

Data Access Policy for TwinsUK

Department of Twin Research (DTR)

King's College London

Management and data access/sharing policy of the Department of Twin Research & Genetic Epidemiology (DTR) and the TwinsUK resource.



Introduction

This document summarises the management and data access/sharing policy of the Department of Twin Research & Genetic Epidemiology (DTR) and the TwinsUK resource. The DTR is part of King's College London, and it is based at St. Thomas' Hospital campus, part of Guy's and St Thomas' NHS foundation Trust (GSTT).

The DTR encompasses the largest adult twin registry in the UK: TwinsUK.

TwinsUK is a national longitudinal cohort study founded in 1992, which now comprises 16,000 male and female, identical and non-identical twin participants aged 18-100.

TwinsUK aims to investigate the genetic and environmental basis of a range of complex diseases and conditions. The overarching aim of the study is to promote physical, mental, and social well-being through providing understanding of the genetic and environmental epidemiology of health and disease across the adult life course. The cohort has extensive biological and questionnaire data available on several hundred phenotypes and it is one of the most genotyped and clinically detailed in the world.

Further details on the DTR and the TwinsUK cohort can be found on our website (<http://www.twinsuk.ac.uk>)

TwinsUK is currently set up as a **supported access resource** for the research community. All data access requests are overseen by the TwinsUK Resource Executive Committee (TREC).

Data Access– Overview

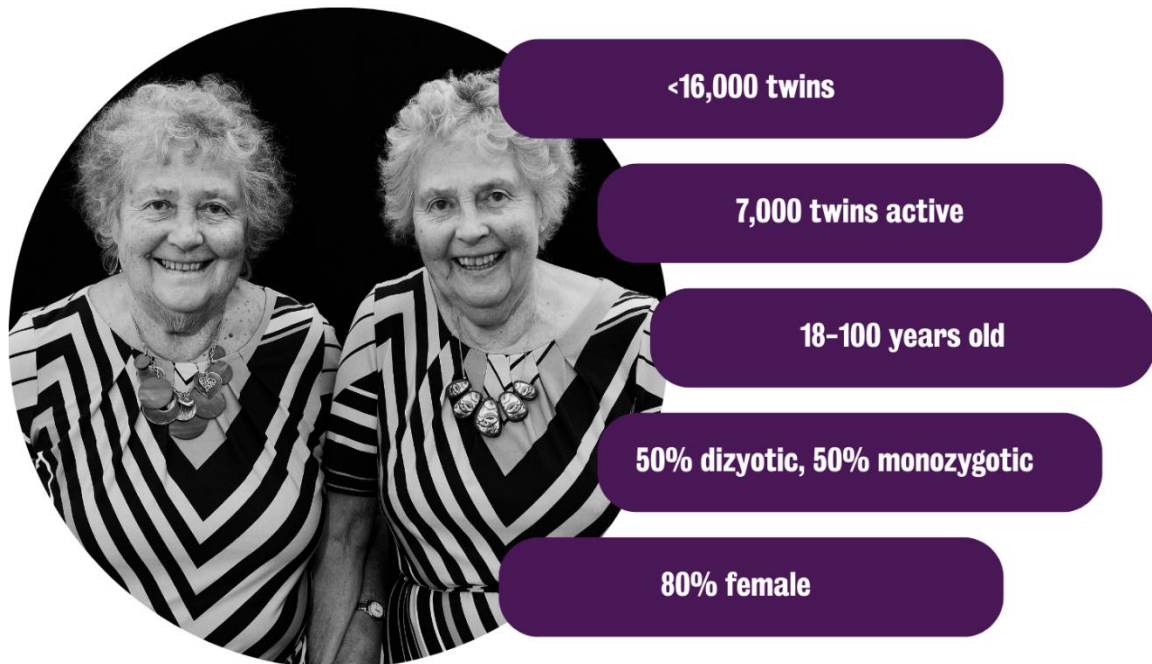
The Department is looking to facilitate and encourage the access and sharing of data and material with the world scientific community to promote and contribute to further scientific research and generate new ideas and knowledge.

The Department of Twin Research is committed to supported access of data and materials, under the oversight of the TwinsUK Resource Executive Committee (TREC). It does this by:

- I. Facilitating multidisciplinary research and general access to phenotypic, genotypic and omics data for a diverse set of clinically related traits
- II. Publishing in peer reviewed journals as well as creating an open forum to interact with the scientific community

Cohort Description

Our cohort of research volunteers, currently comprises 16,000 twins registered on our database, aged 18 to 100 with approximately equal numbers of identical (MZ) and non-identical (DZ) twins. We have data and material on around 11,000 twins. The cohort is mostly female (80%) for historical reasons.



Our cohort provides detailed collection of physical, physiological, behavioural and lifestyle data as well as biological sampling carried out via:

- Twin visits to the DTR clinic run by a highly experienced clinical team.
- Self-administered questionnaires sent to the twin volunteers via email and post every six to twelve months.
- Self-collected biological samples via postal kit

At the Health Research Authority's (HRA) advice, TwinsUK is now set up and governed as a Research Tissue Bank (IRAS: 258513, REC reference 19/NW/0187), with broad ethics and consent covering all future studies, sample collections and sharing of data for health research. This allows us to be flexible in setting up new studies.

Types of Data

We have collected extensive **biological, clinical and questionnaire** data on several hundred phenotypes related to common diseases or intermediate traits. New data are collected and made available to the researchers regularly. We are also developing a program where we will link to data made available by various third parties.

Questionnaire Data & Clinic Data

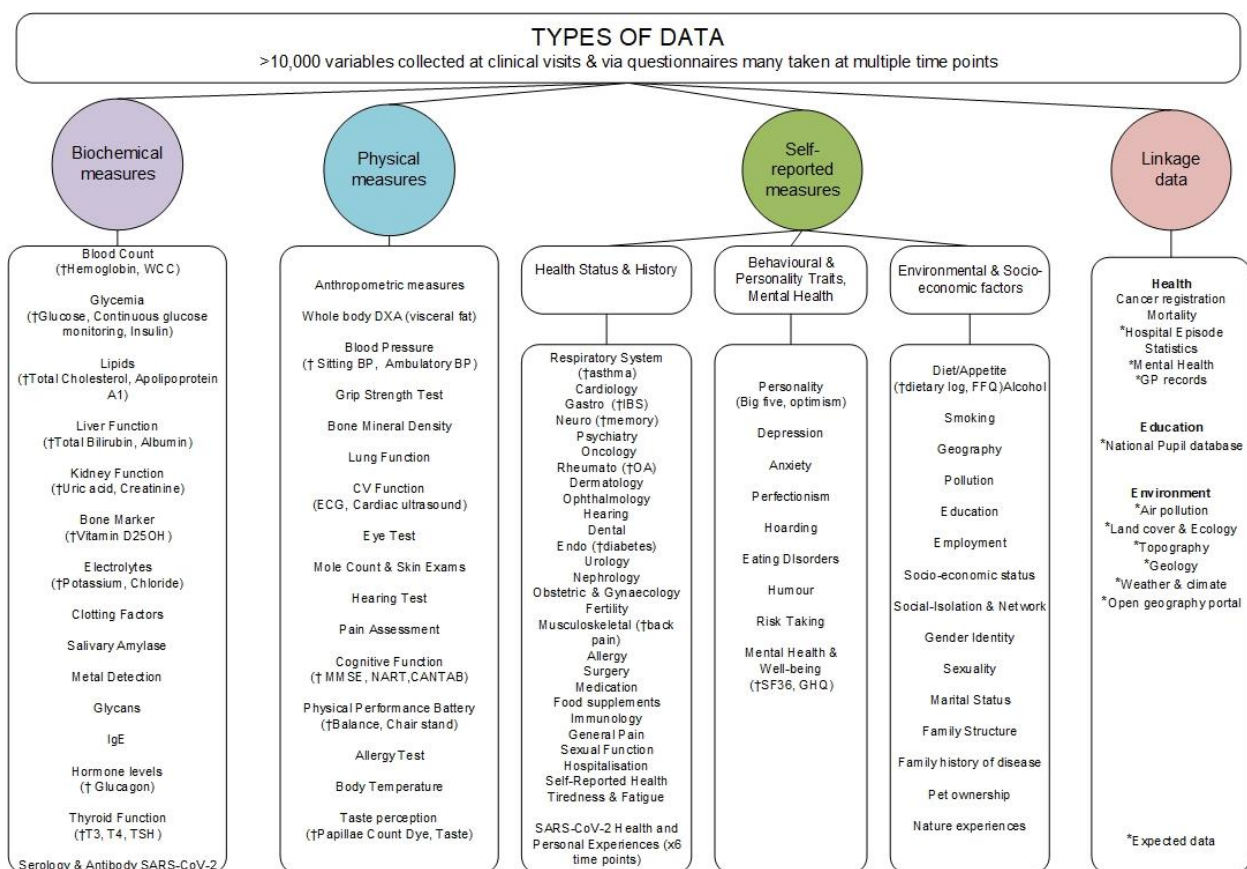


Figure 1 Types of data collected

Applicants can access a list of phenotype descriptions held at the DTR via our website <https://twinsuk.ac.uk/resources-for-researchers/our-data/>

Nutrition Data (Predict Studies)

For access to our Predict studies data on:

- Berry et al, 2020 (<https://www.nature.com/articles/s41591-020-0934-0>)
- Ascinar et al, 2021 (<https://pubmed.ncbi.nlm.nih.gov/33432175/>) (PREDICT studies)

Please contact the **Data and Access Collaboration Manager** and include the following information:

- A brief aim of the project (1-2 lines)
- Summary of the data required (high level)

The committee will first assess if the data is available in a suitable linked format to address your hypothesis prior to completion of a full proposal.

Linkage Data

To further strengthen the resource, the DTR is linking the TwinsUK data with external data sources on as health, education, and environmental records. This is what we call linkage data. These linkages are in development.

The rules governing access to linkage data will differ between datasets and might change over time.

Samples & Omics

TwinsUK holds a wide range of samples & omics datasets.

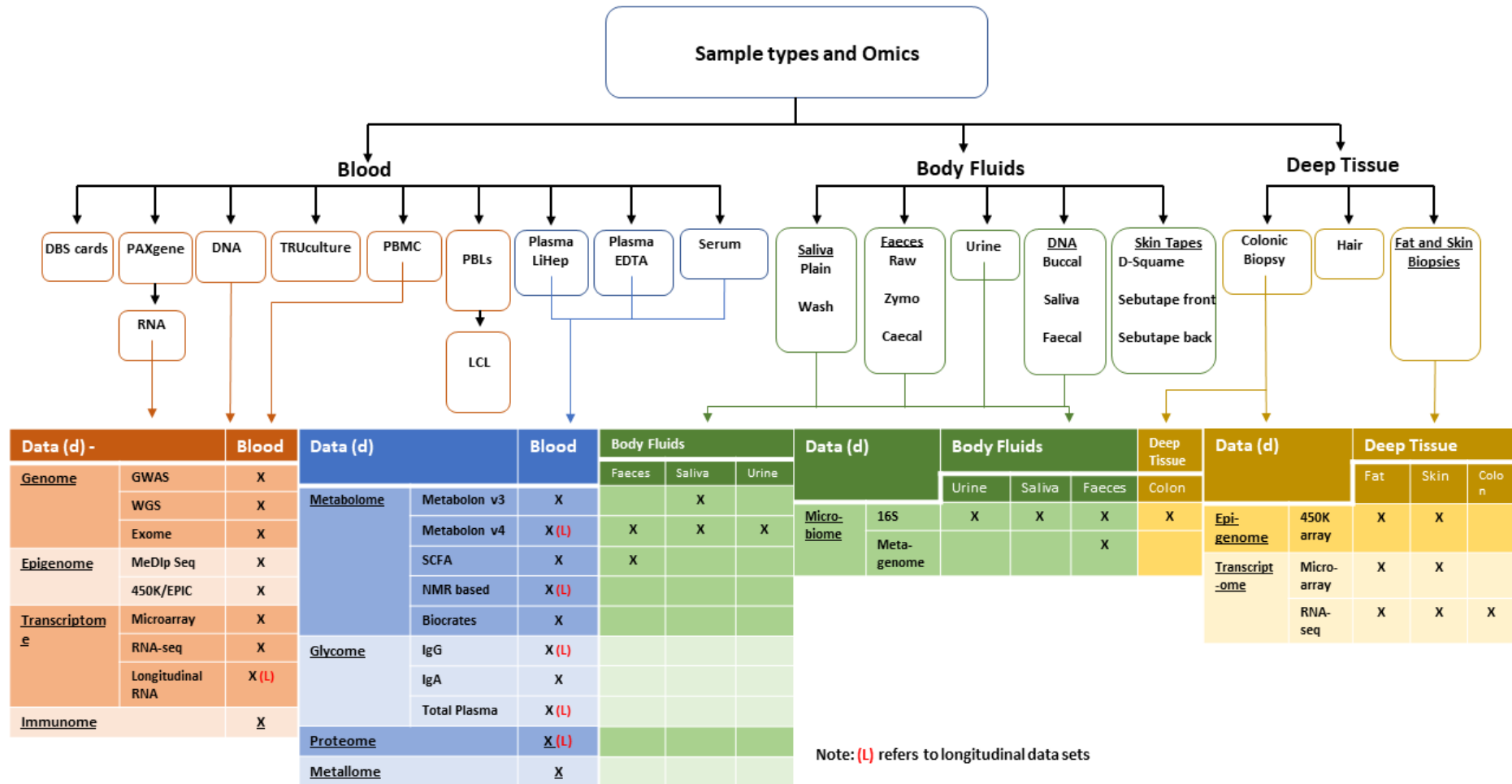


Figure 2 Multi Tissue Samples and Omics datasets

Note that whole or subset of some omics datasets are deposited on archived platforms such: as

- Array Express
- EBI
- EGA
- GEO
- ReShare UK Data Service

An overview of the samples and Omics held at the DTR is available via our website

Sharing of Data

The DTR would like to state that it cannot guarantee data exclusivity.

If there is a possibility of projects overlapping or other investigators/groups working on related topics, the TwinsUK Resource Executive Committee (TREC) may advise you to contact these groups and invite you to discuss your ideas before you proceed with your request.

Where the applicant funds the collection of specific data, we will establish an embargo period for the applicant to work exclusively on the data for either twelve months or after the first publication is accepted (whichever comes first). At this point the data will be made available to other applicants.

Access to linkage data will be made available only to researchers at King's College London via a secure platform.

Once Access to the data is granted, the data cannot be shared with third parties, including journals, without prior permission from the TwinsUK Resource Executive Committee.

Management

Professor Tim Spector is the **Scientific Director** of TwinsUK at DTR, and management of the resource is the responsibility of the **TwinsUK Resource Executive Committee (TREC)**.

The committee members are elected from the academic, operational and laboratory fields as well as members of the twin cohort.

The TwinsUK Resource Executive Committee meets twice monthly to consider submission of data request proposals and papers.

Decisions on the most complex proposals will be escalated to the Consultancy and Service Contracts Lead, Research Management & Innovation Directorate at King's as will potential conflicts of interest and commercialisation opportunities.

We may also consult our Volunteer Advisory Panel (VAP).

The VAP is a consultative body formed of a minimum twelve TwinsUK participants representative of the cohort who meet twice a year and respond to email consultations throughout the year. A further twelve TwinsUK participants (minimum) are consulted over email throughout the year only ("eVAP").

The VAP and eVAP represent the views of the cohort and offer advice regarding study design, protocols, sample collection, participant communications, governance, ethical issues and more.

Further information on management of this committee can be found on: <https://twinsuk.ac.uk/media-and-engagement/twins-engagement/>

Submission Procedure

Requests for collection of new or existing data/material should be processed by submitting a completed online **DTR Data/Material Access Proposal Form** describing your scientific proposition to the TREC.

You can access the form via our website: https://twinsukapps.kcl.ac.uk/data_request

The application process is as follows:

- Register (email and password) and verify the email address
- Provide your details including full name, position, institution, and contact information. Always enter your institution email. Requests from personal email accounts will not be accepted.
- Ensure that you enter full details of your co-applicants and ensure that they agree with your application.
- Provide the data/sample request details, ensuring that you provide as much information as possible.
- You will need to read and sign the Terms of Agreement section

When you have completed your proposal and uploaded all the required documents, you can then submit your proposal.

The proposal must fully describe in detail:

- Hypothesis or research question
- Aims of the proposed research
- Data and/or samples required to meet the aims of your proposal
- Individual variables with an appropriate justification describing how they will meet the aim/hypothesis of the project for which the data is requested.
- All phenotypes and potential confounders that you will require for your research and analysis. Subsequent requests for additional data will require you to submit a new Data Access Application
- Ethics information
- Funding information

On completion and submission of your data access application the Collaborations and Data Access Manager will submit the proposal for discussion at the next available TREC meeting.

All proposals will be submitted to TREC within four weeks from submission.

Data Access – Proposal Form

All the forms should have a comprehensible and detailed description of the below areas:

Science

Describe clearly the hypothesis, rationale, methods and aims of the scientific project for which the data is required

Data

List and describe clearly the data and samples required indicating if you require new or existing data, type of data, phenotypic or genotypic data and if so from which subset. Individual variables need to be specified.

Funding

Please indicate the source of funding for the project. When considering applying for a grant, please specify if you require a member of the DTR to be a co-applicant.

Ethical approval

Please indicate if you have ethical approval and provide us with the name of the Ethics Committee. If relevant, a copy of the ethical approval letter will need to be submitted with the proposal. If necessary, the DTR will obtain ethical approval from their Research Ethics Committee. Some proposed projects may already be covered by the TwinsUK ethics approval.

DTR resources required

Please indicate if additional DTR resources will be required in the form of statistical or logistical support and specify the level of support needed.

Data Access Request – Outcome

Upon submission of the Data/Material Access Proposal form, the TREC will review the proposal at their next meeting.

Information on the outcome will be provided (approved, rejected or pending) within one month from discussion.

You will then be informed of the feasibility of the proposal, and advice on the next stage will be promptly provided. The committee will take different factors into account when deciding the outcome of the data / access request:

1. Scientific and practical feasibility of proposed project
2. Availability of data
3. Impact of the project versus data/material depletion
4. Costs that the project may entail (see 'Costs' below)
5. Ethical approval for this project

Proposal Approved

Once the request is approved by the TREC you will be informed promptly by our Collaborations and Data Access Manager.

As part of our cost recovery process a project management fee will be applied, and you will be asked to confirm agreement (please see further information on our COSTs section).

Bespoken charges may be applied for the collection of biological samples, new data samples and complex data analysis.

You will be asked to provide the financial information listed below, so an invoice can be expedited.

- Full name of your institution
- Company registration number
- Full address of your central finance dept
- Telephone of the central finance dept
- Email of the central finance department
- URL web address
- Purchase Order (PO) Number if your institution has a compulsory PO policy
- VAT number (for UK institutions only)

Transfer of data and materials may require a data or material transfer agreement (MTA) between King's College London (KCL) and the applicant's institution.

Please note – at this stage, negotiations will not be taking place between you and TwinsUK, but between the appropriate departments within KCL and your organisation. The agreement will be signed by these

departments on behalf of you and TwinsUK. Once the agreement has been signed, KCL Enabling agreements will send a copy to TwinsUK and the data/material will be released.

Depending on the nature of the data/material access request this contract will refer to:

- (i) **Transfer agreement for new or existing data**
- (ii) **Transfer agreement for new or existing biological samples**

Once the agreement is in place and the invoice has been expedited, a representative from our Data Team / Laboratory Team will contact you to grant access to the data and/or material requested so you can start working on your project.

New data collection funded by the applicant is subject to a twelve-month protected period from the point at which the relevant dataset is provided for use. After that period, other investigators will be free to apply to use these data.

Linkage Data will be available to King's College London researchers on request. An agreement with NHS Digital will be implemented.

Once the agreement is in place, access will be provided via a King's College London secure platform. The data will be provided deidentified and deanonymized.

Applicants must return a copy of the final dataset used in their analyses along with derived variables and descriptions of these variables on completion of the project or after submission of first publication. **The data cannot be used for any other project or shared with third parties, including journals without prior consent from the TREC.**

Proposal Rejected

The DTR always strives to facilitate the sharing of data within the scientific community. However, in a few cases data requests or proposals may be rejected or deferred.

Potential reasons for submission rejections are listed below:

- I. The DTR does not store the data that the applicant requests
- II. The DTR may reject requests for depletable resources when the scientific aims of the project do not justify this. This will apply to biological samples requests
- III. The request of data may be deferred, not rejected, to comply with the embargo period awarded to another previous request for the same data
- IV. Collection of the requested data/samples conflicts with our duty of care to not overburden the twins. When applicable, requests may be deferred to the HRA NHS Research Ethics Committee for their input and/or our Twin Advisory Panel (VAP)

Proposal Pending

You may be asked to provide further information on your project so a decision can be reached if the initial information proved insufficient.

When there is a possibility of overlap with other investigators or other groups are working on related topics, then the TREC may put you in touch with these groups and invite you to discuss your ideas with them before you proceed with your request.

Proposal - Proposed Audit

On project completion it is compulsory for every scientist using our data to send back to the DTR the list of new derived variables resulting from the project and a list of publications emanating from the research.

The DTR will contact you twelve months after approval of your submission and ask you to fill in an audit form. This form would aim to ascertain if the project reached the targets listed in the initial proposal, expected date of completion, list of publications, and a request to submit new variables resulting from the project to the DTR as well as scientific abstracts and press releases.

If the project has not been completed, a further audit form will be sent to you every six months until completion.

The Audit Forms will be sent to you by the Collaborations and Data Access Manager.

Types of Data Requests

Analysis of Existing Data

Our existing data covers:

- Phenotypes: self-reported data from questionnaires / physical and clinical measures / biochemical
- Omics: genetic data, metabolomics, genetic, proteomics, microbiome, transcriptomics (expression data), methylation

Once the request for access to specific existing data has been approved, and the necessary contract/s signed, the relevant data will be released.

If analysis help is needed after data transfer, this will need to be approved by the committee who will discuss feasibility and costs.

For small datasets (less than 1 gigabyte), the data team will contact the applicant via email and provide a secure link to the data. The data will be available for access for one month following the initial email. If access is required for longer, the applicant must request an extension from the data team lead.

For larger datasets, where email is not an option due to size and bandwidth, files will be made available via various methods depending on the size, type, and sensitivity of the data.

Methods of sharing may include but are not limited to:

- Secure File Transfer Protocol (SFTP)
- MS SharePoint
- Globus File Transfer Service
- Public file sharing stores such as The European Genome-Phenome Archive (EGA)

The Password is not transferable, and breach of this condition will result in termination of the relevant agreement. It is the responsibility of the applicant to inform the DTR of any changes to individuals working on the data.

Collection of New Data

The DTR study team collects new data from the twins via:

1. Self-completed questionnaires, posted or emailed to the twins
2. Clinical twin visits at the DTR
3. Self-collected biological samples via postal kit

The implementation of questionnaires, sample collections and twin visits will require **funding** (see **Costs** section below) and a summary of costs will need to be revised and agreed by both parties before the new data collection can commence.

Collection of new data requests may require consultation with the TwinsUK Volunteer Advisory Panel (VAP) and eVAP.

Collection of new data via a questionnaire or clinical test will be coordinated by the DTR data team to guarantee the confidentiality of our twins.

Further information on questionnaire guidelines and twins' visits guidelines can be found at <https://twinsuk.ac.uk/resources-for-researchers/access-our-data/>

Assays on Biological Samples and Genotyping

To request biological samples or to carry out genotyping on DTR DNA you will need to provide us with details of:

- Type of sample
- Quantity
- Minimum Concentration (DNA) needed

We will be unable to process requests which do not supply this information.

Decisions on the use of biological samples will consider:

- amount of the stored sample required
- amount TwinsUK holds in storage
- perceived scientific value of the proposed study.

It is our general policy that when DNA samples are requested for a study, TREC requires that the study is extended to the full 6,000 samples stored to enhance the relevance and utility of data to the wider community. We want to maximise the use of what is a depletable resource

If you would like further information about samples or laboratory procedures, please contact the TREC via the Collaborations and Data Access Manager.

Linkage Data

Once your request for linkage data has been approved and an agreement with NHS Linkage implemented, the data will be placed de-anonymised and de-identified in a secure platform within the King's College London environment. Access outside this platform will not be provided.

It is our general policy that when DNA samples are requested for tests TREC requires that the test is extended to the full 6,000 samples stored in order to enhance the data to be of relevance to the wider community rather than a unique project and maximise the use of what is a depletable resource

All data are collected via standard SOPs which are adhered to by all staff involved in the data collection. This includes staff at the DTR as well as outside collaborators who may be involved in the data collection

for a particular project. Information on phenotype collection is available from our existing publications which are available from our website or additional information is available on request.

a. Intellectual property

Currently the IP rights for existing data from TwinsUK rest with KCL and Guys and St Thomas' NHS Foundation Trust.

Protection of Research Participants

To ensure confidentiality of our study participants is always protected, we will only supply the data with relation to the study number and birth month and year. If further information is required, this needs to be requested formally in the proposal form. All our twins are aware that the data collected from them may be shared with other research collaborators but the DTR has agreed to ensure that their personal data is always kept confidential.

b. Long-term preservation and sustainability

All data collected as part of a visit to DTR are identified by the date of visit, study type and the visit month and incorporated into our final databases only when checked and cleaned by our staff.

The DTR envisages that data will be made available to the research community at the earliest opportunity. Should there be any publications arising during the duration of the project, these will be made publicly available on the website as soon as they have appeared in press.

Agreements

Before the data /material is transferred, applicants will need to have agreed to the terms and conditions of transfer as outlined in the data access proposal form or a material/data transfer agreement when applicable.

Material and Data transfer agreements need to be signed by authorised legal signatories both at King's College London and the recipient's institution.

Costs

DTR receives funding from different charities and funding bodies via grants to support core activities.

This funding cannot be used for external individual projects and all applicants will be expected to cover the costs of implementation of their data/material access applications and projects. As part of a cost recovery exercise all projects will be asked to meet a mandatory project management fee.

Potential additional costs may apply depending on the nature of the project and the level of involvement required from the DTR team. These will be determined on a project-by-project basis and will reflect only the true costs to the DTR of providing the resources requested ("cost recovery"). Once we have approved a data request proposal, we will let you know the costs of data transfer and implementation. We will also agree distribution of indirect income at this stage.

On agreement of the costs, our finance department will expedite an invoice. For this applicants need to provide the below information

- Full name of your institution:
- Company registration number:
- Full address of your central finance dept:
- Telephone of the central finance dept:
- Email of the central finance dept
- URL web address:
- Purchase Order Number (PO)

Grants

If you are submitting a grant to cover the costs of an agreed project/data or material, we require that you send the final copy of the grant including the finances for approval at least six weeks before the submission date. Proposals received less than six weeks before the submission deadline will not usually be approved. We are happy to provide a letter of support for the proposed project described in your grant. On approval of the grant a formal data request needs to be submitted for consideration of the TREC. For proposals to collect new data we prefer (unless there is a good reason) that a member of the TREC be a co-applicant so they can act as guarantor for the proposed new data. You should send us a copy of the award letter on receipt, and we will then arrange a start-up meeting followed by annual review meetings to agree the objectives, timetable and staff required to meet the grant commitments.

If funding for the proposal is via a specific grant, the DTR will be set up as a sub-contractor to this grant.

Contact with Twins

We have a duty of care to the twins and it is an ethical requirement that the twins are not overburdened. Therefore, no applicants are allowed to contact twins directly. Only members of the DTR research team will be allowed to contact study participants.

Direct contact from the applicant will only be approved in exceptional circumstances by TREC.

Confidentiality

Protecting the confidentiality of the twins is a primary concern of the TREC and the DTR research team. Therefore, no contact or tracing attempts will be made to identify or contact the twins.

All data from the twins will be provided to third parties with no identifiable details.

Authorship and Publication

a. Any publications such as papers, abstracts, and posters, which have arisen from the project for which data has been granted will need to be sent to the TREC for approval, at least 30 days prior to the deadline of submission for papers - with a minimum of 7 working days for abstracts or posters.

We request that the below information is provided:

- Draft of the final version prior submission of the paper
- Project title and reference used to access the data used for the paper
- Journal the paper is being submitted to
- Deadline for submission

We expect to process all papers within two weeks of receipt.

We review all papers to:

- check confidentiality is protected
- ensure that the paper will not bring the Department into disrepute
- identify overlap with other papers published or in preparation.

We also provide advice and feedback to authors where we feel this may be helpful, but our role is not primarily to provide formal peer review.

b. The DTR has the right to withhold permission of publication of the above and will give reasons for doing so.

c. Authorship on papers should follow standard practice. All publications should reflect the involvement of the DTR and if there has been a designated DTR member of staff involved in the project, they should be included as a co-author.

- c. Specific acknowledgement within the text is required. We have agreed a standard acknowledgements section that should be included as is or in a modified form to fit the journal requirements for all papers:

TwinsUK is funded by the Wellcome Trust, Medical Research Council, Versus Arthritis, European Union Horizon 2020, Chronic Disease Research Foundation (CDRF), Zoe Ltd, the National Institute for Health and Care Research (NIHR) Clinical Research Network (CRN) and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London.

Applicants should send us copies of the final submitted draft and subsequent revised drafts as well as inform us when a paper is accepted. An electronic copy of the final published version will be requested for our records.

It is the author's responsibility to ensure papers are freely available for research funded by the Wellcome Trust and other funding bodies that require open access to publications arising from their funding.

A list of DTR publications to date can be found on the DTR website

PR Policy

All press releases on research arising from the study should be seen and approved by the TREC.

We expect the lead author on the paper to agree the press release with the KCL public relations team via the Collaborations and Data Access Manager, and to be available to deal with media enquiries and interviews.

We may also ask authors to prepare a précis of important papers to include in reports to funders and future applications for future core support.

Intellectual Property

Intellectual property rights belong to the DTR, King's College London. We will consider dividing intellectual property rights where applicants will be making a particular contribution. Any such division must be considered and agreed before the project starts via a collaboration agreement.

Feedback

This policy was last updated in September 2022.

We welcome feedback, comments, and suggestions. Please send your comments to victoria.vazquez@kcl.ac.uk.

Data Access Application Process Flowchart

