# Introduction

This Annex lists the data-fields available from the Airwave Study Tissue Bank screening protocol. The reader should consult the main data-dictionary for a description of the screening protocol itself, explanations of formatting conventions, missing values, and other technical information.

## Document Configuration

| | |
|---|---|
| Title | Data Dictionary, Annex A |
| Subject | Interpretation of Data Fields Derived from the Screening Protocol |
| Version | 2.0 |
| Status | Draft |
| Author | Heard, Andy H, Database Manager at Imperial College London. |
| Published | 11/05/2022 |
| Filename | annex-a.docx |

# Changes since Previous Version

The following are the main structural changes since version 1. In addition, there have been a number of corrections made to individual values where further investigations have revealed errors.

1. Follow-up screens have been included, meaning that some participants now have longitudinal screening results.
2. The change in assay for glycosylated haemoglobin has resulted in three new labels and, for clarity, renaming of three existing labels.
3. The medication (treatments) section has been removed to a separate extract.
4. All the ECG data has been removed to a separate extract.
5. The computation of sitting-height has changed following a review of the height of the stools used. We have also corrected a small number of results where the height and sitting-height were switched around during data entry.
6. Values of certain assays that were reported by the laboratory as "< x" were exported as "x" but are now exported as zero (0).
7. Hip values are now reported when available for pregnant participants.
8. A fault in hours_since_eat_blood has been corrected where the individual had a rebleed.
9. The method of determining the contingency code for missing values more reliably detects when a value is missing because it was not a part of the protocol at the time.

# Clinic Results

These data are derived from participants' visit for a health screen, and have been divided into functional groups.

## Identifiers

These fields identify the participant and the screening protocol used.

| Label | Data Type | Description |
|---|---|---|
| barcode | NUMBER (5) | Pseudonymous identifier for one complete screen for a single participant. It is unique within the extract. Participants returning for a second screen (for example, as part of the follow-up protocol) will be assigned a new barcode for later visits. Join visits together using part_id, below. |
| part_id | NUMBER (7) | Pseudonymous identifier that is unique for each participant. Because the screening extract includes follow-up visits, some part_id values are associated with two or more barcodes. |
| protocol | STRING (3) | The protocol used for the visit:<br>• PILOT: The first screening protocol deployed from June 2004 until August 2006.<br>• MAIN: The bulk of the participants were recruited using the MAIN protocol, which succeeded the PILOT.<br>• HYBRID: A small number of baseline screens inadvertently contained elements of both PILOT and MAIN.<br>• FOLLOWUP: A recall protocol for participants with a baseline health-screen.<br>• REACT and REACT-SIEMENS: COVID-19 studies based on Airwave Study participants whose relevant data has been incorporated into this extract. |
| nurse_id_first | NUMBER | An anonymised identifier for the nurse carrying out the initial clinic-based part of the protocol. |
| when_screened | DATE | When the clinic appointment began. |

## General Questions asked by the Nurse

These are the answers to questions put by the nurse.

| Label | Data Type | Description |
|---|---|---|
| repeat_declared | NUMBER (1) | Please read the paragraph about **Error! Reference source not found.** for background on this data-field.<br>The values of this label are:<br>• 0: no repeat declared; no repeat found.<br>• 1: repeat declared but not found.<br>• 2: no repeat declared, but at least one found.<br>• 3: repeat declared and found. |
| age_when_screened | NUMBER (2) | Participant's age in years at the time of the screen. |
| force_region | STRING | A description of the geographic region of the force employing the participant at screening time. |

| Label | Data Type | Description |
|---|---|---|
| is_urine_sample_given | YESANY | Whether the participant provided a urine sample. |
| hours_since_eat_blood | NUMBER | Self-reported hours that passed since the participant ate or drank anything other than water.<br>Until mid-2006, this field was collected by asking the participants for a time (9am, 10pm etc.), which we converted into a duration by comparing it to the clock on the nurse's laptop. Later, we replaced it with a direct question on elapsed hours. |
| hours_since_last_urine | NUMBER | Self-reported number of hours since the participant last passed urine previous to giving this sample. This question was dropped after the Pilot. |
| food_diary_received | YESANY | Whether a completed food-diary was returned at the clinic. Some diaries were returned after the clinic appointment, so a "no" to this question does not necessarily mean there is no food-diary. |
| airwave_diary_typical | STRING (4) | Is the usage reported in the Airwave Usage Diary typical of an average week?<br>• YES "Yes"<br>• MORE "No. I usually use the radio more"<br>• LESS "No. I usually use the radio less"<br>• NONE "Diary not completed"<br>• NOTU "Diary not completed (not a user)" (only available since July 2011). |
| is_menses | YESNO | Whether the participant is menstruating today. |
| is_pregnant | YESNO | Whether the participant is pregnant today. |
| postcode_district | STRING (4) | The outbound portion of the postcode given by the participant as their home address. |

## Lifestyle Questions

These were asked by the nurse in the pilot phase, before being moved to the self-administered questionnaire.

| Label | Data Type | Description |
|---|---|---|
| is_smoker | YESNO | Whether the participant smokes. |
| cigarettes_per_day | NUMBER | Quantity of cigarettes normally smoked per day, ignoring the very occasional cigar or pipe-tobacco. |
| smoking_age_started | NUMBER | Age participant started smoking (years). |
| has_drunk_within_24hours | YESANY | Whether the participant has drunk any alcohol in the last 24 hours. |
| beer_drunk | NUMBER | Quantity of beer drunk during a typical week. |

| Label | Data Type | Description |
|---|---|---|
| beer_type | STRING (1) | Units of beer_drunk: P (pints) or U (standard units) |
| wine_drunk | NUMBER (3) | Quantity of wine drunk during a typical week. |
| wine_type | STRING (7) | Units of wine_drunk: B (bottles) or U (standard units) |
| other_alcohol_units | NUMBER (3) | Quantity (standard units) of alcohol drunk during a typical week other than beer or wine. |

## Blood Pressure

Participants normally have three blood pressure readings, taken consecutively, each of which records systolic and diastolic blood pressures (mm Hg) and pulse (beats /min). Participants reporting themselves as diabetics may have two sets of readings, one standing and one sitting; non-diabetics have the sitting measurement only.

| Label | Data Type | Description |
|---|---|---|
| bp_pulse_sitting_1 | NUMBER (3) | First sitting pulse measurement (beats / min). |
| bp_systolic_sitting_1 | NUMBER (3) | First sitting systolic blood pressure (mm Hg). |
| bp_diastolic_sitting_1 | NUMBER (3) | First sitting diastolic blood pressure (mm Hg). |
| bp_pulse_sitting_2 | NUMBER (3) | Second sitting pulse measurement. |
| bp_systolic_sitting_2 | NUMBER (3) | Second sitting systolic blood pressure. |
| bp_diastolic_sitting_2 | NUMBER (3) | Second sitting diastolic blood pressure. |
| bp_pulse_sitting_3 | NUMBER (3) | Third sitting pulse measurement. |
| bp_systolic_sitting_3 | NUMBER (3) | Third sitting systolic blood pressure. |
| bp_diastolic_sitting_3 | NUMBER (3) | Third sitting diastolic blood pressure. |
| bp_pulse_sitting | NUMBER (3) | Mean sitting pulse measurement. |
| bp_systolic_sitting | NUMBER (3) | Mean sitting systolic blood pressure. |
| bp_diastolic_sitting | NUMBER (3) | Mean sitting diastolic blood pressure. |
| bp_pulse_standing_1 | NUMBER (3) | First standing pulse measurement (beats / min). |
| bp_systolic_standing_1 | NUMBER (3) | First standing systolic blood pressure (mm Hg). |
| bp_diastolic_standing_1 | NUMBER (3) | First standing diastolic blood pressure (mm Hg). |
| bp_pulse_standing_2 | NUMBER (3) | Second standing pulse measurement. |
| bp_systolic_standing_2 | NUMBER (3) | Second standing systolic blood pressure. |
| bp_diastolic_standing_2 | NUMBER (3) | Second standing diastolic blood pressure. |

| Label | Data Type | Description |
|---|---|---|
| bp_pulse_standing_3 | NUMBER (3) | Third standing pulse measurement. |
| bp_systolic_standing_3 | NUMBER (3) | Third standing systolic blood pressure. |
| bp_diastolic_standing_3 | NUMBER (3) | Third standing diastolic blood pressure. |
| bp_pulse_standing | NUMBER (3) | Mean standing pulse measurement. |
| bp_systolic_standing | NUMBER (3) | Mean standing systolic blood pressure. |
| bp_diastolic_standing | NUMBER (3) | Mean standing diastolic blood pressure. |
| handedness | STRING (5) | Whether the participant is LEFT or RIGHT handed?" |
| bp_arm_used | STRING (5) | The arm used for the blood pressure measurement: LEFT, RIGHT or UNKNOWN. |
| bp_cuffsize | STRING (7) | Cuff size used for blood pressure: REGULAR, LARGE or UNKNOWN. |

## Anthropomorphic Measurements

Most participants have two measurements of height, weight, waist-girth, hip-girth and sitting-height (height measured whilst sitting on a stool of known height). Pregnancy is a contraindication for waist-girth, and sometimes for hip-girth. From these measurements we have derived a waist-hip ratio, body-mass-index and sitting-height ratio.

Body Mass Index (BMI) is computed for each trial as (weight ÷ height$^2$). Its mean is computed from the individual BMI values.

We show the individual measurements and computed mean for each measurand.

### Sitting Height

Sitting-height measurements are reported with the height of the stool already subtracted. In the 2013 extract, the stool was stated to be 60.7 cm high; however, further measurements of several stools show that 60.4 cm to be more accurate. Some cross-clinic variation in the dimensions of equipment used does seem to have occurred.

Sitting-height-ratio is defined as (sitting-height ÷ height).

| Label | Data Type | Description |
|---|---|---|
| height_1 | NUMBER | First measurement of height (cm). |
| hip_girth_1 | NUMBER | First measurement of hip girth (cm). |
| weight_1 | NUMBER | First measurement of weight (kg). |
| sitting_height_1 | NUMBER | First measurement of sitting height (cm). |
| waist_girth_1 | NUMBER | First measurement of waist girth (cm). |
| waist_hip_1 | NUMBER | Computed waist-hip ratio for first measurements. |

| Label | Data Type | Description |
|---|---|---|
| body_mass_index_1 | NUMBER | First computed body mass index (kg / $m^2$). |
| sitting_height_ratio_1 | NUMBER | First computed sitting-height-ratio (dimensionless). |
| height_2 | NUMBER | Second measurement of height (cm). |
| hip_girth_2 | NUMBER | Second measurement of hip girth (cm). |
| weight_2 | NUMBER | Second measurement of weight (kg). |
| sitting_height_2 | NUMBER | Second measurement of sitting height (cm). |
| waist_girth_2 | NUMBER | Second measurement of waist girth (cm). |
| waist_hip_2 | NUMBER | Computed waist-hip ratio for second measurements. |
| body_mass_index_2 | NUMBER | Second computed body mass index (kg / $m^2$). |
| sitting_height_ratio_2 | NUMBER | Second computed sitting-height-ratio (dimensionless). |
| height | NUMBER | Mean height (cm). |
| hip_girth | NUMBER | Mean hip girth (cm). |
| weight | NUMBER | Mean weight (kg). |
| sitting_height | NUMBER | Mean sitting height (cm). |
| waist_girth | NUMBER | Mean waist girth (cm). |
| waist_hip | NUMBER | Mean waist-hip ratio. |
| body_mass_index | NUMBER | Mean body mass index (kg / $m^2$). |
| sitting_height_ratio | NUMBER | Mean sitting-height-ratio (dimensionless). |

## Body Composition Analysis

These measurements were made using a Tanita analyser and the results entered by hand.

Our understanding is that the impedance measurements are used by the machine to derive fat-percentage and total-body-water. However, we were unable to persuade the supplier to provide us with the algorithms used or any references to independent evaluations of their methods. We understand (but haven't tested) that the method differentiates between athletic and standard body-types.

### Contraindications

There are three main contraindications for Tanita: pregnancy (is_pregnant) or the presence of metal objects in the body, especially a pacemaker.

| Label | Data Type | Description |
|---|---|---|
| takes_intense_exercise | YESNO | Whether participant reports doing ten or more hours of intense exercise per week. |

| Label | Data Type | Description |
|---|---|---|
| body_type | STRING (8) | The nurse's subjective judgement of whether the participant is a STANDARD or ATHLETE body-type. |
| fat_percentage | NUMBER | Percentage body fat (%) |
| total_body_water | NUMBER | Total body water (kg) |
| impedance_of_body | NUMBER | Impedance of the whole body (Ω) |
| impedance_of_right_leg | NUMBER | Impedance of the right leg (Ω) |
| impedance_of_left_leg | NUMBER | Impedance of the left leg (Ω) |
| impedance_of_right_arm | NUMBER | Impedance of the right arm (Ω) |
| impedance_of_left_arm | NUMBER | Impedance of the left arm (Ω) |
| imp_contra_metal_implant | YES | YES if contraindicated because of a metal implant. |
| imp_contra_pacemaker | YES | YES if contraindicated because of a pacemaker. |
| imp_contra_other | YES | YES if contraindicated for any other reason. |

### Grip Strength

For a small number of most recent participants we measured grip-strength using the Jamar device. The results are based on two measures for each hand.

| Label | Data Type | Description |
|---|---|---|
| grip_right_1 | NUMBER | Grip strength of right hand (kg) |
| grip_left_1 | NUMBER | Grip strength of left hand (kg) |
| grip_right_1 | NUMBER | Grip strength of right hand (kg) |
| grip_left_1 | NUMBER | Grip strength of left hand (kg) |
| grip_right | NUMBER | Mean grip-strength of right hand (kg) |
| grip_left | NUMBER | Mean grip-strength of left hand (kg) |

## Laboratory Results

These are the assay results carried on the blood collected at the clinic.

We used a variety of laboratories and analysers during the project. Most of the baseline results (2004 – June 2015) were assayed at Northwick Park Institute of Medical Research (NPIMR). During the Pilot, we used the laboratory's existing elderly equipment: COBAS Mira (clinical chemistry), H1E (haematology) and ACL-300 (coagulation). For the Main Study, we upgraded to Ilab 350 (chemistry); Advia 2120 (haematology); ACL-8000 (coagulation), and an Elisa Plate Reader for C-peptide. Charing Cross Hospital carried out assays for us from November 2015 until 2021. From 2022 we are planning to use third party commercial laboratories.

Up until 2022, most samples were collected in the field and transported overnight to the laboratory in a thermoporter at 2 – 4 °C where the haematology and (usually) clinical chemistry was carried out on the day of arrival.

## Laboratory Profiles

There are three distinct "profiles" (the set of measurements that we expect to be returned for each barcode) in the cohort, for the Pilot; for the Main Study, and for the follow-up. For the Main Study, measurands were added to clinical chemistry because they were deemed interesting, and because the new chemistry machine could perform these measurements. We gained some extra results from haematology because these came from the new machine without extra effort, although strictly speaking they are not a part of the profile. To balance the budget, we had to drop some of the measures from the original profile.

The boundary between the two profiles is far from distinct however. For example, for several months we ran both profiles in parallel to ensure the new machines were working correctly.

## Glycosylated Haemoglobin

Two sets of variables record the glycosylated haemoglobin (HbA1c) assay. From the beginning of the research until May 2014, we used the Diabetes Control and Complications Trial (DCCT) method. DCCT was then superseded by the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) method. We understand that this change results from the National Glycohaemoglobin Standardisation Program, which was tackling the presence of too many false positives for elevated HbA1c. We were unable to use DCCT for the whole cohort because the manufacturer withdrew support in 2014.

Each sample was measured using one method or the other, with a rare exceptional case where both assays are reported.

### *Mapping Results across Methods*

The two methods both measure the proportion of HbA1c relative to total haemoglobin. However, the DCCT method is reported in units of g/dl, whereas IFCC uses mmol/mol. The reference range of values differs between the two methods in order to ensure that, during the changeover, practitioners would not mistake the method used.

Because of the differences in method, the two series are not directly comparable. However, conversion formulae and reference ranges are published and, according to advice from NPIMR (May 2014), the "normal" ranges for the IFCC method are:

- 20-42 mmol/mol (4-6% in DCCT units) in non-diabetics
- 42-64 mmol/mol (6-8% in DCCT units) in controlled diabetics
- 64-up to 195 mmol/mol (8-20% in DCCT units) in uncontrolled diabetics.

The conversion formulae below were attributed to the IFCC organisation, though an article on Wikipedia (11th January 2017) uses slightly different coefficients.

$$\text{DCCT}(\%) = 0.09148 \times \text{IFCC}(\text{mmol/mol}) + 2.152$$

$$\text{IFCC}(\text{mmol/mol}) = 10.93 \times \text{DCCT}(\%) - 23.50$$

A comparison between the methods using a test sample resulted in IFCC values a little lower than the translated DCCT value, although no test of statistical significance was carried out.

### *Population Distribution for IFCC Assay Methods*

Within the IFCC values, the assays were performed using one of two different methods, depending on the laboratory: a turbidimetric method (NPIMR); or an ion-exchange / HPLC method (Charing Cross). We report this choice as hba1c_method.

A complication is that the distribution of values for the two IFCC methods do not coincide. Although we did not carry out a method comparison study when moving laboratories, a brief analysis in 2018 showed that the population mean using the ion-exchange method is roughly one standard-deviation greater than the mean found using the turbidimetric method. Based on advice from a consultant in Metabolic Medicine & Honorary Research Fellow at Imperial College, "[it]…*will not be possible to derive a population mean/SD as these will always be dependent on the method used… Although the methods correlate there will be absolute differences due to the methods used."*[1]

## Haematology

These data include general haematology assays and differential white-cell analysis. Note that, for haemoglobin, the value reported to participants is that measured in clinical chemistry.

| Label | Data Type | Description |
|---|---|---|
| red_blood_cell_count | NUMBER | Red blood cell count ($10^6$/µL) |
| white_blood_cell_count | NUMBER | White blood cell count ($10^3$/µL). |
| haemoglobin | NUMBER | Haemoglobin (g/dL). |
| haematocrit | NUMBER | Haematocrit (%) |
| mean_red_cell_volume | NUMBER | Mean red cell volume (fl) |
| mean_cell_haemoglobin | NUMBER | Mean cell haemoglobin (pg). |
| mean_cell_haemoglobin_conc | NUMBER | Mean Cell Haemoglobin Concentration (g/dL). |
| platelets | NUMBER | Platelets (103/µl) |
| red_cell_distribution_width | NUMBER | Red cell distribution width (%). |
| corpuscular_haem_conc | NUMBER | Corpuscular haematology concentration (g / dL). |
| mean_platelet_volume | NUMBER | Mean platelet volume (fl). |
| haemoglobin_distribution_width | NUMBER | Haemoglobin concentration distribution width (g/dL). |
| when_haematology_first | DATE | When the assay was first carried out. |
| when_haematology_last | DATE | When the assay was last carried out. |

[1] Email: Busbridge – Heard, 11th February 2019

## Differential White Cell Counts

This is an analysis of the different types of white cell in the blood. The value assayed is the percentage of each cell-type as a percentage of total white cells. In aggregate, therefore, the sum should equal 100%. In practice, the values may sum to more or less than 100%, which we consider to be an artefact of the method.

The clinically significant number, however, is not the proportion of cells of each type but their absolute number (count). Here we report the computed counts, defined as the product of cell-percentage and total white cells. The haematology machine itself carries out the computation of counts, which we report where available. There is often a discrepancy in the last decimal place between the value reported and the product of white-cell-count and percentage because of value rounding.

| Label | Data Type | Description |
|---|---|---|
| neutrophils_count | NUMBER | Neutrophils ($10^3$/µl) |
| lymphocytes_count | NUMBER | Lymphocytes ($10^3$/µl) |
| monocytes_count | NUMBER | Monocytes ($10^3$/µl) |
| eosinophils_count | NUMBER | Eosinophils ($10^3$/µl) |
| basophils_count | NUMBER | Basophils ($10^3$/µl) |
| large_unstained_cells_count | NUMBER | Large unstained cells ($10^3$/µl). |
| nrbca | NUMBER | Nucleated red blood cells per unit volume ($10^3$/µl). |

**Table 1: Differential White Cell Counts**

## Elisa Plate Reader

A DTX 800 Elisa Plate Reader was used to obtain C-Peptide results.

| Label | Data Type | Description |
|---|---|---|
| c_peptide | NUMBER | C-Peptide (pmol / L). |
| when_cpeptide_first | DATE | When the assay was first carried out. |
| when_cpeptide_last | DATE | When the assay was last carried out. |

## Coagulation Analysis

Coagulation analysis was performed to measure fibrinogen and prothrombin time.

| Label | Data Type | Description |
|---|---|---|
| fibrinogen | NUMBER | Fibrinogen (g/L). |
| prothrombin_time | NUMBER | Prothrombin Time (seconds). |
| when_coagulation_first | DATE | When the assay was first carried out. |

| Label | Data Type | Description |
|---|---|---|
| when_coagulation_last | DATE | When the assay was last carried out. |

## Clinical Chemistry

Clinical chemistry assays were performed to obtain these results.

| Label | Data Type | Description |
|---|---|---|
| c_reactive_protein | NUMBER | C-reactive protein (mg/L). |
| glucose | NUMBER | Glucose (mmol/L) |
| total_protein | NUMBER | Total Protein (g/L) |
| albumin | NUMBER | Albumin (g/L) |
| calcium | NUMBER | Calcium (mmol/L) |
| sodium | NUMBER | Sodium (mmol/L) |
| potassium | NUMBER | Potassium (mmol/L) |
| alanine_aminotransferase | NUMBER | Alanine Aminotransferase (U/L) |
| alkaline_phosphatase | NUMBER | Alkaline Phosphatase (U/L) |
| bilirubin | NUMBER | Bilirubin (µmol / L) |
| creatinine | NUMBER | Creatinine (µmol / L) |
| total_cholesterol | NUMBER | Total Cholesterol (mmol/L) |
| hdl | NUMBER | High Density Lipoprotein (mmol/L) |
| ldl | NUMBER | Low Density Lipoprotein (mmol/L) |
| triglycerides | NUMBER | Triglycerides (mmol/L) |
| gamma_gt | NUMBER | Gamma GT (U/L) |
| apolipoprotein_a | NUMBER | Apolipoprotein a1 (g/L) |
| apolipoprotein_b | NUMBER | Apolipoprotein b (g/L) |
| urea | NUMBER | Urea (mmol/L) |
| haemoglobin_dcct | NUMBER | Haemoglobin measured according to the DCCT method (g/dL). This was previously labelled haemoglobin_clinchem. |
| hba1c_conc_dcct | NUMBER | Glycosylated Haemoglobin (g/dL). This was previously labelled hba1c_conc. |

| Label | Data Type | Description |
|---|---|---|
| hba1c_percent_dcct | NUMBER | Glycosylated Haemoglobin as percentage of total haemoglobin (%). This was previously labelled hba1c_percent.<br>We report the values computed by the analyser or, when not available: hba1c_conc_dcct ÷ haemoglobin_dcct ×100% |
| haemoglobin_ifcc | NUMBER | Haemoglobin measured according to the IFCC method (mmol/mol). |
| hba1c_conc_ifcc | NUMBER | Glycosylated Haemoglobin (mmol/mol). |
| hba1c_ratio_ifcc | NUMBER | Glycosylated Haemoglobin according to the IFCC method (dimensionless).<br>We report the values computed by the analyser or, when not available: hba1c_conc_ifcc ÷ haemoglobin_ifcc * 1000 |
| hba1c_method | STRING | The assay method used:<br>• D: DCCT result.<br>• T: IFCC result using a turbidimetric method.<br>• I: IFCC result using ion-exchange. |
| when_clinchem_first | DATE | When the assay was first carried out. |
| when_clinchem_last | DATE | When the assay was last carried out. |