of the genetic risk score (GRS), is a poor measure of variability in the outcome of interest in diverse environments. The second concern relates to the use of potentially endogenous measures of the environment, which would lead to biased estimates. As the commentary notes, estimating cross-sectional GxE effects at the population level is extremely

challenging, and there is an apparent need for natural experiments to examine the GxE effects.

2. GxE effects in the Young Finns Study

We address specific issues related to the Young Finns Study below—first, the measurement of genetic contributions and then the endogeneity of the environment.

The GRSs obtained using gene variants identified by meta-analyses of multiple cohorts across time and places predict only the mean level of an outcome. As the commentary notes, this may not be optimal way to identify the role of genetic background on health outcomes when individuals are exposed to different environments. It is possible that from this perspective individual SNPs would provide an interesting insight on the GxE interactions. However, there is an apparent trade-off. We use the GRS to improve statistical power and to reduce the scope of potential confounders (Palmer et al. 2012; Belsky and Isreal 2014).

There is an obvious risk that the GRS is endogenous either because of confounding the genotype with the environment through population stratification or because of non-random genetic assignment. We stress three points. First, in year 1980, when the sample selection of the YFS took place, the Finnish population was ethnically highly homogenous. This mitigates, although does not rule out the possibility that stratification drives the results. Second, we account for gender and regional effects. Our estimates are also robust to the inclusion of additional covariates. Third, our study compares the GRS distributions for the high/low parental SES. These results do not imply that the genetic assignment would be non-random.

The commentary usefully illustrates the consistency of OLS estimates in the context of the study by Bet et al. (2009), which explores GxE interactions between common variants of the glucocorticoid receptor and self-reported measures of childhood trauma/abuse on depressive symptoms in old age. Two potential contributors to the possible omitted variable bias are specifically highlighted: parental genotype and unmeasured behavior in adolescence. Non-random measurement error may emerge if depressed individuals under- or over-report past trauma. Simultaneity bias, in turn, may emerge through self-selection into riskier peer groups.

Using linked YFS-FLEED data, we find these problems to be less severe for several reasons. First, possible confounding is accounted for by the covariates that describe participants' initial health endowment (birth weight), health-related behavior in adolescence (physical activity and diet), and human capital in early adulthood (education level). Second, the measure of parental resources (income and education) is not self-reported by participants; instead, the information is drawn from comprehensive administrative data compiled by Statistics Finland. Third, the simultaneity bias (reverse causality) emerging from the connections between participants' adult health outcomes and their parents' resources—measured approximately 25 years earlier—is possible but unlikely. However, the problem should be less severe in our research setting than in the designs in which the later outcome and the earlier environment refer to same individual. Fourth, our experiments with instrumental variable estimation provide tentative support for the conclusions based on the OLS estimation. In future research, school reforms can be used to examine the effects of truly exogenous variation in the environment on the outcomes of interest, as suggested in the commentary. However, doing so requires the use of more comprehensive genetic data on the Finnish population (see http://www.biopankki.fi/en/) to identify statistically and economically significant effects.

3. Summary

Belsky et al. (2013) outline three testable models of how genetic factors and environmental risks contribute to the social gradients in health. In the models, (i) both factors make independent contributions (G+E); (ii) gradients stem from the concentration of genetic risks in socially disadvantageous groups (rGE); or (iii) gradients emerge from synergies between genetic and environmental risks (GxE). Although the estimation of these models is inherently challenging in practice, as the commentary correctly discusses, empirical attempts are of great importance. We endorse Belsky and Israel (2014), who conclude that "among the most important contribution to be made from the integration of genetics into social science is improved understanding of the causes and means to treat social gradients in health." Policy-relevant research advances step by step as better data gradually become available. Big data enable the estimation of the effects through more complex research designs.

References

Bet, P.M., Penninx, B.W.J.H., Bochdanovits, Z., et al. (2009). Glucocorticoid receptor gene polymorphisms and childhood adversity are associated with depression: New evidence for a gene-environment interaction. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 150B(5), 660–669.

Belsky, D.W., and Israel, S. (2014). Integrating genetics and social science: Genetic risk scores. Biodemography and Social Biology, 60(2), 137–155.

Belsky, D.W., Moffitt, T.E., and Caspi, A. (2013). Genetics in population health science: strategies and opportunities. American Journal of Public Health, 103(S1), S73–S83.

Palmer, T.M., Lawlor, D.A., Harbord, R.M., et al. (2012). Using multiple genetic variants as instrumental variables for modifiable risk factors. Statistical Methods in Medical Research, 21(3), 223–242.