

Access matters / the return to UK Biobank of derived variables

The purpose of this note is to provide guidance to researchers as to UK Biobank's approach regarding the return and release of specific derived variables. The note does not apply to the generation of simple derived variables (such as BMI) which are returned at the end of the research project but to sample assay data (e.g. the level of LDL cholesterol) and specific key variables derived from raw UK Biobank data (e.g. ventricular function generated from analysis of raw imaging data).

1. In determining when derived variables¹ are a) returned to UK Biobank and b) released to other researchers, UK Biobank's objective is to strike a fair balance between the following two objectives, namely to ensure that:
 - 1.1.1 accurate and consistent derived variables are returned to UK Biobank as soon as practically possible and that these derived variables are made available to other researchers within a reasonable time frame; and
 - 1.1.2 researchers are not discouraged from using UK Biobank by concerns that their research could be comprised (e.g. "scooped") by the early release of key data that they have generated.
- 1.2 There are a number of ways in which researchers can generate derived variables using the UK Biobank resource, principally:
 - 1.2.1 derived data from projects that have requested biological samples to conduct assays; or
 - 1.2.2 derived data from the combination/calculation/formulation of UK Biobank data.
- 1.3 In relation to derived variables, there is a key distinction to draw between the timing of two discrete events, namely:
 - 1.3.1 the timing of return of such derived variables to UK Biobank, where UK Biobank's objective is to ensure that the derived variables are a) consistent, reliable and accurate and b) actually returned to UK Biobank; and
 - 1.3.2 the timing of their release to other researchers, where UK Biobank's objective is to ensure that a balance is struck between the interests of the researcher generating the derived variables and the use of such variables by other researchers.
2. This note sets out the working approaches that UK Biobank will adopt, in relation to the three main sets of circumstances in which derived variables are generated. As ever, UK Biobank's objective is to be fair, transparent and consistent; however, there will be situations where UK Biobank may alter the presumptions or require a different approach and the considerations that UK Biobank will take into account are set out at the end of this note.

¹ Derived variables are data derived from the function or expression of other data or from the assay of samples. In this context, it applies to data derived from a sample (by way of assay or other analysis), data derived from raw data (by way of analysis, interpretation or calculation) or data derived from a combination of other data variables.

3. *UK Biobank (or an external lab on behalf of UK Biobank) conducts sample assays for a researcher*

3.1 The Access Procedures and the MTA (section B4 and clause 5 respectively) already address the situation where UK Biobank conducts sample assays for a specific research project, namely:

3.1.1 the relevant phenotypic data will be released to the researcher at the same time as the assay data and the “exclusive use” period runs from this time.

3.1.2 the period of “exclusive use” of the assay data is typically 3 months (but see below) before they are made available for other researchers to use;

3.1.3 the reason for this provision is to ensure that UK Biobank does not have to wait until the researcher has concluded all of their planned analyses before the assay data can be made available to other researchers.

3.2 In certain circumstances, the 3-month period of exclusive use may be altered by mutual agreement: for example, when the research project is complex or likely to take a long time to complete (see para 6.1 below for the factors that will be taken into account).

4. *The researcher conducts the assays*

4.1 Any assay conducted by a researcher on UK Biobank samples will always be subject to a pilot study on a small number of samples so that UK Biobank can assess the reproducibility and sensitivity of the assay before approving the release of the remaining samples.

4.2 UK Biobank will require the return of assay data prior to releasing other data needed for the research project. This ensures that UK Biobank receives the assay data.

4.3 As above, there is a presumption of a 3-month period of “exclusive use” of the assay data after release of the other data that are required; as above, this period can be extended in certain circumstances (see below in para 6.1 for factors that will be taken into account).

4.4 If a large number of samples are being assayed and/or the assays will take 12 months or more to generate, then UK Biobank may elect to adopt the following approach:

4.4.1 the assay data would be returned by the researcher to UK Biobank periodically (e.g. bi-annually in tranches of 50,000 participants) during the assay period. This periodic return enables UK Biobank to monitor and ensure that the assay is being conducted consistently and reliably; and

4.4.2 the researcher will decide whether to wait until assays on all of the samples have been completed before having the other required data released to them by UK Biobank, or whether they prefer to receive the other data after returning a subset of the participants’ assay data. The first release of the phenotypic data sets the clock running on the period of “exclusive use” for that tranche of assay data.

5. *The researcher generates specific derived variables from raw data*

- 5.1 This approach will be adopted when researchers are generating specific derived variables from raw data (e.g. DICOM imaging files or accelerometry data). UK Biobank will adopt the following approach, which is similar but not identical to the one outlined above (as the raw data are not depletable):
- 5.1.1 the derived variables should be returned by the researcher periodically (bi-annually is the presumed starting point) during the research project;
 - 5.1.2 the return of the derived variable may be linked with the release of the phenotypic data if UK Biobank considers this to be appropriate. If this is the case, then equivalent considerations as with the sample assays above will apply, namely (at the researcher's option) the first release of phenotypic data sets the clock running on the 3-month² period of exclusive use; and
 - 5.1.3 alternatively, if the return of the derived variable is not linked to the release of the phenotypic data then the derived variables will be made available to other researchers within 6 months of completion of the research or on the publication of the first related paper (whichever is earlier).

6. *Factors which UK Biobank will take into account*

- 6.1 UK Biobank considers that the development of consistent and fair principles is integral to its access process; nevertheless, it recognises that not all situations are directly comparable and as such its approach to individual cases will be informed by the following factors (looked at from the perspective of both UK Biobank and the researcher):
- 6.1.1 How have the derived variables been generated and by whom;
 - 6.1.2 How clearly can the derived variables be identified and how closely are they linked to the results of the research project;
 - 6.1.3 Respectively, how much UK Biobank and researcher effort have gone into generating the derived variables, both in intellectual and practical terms;
 - 6.1.4 Do the generation of the derived variables involve the use of proprietary or third party IP;
 - 6.1.5 Is there a clear benefit (such as ensuring ongoing quality control) or detriment, from the perspective of both UK Biobank and the researcher, in requiring an early or phased return of the derived variables;
 - 6.1.6 What are the practical implications, from the perspective of both UK Biobank and the researcher, involved in the actual implementation.

UK Biobank's Access Sub-Committee retains a keen interest in this topic and will keep it under review and amend / update as necessary.

June 2016

² As above, this period can be extended if appropriate.