

UK Biobank data linkage / guidance notes

1. The purpose of this guidance note is to set out how UK Biobank will deal with requests for linkage of the data sets that have been provided for different research projects.
2. When UK Biobank releases data to researchers, these data are released with project-specific randomised ID codes for each participant (i.e. they have been “de-identified”: please see UK Biobank's note on data de-identification [url]).
3. Researchers working on the same research project are entitled to access the same dataset. UK Biobank sends the initial dataset to the Applicant PI who in turn is entitled to distribute this dataset to the named collaborators. As such, all researchers working on the same research projects are working on the same dataset using the same ID codes.
4. If data on the same participant, Mr Smith, are sent to a researcher working on another project, these data will have a distinct randomly generated ID code. Thus, it is not possible for the researchers to link these two data sets on Mr Smith together as they have no way of knowing that the same de-identified participant is involved.
5. UK Biobank recognises that in certain circumstances two or more researcher groups may want to link their datasets together in order to extend the range of their analyses. UK Biobank is prepared to assist researchers in creating linkage between the same participant(s) as long as the proposed linkage:
 - is proportionate, reasonable and technically feasible;
 - is compatible with UK Biobank's Access Procedures; and
 - does not *unacceptably increase* the risk of participant identification.
6. **Linkage with further data required by researchers**
 - 6.1 In the event that researchers require additional data from UK Biobank as part of their original research project then UK Biobank will release the updated/augmented data with the same ID codes if the request is considered to be within the scope of the original project.
 - 6.2 If, however, the request is considered to go beyond the scope of the original proposal then a fresh application will need to be made to UK Biobank with an explanation for the proposed linkage. If the application is successful, UK Biobank will aim to release these data with ID codes that are interoperable.
7. **Linkage with datasets generated by researchers doing similar research**
 - 7.1 *A researcher is undertaking two separate but related hypotheses.* This covers a situation where two separate applications have been made to UK Biobank by the same researcher. If the researcher wishes to link the two datasets it is probable that this linkage request will go

beyond the scope of the original proposals and thus a fresh application will need to be made, including any explanation for the proposed linkage. Where possible, UK Biobank will release the linked dataset with ID codes that are interoperable with the original projects.

7.2 *Two researchers are undertaking similar research with related (or identical hypotheses).* This covers a situation where two separate applications have been made to UK Biobank by two different researchers. If the two researchers wish to link their respective datasets it is probable that the linkage will go beyond the scope of the original proposals and thus a fresh application will need to be made by the researchers to UK Biobank. Where possible, UK Biobank will release the linked dataset with ID codes that are interoperable with the original projects.

8. **Approval mechanism for linkage applications**

8.1 If a new application is required then the procedure and criteria for this application will be the same as for any application made under UK Biobank's Access Procedures.

8.2 Nevertheless, given that the UK Biobank is likely to have already reviewed and approved the research projects between which linkage is sought then the researcher(s) has the option of combining the preliminary and main application into a single application phase.