

We are pleased to share data with you! This document is intended to help standardize the conduct of research among our research group in the interest of smooth communication with students, post-docs and collaborators. Its contents are drawn from the guidelines of the APA, AMA, ASA, and the Dunedin Study. It is important that you read this document before you start working with E-Risk data. We ask that you follow each of the policies and procedures listed below. In some cases, exceptions may be negotiated in advance if you have special circumstances.

Furthermore, please refer to a short editorial on the good practice of registering in advance a statistical analysis plan before beginning a paper using observational data. You can find the paper at the end of this document. Thank you ...

Terrie E. Moffitt  
[tem11@duke.edu](mailto:tem11@duke.edu)  
1-919-684-5831

Louise Arseneault  
[louise.arseneault@kcl.ac.uk](mailto:louise.arseneault@kcl.ac.uk)  
44-(0)207-848-0647

Avshalom Caspi  
[ac115@duke.edu](mailto:ac115@duke.edu)  
1-919-684-5835

### **GETTING STARTED...**

- \* To ascertain the general feasibility of your research project, first discuss your idea with one of the main investigators of the E-Risk Study, Terrie E. Moffitt, Louise Arseneault and Avshalom Caspi. This investigator will become your "sponsoring investigator" and your main contact person for conducting research with E-Risk.
- \* Familiarize yourself with the E-Risk "Policy & Procedures Handbook."
- \* Consult the electronic Data Dictionary and read previously published papers from the study to get information about the data you plan to use.
- \* Take the time to think carefully about your research questions and what variables you will need for your project. Discuss your questions and selection of variables with your sponsoring investigator. Collaborators who continuously contact us for "just one more variable" generate a significant amount of extra work for our research team. At the same time, asking for too many variables without a good justification is not advised. The best proposals are those that carefully outline a set of well justified research questions and request the appropriate number and types of variables required to test the research questions. Requests that do not meet these expectations will be returned for a more careful consideration.
  
- \* Prepare a brief concept paper for your study. A boilerplate for this is at end of this document. A concept paper should contain:
  - a. Your name and affiliation, the name of the investigator sponsoring your work with the data set, the date, a working title for the paper.
  - b. Brief statement of the hypothesis or topic and its significance.
  - c. What analyses you will perform.
  - d. Which variables at which ages you will need. Also specify if the data requested are from one of E-Risk spin-off studies (e.g., SCOPIC, Germfighter, TEDS-Peers).
  - e. A "concept paper response form" that any data caretakers or potential collaborators can fill out.
  
- \* The last sheet has to be signed by all collaborators for permission to use their data. It also offers options that they can check off to signify what sort of role they would like to play (authorship roles:

contribute to design & data, help with analyses, help with writing, critique drafts; or non-authorship role with acknowledgement only). When completed, ask your sponsoring investigator to send the concept paper to Louise who will circulate amongst collaborators.

\* When signed response forms are returned to your sponsoring investigator, you have a project, and the research may proceed.

\* E-risk is **not** a public domain data set. **You may not share your data file with others, and by signing the concept paper you agree to this restriction.**

\* All papers and thesis chapters require an approved concept paper on file, including papers by the active main E-Risk investigators. Researchers who want to start a new project with the same data they used in the past must complete a new concept paper form and request the variables again in case there have been changes or updates.

### **ETHICS TRAINING...**

For research funded in any way by the US NIH, it is a requirement that all investigators have completed the NIH research ethics training course in human subject protection, or equivalent training from their own country. Even if research is not actively funded by NIH, knowledge of human subjects protection is still a requirement of the ethics committees who approve and monitor our E-Risk study. Increasingly journals are also asking for this documentation. Therefore, by accepting data from us to analyze, researchers should understand they are attesting to us that they have completed research ethics training. If you have not yet taken a course, it is easy to take by going to:

<http://www.citiprogram.org/default.asp?language=english>

### **DOING THE ANALYSES...**

\* Working with twin data can be challenging!! The E-Risk Data Dictionary contains instructions on how to work with twin data in the "data structure & usage" section. Please take the time to read carefully and discuss with your sponsoring investigator the challenges and beauty of this data set.

\* Look at the frequency distribution of a variable before you use it and always compare with values reported in the Data Dictionary. They should be the same and if they are not, report quickly to your sponsoring investigator.

\* Also report to your sponsoring investigator any bug or error in a file.

\* Never merge data sets!!!! This is the job of the data managers!!! Ask for a new updated data file that contains all the variables needed for the project.

\* Always keep a detailed record of all the analyses to you run for your E-Risk project. Use "comments" in your command files for data analysis, so that others can retrace your steps later.

\* Save all your printouts in a labeled notebook (electronic or paper). These must be saved for 5 years after the publication of a paper.

### **IF YOU CREATE NEW VARIABLES FROM THE BASIC DATA**

\* We ask that you provide to us any new variables that are published in your paper.

\* When you create a new variable give it the following features:

- variable name ending with a number indicating the assessment age wave
- variable label with data source for the original constituent data  
e.g. Mother or teacher report, DSM4 based diagnosis, Average of ..., Sum of...
- value labels for the coded numeric values (e.g. 0 = no, 1 = yes).

Examples:

harme512 Child Harm at Phases 5, 7, 10, or 12

totexte5 Sum of Mother and Teacher Externalising problems, Phase 5

cdie12 Depression Scale from CDI - Phase 12

It used to be that Variable names were limited to 8 characters and variable labels were only 40 but statistical software is more flexible now. Please assign names & labels that clearly communicate what the variable is. That said, it is helpful to put the most important information at the beginning of the variable label.

A suffix of "518" can refer to the Sum of whatever at those 5 phases or the Average or Any. You need to use the label to make clear what the phase notation refers to.

\* Prepare a system file and documentation for all new variables you created for the study. Include the variable labels and value labels. Documentation includes how you made the variable and its frequency distributions, including the derivation code used to create each variable from the base-data variables themselves (i.e. before any re-codes etc) this makes it easier to see what exactly was done. Send the file and documentation to Antony Ambler, E-Risk Data Manager ([antony.ambler@kcl.ac.uk](mailto:antony.ambler@kcl.ac.uk)).

### **WRITING IT UP...**

\* Don't re-create the wheel; use our existing boilerplates for the "methods" section for consistency across publications. Look at the Data Dictionary and recent submissions for parts to reproduce.

\* Access our reference system if you like, to help construct your reference list.

\* Include the date and the electronic file name, as well as page numbers, on every printed version of the manuscript.

\* The abstract should include the name "E-Risk" so people can track down our work easily.

\* Acknowledge all appropriate individuals, agencies and grants on your manuscript. This is not an option, it is required by our funding agencies. Ask your sponsoring investigator for an up-to-date template of acknowledgements.

\* Suggested basic text could be as follow ...

The authors are grateful to the Study members and their families for their participation. Our thanks to the Nuffield Foundation, the Avielle Foundation, CACI, Inc., and to members of the E-Risk team for their dedication, hard work, and insights. The E-Risk Study is funded by the Medical Research Council (UKMRC grant G1002190). Additional support was provided by National Institute of Child Health and Human Development (grant HD077482) and by the Jacobs Foundation. Louise Arseneault is the ESRC Mental Health Leadership Fellow.

We are grateful to the study mothers and fathers, the twins, and the twins' teachers for their participation. Our thanks to Michael Rutter and Robert Plomin, to Thomas Achenbach for kind permission to adapt the Child Behavior Checklist, and to members of the E-Risk team for their dedication, hard work, and insights.

\* Institutions have become really strict about the way their names appear on publications so they get credit for papers when electronic search engines allocate products to universities and departments. For the E-Risk main investigators, the affiliations are:

*Louise Arseneault:* Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London SE5 8AF, United Kingdom.

*Avshalom Caspi:* Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London SE5 8AF, United Kingdom; Departments of Psychology and Neuroscience, Psychiatry and Behavioral Sciences, and Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina, USA.

*Terrie E. Moffitt:* Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London SE5 8AF, United Kingdom; Departments of Psychology and Neuroscience, Psychiatry and Behavioral Sciences, and Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina, USA.

## **BEFORE YOU SUBMIT FOR PUBLICATION...**

\* All primary authors need to submit the annotated script of their data analysis with the sponsoring investigator **BEFORE** submitting the final draft of the paper, ideally when the co-authors are reading the manuscript. The script will be reviewed by one of the E-Risk data managers and all analyses will be rerun to check against the manuscript. Specifically, here is what we would like to ask all authors to provide:

- a. Script for new variables you've created.
- b. Script for final analyses (in SPSS, STATA, Mplus preferred, but whatever you have used is fine), with documentation included in the script. It would be especially helpful if you could orient us in the script to the manuscript (e.g., analyses for Table 1, etc.).

- \* Your sponsoring investigator must approve the penultimate draft before it is circulated among the co-authors.
- \* Specify the target journal or ask for suggestions.
- \* Ask collaborators if they receive salary support from funding bodies.
- \* Ask collaborators to send written permission to submit the paper if the journal requires it.
- \* Allow all co-authors **3 weeks** from the date of receipt to review the paper. The E-Risk Data managers will check the scripts during that period.
- \* Louise's approval and release of every paper is a condition of our data-sharing agreement.
- \* Sponsoring investigators must have arranged in advance of submission publication costs.

## **SUBMITTING FOR PUBLICATION REVIEW...**

- \* Submit a copy of the paper and letter to the editor to Louise, both dated with the electronic file name on the document.
- \* Ask your sponsoring investigator to inform via email all co-authors about the paper being submitted and to update their cv with the complete reference. It is also good practice to send all co-authors a copy of the paper submitted for their personal record.
- \* E-Risk sends a beginning-of-year thank-you note to all editors who have handled our work – whether papers got accepted or rejected. Be sure your sponsoring investigator passes on the editor's name, journal and address to Louise.

## **DOING REVISIONS...**

- \* Send a copy of the editor's letter and reviews to all co-authors.
- \* Invite your co-authors to comment or contribute to the revision process.
- \* Louise must approve the revised draft before it is sent back to the editor.

## **WHEN THE PAPER IS ACCEPTED FOR PUBLICATION...**

- \* Celebrate!
- \* Give Louise the Word file for the final version for filing in the electronic reprint system.
- \* Your sponsoring investigator should be listed as the corresponding author, unless judged otherwise.
- \* Send a copy of the paper to all co-authors with the citation for their vitae.
- \* If you have permission to use a file of variables for a specific study, return the file to the data manager when you finish. Collaborators and graduates of E-Risk may not take a data file away from the E-Risk office.

## **DURING THE PROCESS TOWARD PUBLICATION...**

\* Your sponsoring investigator will receive the page-proof version. The proofs should always be checked VERY carefully by 2 people, usually the first author and the sponsoring investigator.

#### **ONCE PUBLISHED...**

- \* Notify Louise and send her a PDF copy of the version that was accepted for publication.
- \* Send PDF to all co-authors.

#### **IF THERE IS MEDIA COVERAGE OF THE PROJECT...**

- \* Consult with Louise, Temi or Avshalom about drafting a fact sheet for journalists.
- \* Deliver copies of all media articles, or websites, to Louise.

#### **IF YOU GIVE A LECTURE OR POSTER PRESENTATION ON YOUR WORK**

- \* Notify Louise of the title, date, and the conference's name. We need to report these for our progress reports to funding agencies.
- \* Although oral presentations do not require a concept paper, oral presentations or posters should always relate to a project already approved with a concept paper.
- \* Collaborators can present findings only once a solid draft of the manuscript is written. Expected findings with E-Risk data should **NOT** be submitted for upcoming conferences.

Thank you for reading this document.

We share data with a large number of collaborators, who live in many countries, and who work at many different levels of training and expertise. This memo is intended to instruct students, post-docs and new collaborators who are embarking on their first research project, and to remind senior professors who are embarking on their thousandth research project. We have listed here the practices that help to keep our working relationships comfortable and productive. We hope they work for you too.

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

Proposing Author: Andrea Danese

Author's affiliation, phone, and e-mail address: [andrea.danese@kcl.ac.uk](mailto:andrea.danese@kcl.ac.uk)

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

Proposed co-authors: Moffitt, Caspi, Arseneault, other interested E-Risk investigators, Richie Poulton, Duke Genomic team, Jasmin Wertz, Honalee Harrington, Jessie Baldwin

Provisional Paper Title: The genetics of childhood victimisation

Date: 24/03/2019

Objective of the study and its significance:

This project will conduct primary analyses in the E-Risk cohort twins and mothers, and then attempt replication of key analyses in the Dunedin cohort Study members.

Childhood victimisation is a key risk factor for poor mental and physical health and psychosocial functioning later in life. Associations between childhood victimisation and such outcomes are often interpreted causally. However, because childhood victimisation is not randomly distributed in the population, it is possible that pre-existing features of the child and of the family context the child lives in could affect risk for victimisation as well as later outcomes. There is currently very limited understanding of these alternative non-causal explanations for the outcomes of childhood victimisation. For example, we have shown that the association between childhood victimisation and later cognition is largely explained by pre-existing cognitive vulnerabilities likely reflecting genetic selection [Danese, 2017, Am J Psychiatry](#).

Polygenic risk scores (PRS) offer new ways to test genetic selection in at-risk groups. Of note, early environmental exposures covary with multiple PRS [Krapohl, 2017, PNAS](#) suggesting presence of widespread gene-environment correlation underlying putative environmental effects. In this paper we will test whether victimised children have genetic liability to poor mental and physical health and psychosocial functioning later in life previously associated with childhood victimisation (i.e., psychopathology; cardiovascular disease and associated risk, such as body mass index and C-reactive protein; educational attainment and age at first birth). PRS offers important advantages for testing gene-environment correlation in victimised children compared to discordant-twin comparison methods because too little twin-pair discordance in victimisation is typically observed at young ages.

The significance of these analyses is in their potential to characterise alternative non-causal explanations for the associations of childhood victimisation with poor mental and physical health and psychosocial functioning later in life. In turn, this could highlight vulnerabilities that could help identify children at risk for victimisation and could enable design of targeted interventions for the prevention of victimisation and associated outcomes (based on evocative vs passive mechanisms; see below).

Statistical analyses:

### ***Gene-environment correlations in victimised children***

We will test whether victimised children have greater genetic liability to poor mental and physical health and psychosocial functioning previously associated with childhood victimisation than non-victimised peers. Specifically, we will test group differences for the following PRS in E-Risk children: psychopathology; cardiovascular disease, body mass index and C-reactive protein; educational attainment and age at first birth. We will also use PRS for height as negative control.

### ***Test of evocative and passive gene-environment correlation***

Victimised children have greater genetic liability than non-victimised peers for two reasons. On the one hand, genetic liability for the above outcomes might evoke abusive or neglectful behaviours in parents and peers. On the other hand, genetic liability for the above outcomes might be passively inherited by children along with abusive and neglectful environments. To investigate the relative contribution of passive gene-environment correlation, we will repeat the above analyses focusing on differences in maternal PRS (rather than the children's PRS) between victimised vs non-victimised children.

### ***Sensitivity analyses***

The main analyses will focus on the childhood poly-victimisation variable used in several previous E-Risk papers. As in previous papers [Danese, 2017, Am J Psychiatry](#), we will then test the sensitivity of findings across multiple other measures of childhood victimisation. First, we will test if the findings apply to all component measures of the childhood poly-victimisation variable by re-running the analyses above in children who witnessed domestic violence between the mother and her partner, children who were frequently bullying by peers, or children who suffered physical maltreatment, sexual abuse, emotional abuse, or physical neglect by an adult. Second, because most studies testing the health consequences of childhood victimisation have been based on retrospective recall and retrospective and prospective measures of childhood victimisation identify different individuals [Baldwin, 2019, JAMA Psychiatry](#), we will re-run the analyses above in young adults who retrospectively reported a history of childhood maltreatment based on the Childhood Trauma Questionnaire (CTQ). Third, because of the widespread use of Adverse Childhood Experiences (ACEs) in public health and clinical settings, we will re-run the analyses above in children with increasing number of prospectively measured ACEs.

### ***Replication in Dunedin***

To ensure replicability of the findings in E-Risk to an external cohort, we will repeat similar analysis in the Dunedin Multidisciplinary Health and Development Study. The analyses will test differences in selected PRS (psychopathology; cardiovascular disease, body mass index and C-reactive protein; educational attainment, intelligence, and age at first birth; height as negative control) in children prospectively identified as maltreated, in adults who retrospectively reported a history of childhood maltreatment based on the

CTQ at age 38 years, in children with increasing number of prospectively measured ACEs, and in adults who retrospectively reported ACEs at age 38 years.

Variables Needed at Which Ages (names and labels):

**E-RISK STUDY VARIABLES**

STD DATASET VARIABLES (TWINID, SEX, ...)

CHILDHOOD POLY-VICTIMIZATION,

CHILDHOOD POLY-VICTIMIZATION COMPONENTS,

CTQ at 18,

PROSPECTIVE ACES

'p' at 18

BMI at 18

CRP at 18

NEET at 18

IQ at 18

AGE AT FIRST BIRTH

HEIGHT

**DUNEDIN STUDY VARIABLES**

STD DATASET VARIABLES (ID, SEX, ...)

CHILDHOOD MALTREATMENT,

CTQ at 38,

PROSPECTIVE ACES,

RETROSPECTIVE ACES at 38

'p' at 38



BMI at 38

CRP at 38

EDUCATIONAL ATTAINMENT

IQ at 38

AGE AT FIRST BIRTH

HEIGHT

**VARIABLES FROM BOTH E-RISK STUDY AND DUNEDIN STUDY**

PRS IN E-RISK STUDY MEMBERS (AND THEIR MOTHERS IN E-RISK):

- CROSS-DISORDER PSYCHIATRIC PRS <https://www.ncbi.nlm.nih.gov/pubmed/23453885>;  
<https://www.nature.com/articles/ng.2711>
- SCHIZOPHRENIA PRS <https://www.nature.com/articles/nature13595>
- DEPRESSION PRS <https://www.ncbi.nlm.nih.gov/pubmed/29700475>
- CARDIOVASCULAR DISEASE PRS <https://www.ncbi.nlm.nih.gov/pubmed/26343387>
- OBESITY PRS <https://www.nature.com/articles/nature14177>
- CRP PRS <https://www.ncbi.nlm.nih.gov/pubmed/30388399>
- EDUCATIONAL ATTAINMENT PRS <https://www.nature.com/articles/s41588-018-0147-3>
- INTELLIGENCE PRS <https://www.nature.com/articles/s41588-018-0152-6>
- AGE AT FIRST BIRTH <https://www.nature.com/articles/ng.3698>
- HEIGHT PRS (FOIL) <https://www.ncbi.nlm.nih.gov/pubmed/30124842>

References cited:

Baldwin, J.R., Reuben, A., Newbury, J., Danese, A. Agreement between prospective and retrospective measures of childhood victimization: a systematic review and meta-analysis. *JAMA Psychiatry*, 2019 doi: [10.1001/jamapsychiatry.2019.0097](https://doi.org/10.1001/jamapsychiatry.2019.0097)

Danese, A., Moffitt, T.E., Arseneault, L., Bleiberg, B., Dinardo, P., Gadelman, S., Houts, R., Ambler, A., Fisher, H.L., Poulton, R., Caspi, A. The origins of cognitive deficits in victimized children: Implications for neuroscientists and clinicians. *Am J Psychiatry*, 2017 174(4):349-361 doi: [10.1176/appi.ajp.2016.16030333](https://doi.org/10.1176/appi.ajp.2016.16030333)

Krapohl E, Hannigan LJ, Pingault JB, Patel H, Kadeva N, Curtis C, Breen G, Newhouse SJ, Eley TC, O'Reilly PF, Plomin R. Widespread covariation of early environmental exposures and trait-associated polygenic variation. *Proc Natl Acad Sci U S A*. 2017 Oct 31;114(44):11727-11732. doi: [10.1073/pnas.1707178114](https://doi.org/10.1073/pnas.1707178114)

## Data Security Agreement

Provisional Paper Title	The genetics of childhood victimisation
Proposing Author	Andrea Danese
Today's Date	24/03/2019

### ***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:** .....

**CONCEPT PAPER RESPONSE FORM**

**A.** To be completed by the proposing author

Proposing Author: Andrea Danese

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

Provisional Paper Title: The genetics of childhood victimisation

Potential co-authors: Moffitt, Caspi, Arseneault, other interested E-Risk investigators, Richie Poulton, Duke Genomic team, Jasmin Werts, Honalee Harrington, Jessie Baldwin

Potential Journals: Nat Hum Behav, PNAS, Mol Psychiatry

Intended Submission Date (month/year): 01/10/2019

***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:** .....