

**ENVIRONMENTAL-RISK (E-RISK) LONGITUDINAL TWIN STUDY
CONCEPT PAPER FORM**

Proposing Author: Jasmin Wertz

Author's affiliation, phone, and e-mail address: Duke University, +1 (919) 613-0352
jasmin.wertz@duke.edu

Sponsoring Investigator (if the proposing author is a student, a post-doc or a colleague): Terrie Moffitt, Avshalom Caspi

Proposed co-authors: Terrie Moffitt, Avshalom Caspi, Dan Belsky, Louise Arseneault, Duke Genomics Team, Leah Richmond-Rakerd, Jess Agnew-Blais, Tim Matthews

Provisional Paper Title: Testing 'genetic confounding' and 'genetic nurture'

Date: 08/31/2018

Objective of the study and its significance:

Parents devote a great deal of time and effort to ensuring their children's educational success. They read to their children, buy educational toys, monitor their children's schoolwork and take them to enriching classes and extracurriculars. Such parental investment is partly motivated by the belief that what parents do is crucial for children's educational success. However, the extent to which parenting matters for children's educational success has been subject to considerable debate, fuelled by three lines of evidence from behavior-genetics research. First, genetic influences have been documented for all traits and behaviors, including children's educational attainment (Polderman et al., 2015; Shakeshaft et al., 2013). Second, children's genetics influences the parenting they receive, as indicated by greater similarity in received parenting among genetically identical versus non-identical twin children (Avinun & Knafo, 2014). Third, parents' genetics influence the parenting they provide, as indicated by greater similarity in how identical versus non-identical adult twins parent their own offspring (Klahr & Burt, 2014).

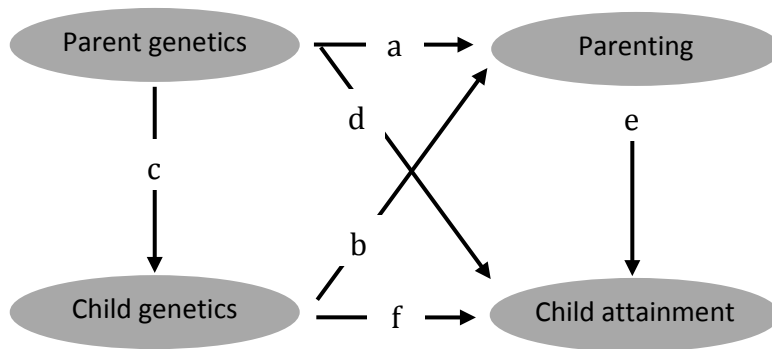
Gene-environment correlations in child development complicate the interpretation of all socialization research (Scarr & McCartney, 1983). In particular, they raise the possibility that genetic influences confound associations between parenting and children's educational attainment. This would be the case if genes that influence educational attainment also influence parenting. In this scenario, parenting may merely be a marker of children's or parents' genetic predisposition, without any effects of children's educational attainment of its own. This possibility can be summarized as 'genetic confounding'. However, gene-environment correlations do not need to imply confounding. Another possibility is that the portion of parenting that is genetically influenced still affects children's educational attainment. This would be the case if genetics influence parenting, and parenting subsequently affects children's educational attainment through environmental ways. Recent research supports this possibility, by showing that genes of parents influence their children's educational outcomes even if those genes are not passed on from parent to child (Kong et al., 2018). This research ruled out genetic confounding by isolating the effects of parents' genes that were not passed on to children. The findings suggest that parents' genetics influence children's educational outcomes via environments parents create; a possibility referred to as 'genetic nurture' (Kong et al., 2018). Thus, treating genetics as only a confounding influence may leave us with an incomplete account of parenting effects on child development.

Here we propose to test ‘genetic confounding’ and ‘genetic nurture’ in E-Risk, using genome-wide polygenic scores, derived from the most recent genome-wide association study for educational attainment (Lee et al, 2018). We plan to match mothers’ and children’s molecular-genetic data with the extensive measures of mothers’ parenting that were collected during the children’s first 12 years of life. We propose to focus on aspects of parenting that have been shown to predict children’s educational attainment: cognitive stimulation; warm, sensitive parenting; low household chaos; and a safe, clean home.

We propose to test genetic confounding by testing whether mothers’ education polygenic scores predicted the parenting they provided (Figure 1, Path a) and whether children’s polygenic scores predicted the parenting they received (Figure 1, path b). We would test these associations controlling for the fact that mothers’ and children’s genes are correlated because mothers pass on genes to their children (Figure 1, path c). We propose to focus on aspects of parenting that have been shown to predict children’s educational attainment: cognitive stimulation; warm, sensitive parenting; low household chaos; and a safe, clean home.

We propose to test genetic nurture by testing whether mothers’ polygenic scores predicted their children’s educational attainment (Figure 1, path d). We would test this association controlling for children’s own polygenic score, because mothers’ polygenic score may predict children’s attainment simply because mothers pass on genes to their children (Figure 1, path c). Based on the above-cited recent research, we expected a positive association, i.e. that mothers’ polygenic scores would predict child attainment net of children’s own polygenic score (Kong et al., 2018; Bates et al., 2018). Given a positive association, we then propose to test whether the observed parenting mothers provided could explain associations between mothers’ polygenic scores and their children’s educational attainment, net of children’s polygenic scores (Figure 1, paths a and e).

Figure 1. Testing genetic confounding and genetic nurture.



Note: Genetic confounding may be present if parents’ education-associated genetics predict the parenting they provide (path a), or children’s education-associated genetics predict the parenting they receive (path b). Genetic nurture would be indicated if parents’ education-associated genetics predict child attainment net of children’s own genetics (path d); this effect may be mediated through the parenting parents provide (paths a and e). The remaining paths are included for completeness in our analyses, but are not the focus of this study; they indicate genetic transmission from parent to child (path c) and effects of child genetics on child attainment (path f).

Statistical analyses:

We would use linear regression models in all analyses. To test genetic confounding, parenting would be predicted from mothers’ and children’s education polygenic scores, first each separately, then both together in the same model. To test genetic nurture, child educational attainment would be predicted from mothers’ education polygenic scores, and children’s polygenic scores would be included in the model to test effects of mothers’ polygenic scores net of child polygenic scores. We would then added the parenting variables to this genetic-nurture model as mediators. Each parenting variable would initially be

tested separately, and then all would be entered together into the same model. Significance of the mediation effect would be tested using formal mediation analyses, as implemented in the binary_mediation package in Stata, with 95% confidence intervals obtained from 500 bootstrap replications. Non-independence of observations of children within families would be accounted for by clustering standard errors at the family level.

Variables Needed at Which Ages (names and labels):

Study: E-Risk

Age 5:

Variable name	Description
familyid	Family ID
atwinid	Twin1 ID
btwinid	Twin2 ID
rorderp5	Random order
sampsex	Sex
zygosity	Zygosity
actvm5	Total activities with mother
warme5	Warmth towards elder twin
disse5	Dissatisfaction/Negativity towards elder twin
NEW State of the home age 5	Average of items bp12m5 and bp13m5
hiedm5	Highest educational qualification mother

Age 7:

Variable name	Description
chstim7	Child Stimulation
papse7	Positive parenting – elder twin
pange7	Negative parenting – elder twin
NEW Chaos7	Average of bp20m7; bp26m7 (reverse-coded); bp25m7 (reverse-coded)

Age 10:

Variable name	Description
chstim10	Child Stimulation
papse10	Positive parenting – elder twin
pange10	Negative parenting – elder twin
warme10	Warmth towards elder twin
disse10	Dissatisfaction/negativity towards elder twin

Age 12:

Variable name	Description
chstim12	Child Stimulation
chaossm12	Chaos Scale- Mother
chaosec12	Chaos Scale – Child
homem12	State of the home – mom interviewer
homec12	State of the home – twin interviewer

Age 18:

Variable name	Description
educachve18	Educational attainment age 18

Genetic data:

Variable name	Description
zrpgsEA3	Residualized, standardized education polygenic score - children
mom_zrpgsEA3	Residualized, standardized education polygenic score - mothers

References cited:

Avinun, R., & Knafo, A. (2014). Parenting as a reaction evoked by children's genotype: A meta-analysis of children-as-twins studies. *Personality and Social Psychology Review*, 18(1), 87-102.

Klahr, A. M., & Burt, S. A. (2014). Elucidating the etiology of individual differences in parenting: A meta-analysis of behavioral genetic research. *Psychological Bulletin*, 140(2), 544.

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsen, B. J., Young, A. I., Thorgeirsson, T. E., ... & Gudbjartsson, D. F. (2018). The nature of nurture: Effects of parental genotypes. *Science*, 359(6374), 424-428.

Lee, J. J., Wedow, R., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T. A., & Cesarini, D. (2018). Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. *Nat Genet*, in press.

Polderman, T. J., Benyamin, B., De Leeuw, C. A., Sullivan, P. F., Van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature genetics*, 47(7), 702.

Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype→environment effects. *Child development*, 424-435.

Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Rimfeld, K., Krapohl, E., Haworth, C. M., ... & Plomin, R. (2013). Strong genetic influence on a UK nationwide test of educational achievement at the end of compulsory education at age 16. *PloS one*, 8(12), e80341.

Data Security Agreement

Provisional Paper Title	Testing 'genetic confounding' and 'genetic nurture'
Proposing Author	Jasmin Wertz
Today's Date	08/31/2018

Please keep one copy for your records

(Please initial your agreement)

- I am familiar with the King's College London research ethics guidelines (<https://www.kcl.ac.uk/innovation/research/support/ethics/about/index.aspx>) and the MRC good research practice guidelines (<https://www.mrc.ac.uk/research/policies-and-guidance-for-researchers/good-research-practice/>).
- My project has ethical approval from my institution.
- I am familiar with the EU General Data Protection Regulation (<https://mrc.ukri.org/documents/pdf/gdpr-guidance-note-3-consent-in-research-and-confidentiality/>), and will use the data in a manner compliant with its requirements.
- My computer is (a) encrypted at the hard drive level, (b) password-protected, (c) configured to lock after 15 minutes of inactivity, AND (d) has an antivirus client which is updated regularly.
- I will treat all data as "restricted" and store in a secure fashion.
- I will not share the data with anyone, including students or other collaborators not specifically listed on this concept paper.
- I will not merge data from different files or sources, except where approval has been given by the PI.
- I will not post data online or submit the data file to a journal for them to post. Some journals are now requesting the data file as part of the manuscript submission process. The E-Risk Study cannot be shared because the Study Members have not given informed consent for unrestricted open access. Speak to the study PI for strategies for dealing with data sharing requests from Journals.
- Before submitting my paper to a journal, I will submit my draft manuscript and scripts for data checking, and my draft manuscript for co-author mock review, allowing three weeks.
- I will submit analysis scripts and new variable documentation to project data manager after the manuscript gets accepted for publication.
- I will delete the data after the project is complete.
- For projects using location data:** I will ensure geographical location information, including postcodes or geographical coordinates for the E-Risk study member's homes or schools, is never combined or stored with any other E-Risk data (family or twin-level data)
- For projects using genomic data:** I will only use the SNP and/or 450K data in conjunction with the phenotypes that have been approved for use in this project at the concept paper stage.

Signature:J.Wertz.....

CONCEPT PAPER RESPONSE FORM

A. To be completed by the proposing author

Proposing Author:

I have read the E-Risk data-sharing policy guidelines and agree to follow them

Provisional Paper Title: Testing 'genetic confounding' and 'genetic nurture'

Potential co-authors: Terrie Moffitt, Avshalom Caspi, Dan Belsky, Louise Arseneault, Duke Genomics Team, Leah Richmond-Rakerd, Jess Agnew-Blais, Tim Matthews

Potential Journals:

Intended Submission Date (month/year): December 2018

Please keep one copy for your records and return one to Louise (louise.arseneault@kcl.ac.uk)

B. To be completed by potential co-authors:

Approved Not Approved Let's discuss, I have concerns

Comments:

Please check your contribution(s) for authorship:

- Conceptualizing and designing the longitudinal study
- Conceptualizing and collecting one or more variables
- Data collection
- Conceptualizing and designing this specific paper project
- Statistical analyses
- Writing
- Reviewing manuscript drafts
- Final approval before submission for publication
- Acknowledgment only, I will not be a co-author

Signature: