

Nuffield Department of Population Health
Cancer Epidemiology Unit

**Data Access Policy for the
Million Women Study**

University of Oxford

V2 November 2018

1. Overview

The Million Women Study is a prospective study of 1.3 million UK women. The aims of the Million Women Study are to study the relevance of lifestyle, environmental, and genetic factors for major chronic diseases (e.g. stroke, heart disease, cancer, dementia), to help improve risk prediction and prevention of these diseases.

Participants were recruited from 1996-2001, completing a postal recruitment questionnaire, and gave permission for follow-up. Electronic linkage, using each woman's unique National Health Service (NHS) number, to NHS Central Registers provides virtually complete continuing follow-up for deaths, cancer registrations and hospital admissions. Study participants are sent postal re-survey questionnaires every 3-5 years. For subsamples of study participants, online surveys have been done and blood samples collected.

The Million Women Study is run by the Cancer Epidemiology Unit in the Nuffield Department of Population Health (NDPH) in the University of Oxford. The study is currently funded largely by Cancer Research UK and by the Medical Research Council.

At recruitment, as was consistent with standard practice at the time, participants were not asked specifically for consent to data sharing with outside bodies. Availability of biological material is limited by the small volume of the samples collected and by the incomplete DNA extraction and genotyping that has been done to date.

Within the above constraints, the Million Women Study welcomes proposals for access to Million Women Study data for health-related research, either for collaborative projects or for other forms of data access to help achieve the study's aims. This document describes the Million Women Study access policy and procedures. It has been developed in concordance with the general principles of data sharing promoted by various research organizations in the UK, and elsewhere, and the Data Access and Sharing Policy of the Nuffield Department of Population Health, University of Oxford.

2. Terminology

Data	Any Million Women Study dataset, including summary datasets, recruitment survey data, re-survey data, linked follow-up information, blood assay results, genotyping data. Any additional datasets (eg meteorological, environmental) from project partners which Million Women Study has stored in its data repository.
Data Access Agreement	Agreement covering the terms of data access to a Requestor of Open Access Data.
Collaboration Agreement	Agreement covering the terms of data access to a Requestor working with a member of the Million Women Study team.
Open Access Data	Data being made available to external <i>bona fide</i> researchers through the Data Access policy.
Restricted Data	Data stored in the Million Women Study data repository which has limitations placed on its use or wider distribution.
Requestor	An individual or group of researchers seeking access to data and/or samples from the Million Women Study.
Data User	An individual or group of researchers that has been granted access to data and or samples from the Million Women Study.

2. Principles of data sharing

As the Million Women Study has information on many different exposures and many different health outcomes over a period of many years, the involvement of a wide range of investigators helps to maximize the value of study data. As data custodian, the Million Women Study research group must maintain the integrity of the database for future use and regulate data access. Data can be released outside the Million Women Study research group only with appropriate security safeguards (i.e. IG Toolkit, ISO 27001 certification or System Level Security Policy assessment) and approvals. The policy on data access is based on the need to:

- Protect participants and act within the scope of their signed consent;
- Ensure compliance with UK legal and regulatory requirements (eg, the UK Data Protection Act, 2018 and the EU General Data Protection Regulation, 2018);
- Ensure that the data security and participant confidentiality is maintained.

3.1. Key components of this data access policy:

- **Open Access Data Availability:** Before data is approved for any analysis, relevant members of the Million Women Study team responsible for generating the data must first undertake required cleaning, processing, quality control, integration and imputation. Where additional data is generated as a result of a specific research award or collaboration, sufficient exclusive access for the investigators and/or their collaborators may be reserved in order to comply with any pre-specified constraints. Once the datasets are made available for sharing, the presumption is that all reasonable requests for data from *bona fide* researchers will be granted. Details of the currently available data, and a timeline for future data releases, are given on the website. Approval(s) which may be required from other bodies (eg NHS Digital) for the sharing by the Million Women Study of linked health data are the responsibility of the requester. Applications should be made in a separate request to the data provider (e.g. NHS Digital), once approval for data access has been obtained from the Million Women Study.

- **Collaborations:** The Million Women Study research group will actively seek and respond to requests for scientific collaborations on specific projects. From time-to-time calls for specific project proposals or collaborations in areas of strategic importance and/or major scientific interest will be published. This model of facilitated collaboration with external researchers will be adopted where it can increase the value and quality of the data. Such collaboration will be governed by a separate Collaboration Agreement. Collaborative Agreements will: (i) identify a dedicated project lead from within the Million Women Study group; (ii) detail arrangements for co-authorship or papers; (iii) cover intellectual property issues; (iv) detail financial commitments where appropriate.
- **Independent Oversight of Access:** Initial decisions on data access are the responsibility of the study team, with advice from the Million Women Study Advisory Committee, which includes independent members (see Annex A for current membership). The Nuffield Department of Population Health Data Access Oversight Committee (formerly known as the Richard Doll Archive Oversight Committee) will provide further scrutiny and governance advice on data sharing for all studies in the Nuffield Department of Population Health. A Requestor can appeal to this committee if they disagree with a study decision on access.
- **Protecting the Identity of Participants:** Safeguards will be maintained to ensure the anonymity and confidentiality of participants' data. Researchers will enter a legal agreement not to make any attempt to identify participants, and the data provided to researchers will not contain any personally identifiable variables (i.e. every data set provided will be "pseudonymised" with uniquely encrypted participant identifiers [PIDs]).
- **Data Security:** All Million Women Study data are held on secure servers in a central data repository that is compliant with internationally recognised information governance standards. A data management team act as gatekeepers and ensure that any shared data is delivered through a secure data delivery system and that any usage of restricted data held in the repository is handled appropriately.
- **Sample Preservation and Access:** 10ml blood samples were collected for 53,000 participants and stored as aliquots of plasma and buffy coat. DNA has been extracted for some 20,000 samples, for which only limited genotyping has been done. Requests for Million Women Study samples will be considered in the context of the limited and depletable nature of this resource but because of their depletable nature will be subject to more stringent access criteria and strongly preferred would be applications for sample assay for the whole cohort using proven assay technology generating useable data. Investigators are currently seeking funding to transform the samples into accessible data.
- **Fees for data access:** Researchers will incur an Access Charge for each approved data request (currently £750 GBP). This is determined on a cost recovery basis and will contribute to the administrative costs incurred in managing and reviewing the application, and in preparing the individual datasets. Researchers may also be required, where appropriate, to cover any additional costs of administering the data sharing (including legal fees if applicable), retrieving, processing and sending the data or samples. Estimated costs for a particular request will be provided during the development of the project proposal.

4. Data access process

Potential Collaborators and data Requestors should first contact Million Women Study investigators or review the Million Women Study website to gain an understanding of the available study data and of projects that have previously been completed and are currently being undertaken.

4.1 Eligibility

Requestors for health-related research applications should be *bona fide* researchers, employed by a recognised academic institution or health service organisation, with experience in health-related research.

4.2 Collaboration requests

Approved researchers who are interested in collaborating on projects with Million Women Study researchers are encouraged to approach the Million Women Study group informally in the first instance by email to millionwomenstudy@ndph.ox.ac.uk or to contact relevant Million Women Study investigators to discuss research ideas and feasibility. Formal enquiries should include a project title and brief outline of the research project and the relevant data of interest. Each project requires a co-investigator from within the Million Women Study group who has a common interest in the project and relevant or complementary research expertise. Once identified, the collaborator and the co-investigator will co-develop a research proposal which will then be reviewed by the Million Women Study Advisory Committee.

4.3 Open Access Data requests:

Submission of a Data Request

Data requestors should complete the Million Women Study Data Request Form. Required information includes: project title and abstract; scientific rationale / methodology; anticipated outputs and project timeline. Additional questions cover ethical issues; collaborators / research team; funding support and data security. Applications should: (i) have clearly defined objectives; include a viable methodology that is likely to generate meaningful results; (ii) be based on an appropriate and available selection of data; (iii) have clearly defined timelines and outputs (e.g. 1-2 papers in peer-reviewed journals).

Review of a data request

Open Access Data requests will be regularly reviewed by the Million Women Study team. The Million Women Study Advisory Committee will review any requests for access that raise particular issues (such as those relating to the use of samples or with complex ethical considerations).

To avoid duplication of effort, where there is substantial overlap between separate proposals submitted at the same time we may suggest that researchers collaborate on a project (after seeking appropriate permissions). The Million Women Study will not insist on collaboration; if proposals meet the criteria for approval the same data may be shared with different institutions at the same time.

The Million Women Study Team will aim to review and respond to Data Requests within 4-6 weeks. A Requestor can appeal to the Nuffield Department of Population Health Data Access Oversight Committee if their request is denied and they disagree with a decision.

5. Terms of data access

Once proposals are approved the following conditions and undertakings are required as conditions of access:

- **Data Access Agreement / Collaboration Agreement.** Before any data are transferred a signed transfer agreement must be in place between the Requestor's institution and the University of Oxford and if appropriate any third party data provider (eg. NHS Digital). A Template Data Access Agreement for the University of Oxford is included on the Million Women Study Website. The agreement will include a copy of the approved project proposal as a Schedule.
- **Signing Authority.** Requestors should be acting as members of a recognized academic institution, research organisation or health organisation. Their request should come from a recognized email domain (eg, .ac.uk, .edu.cn). Their organisation should have formal policies and procedures (i.e. IG Toolkit, ISO 27001 certification or System Level Security Policy assessment) to comply with any legal, ethical or data protection constraints and to ensure that the dataset is stored securely and used responsibly.

- **Ethics and Research Governance Approval.** Where applicable Ethics Committee approval for the research is the responsibility of the Requestor. The Requestor, in conjunction with study investigators, may also need to obtain approval from the Research Ethics Committees responsible for the Million Women Study. Research Governance and R&D approvals, if required, are the responsibility of the Requestor. At present, Requestors requiring individual-level linked outcomes data from NHS providers may require separate approval from the data provider: such applications should be made by the Requestor in conjunction with the Million Women Study team and once approval for the data access request has been obtained. All Approvals will need to be in place before any data are transferred.
- **Limitations on Use.** The data will be used for the purposes of health-related research only and within the constraints of the consent under which the data were originally gathered, and of any contractual agreements between the Million Women Study and its funders or external data sources. Data supplied may only be transferred to Requestors named at the time of the original application or in subsequent applications and specified in the Access Agreement or later amendments. Data cannot be transferred to individuals outside the Requestor's research group without formal approval by the Million Women Study Advisory Committee, and cannot be used for wholly commercial purposes.
- **Identifying Data.** The data provided to researchers will not contain any personally identifiable variables. Data sets will be "pseudonymised" with uniquely encrypted participant identifiers (PIDs). The Access Agreement will contain confidentiality undertakings to further safeguard participants' privacy. Recipients must agree not to link the pseudonymised data provided with any other data set without permission. Recipients must not attempt to identify any individual from the data provided. Should recipients believe that they have inadvertently identified any individual, they must report the incident to the data originators and must not record the identity, share the identification with any other person or attempt to contact the individual.
- **Intellectual Property.** All Intellectual Property Rights in the Data are and shall remain at all times the property of the University of Oxford. All Arising Intellectual Property shall vest in and be owned by the Requestors, and the Requestors shall be encouraged to publish their findings and provide their results (which justify the findings) back to the University of Oxford, along with a suitable license, whereby the University will be granted rights to use all such results for academic and research purposes, including research involving projects funded by third parties.
- **Payment of Access Charges.** Data Requestors are expected to pay Access Charges to contribute to the administrative cost to the study of reviewing the application and preparing data for sharing, etc. Where these are applied, no data will be provided to the Data Requestor until or unless the Access Charges are received in full.
- **Data Release and Delivery.** Once the proposal is approved and the Access Agreement signed, the data and its documentation will be generated in CSV (or any other pre-specified) format, encrypted and released in a secure manner via the Data Access System or by using encrypted physical media.
- **Publicity and Dissemination.** The Million Women Study team reserves the right to publish the title, the names(s) and affiliations(s) of the Chief Investigator(s), a lay summary and a scientific abstract of each piece of collaborative research for which access to the resource has been granted, before identification or publication of results. Requestors who do not wish details of their study to be openly available need to state this in their data request and give the reason. The Requestor shall not use the name or any trademark or logo of the University in any press release or product advertising, or for any other commercial purpose, without prior written consent.
- **Authorship and Approvals.** Collaboration Agreements and Access Agreements will specify expectations regarding authorship and acknowledgements on research outputs. Collaborations require at least one co-author from the Million Women Study

group. For Open Access Agreements no authorship from the Million Women Study team is required. The Million Women Study should be acknowledged in accordance with the Access Agreement. Requestors are asked to send proposed publications to the Million Women Study team not less than 30 days in advance of submission for publication but approval from the study team is not required prior to submission for publication.

- **Publications and Open Access.** All publications of the Results in a peer-reviewed journal, or as a scholarly monograph or book chapter, must be made available from PubMed Central and Europe PubMed Central as soon as possible and no later than six months from the date of final publication. All journal requirements for data release and deposition that are attached to publication should be complied with in full.
- **Integration of the Data.** After completion of work using released Million Women Study data, the original dataset as well as any derived dataset and/or variables generated during the research must be returned to the Million Women Study central data repository for archiving and/or merging with the main database for future use. If considered appropriate, the Million Women Study staff may carry out independent checks and/or validation of the data and results to ensure the continued data integrity and reliability of the study findings.
- **Monitoring and Accountability.** The Data User shall be required to submit annual reports and any other information reasonably requested to evidence the work undertaken by the Data User in connection with the proposed project. If there is substantial delay or difficulty in completing the planned research, the Million Women Study team will have the right, after consultation with the Advisory and Oversight Committees, to terminate the work if in its view there is little chance that the problem will be rectified. If there is substantial deviation or change in the planned use of the data, further approval will be needed.

Annex A: Membership of the Million Women Study Advisory Committee

Chair	Professor Carol Dezateux, University College London
CEU members	Professor Dame Valerie Beral, Million Women Study Professor Jane Green, Million Women Study Professor Julietta Patnick, University of Oxford
External members	Professor Emily Banks, Australian National University; 45 & Up Study Dr Lucy Carpenter, Lay Member Professor Sir Richard Peto, University of Oxford Professor Cathie Sudlow, University of Edinburgh; UK Biobank
Secretariat	PA to Professor Valerie Beral