

UK Biobank / 3rd party resources / treatment of common participants

1. The purpose of this guidance note, which should be read in conjunction with the de-duplication note [url], relates to the identification and analysis of common participants within UK Biobank and another resource. It sets out, in the context of an access application to UK Biobank from the 3rd party resource, the extent to which data fields on common participants within two separate resources - UK Biobank on the one hand and a resource (such as Million Women Study, Genomics England or 23andMe) on the other hand - can be linked and analysed.
2. The reason for this guidance is that UK Biobank considers that the combination of a relatively small number of phenotype/genotype data fields would enable a 3rd party resource to effectively link and re-identify a common participant (if the 3rd party resource retains the participant's identifiers). The purpose of the guidance is thus to minimise this risk and to set out the parameters of what is and what is not permitted.

Flagging of common participants

3. A 3rd party resource may make a request to UK Biobank to retain a flag (as a discrete data field) indicating that a participant is a common participant of both resources.
4. UK Biobank will retain this flag in its own resource, as a separate data field, indicating that the UK Biobank participant is a common participant in the 3rd party resource. If the 3rd party resource performs the matching itself, then this will need to be done using encrypted identifiers and the flag denoting common participants will be made available to both UK Biobank and the 3rd party resource.
5. UK Biobank will not make this flag publicly available on its Showcase database, but it will be available on request to researchers who make a suitable research request and present an appropriate justification (i.e., it will be a 'restricted' data-field). The 3rd party resource is also entitled to retain a comparable flag in its resource on the same basis.

Applications by 3rd party resources to analyse combined data fields in research

6. UK Biobank will consider, on a case-by-case basis, a research request from a 3rd party resource to conduct an analysis which combines, for common participants, data fields from UK Biobank with data fields from the 3rd party resource.
7. In considering such requests, UK Biobank shall exercise its reasonable discretion over the number and nature of the selected fields requested by the 3rd party resource and agree with the 3rd party resource on any potential derived variables (from the analysis) to be included in the resource. UK Biobank may determine, in light of the resources that it needs to deploy, that a reasonable fee should be paid by the 3rd party resource (which fee shall be agreed in advance).
8. For the flagged common participants, UK Biobank will provide the selected UK Biobank data fields to the 3rd party resource using new pseudo-identifiers.
9. On completion of the analysis of the combined data fields by UK Biobank the following protocol will be adopted:
 - The 3rd party resource is permitted to incorporate the agreed derived variables alongside the common participant's existing data fields in the 3rd party resource;
 - The agreed derived variables shall also be provided back to UK Biobank and incorporated alongside the common participant's data fields in UK Biobank; and
 - UK Biobank will require the 3rd party resource to delete the remaining common participant UK Biobank data fields.

Assimilation of data fields

10. Save for the agreed derived variables the 3rd party resource is not permitted to retain or make available any UK Biobank data fields in relation to a common participant. Equally, and save for the agreed derived variables, UK Biobank does not expect and will not seek to incorporate any other data fields within UK Biobank from a 3rd party resource in relation to a common participant.
11. In addition to the 3rd party resource retaining the agreed derived variables, it is also reasonable for the 3rd party resource to incorporate, as necessary, a suitable narrative juxtaposed to the relevant data field along the lines of “verified/confirmed using UK Biobank data”.