

Environmental Risk (E-Risk) Longitudinal Twin Study

Data Sharing Protocol

(updated August 2024)

Brief overview of the study

The MRC-funded Environmental Risk (E-Risk) Longitudinal Twin Study began in 1998 when a nationally representative 2-year birth cohort of 2,232 same-sex twins born in England and Wales in 1994-1995 was constructed by Professors Terrie Moffitt and Avshalom Caspi. At age 5, 1,116 families (93% of those eligible) participated in home-visit assessments and the sample comprised 56% monozygotic (MZ) and 44% dizygotic (DZ) twin pairs; sex was evenly distributed within zygosity (49% male).

Families were recruited to represent the UK population with newborns in the 1990s, to ensure adequate numbers of children in disadvantaged homes and to avoid an excess of twins born to well-educated women using assisted reproduction. More information about recruitment is available here: <https://acamh.onlinelibrary.wiley.com/doi/10.1111/1469-7610.00082> The resulting study sample represents the full range of socioeconomic conditions in Great Britain; the families' distribution is reflected on a neighbourhood-level socioeconomic index (ACORN) that very closely matches the national distribution. This national-representativeness has been maintained over the years, with the E-Risk families' addresses at age 18 being a near-perfect match to the deciles of the UK's 2015 Lower-layer Super Output Area Index of Multiple Deprivation (IMD) which averages 1,500 residents; approximately 10% of the cohort fills each of IMD's 10% bands for the UK population.

This longitudinal study has concentrated on three developmental stages to date: (1) the transition to formal schooling, with assessments at ages 5 (1999-2000) and 7 years (2001-2002), retaining 98% of the cohort, (2) the transition to secondary school, with assessments at ages 10 (2004-2005) and 12 (2006-2007), retaining 96% of the cohort, and (3) the transition to young adulthood, with assessments at age 18 (2012-2014), retaining 93% of the cohort. Home visits to participants before age 18 included assessments with primary caretakers as well as the participants whereas the visits at age 18 included interviews only with participants. You can find out more about the E-Risk Study on our website: <https://eriskstudy.com/>

Available data

Through the years, data have been collected about many different topics, including mental health, obesity, asthma, school performance, criminal offending, violence victimisation, neighbourhood conditions, the family environment, and also biomarkers to investigate inflammation, gene expression, epigenetic DNA methylation, polygenic scores, telomeres, and neuropsychological functions. Currently data are available to access for the assessments conducted at ages 5, 7, 10, 12 and 18 years. Information about the mental health measures available for E-Risk twins at different ages is provided in the Catalogue of Mental Health Measures (part of DATAMIND): <https://www.catalogumentalhealth.ac.uk/?content=study&studyid=ERisk> For detailed information about the other measures available, please contact the E-Risk Study team on eriskstudy@kcl.ac.uk Note, we have recently received funding from the MRC to collect a new wave of data when the twins are aged 30 (in 2024-2025) and this data will become available for researchers to access by the end of February 2027.

Data access

E-Risk data is free to access by researchers from all over the world (and students via their academic supervisor) who are based at universities or research organisations. The following pages of this document explain how to obtain access to this data and the procedures that must be adhered to when using the data and publishing your subsequent findings. If you have any questions at any point during the process, then please email the E-Risk Study team on eriskstudy@kcl.ac.uk and we will be more than happy to help.

Overview of data access process



* The E-Risk data manager will aim to provide a dataset for the applicant's project within 1 month of a formal data use agreement being signed (if one is required), but this will increase to within 2 months during a data collection phase as this is a particularly busy period for this member of staff.

Full details of the data access process are provided on pages 4-5 below.

Overview of data use procedures



* A reproducibility check is only mandatory where the first or senior author is an E-Risk co-investigator.

Full details of the data use procedure are provided on pages 5-7 below.

DATA ACCESS PROCESS

STEP 1: GETTING STARTED...

Please read the entirety of this data sharing protocol. The purpose of this document is to provide information about how to access and work with data from the E-Risk Study. We ask that everyone reads and follows all policies and procedures listed below. If you have any questions at any point during the process, then please email the E-Risk Study team on eriskstudy@kcl.ac.uk and we will be more than happy to help.

Before proceeding to the next step, please check the existing approved projects that utilise E-Risk data, which are available on this website: <https://sites.duke.edu/moffittcaspiprojects/> to ensure your proposed idea is not already being conducted by someone else. Projects which propose analyses that substantially overlap with a project that has previously been approved will not be considered. If in doubt, please contact Professor Helen Fisher (helen.2.fisher@kcl.ac.uk) who can advise.

Next, please contact Prof Helen Fisher (helen.2.fisher@kcl.ac.uk) to discuss the feasibility of your idea and to arrange for an E-Risk co-investigator to sponsor your project (if needed). Due to the type of informed consent we have previously obtained from Study participants, researchers who are not co-investigators of the E-Risk Study will require a sponsor who is an E-Risk co-investigator. The E-Risk sponsor takes responsibility for all activities of the project with respect to the Study, and this is a serious responsibility. Prof Fisher will liaise with E-Risk co-investigators on your behalf to ascertain who would be most suitable to sponsor your project and has the capacity to undertake this role (given the number of projects they are already sponsoring).

STEP 2: CONCEPT PAPER

Prepare a concept paper briefly outlining the rationale for your project, your research questions and hypotheses, the significance of the topic, your proposed analyses, and the variables required and at which age(s). It also includes our confidentiality and data security policies that you must agree to adhere to. This concept paper must be completed via our online form: <https://redcap.link/ERiskConceptPaperForm> You can find a copy of the fields you are required to complete at the end of this document and you can also download an offline Word template for drafting your responses from our website: <https://eriskstudy.com/data-access/>

When completing the concept paper form, please take the time to carefully consider which variables you will need (and discuss this with your E-Risk sponsor where applicable). 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 those questions. You can find information about the mental health measures available for E-Risk twins on this website: <https://www.catalogumentalhealth.ac.uk/?content=study&studyid=ERisk> For detailed information about the other measures available, please consult previous E-Risk publications (a full list is available on the E-Risk website: <https://eriskstudy.com/publications/>) and for specific variables you can search the E-Risk electronic data dictionary (available via your E-Risk sponsor where applicable).

Think carefully about the statistical analyses you will perform. We suggest you read this useful paper: Thomas, L. & Peterson, E.D. (2012). The value of statistical analysis plans in observational research: defining high-quality research from the start. *JAMA*, 308 (8), 773-774. doi: 10.1001/jama.2012.9502.

<https://jamanetwork.com/journals/jama/article-abstract/1352120>

Prof Fisher will circulate your concept paper within 5 working days of receipt to the E-Risk steering committee (which comprises Prof Helen Fisher [current PI of E-Risk], Prof Louise Arseneault [PI of E-Risk phase 18], Prof Terrie Moffitt [co-founder of E-Risk] & Prof Avshalom Caspi [co-founder of E-Risk]) and named co-authors for review. You will receive a decision via email within one month of circulation along with the signed approval forms and any comments. The proposing author is responsible for keeping all the signed copies of the concept paper approval forms. You will need them to know who should do a mock review when the manuscript has

been drafted. If any concerns are raised in the concept paper response forms, then Prof Fisher will discuss revising the concept paper with you and send it back out for review if needed.

Once approved, your concept paper will be made public on this website:

<https://sites.duke.edu/moffittcaspi/projects/> and this constitutes pre-registration.

STEP 3: RECEIVING THE DATA

Prior to receiving the data, you must have completed relevant research ethics training. You are required to confirm this in the concept paper form. For research funded by the US NIH (which includes E-Risk projects) it is a requirement that all investigators have completed a research ethics training course in human subjects' protection, responsible conduct of research, or equivalent. By accepting data from us to analyse, the researcher 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 do so by going to: <https://about.citiprogram.org/>

If you or any of the people who will need to access the data are **not** a member of staff, a student or affiliate of King's College London, then a formal data use agreement will be required between King's College London and the university or research organisation that employs you/any collaborator having access to the data. This needs to be signed by both universities/organisations before data access can be granted. Professor Fisher will be in touch with you to organise the relevant paperwork if this is required.

The E-Risk data manager will aim to provide a dataset for your project within 1 month of a formal data use agreement being signed (if one is required), or of the concept paper being approved (if a data use agreement is not required), but this will increase to within 2 months during a data collection phase as this is a particularly busy period for this member of staff. The data will be sent to you via a secure data transfer system. Please note, the E-Risk Study is not a public-domain dataset. You may not share your data file with others, except with those indicated in the concept paper as requiring access, and by submitting the concept paper you agree to this restriction.

Prior to downloading data onto your computer, you must ensure that your 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.

DATA USE PROCEDURE

STEP 4: ANALYSES

Upon receiving a data file, check the frequency distributions (and missing values) for all variables before conducting your proposed analyses. Report any bugs or concerns via email to eriskstudy@kcl.ac.uk and ask the team about ways to handle any missing data.

Please annotate your code-scripts and command files, so that others can retrace your steps later. Save all code and output files.

If you create a new variable from the raw data, please 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).

Example: **dxmde18** Major depressive episode, dsm4, age 18

STEP 5: WRITING IT UP

Don't re-create the wheel; use our existing boiler-plates for the "methods" section for consistency across publications and acknowledge all appropriate individuals, agencies and grants on your manuscript. This is not an option; it is required by our funding agencies. You can obtain the methods and acknowledgements boiler-plates from your E-Risk sponsor or eriskstudy@kcl.ac.uk. Be sure to obtain funding information (e.g., fellowships, training grants) from all co-authors and include this in your manuscript.

We strongly encourage you to complete a STROBE checklist as this is best practice for reporting of observational studies. On the checklist, please report the page number where each item has been fulfilled. The checklist can be found here: <https://sites.duke.edu/moffittcaspiprojects/strobe-checklist/>

Include page numbers on every version of your manuscript. Please name the manuscript file with your surname, a short handle for the project, and the date, spelling out the month (because the middle number can be day in the USA and month everywhere else). For example: Beckley_Victimoffender_4Sept2017.doc. When published add the journal name. For example: Beckley_Victimoffender_JDLCC_10April2018.doc.

STEP 6: MOCK REVIEW

Once the manuscript draft has been agreed between the first and senior authors (plus the E-Risk sponsor where applicable) then please circulate it via email to all co-authors and Prof Fisher (helen.2.fisher@kcl.ac.uk) for review and approval (allowing at least 3 weeks).

STEP 7: REPRODUCIBILITY CHECK (if applicable)

Once you have incorporated all mock reviewer comments, please send the revised version of your manuscript and code files to eriskstudy@kcl.ac.uk and we will arrange a reproducibility check of your statistical analyses (if applicable). All manuscripts led by an E-Risk co-investigator (as first or senior author) must undergo a reproducibility check. This is not mandatory for manuscripts led by researchers external to the E-Risk study though we would strongly encourage you to arrange for this to be done within your own teams. Checking will be carried out by an independent checker, using a fresh copy of the original dataset, separate from the dataset with which you have been working. The checker will attempt to create your code from scratch, by reading a copy of your manuscript. The goal is to re-produce all the numbers in the paper, without looking at your code. Please review the document "Reproducibility Stat Check for Dunedin and E-Risk Papers," which can be found here: <https://sites.duke.edu/moffittcaspiprojects/reproducibility-check/>

Where applicable, the independent data checker will email you regarding any discrepancies or inaccuracies in your analyses and the figures within your manuscript within 2 months of receiving your manuscript. Please then update your manuscript accordingly.

STEP 8: SUBMITTING FOR PUBLICATION REVIEW

Once you have made any further changes, please send the final version of your manuscript to Prof Fisher (helen.2.fisher@kcl.ac.uk) for a final check and approval *before* you submit it. She will let you know if any further changes are needed (e.g., to study description, funding sources, affiliations, etc) and approve it within 5 working days.

Consider running your abstract through AI software such as Elsevier journalfinder, Springer journalsuggester, or Wiley find-a-journal. Editors have dashboards that use AI to check fit to the journal (and that also check for grammar and spelling mistakes), often resulting in desk rejection.

Upon submission, send the email confirmation and all submitted files (including covering letter to the Editor) round to the co-authors and send these plus the code files, any newly created variables, and the completed STROBE checklist to eriskstudy@kcl.ac.uk so that we can keep a record of them.

Before uploading your manuscript to a preprint server, please discuss with Prof Fisher (helen.2.fisher@kcl.ac.uk).

STEP 9: DOING REVISIONS

Send the editor's letter and reviews to all co-authors and invite them to comment or contribute to the revision process. The final version of the response letter and revised manuscript must be sent to Prof Fisher (helen.2.fisher@kcl.ac.uk) for approval before resubmission (she will aim to provide comments and approval within 5 working days). We do not ordinarily distribute revisions to all co-authors, unless co-authors request to see these, or if the revisions are extensive and contain new or different information.

STEP 10: WHEN THE PAPER IS ACCEPTED FOR PUBLICATION

Celebrate! Upon acceptance, inform all co-authors and forward a copy of the acceptance email along with the word version of the accepted manuscript plus code files and outputs (including any newly derived variables) to eriskstudy@kcl.ac.uk and we will advise about making the paper open access and depositing it in the appropriate repositories.

You must now delete the data file. You may not retain the data file. The data remain the property of the E-Risk Study and cannot be used for further analyses. The data manager will archive your file for you, should you need it again in future. If you wish to run new/further analyses, then you must submit a new concept paper and go through the process above again to receive a new dataset.

STEP 11: DURING THE PROCESS TOWARD PUBLICATION & BEYOND

When you receive the copy-edited proofs of the manuscript, review this with your senior author (and E-Risk sponsor where applicable). Please check the proofs carefully. This is the last chance you will have to catch mistakes. You don't want to have to retract your paper because of a copy-editor's mistake.

Once published, please send a PDF of the paper to all co-authors and Prof Fisher (helen.2.fisher@kcl.ac.uk), providing the full citation for their CVs in the email.

If media coverage is expected, consult with Prof Fisher (helen.2.fisher@kcl.ac.uk) about drafting a fact sheet for journalists. Deliver copies of all media articles, or websites, to eriskstudy@kcl.ac.uk, as these need to be reported to our funding agencies.

If you give a lecture or presentation about your work, please notify eriskstudy@kcl.ac.uk of the title, date, and context of the presentation. We also need to report these to our funding agencies.

Thank you for reading this document. We share data with a large number of researchers, who live in many countries, and who work at many different levels of training and expertise. The purpose of this memo is to instruct students who are embarking on their first research project, and to remind senior professors who are embarking on their nth 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.

Yours sincerely, Prof Helen Fisher, Prof Terrie Moffitt, and Prof Avshalom Caspi

E-Risk Study Concept Paper offline template

Note, this is an offline version of this online form: <https://redcap.link/ERiskConceptPaperForm> You must complete the online form to submit your concept paper to the E-Risk team for consideration.

Please read the E-Risk data sharing protocol (available here: <https://eriskstudy.com/data-access/>) and then complete all the fields below. If this data request is approved, this form will be posted publicly on this website: <https://sites.duke.edu/moffittcaspiprojects/> alongside previously approved concept papers. If you have any questions, please email us at: eriskstudy@kcl.ac.uk

1. Collaborating researchers

Please note:

- *Once approved, a formal data use agreement will be required between King’s College London and the university or research organisation that employs any collaborator having access to the data if they are not a member of staff, a student or affiliate of King’s College London. This needs to be signed by both universities/organisations before data access can be granted.*
- *For projects carried out by a student (e.g., MSc/MA, MPhil/PhD, clinical doctorate), the lead applicant should be the student’s supervisor at the same university, and the student should be named as the student collaborator requiring access to the data.*
- *If you have additional collaborators, please name them below and indicate whether they need to have access to the data. It would be common, for instance, for other researchers to see summary results of analyses and act as co-authors on your paper without having access to the data. You will not be permitted to share the dataset except with those indicated in the table as requiring access.*

Category	Name	Email address	University / organisation	Needs access to data for analysis? (yes/no)
Applicant <i>(lead researcher)</i>				
Student collaborator <i>(if data is for their dissertation/thesis)</i>				
E-Risk Sponsor <i>(if applicant is not an E-Risk investigator)</i>				
Other collaborator				
Other collaborator				
Other collaborator				
Other collaborator				
Other collaborator				
Other collaborator				
Other collaborator				

2. The project proposal

Note: Please provide sufficient detail to enable the committee to review your proposal. Please be as specific as possible about the project aims and analysis methods as once approved this concept paper will be posted publicly and thus will act as a form of pre-registration of your project. Expand boxes as required.

Title of project	
Background and rationale for project: (approx. 300 - 1000 words)	
Project aims / objectives:	
Brief statement of your hypothesis:	
Data analysis methods to be used: (approx. 100 - 500 words)	
Significance for theory, research methods, or clinical practice:	
References cited:	

3. Expected project outcomes

Please note:

- *The stated end date must be within 24 months of the date when this form is submitted. This end date will form part of the formal data use agreement and on this date you should delete the dataset. Therefore, it must be a realistic date for completion of the project including all analysis, writing a manuscript, review of the manuscript by all collaborators, submission, revisions, and acceptance of a paper for publication.*
- *If you require an extension to the end date of the project, then you should contact Prof Fisher (helen.2.fisher@kcl.ac.uk) to discuss this. If you have signed a formal data use agreement, you will need to complete a form to request a licence extension. In some cases, we may also ask you to complete a new concept paper form if there have been substantial changes to the project or a long period of time has elapsed (e.g., greater than a year since the end date of the original project).*
- *If the objective of the project is not a journal publication, please suggest an end date within 12 months instead of 24 months, and state a measurable, concrete outcome. If the objective of the project is a student dissertation, then the expected end date should be the deadline for submission of the dissertation; dissertation projects will only be accepted on agreement that they are strictly not for publication.*

Date form submitted	
End date for the project	
Do you expect to publish your results in a journal? (yes/no)	
If yes, please provide a provisional list of author names	
If yes, please provide a provisional list of journals	
If no, please state the expected outcome if the project is successful (e.g., dissertation, thesis chapter, report)	

4. List of variables required

Please note:

- *When specifying variables, please be unambiguous. For each variable, specify the name of the measure, twin age, informant, and if you want specific subscales/derived categories (e.g., Depression from interview with twin at age 18; both number of symptoms and DSM-IV diagnosis). Alternatively, for maximum clarity, give actual variable names (e.g., MDESXE18 - MDE Symptom scale - P18 – Elder; DXMDEE18 - Major depressive episode, dsm4 – P18 – Elder).*
- *By default, the dataset will usually include twin and family IDs, the “random” and “true” twin order variables, the cohort the twin is from (1994 or 1995), twin sex, ethnicity and zygosity variables, and family socioeconomic status at age 5. These routine background variables are listed in the table below. If you require further background variables, please specify them in your list.*
- *Access to some parts of the dataset are restricted, namely identifiable data (e.g., postcodes, video recordings, individual-level genotypic and epigenetic data) which will not be shared outside King’s College London, and linked administrative data which is only accessible via the UK Longitudinal Linkage Collaboration’s Trusted Research Environment (this requires a separate formal data access agreement).*

Variable name	Description
<i>Background variables that will be included by default:</i>	
FAMILYID	Unique family identifier
ATWINID	Twin A ID (ex chkdg)
BTWINID	Twin B ID (ex chkdg)
RORDERP5	Random Twin Order
TORDER	True Twin Order
RISKS	Sample Groups
COHORT	Cohort
SAMPSEX	Sex of Twins
ZYGOSITY	Zygosity
SETHNIC	Ethnicity of Twins
SESWQ35	Social Class Composite
<i>Age 5 variables:</i>	
<i>Age 7 variables:</i>	
<i>Age 10 variables:</i>	
<i>Age 12 variables:</i>	
<i>Age 18 variables:</i>	
<i>Age 26 variables:</i>	
<i>Age 30* variables:</i>	
Are you requesting access to identifiable or linked data? (yes/no)	
Which format(s) do you require the data in? (e.g., .csv, SPSS, Stata, etc)	

* Note these variables will not be available until early 2027.

5. Data security agreement and signature

Please click in each box to indicate that you will adhere to each of the points listed below.

<input type="checkbox"/>	I am current on Human Subjects Training (CITI (www.citiprogram.org) or equivalent)
<input type="checkbox"/>	My project is covered by the King's ethics committee OR I have /will obtain ethical approval from my home institution.
<input type="checkbox"/>	I will treat all data as "restricted" and store in a secure fashion. My computer or laptop is: a) encrypted (recommended programmes 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.
<input type="checkbox"/>	I will not "sync" the data to a mobile device.
<input type="checkbox"/>	In the event that my laptop with data on it is lost, stolen or hacked, I will immediately contact Prof Helen Fisher (helen.2.fisher@kcl.ac.uk), PI of the E-Risk Study.
<input type="checkbox"/>	I will not share the data with anyone, including my students or other collaborators not specifically listed on this concept paper as requiring access to the data.
<input type="checkbox"/>	I will not post data online or submit the data file to a journal for them to post. <i>Some journals are now requesting the data file as part of the manuscript submission process. Study participants have not given informed consent for unrestricted open access, so we have a managed-access process. Speak to Prof Helen Fisher (helen.2.fisher@kcl.ac.uk) for strategies for achieving compliance with data-sharing policies of journals.</i>
<input type="checkbox"/>	I will delete all data files from my computer after the project is complete or by the end date specified on this form, whichever is earliest. Collaborators and students may not take a data file away from the office. <i>This data remains the property of the Study and cannot be used for further analyses without an approved concept paper for new analyses.</i>
<input type="checkbox"/>	I have read the Data Sharing Protocol and agree to follow the instructions.

By submitting this document, the applicant confirms that they have read and understood their responsibilities as described in the E-Risk data sharing protocol (see: <https://eriskstudy.com/data-access/>) and that they agree to comply with them if this data request is approved. The applicant is asked to type their full name below and select today's date.

Type your full name:	
Date:	

Note, you will need to copy and paste the information above into this online form:
<https://redcap.link/ERiskConceptPaperForm> in order to submit it to the E-Risk Study team.

Assuming no additional information or amendments are required to process your form then the E-Risk Data Sharing Committee will aim to provide you with a decision within one month of submission. If approved, we will then securely provide access to the requested data within 1-2 months of a formal data use agreement being signed (if one is required).