

# UK Biobank

## Repeat Assessment Data

---

### September 2013

Version 1.0

<http://www.ukbiobank.ac.uk/>



## Contents

|                                       |   |
|---------------------------------------|---|
| 1. Introduction                       | 3 |
| 2. Repeat assessment visit            | 3 |
| 3. Regression dilution bias           | 3 |
| 4. Repeat assessment data in Showcase | 4 |
| 5. Repeat assessment data in datasets | 5 |
| 6. Requesting repeat assessment data  | 5 |

## 1 Introduction

UK Biobank holds an unprecedented amount of data on half a million participants aged 40-69 years (with a roughly even number of men and women) recruited between 2006 and 2010 throughout the UK.

The data Showcase (available at <http://www.ukbiobank.ac.uk>) aims to present the data available for health-related research in a comprehensive and concise way, and to provide technical information for researchers considering applying to use the resource.

Showcase is updated on a regular basis as health outcomes data are received from external data providers (such as the death and cancer registry) and as data from 'enhancement' projects, carried out by UK Biobank, are collected.

This manual aims to provide an introduction to the **repeat assessment data**, which was collected on a subset of 20,000 participants, who were invited to reattend an assessment centre to undergo repeat baseline measures.

## 2 Repeat assessment visit

A repeat assessment of 20,000 participants was carried out between August 2012 and June 2013 at the UK Biobank Co-ordinating Centre, Stockport, UK. Participants who lived within a 35 km radius of the assessment centre were invited via email or letter, with an overall response rate of 21%.

Participants underwent a repeat of the baseline assessment visit, which included obtaining information on a participant's health and lifestyle, hearing and cognitive function, collected through a touchscreen questionnaire and brief verbal interview. A range of physical measurements were also performed, which included: blood pressure; arterial stiffness; eye measures (visual acuity, refractometry, intraocular pressure, optical coherence tomography); body composition measures (including impedance); hand-grip strength; ultrasound bone densitometry; spirometry; and an exercise/fitness test with ECG. Samples of blood, urine and saliva were also collected.

Repeat assessments are scheduled to be carried out every 2-3 years during follow up. The data collected not only enriches the UK Biobank resource with information on participant health and lifestyle over time, but also provides an opportunity to calibrate existing measures and account for regression dilution bias (section 3).

## 3 Regression dilution bias

Variation in the measurements captured at baseline may lead to underestimated associations between risk factor and outcome – known as regression dilution bias.

Regression dilution bias can be introduced through a variety of different ways, including, through measurement error, through short-term biological variability (e.g. diurnal or seasonal variation) or through longer-term within-person variability (e.g. changes to diet and physical activity or medication use, which could modify risk factor values at the time of measurement).

The information collected through re-assessment of a large subset of participants can be used to adjust for regression dilution bias and produce more precise measures for the entire cohort.

## 4 Repeat assessment data in Showcase

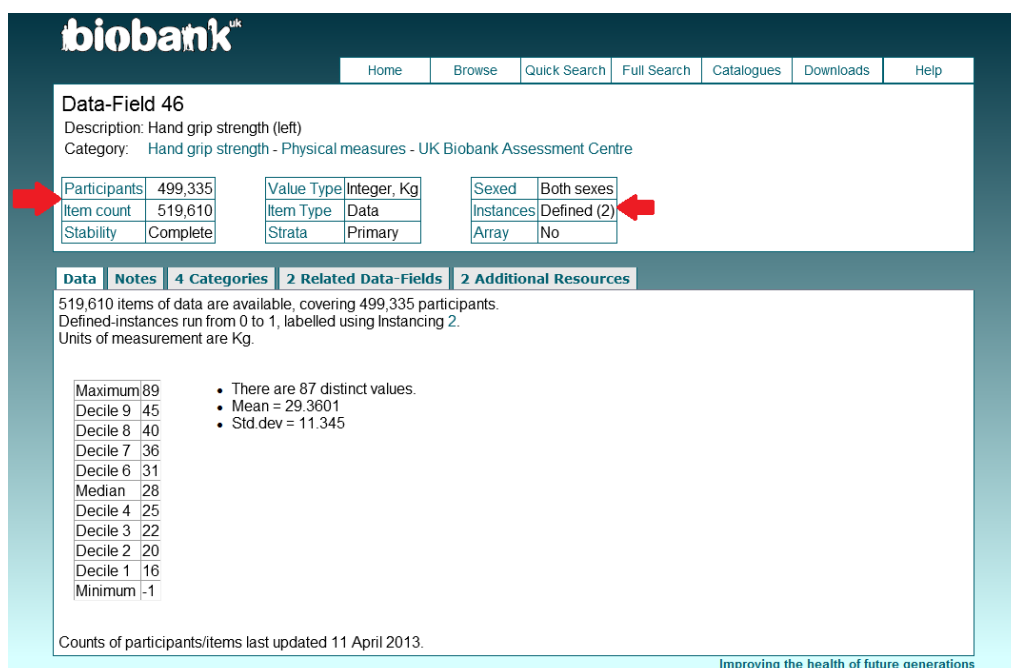
Repeat assessment data is presented in Showcase in the usual format - for further guidance on navigating Showcase; please consult the [Showcase User Guide](#).

The panel in the top-half of the data-field screen provides a brief description and the category location of the data-field (Figure 1). It also includes more detailed technical information about each data-field. This includes information on:

- **Participants:** the number of participants that have the data item
- **Item Count:** the number of data items available
- **Stability:** whether the data-field is complete or changes over time
- **Value type:** the format and units of the data-field
- **Item type:** whether the data-field is a simple data point, relates to an inventory of biological samples, or is a large data object
- **Strata:** the likely relevance to researchers of the data-field
- **Sexed:** whether the data-field is available for both sexes
- **Instances:** how many occasions participants have this measurement performed
- **Array:** whether there are multiple data items for each instance.

For example, Figure 1 shows that data on hand grip strength has 2 instances, because the measurement has been performed on two occasions i.e. at baseline and at repeat assessment. Comparing the item count with the participant count indicates that there have been 519,610 measurements taken on 499,335 participants.

**Figure 1:** illustration of a data field in Showcase



## 5 Repeat assessment data in datasets

All data fields are labelled with the format “F\_I\_A”, where F is the field ID, I is the instance index and A is the array index.

All indices run from the value 0, hence 46\_0\_0 corresponds to Field 46 (Hand grip strength (left)), Instance 0 (baseline measurement), Array 0 (first measurement taken), while 46\_1\_0 corresponds to Field 46 (Hand grip strength (left)), Instance 1 (repeat assessment measurement), Array 0 (first measurement taken).

Figure 2 shows how the data on hand grip strength is presented in a downloaded dataset. Data fields appear in instance and array order i.e. all the measurements taken on the first instance appear first, followed by the second, third, fourth instance and so on.

**Figure 2:** illustration of a data field in a dataset

|    | A      | B      | C      | D      | E      | F      | G      | H      |
|----|--------|--------|--------|--------|--------|--------|--------|--------|
| 1  | 46_0_0 | 46_1_0 | 47_0_0 | 47_1_0 | 48_0_0 | 48_1_0 | 49_0_0 | 49_1_0 |
| 2  | 38     | 38     | 44     | 40     | 88     | 97     | 100    | 105    |
| 3  | 40     | 30     | 45     | 32     | 76     | 80     | 103    | 101    |
| 4  | 27     | 20     | 26     | 20     | 74     | 78     | 87     | 91     |
| 5  | 24     | 19     | 34     | 20     | 74     | 72     | 97     | 95     |
| 6  | 46     | 43     | 48     | 40     | 87     | 91     | 96     | 96     |
| 7  | 38     | 36     | 30     | 36     | 105    | 100    | 109    | 110    |
| 8  | 20     | 22     | 22     | 21     | 80     | 91     | 99     | 98     |
| 9  | 42     | 35     | 40     | 41     | 83     | 91     | 101    | 103    |
| 10 | 48     | 46     | 42     | 46     | 105    | 104    | 108    | 114    |
| 11 | 24     | 24     | 23     | 24     | 84     | 85     | 102    | 110    |

## 6 Requesting repeat assessment data

Repeat assessment data can be requested via the [Access Management System](#), which is the standard route for obtaining UK Biobank data.

Researchers are required to register and submit an application, along with a Showcase ‘basket’, which contains the data fields to be included in the dataset. By default, researchers will receive both the data from the baseline and the repeat assessment visits. Researchers who do not wish to receive repeat assessment data should make this explicitly clear in their application.

Further details about registering and applying to use the resource can be found in the [Scientists Section](#) of the UK Biobank website.