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

Proposing Author: Temi Moffitt

Author's affiliation, phone, and e-mail address: tem11@duke.edu

Proposed co-authors: Richie Poulton (for Dunedin), Louise Arseneault (for E-risk), Terrie Moffitt (for both), Avshalom Caspi and Daniel Belsky (these 2 will do analyses)

Provisional Paper Title: SSGAC polygenic score validation. Actual title of paper to be determined by Dr. Aysu Okbay, who leads this multi-cohort project.

Date: 16 May 2018

Objective of the study and its significance:

The US-based Social Science Genetics Consortium, known as SSGAC (https://www.thessgac.org/) is a group who coordinates very large-sample GWAS analyses of phenotypes of interest to social scientists. The SSGAC group achieves very large samples to improve the quality of GWAS work, for example by combining samples from multiple cohorts, UK Biobank, and 23andMe. They are most well-known, perhaps, for their GWAS of educational attainment, and for the polygenic scores generated from this GWAS. The education polygenic score has been calculated for Dunedin and E-Risk, and has resulted in several strong papers. However, neither Dunedin nor E-risk have ever contributed data to the SSGAC GWAS discovery work.

This year, the SSGAC group has achieved a number of new polygenic scores, for phenotypes beyond educational attainment, and Dan Belsky has negotiated with SSGAC a MOU to share those scores with Dunedin and E-risk for our use. These scores, when obtained later this year, will be added to our Study data sets and data dictionaries.

In the shorter run, the SSGAC investigators, specifically Aysu Okbay, plan to publish a paper reporting how their new polygenic scores correlate with the respective phenotypes in a goodly set of multiple samples who are independent of the discovery GWAS. This paper will establish validation of the GWAS-derived polygenic scores. If this concept paper is approved, we have the opportunity to contribute to this publication. The publication will, of course, have numerous authors, and as is customary with multi-cohort papers of this type, each cohort can be represented by only a couple of authors who are study leaders. Our contribution will not be to the discovery GWAS, but will be to the confirmatory component of the analysis. That is, once GWAS for new phenotypes are conducted, and after polygenic scores for these new phenotypes are constructed, we will be test how they relate to the phenotypes in our cohort studies.

The pros of this are, repaying SSGAC's generosity for access to the polygenic scores by supporting their validation paper, co-authorship, sharing data for a good cause, and, of course, the clear scientific value of validating the new polygenic scores. Note that if we do not take part in this publication, the polygenic scores will still ultimately be available to us.

The biggest con might be that a correlation for the association between a certain polygenic score and a certain phenotype would be published by SSGAC, before we published it ourselves. However, we tend to use polygenic scores to tell rather more elaborated developmental stories, using phenotypes from multiple

waves and multiple sources, which Dunedin and E-risk team members could still do of course.

Statistical analyses:

All analyses would take place at Duke.

How it would work is that SSGAC would send us the set of new polygenic scores in late June, and we would test how the scores correlate with basic phenotype measures in Dunedin and E-risk. We would send back to them tables of output. The statistics would appear in their paper on tables reporting heaps of correlations, for many different samples and cohorts.

They have sent us quite a long shopping list, some of which we have, some of which we don't:
Anthropometric: Height, BMI
Personality: Big 5, Morning person, Adventurousness, Religiosity, Recharge by socializing, Risk preferences, Narcissism
Cognition: Alzheimer's, Educational attainment, Intelligence, Math ability, Age started reading
Well-being: Subjective well-being, Family relationship satisfaction, friendship satisfaction, Work satisfaction, Financial satisfaction, Loneliness/Isolation
Psychiatric: Depressive symptoms, Schizophrenia
Health behaviors: Ever smoker, Cigarettes per day, Physical activity, Ever drink, Drinks per day, Self-rated health
Reproductive: Age at first birth, Number of children ever born
Health outcomes: COPD, Allergies, Asthma, T2 Diabetes, Migraine

We would report the association between the polygenic score for each particular phenotype, and that phenotype in our cohorts.

Variables Needed at Which Ages (names and labels):

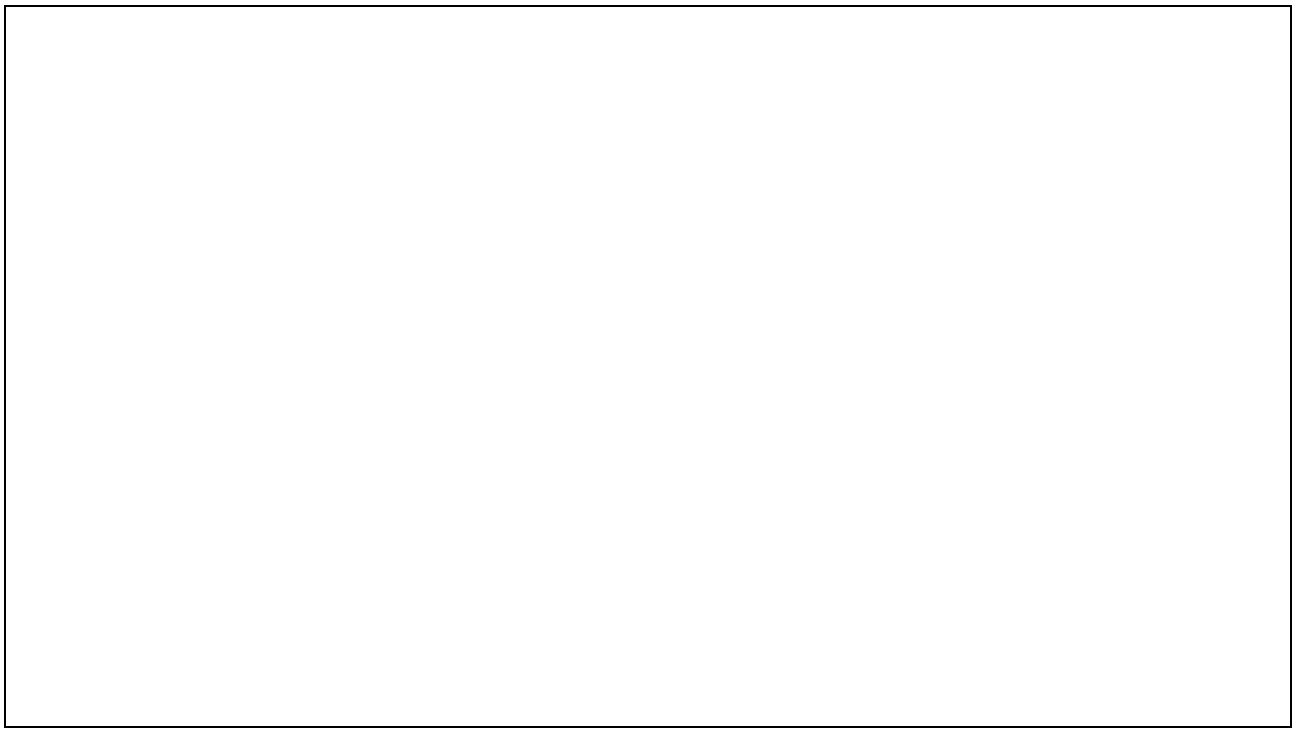
Avshalom and Honalee will work this out, following the specific request from SSGAC (as above). We want to keep it simple.

References cited:

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., …Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*, 539–542. https://doi.org/10.1038/nature17671

Belsky, D. W., Moffitt, T. E., Corcoran, D. L., Domingue, B., Harrington, H., Hogan, S., … Caspi, A. (2016). The genetics of success: How single-nucleotide polymorphisms associated with educational attainment relate to life-course development. *Psychological Science*, *27*, 957–972. https://doi.org/10.1177/0956797616643070

Wertz, J., Caspi, A, Belsky, D.W, Beckley, A.L., Arseneault, L., J. C. Barnes, D. L. Corcoran, S. Hogan, R. Houts, N. Morgan, C. L. Odgers, J. Prinz, K. Sugden, B. S., Williams, R. Poulton, TE Moffitt (2018 online). A polygenic score for educational attainment also predicts criminal offending: Replicated evidence from two birth cohorts. Psychological Science.



Data Security Agreement

Provisional Paper Title	SSGAC polygenic score validation
Proposing Author	Temi Moffitt
Today's Date	16 May 2018

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

_x___ 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/)

__x__ My project has ethical approval from my institution.

_x___ 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.

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

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

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

__x__ 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.

__x__ 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.

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

__x__ **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)

__x__ **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: .....Temi Moffitt.....

CONCEPT PAPER RESPONSE FORM

A. To be completed by the proposing author

Proposing Author:

xxx I have read the E-risk & Dunedin data-sharing policy guidelines and agree to follow them

Provisional Paper Title: SSGAC Polygenic Score validation

Potential co-authors: Louise Arseneault, Daniel W Belsky, Avshalom Caspi, Richie Poulton, Temi Moffitt will be added to a long list of authors, headed by SSGAC

Potential Journals: ??

Intended Submission Date (month/year): August 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: .....