

## Introduction

This Annex lists the data-fields available from the Airwave Study Tissue Bank treatments and medications export.

### Document Configuration

Title	Data Dictionary, Annex F
Subject	Sample Use Matrix
Version	1.1
Status	Draft
Author	Heard, Andy H, Database Manager at Imperial College London.
Filename	annex-f.docx

## Overview

The Sample Use Matrix is a single file, each record of which states which assays have been carried out on the sample. We also include columns coding for other data collection exercises that are not sample related.

### Changes Since Version 1.0

Clarification made for the values of aw1\_npc\_blood\_urine and aw2\_metabolon\_npc. Corrected heading errors.

## Data Labels

The output file described below is tab-separated and includes a header record of column names.

Label	Data Type	Description
barcode	NUMBER (5)	Health-screening identifier.
part_id	NUMBER (7)	Participant identifier.
barcode_seq#	NUMBER (1)	A participant's first screen is 1, the second 2, etc.
age_when_screened	NUMBER (2)	Participant's age (years) when the sample was collected.
project_stage	STRING	The project stage when collected. <b>BASELINE</b> (Pre June-2015); or, <b>FOLLOW-UP</b> (After June-2015).
gwas	CHAR	Whether GWAS has been obtained. <b>S</b> = Genome obtained successfully. <b>P</b> = Partially successful (call-rate < 95%). <b>F</b> = Failed to obtain genome.
exome	CHAR	Number of exome datasets obtained.
methylation_2018	NUMBER (1)	Number of methylations datasets obtained.

Label	Data Type	Description
aw1_npc_blood_urine	NUMBER (1)	Number of metabolomics datasets obtained from the "Airwave 1" assay, which was carried out at the National Phenome Centre (NPC) and provided data on paired aliquots of blood and urine taken from the same sample. Most results are "2", meaning results exist for blood and urine. "1" means only blood or urine is available.
aw2_metabolon_npc	NUMBER(1)	Number of metabolomics datasets obtained from the "Airwave 2" assay, which was an analysis carried out by Metabolon and for a subset of the same samples at the NPC. Here, "2" means that both NPC and Metabolon assays exist; "1" means only 1 result exists, most likely Metabolon.
aw3_baseline_fup	NUMBER (1)	Number of metabolomics datasets obtained from the "Airwave 3" metabolomics assay. This was carried out by the NPC on participants' paired (baseline and follow-up) samples.
nightingale_2022	NUMBER (1)	Number of metabolomics datasets obtained from the "Nightingale" metabolomics assay in 2022.
olink_2023	NUMBER (1)	At present (January 2023) this reflects the number of aliquots selected for the OLink proteomics assay.
cognitive	NUMBER (1)	Number of cognitive tests carried out.
food_diary	CHAR	Whether a food diary is available. C = Coded. U = A food diary believed to exist, but is not yet coded.
diabetes_inferred	CHAR	An assessment of the likelihood that the sample was donated by a participant with diabetes:  H = High degree of confidence M = Medium confidence L = Low confidence N = Not likely

## Version History

VERSION	Filename	Date Exported	Total Rows	CRC-64
1	sample-use-matrix-v1.tsv	26 <sup>th</sup> January 2023	69,225	969EE4E153BF0962