

Nuffield Department of Population Health
Cancer Epidemiology Unit
University of Oxford

Data Access Policy for the Million Women Study

VERSION S3.2 DECEMBER 2022

1. Overview

The Million Women Study (HRA Ref: 171497) is a prospective study of 1.3 million UK women. The main aim of the Million Women Study is to provide reliable evidence on the relevance of lifestyle, environmental, and genetic factors for major chronic diseases (e.g. stroke, heart disease, cancer, dementia), to help improve prevention of these diseases.

Participants were recruited from 1996-2001, completing a postal recruitment questionnaire, and gave permission for follow-up. Electronic linkage to NHS Databases using each woman's unique National Health Service (NHS) number provides virtually complete continuing follow-up for deaths, cancer registrations and other health outcomes such as hospital admissions. Study participants are sent postal re-survey questionnaires every 3-5 years. For subsamples of study participants, specific postal and online surveys have been done and blood samples collected. This Policy applies to all data held by the study investigators on MWS participants, through the main Million Women Study (MWS), the MWS: Disease Susceptibility in Women (HRA Ref: 300001, blood collection).

The Million Women Study is run by the Cancer Epidemiology Unit based in the Nuffield Department of Population Health (NDPH) in the University of Oxford. Since its inception, the study has mainly been funded by Cancer Research UK and 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.

Within the above constraints, the Million Women Study welcomes proposals for access to Million Women Study data for health-related research, for collaborative projects, or for other forms of data access to help achieve the study's aims. This document has been developed in concordance with the general principles of data sharing promoted by various research organisations in the UK, and elsewhere (e.g. [making your research data open – UKRI](#)), and the Data Access and Sharing Policy of the Nuffield Department of Population Health, University of Oxford (<https://www.ndph.ox.ac.uk/data-access>).

Data use within NDPH by study investigators (alone or with NDPH colleagues) as set out in the study Protocol is outside the scope of this policy.

2. Terminology

| | |
|-------------------------|---|
| Data | Any Million Women Study dataset, including summary datasets, recruitment and re-survey data, linked follow-up information, blood samples and assay results. |
| Data Use Agreement | Agreement covering the terms of data access to a Requestor of Open Access Data. |
| Collaboration Agreement | Agreement covering the terms of a collaboration for a Requestor working with a member of the Million Women Study team. |
| Open Access Data | Data being made available to external bona fide researchers through the Data Access policy. |
| Restricted Data | Data stored in the Million Women Study data repository which have limitations placed on use or wider distribution. |
| Requestor | An individual or group of researchers seeking access to data from the Million Women Study. |
| Data User | An individual or group of researchers that has been granted access to data from the Million Women Study. |

3. Principles of data sharing

As the Million Women Study has information on many different exposures and 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 (NHS Data Security and Protection NHS IG Toolkit, ISO 27001 certification or System Level Security Policy which has been approved by the Requestor's organisation) and Million Women Study Data Access Applications Review Panel approval. The policy on data access is based on the need to:

- Maximise the value of study data for health-related research
- Protect participants and act within the scope of their signed consent
- Ensure compliance with UK legal and regulatory requirements (e.g., the UK Data Protection Act, 2018 and the EU General Data Protection Regulation, 2018)
- Ensure that data security and participant confidentiality are maintained
- Provide academic return and training for the investigators developing the study; in particular for doctoral students and early career researchers who are developing their scientific skills while working on the cohort.

Key components of this data access policy:

Open Access Data Availability: Before data are 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 are 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 data are made available for sharing, the presumption is that all reasonable requests for data from bona fide researchers will be granted, as long as the proposed project does not overlap significantly with projects currently being conducted by the study investigators. Details of the currently available data, and a timeline for future data releases, are given on the [study website](#).

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 researchers outside the study team will be adopted where it can increase the value and quality of the data.

Collaboration with researchers from outside NDPH is subject to the data access procedures outlined in this document. Such collaboration will be governed by a Collaboration Agreement. Collaboration 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. External researchers visiting NDPH are required to hold an academic visitor agreement between the University of Oxford and the visitor's institution.

Independent Oversight of Access: Initial decisions on data access are the responsibility of the Million Women Study Data Access Applications Review Panel (which includes the study PIs and other members of the MWS research team). Further advice will be sought (if required) from the Million Women Study Advisory Committee, which includes independent members. The NDPH Data Access External Oversight Committee provides further scrutiny and advice on data sharing for all studies in NDPH. 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 confidentiality of participants' data. Researcher's institutions will enter a legal agreement (Data Use Agreement) not to make any attempt to identify participants, and the data provided to researchers will not contain any personally identifiable variables (every data set provided will be "pseudonymised" with uniquely encrypted participant identifiers [PIDs], and only the study investigators will hold the 'key' allowing the PIDs to be linked to identity).

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 around 52,000 participants and stored as aliquots of plasma and buffy coat. To date, only limited genotyping information is available on these samples. It is generally expected that requests for direct access to study samples will not be approved, due to the limited and depletable nature of this resource.

Fees for data access: Researchers will incur a basic access charge for each approved Open Access Data request (currently £2500 GBP including VAT). 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 data Requestors and Collaborators may wish to review the Million Women Study website to gain an understanding of the available study data and of the types of research projects that have, or are currently being, undertaken.

Eligibility: We invite Open Access Data requests for health-related research from bona fide researchers, employed by a recognised academic institution or health service organisation, with experience in health-related research. They should be able to clearly demonstrate, through their peer-reviewed publications in the area of interest, their ability to conduct independent research. We do not accept Open Access Data requests from commercial organisations. Potential collaborators will generally be subject to the same eligibility criteria as Open Access Data Requestors. However, we will consider collaborations with commercial organisations involved in health-related research, where they will contribute valuable specialist expertise, and the proposed research is likely to have clear benefits to health and/or social care.

Open Access Data requests: Data Requestors will first need to complete the Million Women Study Data Access Application Form and return it to mws.access@ndph.ox.ac.uk. This form requires the Requestor to provide: a project title and abstract; scientific rationale, methodology; anticipated outputs and project timeline; and to specify each individual variable they would like to request. Additional questions cover ethical issues, collaborators/research team, funding support and data security. Applications which do not contain all of the requested application material will be returned to the Requestor for amendment.

Review of an Open Access Data request: Open Access Data requests will initially be assessed by the Million Women Study Data Access Applications Review Panel. The Million Women Study Advisory Committee will review requests for access that raise particular issues (such as those relating to the use of samples or with complex ethical considerations). Approved projects will: (i) have clearly defined objectives; (ii) include a sound methodology that is likely to generate meaningful results; (iii) be based on an appropriate and available

selection of data; (iv) have clearly defined timelines and outputs (e.g. 1-2 papers in peer-reviewed journals); (v) be able to demonstrate a clear potential benefit to health and social care. Projects that overlap significantly with projects already being conducted by the study investigators may be rejected.

The Million Women Study Data Access Applications Review Panel will aim to review and respond to Data Requests at a review panel meeting held approximately every 6 weeks. Applications received less than 5 working days before the next meeting will not be considered until the following meeting. A Requestor can appeal to the NDPH Data Access External Oversight Committee if their request is denied.

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.

Collaboration requests: Potential collaborators should contact the Million Women Study investigators at mws.access@ndph.ox.ac.uk for a preliminary discussion of feasibility of data access and collaboration. Projects will require 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. Any data sharing required by the collaboration will be subject to the same terms of use as those obtained under Open Access Data Requests.

5. Terms of data access

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

Data Use Agreement: Before any data are transferred a signed agreement must be in place between the Requestor's institution and the University of Oxford.

Collaboration Agreement: Where data sharing is part of the collaboration an agreement covering the terms of the collaboration and Data Use Agreement will be required

Signing Authority: Requestors should be acting as members of a recognised academic institution, research organisation or health organisation. Their request should come from a recognised email domain (e.g., .ac.uk, .edu.cn) to comply with any legal, ethical or data protection constraints and to ensure that the dataset is stored securely and used responsibly.

Data security requirements: In order to share the Million Women Study data, we require one of the following formal data security policies and procedures ([Data Access Request Service \(DARS\): guidance notes on security - NHS Digital](#)):

- NHS Digital Data Security and Protection Toolkit Registration
- ISO 27001 certification
- A documented institutional System Level Security Policy (SLSP) detailing all data security procedures within the Requestor's organisation. This must be a recently reviewed policy and be signed off by the Requestor's organisation. We expect the SLSP to adhere to the guidance provided by our data providers.

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. 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 (University of Oxford) and its funders or external data sources. Data supplied may only be transferred to Requestors named at the time of the original application, and specified in the Data Use Agreement or later amendments. Data cannot be transferred to individuals outside the Requestor's research group without formal approval by the Data Access Applications Review Panel, and cannot be used for wholly commercial purposes.

Identifying Data: The data provided to researchers will not contain any personally identifiable variables. Datasets will be pseudonymised with dataset-specific uniquely encrypted participant identifiers (PIDs). The Data Use 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. 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 (the Million Women Study team) 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 Data Use 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, for example via an encrypted, secure file transfer system.

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 research for which access to the resource has been granted, before identification or publication of results. 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 Data Use Agreements will specify expectations regarding authorship and acknowledgements on research outputs. Research outputs generated through a Collaboration require at least one co-author from the Million Women Study group but for those generated as a result of an Open Access Data request, no authorship from the Million Women Study team is required. The Million Women Study should be acknowledged in accordance with the Data Use 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. Requestors must notify the Million Women Study team of any publications that have arisen from the Data Use Agreement.

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

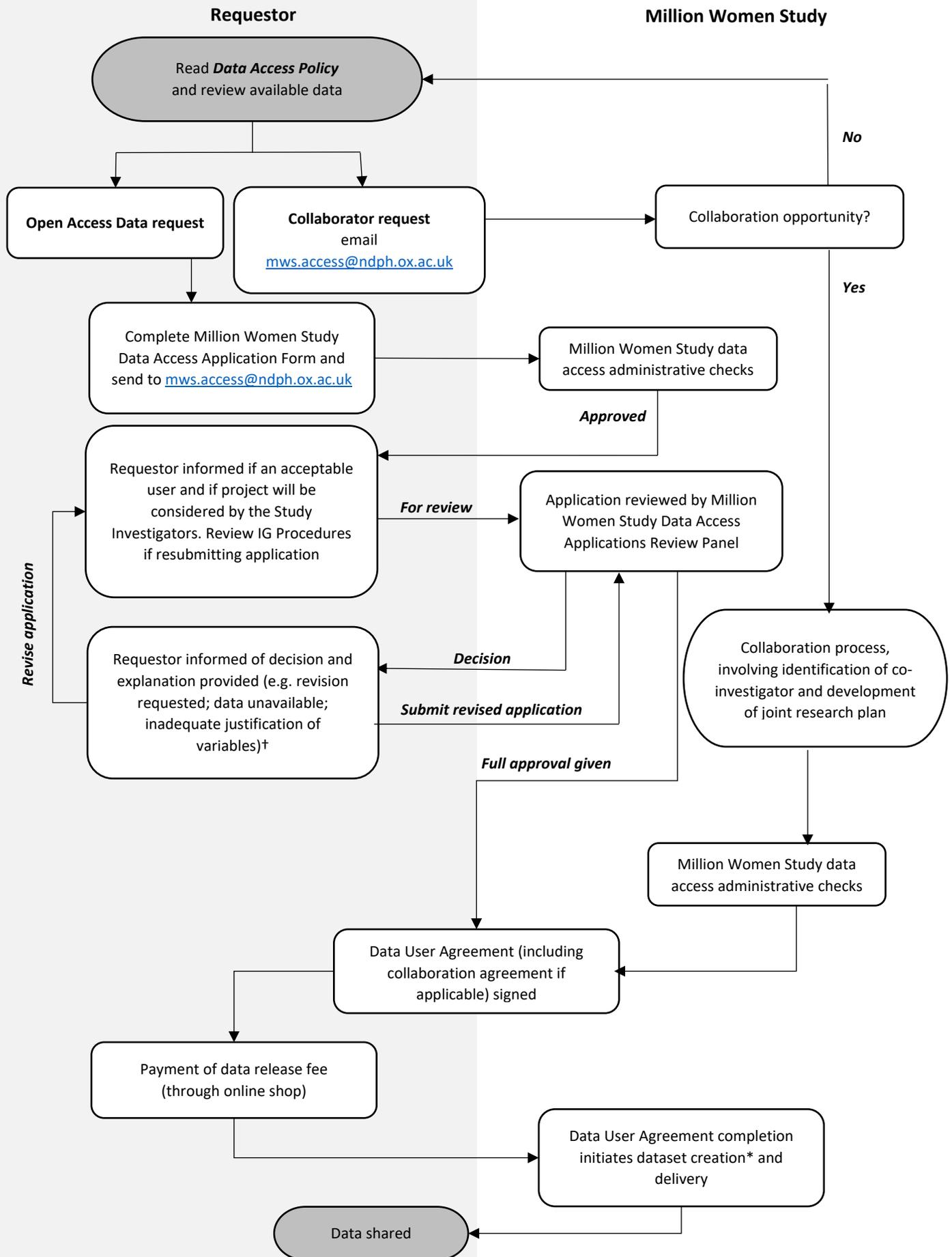
and Europe PubMed Central as soon as possible and no later than six months from the date of final publication. All journal requirements for data release and deposition that are attached to publication should be complied with in full.

Integration of the Data. After completion of work using released Million Women Study data, the original variables and 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 the project is delayed beyond the agreed end date, the Data User shall apply to The Million Women Study for a period of extension for use of the data. 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. Under such circumstances, all Data that have been provided must be deleted and a deletion certificate provided.

Data Destruction. When the project has been completed all Data that have been provided must be deleted by the data user and a deletion certificate provided to the Million Women Study Investigators.

Appendix 1: Flow diagram of Million Women Study data request process for sublicensed data



† The Requestor may appeal the decision as described in the NDPH Data Access Policy.

* All data sets pseudonymised with unique participant identifiers, download uses AES-256s secure encryption.

| | | | |
|------------|----------------|--|------------|
| 1.0 | June 2015 | Aligned with RDCA | Jane Green |
| 1.1 | September 2015 | Updated in line with MRC policy | JG |
| 1.2 | September 2016 | Updated in line with NDPH policy | JG |
| 1.3 | May 2017 | Minor revisions | JG |
| 1.4 | August 2017 | Revised advisory committee list | JG |
| 1.5 | May 2018 | Updated in line with NDPH policy | JG |
| 2.0 | November 2018 | Revised for NDPH | GR, JG |
| 2.1 | July 2020 | Minor revisions | JG, GR |
| 3.0 | May 2022 | Revised for Sublicence applications | GR, SF |
| 3.1 | September 2022 | Revised for Sublicence application queries | GR, SF, KB |
| 3.2 | December 2022 | Revised for Sublicence application queries | GR, SF, KB |