Whitehall II Policy on Data Sharing

Information for researchers interested in analysing existing Whitehall II data

Whitehall II Study
Department of Epidemiology & Public Health
University College London
1-19 Torrington Place
London WC1E 6BT
email: a.sanchez@ ucl.ac.uk

August 2017

Contents

1. The Whitehall II Study

1.1 Background

The Whitehall II (WHII) study was established to explore the relationship between socio-economic status, stress and cardiovascular disease. A cohort of 10,308 participants aged 35-55, of whom 3,413 were women and 6,895 men, was recruited from the British Civil Service in 1985.

Since this first wave of data collection, self-completion questionnaires and clinical data have been collected from the cohort every two to five years with a high level of participation. As a result, the WHII study is currently very well placed to answer questions about how previous and current circumstances affect health and quality of life in an ageing cohort.

The WHII study continues to collect data in order to examine the interrelationships between socioeconomic biological, psychosocial, and behavioural factors in the ageing process, and identify key determinants of late life depression, cognitive decline, chronic disease, and physical functioning. Data collection is intended to continue until 2030.

Summary of the data collection phases of the WHII study

Phase	Dates	Type	Number of participants
1	1985-1988	Screening / questionnaire	10,308
2	1989-1990	Questionnaire	8,132
3	1991-1994	Screening / questionnaire	8,815
4	1995-1996	Questionnaire	8,628
5	1997-1999	Screening / questionnaire	7,870
6	2001	Questionnaire	7,355
7	2002-2004	Screening / questionnaire	6,967
8	2006	Questionnaire	7,173
9	2007-2009	Screening / questionnaire	6,761
10	Feb-Mar 2011	Screening/Questionnaire	277 (Pilot*)
11	2012-2013	Screening / questionnaire	6,318
12	2015-2016	Screening / questionnaire	5632

* A small group WHII participants were selected randomly and invited to attend our Phase 10 clinic. This enabled us to successfully pilot new measures introduced at Phase 11.

Please refer to the Whitehall II website for more information (www.ucl.ac.uk/whitehallII).

The study is approved by the Joint UCL/UCLH Committees on the Ethics of Human Research (Committee Alpha). The WHII data management, data sharing and data preservation protocols are in line with MRC guidelines.

1.2 Funding

Since it began, Whitehall II has been supported by over 60 grants from different funding bodies including the Medical Research Council (MRC, MR/K013351/1; G0902037); Economic and Social Research Council (ESRC, ES/J023299/1); British Heart Foundation (BHF, RG/07/008/23674; RG/13/2/30098; PG/11/63/29011); Health and Safety Executive; Department of Health; National Heart Lung and Blood Institute (NHLBI; R01HL036310), US, NIH: National Institute on Aging (NIA; R01AG013196; R01AG034454), US, NIH; Agency for Health Care Policy Research (HS06516); and the John D and Catherine T MacArthur Foundation Research Networks on Successful Midlife Development and Socio-economic Status and Health.

Currently, the WHII study is supported by grants from the Medical Research Council (MR/K013351/1; G0902037) and the British Heart Foundation (RG/13/2/30098).

1.3 Access to existing WHII data

The WHII research data are available as a resource for the scientific community to maximize the value of the data for research and eventual patient and public benefit. WHII data and samples cannot be used for commercial purposes.

Our data sharing model follows a gated-access approach, which is intermediate between the free download of all data ever collected in the study (open access) and a strict model where only a small number of close collaborators are allowed to use the data (restricted access).

Data sharing at WHII involves an application process for *bona fide* researchers with an established scientific record. This approach ensures the reputations of the funding bodies, the WHII team and its participants are not compromised through unethical, premature or opportunistic data analysis. We follow the MRC definition of *bona fide* research:

www.mrc.ac.uk/news-events/publications/mrc-policy-and-guidance-on-sharing-of-research-data-from-population-and-patient-studies/ (page 24)

In line with this definition, we will accept applications from *bona fide* researchers who:

i) conduct *bona fide* research. This involves high quality, ethical projects for research purposes using rigorous scientific methods. There must be an intention to publish the research findings for wider scientific and eventual public benefit, without restrictions and with minimal delay.

ii) have a formal relationship with a *bona fide* research organisation, which is an established academic institution, research body or organisation with the capability to lead or participate in high quality, ethical research. It is not a requirement that research is the primary business of that organisation, or that the organisation is publicly financed. In this context, a public-private partnership may qualify as a *bona fide* research organisation.

An independent advisor, currently Professor Tom Meade from the London School of Hygiene and Tropical Medicine, oversees the WHII data sharing process on a regular basis.

This policy document explains the application process for researchers interested in analysing existing WHII data and the terms of conditions that must be agreed before data can be used.

1.4 WHII data available for sharing

The data collected at one wave will be available to the wider community as soon as the data collected at the next wave have been collected, curated and ready for internal analysis. For instance, the data from phase 11 were available for data sharing in 2017, which is when the data from phase 12 were curated and ready for internal analysis.

Currently WHII data from phases 1 to 11 are available for data sharing. These are summarised in the tables below. For more detailed information, please refer to the data sharing section of the WHII website (http://www.ucl.ac.uk/whitehallII/data-sharing), which contains:

- A data dictionary with variables from data collection phases 1 to 11 (Excel)

- Copies of the WHII Health Survey Questionnaires (PDF)

On request we make available documentation not available online, such as copies of other WHII questionnaires (FFQ, MMSE, GKQ, cortisol diaries, etc), as well as other study documentation such as clinical protocols or syntax of derived variables.

From the 1st September 2017 we charge a modest fee to cover the administrative costs associated with the management and generation of the bespoke datasets.

The WHII data management and data preservation protocols are in line with MRC recommended guidelines (www.mrc.ac.uk/research/research-policy-ethics/data-sharing/policy/).

1.4.1 Data from self-report questionnaires

These data are available from phases 1 to 11:

Family history

Family: health characteristics

Family: socioeconomic characteristics

Health

ADL

Birth weight

CHD

Diabetes

General health

Hospital admissions

Longstanding illness

Medication

Stroke

Women's health

Health behaviours

Diet (FFQ, TFEQ)

Drinking

Physical activity

Smoking

Personality, Mental Health & Well Being

      CASP

      CESD

      GHQ

      Life events

      Psychological states

      Satisfaction

      SF-36

      Traits

Psychosocial / Socioeconomic

      Education

      Employment status (grade)

      Retirement

      Sickness absence

      Social activities

      Social support

      Socio-economic characteristics

      Work

1.4.2 Clinical examinations data

These data are available from collection phases 1, 3, 5, 7, 9 and 11:

Anthropometry

Blood analysis

      Biomarkers

      Blood cells count

      Carbohydrate metabolism

      Lipids

      Vitamins

Salivary and hair Cortisol

Cardiovascular measures

      Blood pressure

      Carbohydrate metabolism, lipids

      Electrocardiograms

      Heart Rate Variability

      Pulse Wave Velocity

Cognitive function

      AH4, Mill Hill, memory, verbal fluency

      Mini Mental State Examination (MMSE)

Mental Health (CIS-R)

Physical functioning

    Balance and Chair Rises

    Walking speed

    Finger Tapping Test, Grip Strength

    Lung function

1.4.3 Validated clinical events

Self-reported clinical events have been validated using GP and hospital information. Prevalence, incidence and person-years are available up to phase 11:

Cancer

CHD

    Non-fatal MI

    Fatal CHD

    CHD

    CHD (excluding Rose Angina)

    Angina

Stroke

    All Stroke

    Haemorrhagic Stroke

    Ischaemic Stroke

    TIA

1.4.5 Environmental data

The Whitehall II air pollution concentrations were generated by the Environmental Research Group, King's College London. Additional markers of the physical environment were generated by Triantafyllos Pliakas, London School of Hygiene and Tropical Medicine:

    Aesthetics

    Air Pollution

    Crime

    Deprivation

    Distance to services

    Land Use

    ONS Urban-Rural

    Population

    Public Transport

    Road Network

    Road Traffic Accidents

1.4.6 Genetic data

We have a separate data sharing policy for projects involving genetics. Please contact Prof Meena Kumari (*mkumari@essex.ac.uk*) for more information.

2. Application process

This section describes the process that researchers interested in analysing existing WHII data must follow in order to have access to the data.

2.1 Informal contact

It is recommended that the prospective data user first contacts the WHII Data Manager, Aida Sanchez (*a.sanchez@ucl.ac.uk*), or a WHII researcher to informally discuss the feasibility of the proposed project.

For genetics projects, please contact Dr Meena Kumari (*mkumari@essex.ac.uk*).

2.2 Project Proposal

Researchers should apply for a project ideally aimed at the publication of one or two scientific articles within the first two years. The application requires the submission of two documents for initial approval by the WHII research team:

- WHII Data Sharing Application Form

- List of WHII variables names needed for the project

Application form

The WHII Data Sharing Application Form can be found at the end of this document. The applicant should provide the following information:

- Name, affiliation and main publications of the <u>principal investigator</u>, who should be the main applicant.

- Preferred WHII Contact Researcher, if any.

- List of all researchers involved in the project and their affiliation.

- List of outputs from previous projects where WHII data were used by any of the applicants.

- Details of funding if WHII data are included in the grant application.

- Scientific hypothesis and the objectives of the proposed study, which should be <u>as focused and specific</u> as possible.

- Planned scientific outputs.

List of exact variable names

The variables requested must be consistent with their project proposal. The applicant should download the Excel data dictionary from the WHII website (www.ucl.ac.uk/whitehallII/data-sharing) and highlight the variable names needed for the project.

PhD projects

For PhD projects that do not involve a WHII researcher, the main applicant should be the Principal Investigator or the PhD supervisor. These applications will be considered as a regular data sharing project. Therefore, the same level of support will be offered to these projects (i.e. response to the initial queries on the supplied dataset), and no additional supervision should be expected from any WHII researcher.

Funding

The applicant should specify whether they are seeking funding to carry out the project and if the grant application is partially or totally based in the use of WHII data.

If the WHII data user wishes to apply for funding once the application has been approved, and the grant application references WHII data, the WHII Contact Researcher must be informed about the funding application.

Overall accountability to funding bodies for research projects remains the responsibility of the grant-holders.

2.3 Evaluation of the proposal

The application form and data dictionary should be sent to the data manager. These documents will be evaluated by WHII research team, which meets monthly.

The evaluation of the proposal will be made at the first meeting following receipt of the application form.

Applicants should allow time for the application review process. Applications received less than 5 working days before the meeting will not be considered until the meeting in the following month.

A member of the WHII research team will be selected as the WHII Contact Researcher for each project. If possible, s/he will be the preferred WHII Contact Researcher named on the application form.

The WHII Contact Researcher will inform the applicant of the decision in writing.

Approval from the committee is "in principle", subject to the applicant signing and returning the **Whitehall II Data User's Agreement Form** (see final page of this document) and the payment of the cost associated with the data release (*Whitehall II Data Sharing Fee*).

The WHII Contact Researcher will be the point of contact for the duration of the project.

Further use of the data *outside those specifically stated in the application form in other areas of interest* will require a new application being submitted and approval by the WHII research team.

Payment

From the 1st September 2017 the Whitehall II study charges a fee of £500 + VAT for each data sharing project. This will cover the administrative costs associated with the management and generation of the bespoke dataset.

This fee will need to be paid upon approval of the application. To proceed with the payment, please click on *Whitehall II Data Sharing Fee*. You will need to have your data sharing reference number in hand.

Additional requests might incur further costs depending on the required data management time.

2.4 Data release

The WHII Data Manager will aim to provide the dataset and associated documentation within two weeks of receipt of:

- the Whitehall II Data User's Agreement signed by the main applicant

- the detailed list of variables requested

- the Data Protection registration, if requested (see section 2.4.2).

- payment of the cost associated with the data release

The dataset can be supplied in a variety of formats, such as SAS, SPSS, Stata, etc*.*

We recognise that further variable requests might be necessary once the preliminary analysis of data has started. *For these extra requests or any other queries regarding the data, please contact the WHII Contact Researcher.* Further costs depending on the required extra data management time.

2.4.1 Data anonymisation

The WHII information has been obtained in confidence from study participants. Preserving the confidentiality of the data is of paramount importance.

The WHII database does not contain any individual identifiers such as names, addresses, telephone numbers, NHS numbers and the like, conforming to the UK Data Protection Act. In addition, extra restrictions have been applied to the data released to external WHII data users. We use two basic tools to achieve this:

1. Participants are identified using an anonymised ID that will be *different for each WHII data sharing project*. The purpose is to avoid linkages that could potentially result in the identification of individuals.

2. All datasets will be stripped off specific variables that can create a risk of participant identification (such as complete dates of birth and death) or that have low prevalence rates (such as medical events dates, exact diagnosis codes, job title, Civil Service department, etc), which could potentially enable the identification of subjects with unusual characteristics.

However, given the nature of the data, it is virtually impossible to prevent the identification of specific individuals were one so minded. Therefore, it will be the responsibility of data users to ensure that no participant's identity is disclosed under any circumstances (see section 2.4.2).

2.4.2 Ethics and confidentiality

The approved projects will not require individual ethical approval. Ethical approval from the data user's local ethical committee is the responsibility of the data user.

WHII data can be provided without restriction within the EEA, which comprises all countries in the European Union, Iceland, Liechtenstein, Norway and the States of Guernsey.

WHII data can also be transferred to a country or territory outside the EEA if the applicant and their collaborators provide an adequate level of protection of personal data and operate under the Data Protection Scheme in place in their country. If requested by the WHII team, a copy of this Data Protection registration should be sent to the WHII Contact Researcher.

As mentioned above, it will be the responsibility of WHII data users to ensure that no participant's identity is disclosed under any circumstances. WHII data users must also preserve the confidentiality of the data in outputs and publications (e.g., tables with small numbers might lead to the identification of individuals).

2.5 Re-submission

Should an application be refused by the WHII research team, the applicant is welcome to make the appropriate changes and submit a new application. Reasons to request re-submission of a data sharing application include:

- The applicant requests confidential data or proposes to link data in a way that confidentiality would be compromised.

- Lack of outputs from previous projects based on WHII data

- Concerns that the project is unclear or vague.

- Concerns that the applicants might not be *bona fide* researchers with an established scientific record (please refer to section 1.3)

- Concerns that the project could damage the reputation of the WHII participants, UCL and/or funding bodies through unethical, premature or opportunistic data analysis.

- Concerns that the scope of the project largely exceeds the recommended scope of 1-2 publications within the first 2 years.

- The project proposal heavily overlaps with on-going funded WHII projects.

The revised application should identify how the WHII research team's concerns have been addressed.

Should there be any disputes over data availability, confidentiality or exclusivity, the WHII research team reserves the right to make a final decision if an agreement cannot be reached.

3. Terms and Conditions

The main applicant and their collaborators must abide by these terms and conditions, which are in line with the MRC data sharing guidelines. Failure to comply with them will result on any future data sharing applications from the main applicant and/or their collaborators being refused by the WHII team.

3.1 Data application

- Only the main investigator of the project is entitled to ask for data.

- The applicant and their collaborators must be *bona fide* researchers with an established scientific record, who must conduct high quality, ethical research when using WHII data. The aim of the research should be *bona fide* research with eventual public benefit.

- The description of the project, analysis and publication of data will be as specific and focused as possible.

- A precise list of the variables required for the project must be provided to the WHII team. This list will be compiled using the Whitehall II data dictionary available on the study website.

- Major changes to the project will require a new data sharing application.

- The data will be released upon receipt of the signed data sharing agreement and payment of the fee.

- If the WHII data user applies for funding once the application has been approved, and the grant application is based on WHII data, they must inform the WHII Contact Researcher.

3.2 Data usage and security

- WHII Data, in whole or in part, must be used only for the specified research project. Major changes require a new application.

- Once the project has been approved, data users should work under the Data Protection Scheme that operates in their country. If requested, a copy of this Data Protection registration should be sent to the WHII Contact Researcher.

- The WHII team should be informed of any new collaborators in the project.

- Data cannot be transferred to persons not working on the application project.

- Data users will *not* have sole and exclusive access to their required set of data. The WHII team cannot guarantee that the data used in their application will be for exclusive use of the applicant and their research team.

- Analyses are allowed only according to the protocol described in the application;

- Data errors must be notified to the WHII Contact Researcher.

- Secure data access, such as passwords, firewalls, etc., must be in place to ensure that the data are kept secure. WHII data users should access the provided dataset using a network drive set up by the academic organization where they work, and avoid keeping datasets on their own PC/laptop, unless the device is encrypted (minimum AES-256). This network drive should be accessible through password control.

- WHII data users working on the same project should use a shared drive for exchanging files, and avoid the use of memory sticks or attachments in e-mails.

- Users must offer their derived variables to enrich the WHII database, and supply the appropriate documentation.

- On completion of the project, the data must not be used anymore. All electronic copies of WHII data held by the applicant must be deleted.

- There can be no more copies of the data than is reasonable for backing up work.

- WHII data users must be aware that the data may allow individuals to be identified. Therefore, it will be the data user's responsibility to ensure that the participants' identity is not disclosed under any circumstances.

- It is forbidden to match or attempt to match individual records to any other data.

3.3 Publications

- The scope of the project should ideally be aimed at the publication of one or two scientific articles within two years after receipt of the requested dataset.

- The applicant must inform the WHII Contact Researcher and data manager (a.sanchez@ucl.ac.uk) of all publications resulting from their use of Whitehall II data.

- The name of Whitehall II Study must be included in the title or subtitle. This is not necessary in the case of multi-cohort/consortia papers.

- Authorship on papers should reflect the normal convention of including in appropriate order the instigator(s) of the idea of the paper, the person(s) carrying out the research and the writer(s) of the manuscript. If the project has received significant input from the WHII Contact Researcher then s/he may also be included as an author.

- Please add the following sentence to the paper:

  "Whitehall II data are available to bona fide researchers for research purposes. Please refer to the Whitehall II data sharing policy at http://www.ucl.ac.uk/whitehallII/data-sharing"

- The WHII team should receive credit for the amount of work carried out to establish the WHII resource throughout the years. Therefore, a suitable note of **acknowledgement** should be agreed with the assigned WHII Contact Researcher, along the following lines:

  "We thank all participants in the Whitehall II Study, Whitehall II researchers and support staff who make the study possible. The UK Medical Research Council (MR/K013351/1; G0902037), British Heart Foundation (RG/13/2/30098), and the US National Institutes of Health (R01HL36310, R01AG013196) have supported collection of data in the Whitehall II Study."

- If you have used environmental data, please add the following:

  "The Whitehall II air pollution concentrations were generated by the Environmental Research Group, King's College London. Additional markers of the physical environment were generated by Triantafyllos Pliakas, London School of Hygiene and Tropical Medicine."

- If requested, the applicant must prepare a summary of their main WHII findings.

Please note that the WHII team retains the right to modify the contents of this document at any time.

Whitehall II Data Application Form

Please note that a separate application form is required to access WHII Genetics data.

Preferred member of the WHII team to be the WHII Contact Researcher (if any):

----

1. PRINCIPAL INVESTIGATOR (PI)

For PhD projects, the PI on the application must be the supervisor, not the student. Please do not enter the name of the student as the PI.

Title, forename, surname:	
Work Address:	
Telephone:	
Email:	

Please attach the PI's Curriculum Vitae.

2. RESEARCH TEAM / CO-APPLICANTS

Details of each Research Team member involved in the proposed project who will have access to the data.

Name of co-applicant	Employing organisation	Position in organisation	Email address

3. PUBLICATIONS

List of the Research Team main publications.

4. PREVIOUS WHITEHALL II APPLICATIONS

IF you or any of the Research Team members have applied for Whitehall II data, please give details:

Main applicant:

Application date:

Project title:

Resulting publications/outputs:

5. FUNDING

Do you already have funding to carry out this project (Yes/No)?

If you are planning to seek funding to carry out this project and the grant application is to be partially or totally based in the use of Whitehall II data, please give details about the funding application.

Name of funding body:	
Dates:	
Applicant(s):	
Title of grant application:	

6. RESEARCH PROJECT

6.1 PROJECT TITLE

6.2 SUMMARY
A brief summary of up to 200 words describing the aims of the study/research project.

6.3 CONTEXT
Where research is part of a larger programme, please give details.

6.4 PROJECT DESCRIPTION
Full description of the purpose/s for which the data are requested (maximum 4 A4 sides).

- Background

- Objectives. Please list clearly:

    1. **Exposure**

    2. **Outcome**

    3. **Confounders/Adjustment variables**

- Methodology and planned statistical analyses.

- Competing interests

- References (max 10)

6.5 PLANNED SCIENTIFIC OUTPUTS

Intended outputs/publications arising from the use of these data, including abstracts, posters and research papers.

7. DATA REQUESTED

7.1 DATA FORMAT NEEDED

Please indicate what format you will be using the data (SAS, SPSS, Stata, csv):

7.2 DATA REQUIRED FOR ANALYSIS

1) Download the WHII data dictionary from the Whitehall II website (www.ucl.ac.uk/whitehallII/data-sharing). It is an Excel file containing the <u>exact variables names</u> for every collection phase.

2) Highlight all of the variables needed. The list of highlighted items and data collection phases must be consistent with the project proposal.

3) Send the highlighted WHII data dictionary file at the same time as this application form.

Whitehall II Data User's Agreement

To be signed **once the data sharing application has been approved**.

Main applicant

Title, forename, surname: ....

Work Address: ....

....

....

Telephone: ....

Email: ....

Project title: ....

I agree that my project will use the requested Whitehall II data and is to be conducted according to the contents of the document "Whitehall II Policy on Data Sharing".

I agree that both *myself and my collaborators* will abide by the terms and conditions outlined there within, also *summarised in the next page.*

I agree that failure to comply with these terms and conditions will result on any future data sharing applications *from me and/or my collaborators* being refused by the WHII team.

Signature of **main applicant:** ....

Name in block capitals: ....

Date: ....

Please sign, date and return to the WHII Contact Researcher in electronic or paper form.

Summary of the WHII Data Sharing terms and conditions

For the full text, please refer to the document entitled "Whitehall II Policy on Data Sharing", available at www.ucl.ac.uk/whitehallII/data-sharing.

Application

- Only the main investigator of the project can apply for WHII data.

- The applicant and their collaborators must be *bona fide* researchers with an established scientific record, who must conduct high quality, ethical research when using WHII data.

- A precise list of the variables required for the project must be provided to the WHII team.

- Major changes to the project will require a new data sharing application.

- The data will be released upon receipt of the signed data sharing agreement and payment of the fee.

- If the WHII data user applies for funding once the application has been approved, and the grant application is based on WHII data, they must inform the WHII Contact Researcher.

Data usage and security

- Data must be used only for the specified research project.

- Data cannot be used by persons not mentioned in the application or distributed to third parties.

- Users will NOT have sole and exclusive access to their required dataset.

- Data errors must be notified to the WHII Contact Researcher.

- Users must offer their derived variables to enrich the WHII database.

- WHII data users must work under the Data Protection Scheme that operates in their country.

- Secure data access must be ensured by using secure networks, passwords, firewalls and/or highly encrypted devices.

- WHII data users working on the same project should use a shared drive for exchanging files, and avoid the use of memory sticks or attachments in e-mails.

- WHII data users must be aware that the data may allow individuals to be identified. Therefore, it will be the data user's responsibility to ensure that the participants' identity is not disclosed under any circumstances.

- It is forbidden to match or attempt to match individual records to any other data.

- On completion of the project, all electronic copies of WHII data must be deleted.

- There can be no more copies of the data than is reasonable for backing up work.

Publications

- The scope of the project should be one or two scientific articles within the first two years.

- The WHII data user must inform the WHII Contact Researcher and the data manager of all publications resulting from their use of WHII data.

- The name of Whitehall II Study must be included in the title or subtitle of publications.

- A suitable note of acknowledgement should be added. Please refer to section 3.3) of this policy.

- If the project has received significant input from the WHII Contact Researcher then s/he may also be included as an author.