CONCEPT PAPER FORM

Proposing Author: Leah Richmond-Rakerd

Author's affiliation, phone, and e-mail address:
Affiliation: Dept. of Psychology & Neuroscience, Duke University
Phone: 919-613-4538
E-mail: leah.richmondrakerd@duke.edu

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

Proposed co-authors: Dan Belsky, Jasmin Wertz, Richie Poulton, Jennie Connor, Sandhya Ramrakha, Louise Arseneault, HonaLee Harrington, Renate Houts, Duke Genomics Team (Karen Sugden, Benjamin Williams, Joseph Prinz, David Corcoran)

Provisional Paper Title: A polygenic score for age at first birth predicts self-regulation more broadly: Results from two longitudinal cohort studies

Date: 29 March 2018

Objective of the study and its significance:

Reproductive development and behavior comprise multiple related phenotypes, including pubertal timing, sexual debut, and childbearing (Udry, 1979; Wachter & Bulatao, 2003). These processes are associated with a range of important life outcomes, including educational attainment, wealth, and physical and psychological health (Harden, 2014a; Kong et al., 2017; Santelli et al., 2017). Understanding individual differences in reproductive development and behavior is a research goal that cuts across the social, medical, biological, and evolutionary sciences.

Early sexual activity is associated with externalizing problems including delinquency and early substance use (Zimmer-Gembeck & Helfand, 2008), risky sex (e.g., lower rates of contraceptive use; Manlove et al., 2006), and early pregnancy (Heywood et al., 2015). Quantitative genetic research suggests that much of the relation between early reproductive and risk-taking behavior can be attributed to shared genetic factors and aspects of the family environment (Deutsch et al., 2014; Harden, 2014b; Harden et al., 2008; Huibregtse et al., 2011). Molecular-genetic studies have produced consistent results. A recent genome-wide association study (GWAS) uncovered genetic variants associated with age at first birth (AFB; Barban et al., 2016) and observed significant genetic overlap between AFB and a range of associated traits, including reproductive phenotypes (e.g., age at first sexual intercourse) and disinhibitory behaviors (e.g., early-onset smoking). These findings suggest that the genetic variants associated with early childbearing may also increase liability to generalized risk-taking behavior. We propose to address this question in the present study.

A challenge in applying findings from GWASs to developmental research is the very small effect sizes for individual single-nucleotide polymorphisms. However, polygenic risk scores, which aggregate across risk alleles, yield an index of genetic risk with a potentially larger effect size (Belsky & Israel, 2014; Plomin et al., 2009). Prior analyses from our group (e.g., Belsky et al., 2013; Belsky et al., 2016; Wertz et al., 2018) have demonstrated the utility of polygenic prediction methods for understanding the etiology of psychiatric and behavioral outcomes of interest. We propose to test whether a polygenic score developed from a

GWAS of age at first birth predicts individual differences in reproductive development and behavior and disinhibitory risk-taking behavior across the life course. We will examine associations in both the Dunedin and E-Risk cohorts.

This study has four primary aims:

Aim 1) To test whether a polygenic score for age at first birth predicts reproductive behavior.
To show that the polygenic score performs in our datasets as expected, we will first test whether it predicts three important measures of reproductive behavior:

a) Age at first intercourse.
b) Age at first pregnancy.
c) Age at first birth.

Aim 2) To rule out the possibility that the polygenic score comprises genetic influences on reproductive biology.
There are genetic contributions to reproductive viability (Barban et al., 2016; Harden, 2014a) and youth who reach puberty at younger ages than their peers are at greater risk for early intercourse and pregnancy (Baams et al., 2015; Udry, 1979). Therefore, it is possible that the polygenic score comprises genetic influences on reproductive biology, which explain its associations with reproductive outcomes. To rule this out, we will test whether the score predicts:

a) Pubertal timing.
b) Fecundity (independent of the effects of early sexual intercourse and risky sexual behavior).

Aim 3) To test whether the polygenic score is associated with disinhibitory risk-taking behavior across the life course.
We will test whether the polygenic score predicts risky sexual behavior and poor sexual health outcomes, using the following indicators:

a) Contraceptive use.
b) Number of sexual partners.
c) Intercourse while under the influence of alcohol or drugs (available only in E-Risk).
d) History of sexually transmitted infection.

We will test whether the polygenic score predicts additional measures of externalizing problems across the life course, including:

a) Low childhood self-control.
b) Adult antisocial behavior.
c) Adult substance use disorder.

Aim 4) To test whether associations with the polygenic score are explained by contextual factors.
Low socioeconomic status is associated with early reproductive timing (Penman-Aguilar et al., 2013; Wellings et al., 2016) and risk-taking behavior (Mason et al., 2010). In addition, early parenthood tends to run in families (Meade et al., 2008; Sipsma et al., 2010). We therefore need to test whether any effects obtained in Aims 1 -- 3 persist after we account for:

a) Socioeconomic background.
b) Family history of early childbearing (maternal age at first birth).

Statistical analyses:

Our predictor variable will be the polygenic score computed using summary statistics from the latest GWAS of age at first birth (Barban et al., 2016). Because the majority of genetic associations within the GWAS overlapped across men and women, we will use the polygenic score computed using the pooled male-female sample. However, we will also conduct sensitivity analyses using the polygenic scores computed within sex.

Aim 1) Does the polygenic score predict measures of reproductive behavior as it should?
Aim 2) Does the polygenic score comprise genetic influences on reproductive biology?
For Aims 1 and 2, we will use regression to test for associations between the polygenic score and reproductive outcomes. The type of regression model used will depend on the scale of the outcome

variable (e.g., ordinary least squares for continuous data, logistic models for binary data, Poisson or negative-binomial models for count data, Cox/hazard models for time-to-event data).

Aim 3) Does the polygenic score predict disinhibitory risk-taking behavior?
For analyses of risky sexual behavior, we will create a composite measure (scale or factor score) using the above-described indicators.

We will use regression to test whether the polygenic score is associated with risky sexual behavior and other measures of externalizing behavior across the life course. The type of regression model used will depend on the scale of the outcome variable.

Aim 4) Do effects persist after adjusting for family context?
We will re-run the models used in Aims 1 -- 3, adjusting for SES and maternal age at first birth.

Supplemental analyses.
**Although we would like to conduct all of these analyses, some may not be possible due to sample size constraints.*

Within the E-Risk cohort, we will:

a) Conduct within-twin-pair tests for associations between the polygenic score and reproductive and externalizing outcomes (among dizygotic pairs).
b) Predict study members' reproductive and externalizing outcomes from their mothers' polygenic scores for age at first birth.
c) Compute the heritability of and genetic correlations between reproductive outcomes.

Note.
Regression models within the E-Risk data will account for the non-independence of twin observations by clustering standard errors at the family level.

Analyses in which men and women are combined will control for sex.

Described above are the primary, pre-planned analyses. Secondary analyses may be added as suggested through internal review and will be identified as secondary in the manuscript.

Variables Needed at Which Ages (names and labels):

Study: E-Risk

Phase	Variables	Description
	Identifiers	
5	familyid	Family number of twin pair
	atwinid	Twin A ID number
	btwinid	Twin B ID number
	sampsex	Sex of participant
	zygosity	Zygosity – current from October 2016
18	tagee18	Exact age at interview – Ph. 18
	Contextual risk factors	
Birth	Sage1st	Maternal age at first birth
5	Seswq35	SES
	Reproductive history	
18	Sex12a	Age of menarche (female Q)
	Sex2	Ever had intercourse

	Sex4	Age at first intercourse
	Sex13	Ever been pregnant (female Q) / sexual relationships resulted in pregnancy (male Q)
	Sex15	Number of past pregnancies (female Q) / number of times relationship resulted in pregnancy (male Q)
	Parente18	Study member is a parent
	Sexual risk-taking and sexual health	
18	Sex6	Number of lifetime sexual partners
	Sex7	Frequency of contraception use
	Sex8	Frequency of condom use
	Sex11	Frequency of intercourse after alcohol/drug use
	STD1a	Ever diagnosed with Chlamydia
	STD2a	Ever diagnosed with HPV
	STD3a	Ever diagnosed with Gonorrhea
	STD4a	Ever diagnosed with genital herpes
	STD5a	Ever diagnosed with hepatitis
	STD6a	Ever diagnosed with other STD
	Other risk-taking behavior	
10	Lowsc510e	Low self-control, ages 5-10
18	Anycrime18	MoJ – any criminal offence (with updated records)
	Cd1e18 -- cd44e18	Self-report delinquency/offending (computer interview)
	Dxalcdepe18	DSM-IV alcohol dependence
	Dxmarje18	DSM-IV cannabis dependence
	Dxdrugme18	DSM-IV drug dependence (or methadone maintenance)
	Smkdxftnde18	Fagerström diagnosis for nicotine dependence
	Extcf_e	Externalizing psychopathology factor
	Genetic data	
	PRS for age at first birth, original	
	PRS for age at first birth, clumped	
	PRS for age at first birth, residualized for PCAs from White ethnicity subset	
	PRS for age at first birth, residualized, standardized	
	Principal components 1-10	
	Random order for genetic data	
	PRS for mother's age at first birth	Not yet computed

Study: Dunedin

**Phase 38 variables will be updated to incorporate Phase 45 data as they become available.*

Phase	Variables	Description
	Identifiers	
Birth	Snum	Participant ID number
	Sex	Participant sex
18 -- 38	Age	Exact ages at interview
	Contextual risk factors	
Birth	Mumfstbirth	Maternal age at first birth
	Sesav115	SES

	Reproductive history	
18	FstPeriod	Age at menarche, in months
21	Virgin21	Ever had intercourse
	FstSex21	Age at first intercourse
	Age1stpreg	Age at first pregnancy
	Totpreg21	Total number of pregnancies
	Kids1521	Total number of children, 15-21
26	Kdwhen26	Age at first birth **We will select for births prior to age 21.*
38	FstSexMinTemp	Age at first intercourse, minimum of ph.21 & ph.38
	FirstPregEstTemp	Age at first pregnancy
	TotPregEstTemp	Total number of pregnancies, with LHC data
	TotKidsLT38	Total number of children
	Age_at_birth1	Age at first birth, ph. 38 (with data from parenting study)
	Sexual risk-taking and sexual health	
21	Sexyr21	Number of past-year sexual partners
	LiftimOSpartnrs21	Number of lifetime sexual partners
	Condom21	Frequency of condom use
	Rsksex21	Risky sex measure from Caspi et al., 1997
	Dangsex	Risky sex measure from Ramrakha et al., 2000
	Hadstd21	Ever had an STD
	Other risk-taking behavior	
11	Lscuw311	Low self-control, ages 3-11
26	Tax_class	C. Odgers antisocial behavior developmental taxonomy
38	Anyconv38	Any conviction up to 38
	ChrAlc1838	Chronic alcohol dependence, 18-38
	ChrMar1838	Chronic cannabis dependence, 18-38
	ChrSub1838	Chronic substance dependence, 18-38
	ChrMarDrg1838	Chronic cannabis or other drug dependence, 18-38
	Ptot38	Lifetime cigarette consumption
	Ext_cf	Externalizing psychopathology factor
45	SRrisktake45	Self-reported risk-taking
	Genetic data	
	PRS for age at first birth, original	
	PRS for age at first birth, clumped	
	PRS for age at first birth, residualized for PCAs from White ethnicity subset	
	PRS for age at first birth, residualized, standardized	
	Principal components 1-10	

References cited:

Baams, L., Dubas, J.S., Overbeek, G., & van Aken, M.A.G. (2015). Transitions in body and behavior: A meta-analytic study on the relationship between pubertal development and adolescent sexual behavior. *Journal of Adolescent Health, 56,* 586-598.

Barban, N., Jensen, R., de Vlaming, R., Vaez, A., Mendemakers, J., Tropf, F., . . . Mills, M.C. (2016). Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nature Genetics, 48,* 1462-1475.

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

Belsky, D. W., Moffitt, T. E., Baker, T.B., Biddle, A.K., Evans, J.P., Harrington, H., . . . Caspi, A. (2013). Polygenic risk and the developmental progression to heavy, persistent smoking and nicotine dependence: Evidence from a 4-decade longitudinal study. *JAMA Psychiatry, 70,* 534-542.

Belsky, D. W., Moffitt, T. E., Corcoran, D. L., Domingue, B., Harrington, H., Hogan, S., ... & Poulton, R. (2016). The genetics of success: How single-nucleotide polymorphisms associated with educational attainment relate to life-course development. *Psychological Science*, 27, 957-72.

Deutsch, A.R., Slutske, W.S., Heath, A.C., Madden, P.A.F., & Martin, N.G. (2014). Substance use and sexual intercourse onsets in adolescence: A genetically informative discordant twin design. *Journal of Adolescent Health, 54,* 114-116.

Harden, K.P. (2014a). Genetic influences on adolescent sexual behavior: Why genes matter for environmentally-oriented researchers. *Psychological Bulletin, 140,* 434-465.

Harden, K.P. (2014b). A sex-positive framework for research on adolescent sexuality. *Perspectives on Psychological Science, 9,* 455-469.

Harden, K.P., Mendle, J., Turkheimer, E., & Emery, R.E. (2008). Rethinking timing of first sex and delinquency. *Journal of Youth and Adolescence, 37,* 373-385.

Heywood, W., Patrick, K., Smith, A.M.A., & Pitts, M.K. (2015). Associations between first sexual intercourse and later sexual and reproductive outcomes: A systematic review of population-based data. *Archives of Sexual Behavior, 44,* 531-569.

Huibregtse, B.M., Bornovalova, M.A., Hicks, B., McGue, M., & Iacono, W.I. (2011). Testing the role of adolescent sexual initiation in later-life sexual risk behavior: A longitudinal twin design. *Psychological Science, 22,* 924-933.

Kong, A., Frigge, M.L., Thorleifsson, G., Stefansson, H., Young, A.I., Zink, F., . . . Stefansson, K. (2017). Selection against variants in the genome associated with educational attainment. *PNAS.* doi: 10.1073/pnas.1612113114

Manlove, J.S., Terry-Humen, E., & Ikramullah, E. (2006). Young teenagers and older sexual partners: Correlates and consequences for males and females. *Perspectives on Sexual and Reproductive Health, 38,* 197-207.

Mason, W.A., Hitch, J.E., Kosterman, R., McCarty, C.A., Herrenkohl, T.I., & Hawkins, J.D. (2010). Growth in adolescent delinquency and alcohol use in relation to young adult crime, alcohol use disorders, and risky sex: A comparison of youth from low- versus middle-income backgrounds. *Journal of Child Psychology and Psychiatry, 51,* 1377-1385.

Meade, C.S., Kershaw, T.S., Ickovics, J.R. (2008). The intergenerational cycle of teenage motherhood: An ecological approach. *Health Psychology, 27,* 419-429.

Penman-Aguilar, A., Carter, M., Snead, C., & Kourtis, A.P. (2013). Socioeconomic disadvantage as a social determinant of teen childbearing in the U.S. *Public Health Reports, 128,* 5-22.

Plomin, R., Haworth, C.M.A., & Davis, O.S.P. (2009). Common disorders are quantitative traits. *Nature Reviews Genetics, 10,* 872-878.

Santelli, J.S., Song, X., Garbers, S., Sharma, V., & Viner, R.M. (2017). Global trends in adolescent fertility, 1990-2012, in relation to national wealth, income inequalities, and educational expenditures. *Journal of Adolescent Health, 60,* 161-168.

Sipsma, H., Biello, K.B., Cole-Lewis, H., & Kershaw, T. (2010). Like father, like son: The intergenerational cycle of adolescent fatherhood. *American Journal of Public Health, 100,* 517-524.

Udry, J.R. (1979). Age at menarche, at first intercourse, and at first pregnancy. *Journal of Biosocial Science, 11,* 433-441.

Wachter, K.W., & Bulatao, R.A. (eds). Offspring: Human fertility behavior in biodemographic perspective. Washington, DC: National Academies Press; 2003.

Wellings, K., Palmer, M.J., Geary, R.S., Gibson, L.J., Copas, A., Datta, J., . . . Wikinson, P. (2016). Changes in conceptions in women younger than 18 years and the circumstances of young mothers in England in 2000-12: An observational study. *The Lancet, 388,* 586-595.

Wertz, J., Caspi, A., Belsky, D.W., Beckley, A.L., Arseneault, L., Barnes, J.C., . . . Moffitt, T.E. (2018). Genetics and crime: Integrating new genomic discoveries into psychological research about antisocial behavior. *Psychological Science.* doi:10.1177/0956797617744542

Zimmer-Gembeck, M.J., & Helfand, M. (2008). Ten years of longitudinal research on U.S. adolescent sexual behavior: Developmental correlates of sexual intercourse, and the importance of age, gender and ethnic background. *Developmental Review, 28,* 153-224.

Data Security Agreement

Provisional Paper Title	A polygenic score for age at first birth predicts the development of externalizing behavior: Results from two longitudinal cohort studies
Proposing Author	Leah Richmond-Rakerd
Today's Date	29 March 2018

Please keep one copy for your records
(Please initial your agreement)

__LR__ 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/)

__ LR __ My project has ethical approval from my institution.

__ LR __ 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.

__ LR __ I will treat all data as "restricted" and store in a secure fashion.

__ LR __ I will not share the data with anyone, including students or other collaborators not specifically listed on this concept paper.

__ LR __ I will not merge data from different files or sources, except where explicit approval has been given by the PI.

__LR__ 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.

__ LR __ 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.

__ LR __ I will submit analysis scripts and new variable documentation to project data manager after the manuscript gets accepted for publication.

__N/A__ **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)

__ LR __ **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.

LR	I am current on Human Subjects Training (CITI (www.citiprogram.org) or equivalent)
LR	My project is covered by Duke or Otago ethics committee OR I have /will obtain ethical approval from my home institution.
LR	I will treat all data as "restricted" and store in a secure fashion. My computer or laptop is: a) encrypted (recommended programs are FileVault2 for Macs, and Bitlocker for Windows machines) b) password-protected c) configured to lock-out after 15 minutes of inactivity AND d) has an antivirus client installed as well as being patched regularly.
LR	I will not "sync" the data to a mobile device.
LR	In the event that my laptop with data on it is lost, stolen or hacked, I will immediately contact Professor Moffitt or Caspi. (919-684-6758, tem11@duke.edu, ac115@duke.edu)
LR	I will not share the data with anyone, including my students or other collaborators not specifically listed on this concept paper.
LR	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 Dunedin Study Members have not given informed consent for unrestricted open access, so we have a managed-access process. Speak to Terrie or Avshalom for strategies for achieving compliance with data-sharing policies of journals.*
LR	I will delete all data files from my computer after the project is complete. Collaborators and trainees may not take a data file away from the office. The data remains the property of the Study and cannot be used for further analyses without an approved concept paper for new analyses.

Signature: .....L. Richmond-Rakerd.....