Whitehall II
Data Collection Procedure
Phases 11 and 12

Version 2 – March 2015 – Dr Aida Sanchez
With the collaboration of Therese Butler, Beverly Milne, Stephanie Smith

Department of Epidemiology and Public Health,
University College London
1-19 Torrington Place
London WC1E 6BT

Table of contents

W2 Staff

This document outlines in detail all of the tasks undertaken during a Whitehall II (W2) data collection wave.

The following members of staff are responsible for the completion of these tasks. These are not carried out in their totality by them personally, but they are responsible for the completion by others members of the team as and when needed.

Thérèse Butler (TB)

Data Manager/Database programmer

Thérèse is responsible for developing and maintaining a number of administrative and clinical databases used on the project by the study team. She is also responsible for the data management of clinical data.

Mrs Beverly Milne (BM)

Clinical Coordinator.

Beverly is responsible for the co-ordination of the clinic and home visits for the study. This includes developing standardized protocols for all measures and adapting these for the home visits as well as liaising with other members of the team on clinical data and measurement needs.

Dr Aida Sanchez (AS)

Senior Data Manager.

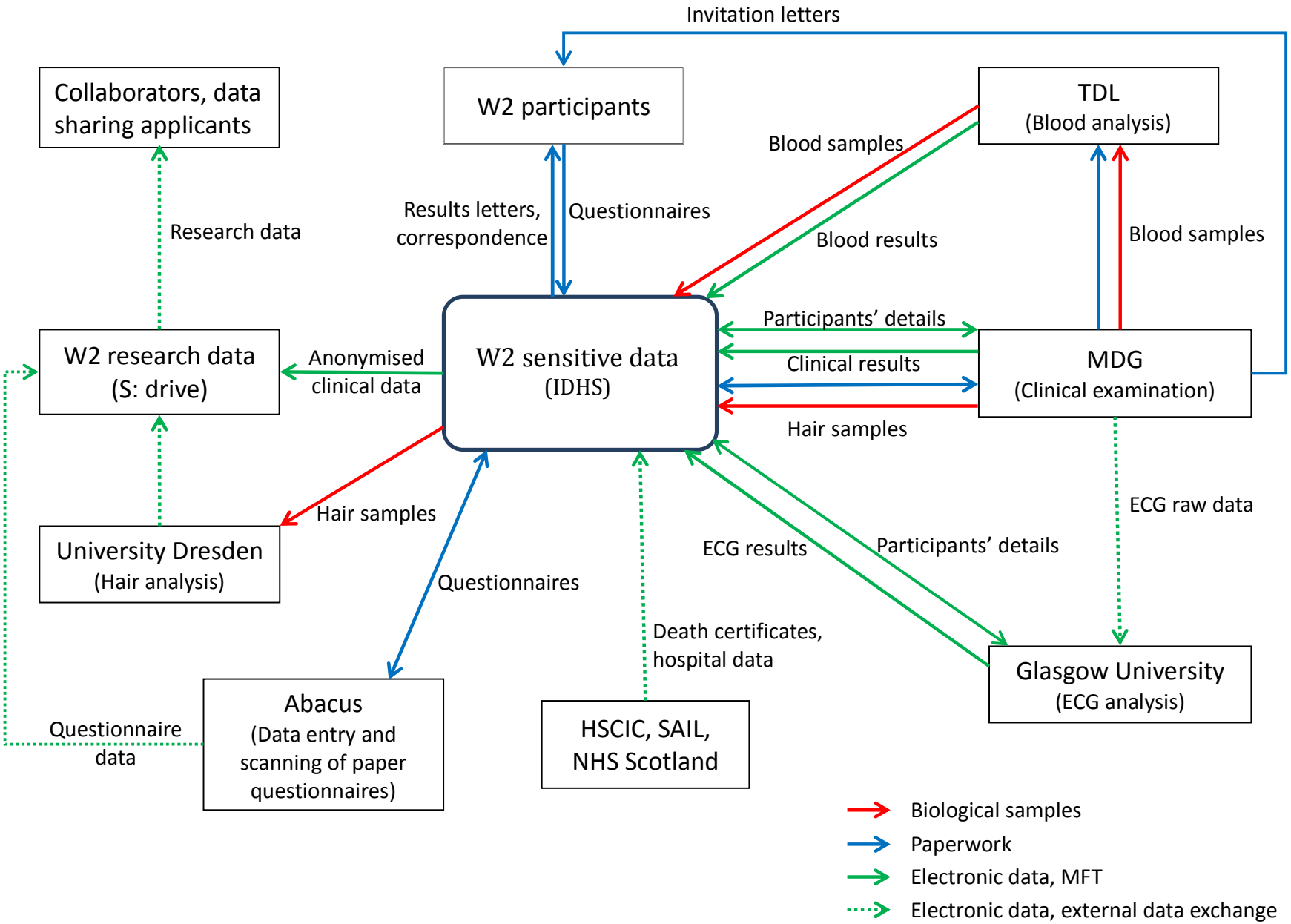
Aida is responsible for Whitehall II research data, data management plan and production of metadata. She is also responsible for the study information governance, implementation of the data sharing policy and the linkage with external sources of health data.

Stephanie Smith (SS)

Study and Data Co-ordinator

Stephanie is responsible for the management and co-ordination of the data collection, questionnaires, communication with participants and administrative activities.

Overview of the W2 data collection



Update of details/GP consent/ Proxy nomination

Preparation

1. Design (SS)
2. Printing (SS)

Data collection

1. Send to participants (SS)
2. Receive from participants (handed in at the clinic or received by post) (SS)
3. Record what forms have been received (SS)
4. Checking forms for completeness and chase participants for missed items (SS)

Database programming

1. Update White1 database to now include audit consent and all proxy information (TB)
2. Update the Consent database used during data collection (TB)
3. Update the main Consent database containing consent information at all phases (TB)

Data entry

1. Single data entry of the Update of Details, in White2 database (SS)
2. Single data entry of GP details, in White2 database (SS)
3. Double data entry of GP and audit consent in the White2 database and in the Consent database (TB)

Data processing

1. Check addresses, eg validity of postcodes before big mailouts (TB)
2. Check GP and audit consent entered in the White2 database against the Consent database and correct discrepancies (TB)
3. Check that GP and audit consents have been entered for all participants at that phase (TB)

Scanning

1. Scan of the GP consent forms by data entry company, but on UCL premises (SS)
2. Filing (SS)

Proxy consent

Preparation

1. Design of the proxy consent forms and consultee declaration forms (BM)
2. Design the information sheets for both the proxy and consultee (BM)
3. Printing (SS)

Data collection

1. Send to participants (SS)
2. Receive from participants (handed in at the clinic or received by post) (SS)
3. Record what forms have been received (SS)
4. Checking forms for completeness and chase proxy for missed items (SS)

Database programming

1. Update the Consent database used during data collection (TB)
2. Update the main Consent database containing consent information at all phases (TB)

Data entry

1. Double data entry of proxy consent in the Consent database (TB)
2. Double data entry of consultee declaration consent in the Consent database (TB)

Data processing

1. Check addresses, eg validity of postcodes before mailouts if required (TB)
2. Check if double data entry records match for proxy and consultee declaration consents, and correct discrepancies (TB)
3. Check that consultee and proxy consents were obtained for adults with incapacity (TB)

Scanning

1. Scan of the GP consent forms by data entry company, but on UCL premises (SS)
2. Filing (SS)

Screening consent forms

Preparation

1. Design & layout of consent forms (BM/Researchers)
2. Printing (SS)

Data collection

1. Further check of form contents in office (SS)
2. Contacting participants to correct forms if necessary (SS)
3. Keep a record of all consent issues (eg dates of form sent back, returned) (TB)

Database programming

1. Update the Consent database used during data collection (TB)
2. Update the main Consent database containing consent information at all phases (TB)

Data entry

1. Organise double data entry in the Consent database for all consent forms (TB)

Data processing

1. Check if double data entry records match, and correct discrepancies (TB)
2. Check screening consent was obtained for all screened participants (clinic and home visits) (TB)
3. Check test/measure not done where consent was denied (TB)
4. Inform the clinic coordinator and the genetic data manager if DNA consent was denied (TB)
5. Inform clinic coordinator if storage of blood samples was denied (TB)

Questionnaires

GKQ = General Knowledge Questionnaire
HSQ = Health Survey Questionnaire
MMSE = Mini-mental Examination
TMT = Trial Making Test

Preparation

1. In-house design of self-completion questionnaires: HSQ, CQ, Proxy (SS/Researchers)
2. In-house design of clinical questionnaires: MMSE, GKQ and TMT (SS/ Researchers)
3. Send HSQ and GKQ for professional finish (SS)
4. Prepare data dictionaries for each questionnaire type (AS)
5. Obtain quoting from the data entry company for double data entry, GKQ coding, scanning and production of PDF database (SS)

HSQ Pilot

1. Recruit pilot participants (SS)
2. Send and receive HSQ questionnaires (SS)
3. Check questionnaires (Researchers)
4. Collect comments from researchers and update the HSQ questionnaire accordingly (SS)
5. Send final version HSQ professional finish (SS)
6. Printing (SS)

Database programming

1. Update Documents despatched database (questionnaires received per questionnaire type) (TB)
2. Update Appointments database to keep a record of what questionnaires are received by post (TB)
3. Update Abacus database to record all documents sent externally for data entry (TB)

Data collection

1. Arrange the printing of all questionnaires (SS)
2. Print labels/barcodes to identify participant and questionnaire type (SS)
3. Send the HSQs to participants (SS)
4. Receive the questionnaires from the clinics/home visits or by post (SS)
5. Chase participants for whom we haven't received a HSQ, or from whom we need to clarify HSQ questions (SS)
6. Carry out telephone interviews if necessary (SS)
7. Carry out proxy interviews if necessary (SS)

Data entry and scanning

1. Send data dictionaries to the data entry company so that they can set up their databases (AS)
2. Generate lists of questionnaires received (using HSQ barcodes and CAPI data) (TB)
3. Check lists against actual questionnaires received and resolve discrepancies if any (TB)
4. Send the questionnaires in batches to data entry company, monthly (SS)
5. Double-entry, external (Data entry company)
6. Scanning (Data entry company)
7. Deal with queries from the data entry company regarding data entry or coding (SS)
8. Receive questionnaires (check what has been received against dispatch sheet) (SS)
9. Use dispatch sheets to enter date of questionnaires returned in Abacus database (TB)
10. File returned questionnaires (SS)

Data coding

1. Prepare coding manuals for HSQ, MMSE and GKQ (SS)
2. Coding of HSQ and MMSE, in house (SS)
3. Coding of GKQ (Data entry company)

Alchemy

1. Organise labels for Alchemy database (SS)
2. Receive scanned images in Alchemy database (SS)

Data Processing

1. Request the data entry company to send the electronic file (csv) for each questionnaire type throughout the collection phase for regular data quality checks (AS)
2. Perform data quality checks (AS/Researchers)
3. Receive final electronic file (.csv) for each questionnaire type at the end of the collection phase (AS)

Generation of Research Data (DVP)

1. Transform the csv file into a SAS file for each questionnaire type (AS):
   - HSQ
   - GKQ
   - MMSE
   - CQ
2. Check study numbers and quality of data (AS)
3. Data cleaning (AS)
4. Write SAS programs to assign labels, format labels and to compute derived variables (AS)
5. Check data quality (Researchers)
6. Further modifications of DVP data if necessary and add to DVP (AS)

Planning of Clinical Data Collection

Protocols and Clinic Flow

1. Identify which researcher responsible for which measures (BM)
2. Agree tests for each phase with Researchers (BM)
3. Develop protocols for each clinical measure, including GK (BM/Researchers)
4. Develop protocols for crucial questions, proxy interview, MMSE and TMT protocols (SS)
5. Develop and agree clinic flow for effective running of the clinic and home visits (BM)

Ethics

1. Create substantial amendment in IRAS for submissions to Ethics committee, using the Minimal dataset already created for Whitehall Study (BM)
2. Confirm that Ethics committee used at previous phase is still functioning as some committees get dissolved and project gets passed onto a new committee. Checks dates of monthly meetings to allow sufficient time for submission (BM)
3. Confirm with the sponsor, Joint UCL/UCLH R&D, who is responsible for signing off the study prior to submission (BM)
4. Submit the substantial amendment electronically to the Ethics committee and deal with any queries raised. Attend meeting if required (BM/Researchers)

Equipment

1. Equipment: Agree, purchase, calibrate and check equipment requirements for each measure (BM)
2. Set up service contracts as required i.e. ECG machines (BM)

Biological Sample Management

1. Agree process with laboratory (TDL) for analysis of bloods and collection or sending of samples (BM)
2. Sample storage: assess storage space for blood samples in the freezers and inform researchers (BM)
3. Hair sample storage: agree sorting and storage of samples with researcher responsible (SS)

Database programming

4. Update Appointments database (TB)
5. Update Results database used for storing CAPI data, blood results, ECG results, result letters (TB)
6. Update ECG serial comparison database (TB)

Data Processes

1. Data queries: agree procedure for dealing with data queries (TB)
2. Bloods- agree data transfer with TDL: file format, frequency and security (TB)
3. Lung Function –agree pre-population method and data format (XML), transfer method and frequency, conversion from to SAS (AS)
4. PWV - agree data format, transfer method and frequency, backup, conversion to SAS (AS)
5. CIS-R - agree data transfer: file format, frequency, security (AS)
6. ECG:
    - Agree with ECG Core Lab in Glasgow the transmission of the ECGs and data transfer (TB)
    - Send Glasgow list of all active participants in Excel with dates of previous ECGs (TB)
    - Prepare file in format required for pre-population of EC Sense database (TB)

CAPI Development

1. Gather and write CAPI requirements as needed for programmers (TB)

2. Liaise with programmers via MDG/TNS to answer programmers' queries (TB)
3. Provide information additional to CAPI requirements if needed, eg warning/error messages (TB)
4. Provide test cases to be used for testing CAPI (TB)
5. Provide scenarios to help test CAPI (TB)
6. Provide feedback to MDG on all testing of CAPI performed by UCL (TB)

External Company

1. Planning – Liaise with external company to ensure they will be able to meet the needs required by the study with regards staffing, accommodation (if conducting a clinic) and coverage for the home visits. Assess what the equipment needs are and ensure the correct equipment is purchased. (SS/BM)
2. Statement of Works (SOW) – Check and edit the SOW to ensure all necessary costs are accounted for and ensure it outlines how the external company will conduct the clinic and home visits. Ensure the timeline is outlined and this meets the requirements for the study i.e. clinical screening is completed in time stipulated in the grant applications. (SS/BM)
3. Contract – Liaise with UCL Contracts, UCL Finance, etc (Project Manager)
4. Meetings - Form working groups and set up meetings between UCL and external companies. (SS)

Training

1. Write training notes (BM/SS)
2. Develop training programme for all clinic and home visit staff (BM)
3. Develop manual for training for all clinical staff (BM)
4. Arrange training for all clinical staff (BM)
5. Arrange training for all admin staff working in house (SS)

Clinical Data Collection

Preparation

1. Identify content of all labels (BM/SS)
2. Generate data and Access reports or barcodes for labels required (paperwork, samples) (TB)
3. Print labels (SS)
4. Design of Screening Form (SS/BM)
5. Set up Appointments database to generate:
   - initial letters to be sent to potential participants (TB)
   - participants' details to send to external company (TB)
6. Contact potential participants (SS)
7. Generate paperwork and labels in batches and send them to external company (SS)

Database programming

1. Maintain all databases as required (Appointments, Results, Abacus, Documents despatched, ECG Serial Comparison) (TB)

Pilot

1. Design a pilot plan to ensure sufficient numbers of participants to test the flow of the clinic (BM)
2. Create advert to recruit pilot participants (SS)
3. Collect information from volunteers wanting to be pilot participants (SS)
4. Update pilot White2 database to record pilot participants' details (TB)
5. Provide a modified version of the Appointments database for generating labels required (TB)
6. Design and print pilot paperwork and labels (SS)

Data Collection

1. Contact participants regarding participating (SS)
2. Arrange appointments (external company)
3. Deal with queries from external company relating to contacting participants (SS)
4. Perform the clinical measurements in clinics or at the participant's home (external company)
5. Data entry (see section below for details)
6. Blood samples:
   - Shipment of samples to TDL, chemical analysis and transmission of results to Whitehall (TDL)
   - Liaise with TDL for any queries related to blood collection and analysis - (BM/TB)
7. ECG:
   - Send ECG data to ECG Core lab in Glasgow, electronically (MDG)
   - Analysis, management and diagnosis of ECG (ECG Core lab in Glasgow)
   - Transmission of ECG results and ECG graphs to Whitehall (ECG Core lab in Glasgow)
   - Liaise with Glasgow ECG Core lab in Glasgow regarding ECG queries (BM/TB)
8. CAPI results: import CAPI data (csv files) into Screening tables in Results database (TB)
9. Lung Function results: receive XML data and transform to SAS (AS)
10. PWV results: receive PWV data and transform to SAS (AS)
11. Deal with queries from clinic staff and participants at the clinic (BM)
12. Deal with queries from participants before and after the clinic visit (SS)
13. Deal with data entry queries, liaising with relevant WII staff as needed (TB)

Data entry

1. Clinic up to P11: double data entry of screening forms, using in-house Screening database (TB)

2. Home visits at P9 & P11: data entry done and validated in CAPI by nurses as the measure is taken (NatCen)
3. Clinic and home visits at P12: data entry done and validated in CAPI by nurses as the measure is taken (external company)

Data processing

1. Protocol checks throughout data collection (BM/SS)
2. Organise regular protocol check timetable for researchers (BM)
3. CAPI: data quality checks on CAPI data throughout data collection. Cleaning of screening data as collection progresses running checks in Access VBA in Screening/Results database (TB)
4. ECG: save data in pXlabdata database; data quality checks run monthly; queries to be resolved with Glasgow (TB)
5. Blood (TDL) results stored in pXlabdata database; data quality checks run monthly; queries to be resolved with TDL (TB)
6. CIS-R: check serial number and date weekly (TB)
7. Lung Function: data quality checks throughout data collection (AS)
8. PWV: data quality checks throughout data collection (AS)
9. Organise data quality check for researchers (AS)

Generation of Research Data (DVP)

1. Access queries run to check data further – editing screening data if necessary (TB)
2. Run queries in Access to prepare Excel files required for DVP as agreed with data manager (TB):
   - CAPI (screening)
   - Bloods
   - CIS-R
   - ECG
   - Hair cortisol
3. Create a final dataset in SAS for each measurement (AS):
   - CAPI (screening)
   - Bloods
   - CIS-R
   - ECG
   -Hair cortisol
   - Lung Function
   - PWV
4. For each measurement do the following: into a SAS file and process data (AS):
   - Check study numbers and quality of data
   - Data cleaning
   - Write SAS programs to assign labels, format labels and to compute derived variables (AS)
5. Check data quality (Researchers)
6. Further modifications of DVP data if necessary and add to DVP (AS)

Clinical Events tracing

Project management

1. Organise and chair meetings to plan strategy (AS)
2. Run queries to compare data between DVP and HES, as needed for planning (Statistician)

Generation of baseline datasets

1. Identification of self-reported cardiac events: creation of baseline dataset using SAS queries using information from HSQ (AS)
2. Summary of ECG data: dataset for the participants who self-report cardiac events (AS)
3. Summary of previous CHD events: dataset from existing CHD events data in DVP (AS)

Application for hospital data

4. Apply for HES data (AS)
5. Apply for hospital data in Scotland (AS)
6. Apply for hospital data in Wales (AS)
7. Apply for MINAP data (AS)

Processing of hospital data

1. Receive hospital data received from the different sources and convert to SAS (AS)
2. Process and format hospital data using SAS (AS)
3. Identify CHD and stroke clinical events (AS)

Generate CHD reports

1. Collation of the CHD event information using all sources available (TB):
    - Self-reported dataset
    - ECG
    - Hospital data
    - Previous events
2. Generate reports using an Access database (TB)

Coding of events

1. Design coding strategy (Researchers)
2. Code all events for each participants using the CHD reports (Researchers)
3. Design of Access database to do data entry (TB)
4. Data entry of codes using Access database (Researchers)

GP letters

1. Identify what events need to be validated by a GP (Annie)
2. Design of GP database to generate letters and store results (TB)
3. Generate GP letters using an Access database (TB)
4. Send and manage the GP letters (BM)
5. Data entry of information returned by the GPs in an Access database (BM)
6. Code information provided by the GPs (Researchers)
7. Data entry of codes from GP data (BM)

Final DVP event file

1. Collation of all CHD event codes (TB)

2. Generation of the new research events dataset using the following information and applying the 28 day rule (TB):
   - existing CHD events on DVP
   - New events
   - mortality data
   - Rose angina (ECG abnormalities)
3. Generation of DVP dataset (Statistician)