



TEDS RESEARCH CENTRE  
Dept. Box No. P083  
Freepost RTSS-XCUX-CHTR  
London SE5 8AF  
Freephone: 0800 317 029  
Email: teds-project@kcl.ac.uk

## Twins Early Development Study: Privacy Policy for use of Twin DNA and Data, 2020

TEDS is an academic research project based in **King's College London (KCL)** and funded by the **Medical Research Council (MRC)**. As such, we aim to comply with the policies of KCL and MRC with respect to the storage, processing and sharing of research data. We comply with the **UK Data Protection Act 2018 (DPA)** and with the **EU General Data Protection Regulation 2016 (GDPR)**. We are sending this updated privacy policy to all TEDS twins because our *legal basis* (see below) for processing participant data has been redefined under these recent regulations. Under GDPR, the term **processing data** refers to everything we do with participant data, including collection, storage, conversion into different forms, and analysis for research.

Over the course of the TEDS study, we have collected a range of **participant materials** from twins and their families. These include contact details, data from questionnaires and tests, DNA from cheek swabs and saliva samples, letters, emails and photos. In this document, we will explain how these materials have been processed.

### What data and materials have we collected from participants?

Most of the data collected by TEDS from participants are for the direct purposes of our research, and are referred to as **research data**. These include contact details, data from questionnaires and tests and data derived from the DNA samples. Research data fall into several categories as described below.

**Administrative data** (a type of *research data*) include names, genders, birth dates, postal addresses, phone numbers and email addresses for twins and their parents, in addition to records of participation in current and previous TEDS studies. We also hold names, genders and birthdates of siblings of twins. The administrative data have been collected since the start of TEDS and contact details are continually updated whenever we make contact with participants. Some of the administrative data are also used in research analysis, for example twin genders, birth dates (for calculating ages at specific time points) and postcodes (for linking with other data, see below). However, the day-to-day use of the administrative data is for management of TEDS studies and contacts with participants.

*Research data* more generally include the data from all activities that we have asked participants to carry out since TEDS started. These activities have included questionnaires (by post and online) and cognitive tests (online, by phone and in person). In the early years of TEDS, these data were all supplied by parents of the twins, and we asked parents to carry out simple tests of intelligence on the twins. We have continued to ask parents to fill in questionnaires about the twins ever since. For selected families at age 4, our own staff made visits in order to test the twins in their homes. We used telephone testing to collect data from twins at age 7; since then, we have asked twins to fill in their own questionnaires and to carry out web tests. We collected data from the teachers of twins (between ages 7 and 14) by asking parents to supply the names and school addresses of the teachers at each stage. We collected similar types of data from younger siblings of twins, using parent questionnaires and telephone testing, when those siblings were aged 2 to 7. All these varieties of *research data* are called **phenotypic data**, to distinguish them from *genotypic data* (see below).

We have also collected twin **DNA samples**. When twins were children, this was done using cheek swabs supervised by parents and with parental consent. More recently, we have collected saliva samples from adult twins with their own consent. In both cases this was done by post. When returned, the cheek swab and saliva samples were processed in our labs to extract DNA samples. The DNA samples have been retained, in frozen form, while the original cheek swabs and saliva have been destroyed. Many of the DNA samples have been **genotyped**, which means that information about each twin's genetic sequence has been derived, in coded electronic form. These data are referred to as **genotypic data**, which is another category of *research data*.

All participant *research data*, including the *administrative*, *phenotypic* and *genotypic data*, are classified as **personal data** under *GDPR*. Some are further classified as **special category** personal data: these include (where provided) data relating to ethnic origin, health, sexual orientation and behaviour, political opinions, religious beliefs and all *genotypic data*. All *participant materials* have been collected in identifiable form but, as explained in the next

sections, the data are ultimately transformed into *anonymous* or *pseudonymous* form before they are used by researchers for analysis.

Some twins and their families have given us photos for specific uses (described below), with their consent. The **twin photos** are essentially used in TEDS as publicity materials and are not classified as *research data*.

### Who uses the TEDS participant materials?

The TEDS **admin team** includes the Co-ordinator, the Data Manager and Research Assistants, all based in the TEDS office in the **SGDP Centre** building at Denmark Hill, London. These staff are responsible for collecting all *participant materials* and for tasks such as data entry and data storage. This necessitates handling the participant materials in identifiable form.

**Lab staff**, also based in the *SGDP Centre*, have processed the cheek swab and saliva samples in order to extract and store the *DNA samples* and subsequently to obtain the *genotypic data*.

**Researchers**, of various types, are provided with *phenotypic* and *genotypic data* in non-identifiable form (see *anonymous and pseudonymous data* below), for the purpose of statistical analysis in order to test their research ideas. Researchers include **TEDS researchers**, employed by KCL as TEDS staff and based in the *SGDP Centre*; and **research collaborators** who may be **KCL collaborators** employed by KCL but not as TEDS staff or **external collaborators** employed at other universities. External collaborators may be temporarily based at KCL for specific research projects, in which circumstances they are called **research affiliates**. In addition to university employees in these various categories of researcher, access to the data may be given to postgraduate research students (typically PhD students) under the direct supervision of such employees. The sharing of the valuable TEDS data with collaborators is encouraged by TEDS, by the *MRC* and by *KCL*, as a way of getting the best possible value in terms of research benefits such as developing scientific understanding and improving public health.

### How and where do we store participant materials?

TEDS has a secure database system for storing participant *administrative data* such as contact details; and a network share drive for storing the *phenotypic data*, the *twin photos*, and other files necessary for administering the study. Both the database and the share drive are part of the KCL computer network; only members of the TEDS *admin team* are authorised to have access to them.

Where participants have sent information by letter or by email, the necessary *administrative data* have been transferred into the database system then the original letter or email has been destroyed or deleted. Similarly, all *twin photos* have been converted into electronic files, and original physical prints have been destroyed.

In a similar way, all *phenotypic data* are ultimately converted into electronic formats. In questionnaires and tests, each participant response is typically converted into a numeric category, or occasionally a numeric quantity (such as height), or more rarely as text, and these are stored as electronic data. Information collected by phone interviews and during home visits were originally recorded on paper questionnaires or score sheets, and again converted into electronic form by similar means. When the accuracy of electronic transfer has been verified, the original media (usually paper) are destroyed.

*Phenotypic* and *genotypic data* are converted into non-identifiable forms (see below) before they are shared with *researchers*. *TEDS researchers*, *KCL collaborators* and *research affiliates* may all be given access to share drives in the KCL network where such data may be stored. Note that these locations are separate from those used by the *admin team* as described above, so researchers do not have access to identifiable *personal data*.

*External collaborators*, including *research affiliates*, are only provided with access to the TEDS data (again in non-identifiable form) under strict conditions that are governed by legal contracts called **data sharing agreements**. These agreements specify, among other things, that the data must be stored securely, not shared with any third parties, are only used for the research purposes agreed in advance, and are deleted when the research project has ended.

The twin *DNA samples* are stored in freezers within the secure labs at the *SGDP Centre*. Access to the labs is tightly controlled and restricted to authorised and qualified *lab staff*. Each DNA sample is identified by ID number, but the lab records do not contain other identifying information.

### How do we use the participant materials?

*Administrative data* such as twin, parent and sibling contact details are used solely for the purpose of making contact in relation to our studies, for example to ask for new research data or to send newsletters. Administrative details

such as names and contact details are not shared with any third parties except for related research studies within KCL and, even then, they are only shared with the explicit consent of participants.

*Twin photos* are occasionally used in academic presentations, publications, related media coverage, our web site and social media; such photos are collected from twins with their explicit consent for the given use. If we retain any twin photo for more than two years for any purpose, for example on our web site, we will seek fresh consent.

Parts of the *administrative data*, namely the postcode, may be linked to information about the area in which a participant lives, for the purpose of statistical research. Where this has been done, the postcode itself is not provided to researchers, and steps are taken to ensure complete anonymity in the research data that are given to researchers for analysis.

The research data that we have collected from participants are used in statistical research projects. These projects are carried out by TEDS researchers and by external collaborators, both within KCL and elsewhere (including EU and non-EU universities). The data are only shared with bona fide academic researchers affiliated to recognised university research departments. The conditions under which we share the data are described on our data access policy web page (<http://www.teds.ac.uk/researchers/teds-data-access-policy>). These conditions are designed to give maximum protection to the way in which the data are used and especially to the confidentiality of participants. All the research, whether in KCL or by collaborators at other universities, involves statistical analysis, and the results are published in peer-reviewed academic journals.

Before TEDS research data are shared, they are made *anonymous* or *pseudonymous* (see below), so that individual participants cannot be identified from the data.

The *DNA samples* are being retained because we plan to use them to extract more detailed genetic information including full-gene sequencing. The DNA samples will not be shared with external collaborators but will only be used by authorised researchers within KCL.

The *genotypic data*, obtained from DNA, form an important part of our ongoing research. The raw genotypic data will be kept and analysed securely at KCL and not transferred externally, because of the sensitive nature of these data and because of risks related to identifiability of the data. Any *external researcher* wishing to analyse the genotypic data may only do so with the involvement of a researcher affiliated to KCL, who may then carry out the analysis of the data at KCL: the summary statistics from analysis, but not the raw genotypic data, may then be shared externally.

Information derived from the genotypic data, called “polygenic scores”, are very valuable for current research. These are shared with external researchers because they do not carry any risk of identifiability – they are decimal numeric scores associated with specific psychological or physiological traits.

### **Anonymous and pseudonymous data**

Each piece of data in its original or raw state from a participant, for example in a paper questionnaire, is identifiable: the TEDS *admin team* needs to know the identity of the participant who supplied it. Each participant is identified by an ID number, and by personal details such as names that are linked to the ID number. Data in this identifiable state is not suitable for sharing with *researchers*, for reasons of confidentiality.

Before passing data to researchers for analysis, the *admin team* take every possible step to ensure that the data are not identifiable to those researchers. We do this firstly by removing direct personal identifiers such as names, birth dates, postcodes and any combinations of rare or unusual responses. One way of doing the latter is to remove verbatim text responses made by participants, and by converting all responses (in questionnaires and tests) into common numeric categories. Secondly, we change the ID numbers so that researchers are unable to link the data back to any identifiable participant. This can be done in two ways: by making the data *anonymous* or *pseudonymous*.

In **anonymous** data, the ID numbers are randomly and irreversibly encrypted: they cannot be changed back to the original IDs, therefore they cannot be linked back to identifiable participants. In nearly all instances (with exceptions explained below), TEDS data shared with *researchers* are made anonymous as this is the best means of protecting confidentiality.

In **pseudonymous** data, the ID numbers are reversibly encrypted: given a suitable **key**, it is possible to decrypt the ID numbers and return them to their original identifiable state. In some circumstances, where it's necessary for researchers to link different TEDS datasets together, *pseudonymous data* are shared with researchers working within KCL (*TEDS researchers, KCL collaborators and research affiliates*). Typically, this is necessary when researchers need to link *phenotypic data* with *genotypic data*. For *external collaborators*, only anonymous data are shared. Those

researchers who use pseudonymous data are not given access to the encryption/decryption *key*, so they cannot identify participants: the *key* is stored where it may only be accessed by the *admin team*.

For research purposes, it is often necessary to link together the data from different sources for the same participant, for example by linking the data from questionnaires at different ages in order to analyse changes over time. This linking is done by the *admin team*, by matching participant ID numbers, prior to making the ID numbers *anonymous* or *pseudonymous*, and before passing the resulting dataset to a *researcher* for analysis.

### **How long will we keep the DNA and data?**

As TEDS is an ongoing longitudinal study, without a fixed end date, the electronic *research data* will be kept for the duration of the study. When the TEDS study eventually ends, the research data will be archived for future research purposes, but in an anonymous form whereby participants are not identifiable. *Administrative data* such as names and contact details will be deleted when the TEDS study ends.

In a similar way, the twin *DNA samples* will be kept indefinitely for future ethically approved research.

### **What is our legal basis for processing your personal data?**

The purpose of TEDS is to conduct scientific research with the aim of improving the public good. This is part of the wider purpose of KCL to conduct research aiming to improve scientific understanding. It is also part of the wider purpose of our funders, the *MRC*, in promoting scientific research for the purpose of improving public health. Hence our official **legal basis** for processing your *research data*, under *GDPR* and the *DPA*, is the performance of a task carried out in the public interest; and the legal condition that we rely upon for processing your *special category personal data* is “processing for scientific and statistical research purposes”.

In the small number of cases where we have retained *twin photos*, our *legal basis* for processing them is the consent given by the respective twins. These photos do not form part of the research dataset as such but are used as publicity materials as described above.

These *legal bases* are separate to, and additional to, issues of ethical approval for TEDS and consent to participate in the study. Whenever we collect data from a participant, we only do so after the study participant has been provided with information about the study and has (explicitly) consented to participate in the study. Prior to any data collection, the conduct of the study is passed by an ethics committee within KCL.

### **What are your rights as participants?**

As a participant, you have the right to **withdraw** from TEDS at any time. If you make this decision, we will delete your contact details and agree not to contact you again. However, under the *legal basis* described above, we would still have the right to continue to use any other *research data* that you had previously provided, after your initial consent to participate in the study and prior to withdrawal.

Because TEDS is a longitudinal study, we will typically contact you repeatedly over a period of years for different TEDS studies, if you provided initial consent for us to do so. You may decide to **opt out** of any given TEDS study, meaning that you may decide not to participate in the study but with the option of participating again in subsequent studies. If you decide to *opt out* in this way, please let us know when we have invited you to take part.

You have the right to receive a copy of all *administrative data* that we hold about you, for example your contact details. You may also ask us to correct or delete parts of the contact details that we hold, such as email or phone details. You may also ask us to correct any errors that you think you made when providing your *research data*.

You have the right to ask us to destroy your *DNA sample*, if you have any objection to us using your DNA for further research in the future.

You have the right to ask us to delete any photos of you, held in our files or published on our web site.

Related to our *legal basis* for processing *research data*, as stated above, there are research exemptions under the *DPA* that mean that, upon your request for erasure of *phenotypic or genotypic data*, or for copies of the same, we may not comply. This is because such requests would be detrimental to the research purposes for which the data were originally collected.

If you have any questions about TEDS, including requests to exercise your rights as stated above, you can contact us directly using the details at the top of this document. If you have a complaint about TEDS, or a query concerning our use of your personal data, you should in the first instance contact the TEDS *admin team* ([teds-project@kcl.ac.uk](mailto:teds-project@kcl.ac.uk)). If your query is not then resolved, you may contact the KCL **Data Protection Officer**, whose contact details are on the

KCL core privacy notice page (<https://www.kcl.ac.uk/terms/privacy>). If you are not satisfied with the response given by the KCL Data Protection Officer, you are entitled to complain to the **Information Commissioner's Office** (<https://ico.org.uk/global/contact-us/>).