

WEST OF SCOTLAND TWENTY-07 STUDY: HEALTH IN THE COMMUNITY

Data Sharing Policy Version 1.2, 1st December 2015

We are committed to maximizing the use of Twenty-07 data to advance knowledge to improve human health and welcome proposals for collaborative projects from bona fide researchers. Our policy reflects the MRC's guidance on data sharing.

<http://www.mrc.ac.uk/research/research-policy-ethics/data-sharing/data-sharing-population-and-patient-studies/>

We have developed this data sharing policy with the aim of making data as widely and freely available as possible while safeguarding the privacy of participants, protecting confidential data, and maintaining the reputation of the study. We welcome comments on this policy.

This policy covers quantitative electronic data from the Twenty-07 Study's main interviews and postal questionnaires, including the physical measures and blood assays conducted. It also covers coded mortality information from death certificates. Twenty-07 respondents can be linked, if they consented, to the Scottish Morbidity Records (coded hospital records); a PAC application is required for external users to access coded hospital records.

This policy does not cover our tissue bank or subsample studies. Please contact us separately if you wish to discuss access to these data.

For further information contact: Anne Ellaway, Study Director, Anne.Ellaway@Glasgow.ac.uk or the Study's Data Scientist Mary-Kate Hannah Mary-Kate.Hannah@Glasgow.ac.uk. A message can be sent to both of us simultaneously via sphsu-twenty07@glasgow.ac.uk

BACKGROUND TO THE STUDY

The West of Scotland Twenty-07 Study: Health in the Community is funded by the Medical Research Council (MRC) and is conducted by researchers at the Council's Social & Public Health Sciences Unit in Glasgow. The study aims to investigate the relationship between social position and health in the Central Clydeside Conurbation. In addition to a 'regional' sample two localities, in Glasgow city with better than average and worse than average health profiles, were selected. The Twenty-07 study is longitudinal in design, involving three age cohorts (aged approximately 15, 35 and 55 at the baseline survey in 1987/88) followed for 20 years with interviews having taken place approximately every five years. The total sample size at baseline was 4,510. In the fifth wave (2007/08) respondents were asked for blood and DNA samples and for permission to access their Scottish hospital records. No further data collection is planned for the main study. However, in addition to the main study, a number of subsample projects have been undertaken, for example in relation to gender, religion and music, and some of these are ongoing.

Further information about the study can be found in an IJE Cohort Profile <http://ije.oxfordjournals.org/content/38/5/1215> or on the Study website. <http://2007study.sphsu.mrc.ac.uk/>

Full variable level information from the study are available on the website through our SWIFT data discovery and access system, which is constantly being improved and developed .

PRINCIPLES OF DATA SHARING

1. Applications from bona fide researchers to use existing Twenty-07 data are welcome. All potential collaborators work with a link person, an experienced Twenty-07 researcher – to support their application and access to the data. It is assumed such collaboration will lead to a joint publication. In line with MRC policy, all publications should be in Open Access journals. The term bona fide researcher is defined in the MRC data sharing policy.

<http://www.mrc.ac.uk/research/research-policy-ethics/data-sharing/data-sharing-population-and-patient-studies/>

2. All collaborators are required to abide by MRC's policies on managing personal data, not pass on the data to a third party or attempt to identify individuals. At the end of the project all Twenty-07 data must be returned to the Unit or destroyed, and any copies destroyed. The confidentiality and potential for statistical disclosure of all data requests are assessed. Where potentially sensitive data are necessary, external collaborators will be required to conduct the analyses at SPHSU. Collaborative projects analysing the existing Twenty-07 data do not require further ethical review; however, if this is a requirement of applicant's institution, then they are required to ensure such approval is gained.
3. The following criteria are employed to consider applications.
 - The research question(s) proposed should aim to develop understanding of how to improve human health and be of good scientific value, ethically and legally sound.
 - The analytical techniques proposed should be appropriate and robust and Twenty-07 data should be able to address the question(s).
 - The skills and experience of the applicant(s) should be appropriate to undertake the project.
 - There should be an intention to publish the research findings for wider scientific scrutiny and benefit in open access journals.
 - The Twenty-07 resources required to create the necessary dataset for the project should be reasonable.
 - If the research question overlaps with an existing (external or internal) project, the researchers concerned will be encouraged to collaborate rather than make a separate application.
4. No charge is made for providing data to external users. If a particularly time consuming request is received, for example, that required inputting and coding of significant amounts of data, then external funding would need to be found to support such work.

GOVERNANCE FOR DATA SHARING

5. Twenty-07 has an external steering committee, set up according to MRC guidelines, i.e. it has formal terms of references; more external than internal members and an independent external chair.

<http://2007study.sphsu.mrc.ac.uk/KeyPeople.html>

6. The Steering Committee has oversight of the data sharing process. Applications to use the data are assessed and approved by an internal data access group. <http://2007study.sphsu.mrc.ac.uk/KeyPeople.html> Once approved internally the application is sent to the Steering Committee Chair for final approval.
7. Applications which might be considered sensitive will be discussed by the full Study Steering Committee. If appropriate additional external advice may be sought. If applicants are unhappy with the decision on their proposal this will also be considered by the full Steering Committee. If applicants are still unhappy with the decision or process then the matter will be referred to the MRC Programme Manager within Head Office.
8. An annual report is made to the Steering Committee of the external enquiries received, applications made and their outcome, progress with internal and external projects, publications produced, etc.

PROCESS OF DATA SHARING

9. A contact email address (sphsu-twenty07@glasgow.ac.uk) is given on both the Unit and MRC gateway websites, while other publicity (e.g. the IJE cohort profile) contains the Study Director's contact details. We endeavour to reply to an initial enquiry within one week.
10. External users are required to complete an application form to access the data. (<http://2007study.sphsu.mrc.ac.uk/Data-Sharing-Form.html>) The researchers concerned and the Head of the Department that will hold the data are required to sign the form to say that they will abide by the terms and conditions of use as set out in the declaration form and associated policies.
11. Once an application has been received, it will be reviewed by the internal data access group, and the Study's Data Scientist will identify any data issues that may need to be considered (e.g. confidentiality, possibility of deductive disclosure). Data are not released externally if they allow the possibility of identification of individuals within the study, potentially sensitive data – e.g. occupations, illnesses, postcodes – are therefore only provided at a highly aggregated level. We aim to complete this review process in a month; any issues will be discussed with the applicants so the application can be revised if necessary.
12. The final decision to release the data is made by the Chair of the Study's Steering Committee.
13. Overall therefore the application process should be completed in six weeks.
14. Once an application has been approved, the application is signed by the Study and Unit Directors and an approval letter is sent to the applicants, which confirms the terms of the agreement.
15. A dataset is then created for the project. Currently we estimate this will take a week per wave of data collected, although more time may be required if complex derived variables are requested which have not previously been constructed. A set of random id numbers is allocated to the project and documented. This ensures that if external researchers have several ongoing projects they cannot combine datasets together in ways that could potentially lead to possible breaches of participant confidentiality.

16. As well as a dataset we also provide external users with notes re key issues with variables requested, e.g. coding information, issues re inconsistencies over time, medications that need to be considered for physical measures and biomarkers, information on inverse probability weights for drop out if appropriate.
17. All data transfers are encrypted.
18. External users with ongoing projects are required to provide an annual progress report once a year to the steering committee. When a project is completed, external users are required to return the dataset and any documentation on derived variables to the Study and destroy all local copies.

DATA DISCOVERY

19. Information about the Twenty-07 Study is publicly available in a number of places at two levels.
20. Descriptions of the overall study and its availability for data sharing as well as information on who to contact to find out more, are available in:
 - IJE cohort profile <http://ije.oxfordjournals.org/content/38/5/1215>
 - on the MRC Data Gateway <https://www.datagateway.mrc.ac.uk> ;
 - on Unit Study website <http://2007study.sphsu.mrc.ac.uk/> .
21. We are currently working towards generating metadata for the main Twenty-07 questionnaires to DDI 3 standards. Once variables directly extracted from the questionnaires are fully documented in SWIFT to the desired standards, we will begin work on documenting core derived variables.
22. The variable level metadata are available on the Unit's Study website in our data access system: SWIFT. Once the MRC Data Gateway, or equivalent, has established a system for importing metadata our variable level information will also be deposited on that website.

Study Website

23. The Study website has four main sections:
 - A general description of the study, written in lay terms, describing its purposes, data collections etc, as well as key findings.
 - Contact area for respondents to update their address details.
 - Study resource area – documentation for study – full copies of working papers describing sampling etc, questionnaires; a full list of all publications arising from the Study.
 - Information on data sharing and access to SWIFT.
24. The study website will be updated at least annually to ensure the most up-to-date data sharing information is provided there.

SWIFT

25. SWIFT is a web-based data discovery and access system. Originally written for the MRC 1946 birth cohort, we have adapted it to suit the structure of Twenty-07 and the access process we wish to adopt. First time users need to register to use SWIFT. Once provided with a login, users can search SWIFT for variable level information by topic, by wave and by cohort. Variable displays provide information on the question from which variable was derived, value codes and labels and links to identical/similar variables at other waves/cohorts.
26. SWIFT can then be used to create baskets of variables, for which SPSS syntax is generated to extract the variables from data files. No actual data are linked to via SWIFT. Only when the Study Director confirms that the

application and approval process have been completed is a dataset compiled and sent to the external user.

Anne Ellaway and Mary-Kate Hannah
Twenty-07 Research Project Director and Data Scientist, 1st December
2015