

UK Biobank policy on dealing with sensitive phenotypes and aggregate numbers in publicly available materials

The aim of this policy is to provide UK Biobank researchers with guidance in relation to the publication of sensitive phenotypes and minimum aggregate numbers in publicly available materials.

1. This guidance covers two topics:

- i) The display of particularly sensitive phenotypes. UK Biobank considers that these are ones which may indicate that a criminal offence has been committed by or against a participant; and
- ii) The minimum numbers to be used in charts, tables and summary fields.

2. The rationale behind this guidance is that:

- i) Particularly Sensitive phenotypes:
 - although the potential risk of re-identification is very low, the consequences of re-identification in conjunction with a highly sensitive phenotype (possible indication of a criminal offence) would be highly problematic (for the relevant participant); and
 - This is not intended to discourage legitimate research work on such topics, just to ensure that the public display of this information is appropriately limited.
- ii) Minimum aggregated numbers:
 - UK Biobank does not permit the publication of individual level participant data – and please see related guidance on geographical data and rare phenotypes – unless this is strictly necessary and commensurate with the prevailing context: for example, the use of a single MRI image is permitted (with suitable safeguards as set out in the guidance) and in any event such data must always be minimised to the maximum extent possible; and
 - UK Biobank recognises that researchers do regularly seek guidance on the minimum numbers that can be published in a chart, table or similar and thus the purpose of this guidance is to provide a consistent approach where, on the basis of a single characteristic or combination of characteristics, there may be an increase in the potential re-identification risk, with 5 being the minimum number which applies to any type of publication.

3. This guidance applies to any data made publicly available by a researcher, including:

- i) Tables and charts in publications; and
- ii) On-line data catalogues including browser tools developed to search and display UK Biobank phenotypes.

Guidance on sensitive phenotypes

Below is the list of fields containing sensitive phenotypes, contained in the Showcase, that should not be publicly displayed.

Showcase Field ID	Category	Description
20488	Traumatic events	Physically abused by family as a child
20490	Traumatic events	Sexually molested as a child
20523	Traumatic events	Physical violence by partner or ex-partner as an adult
20529	Traumatic events	Victim of physically violent crime
20531	Traumatic events	Victim of sexual assault
29077	Adverse life events	Physically abused by family as a child
29079	Adverse life events	Sexually molested as a child
29083	Adverse life events	Physical violence by partner or ex-partner as an adult
29084	Adverse life events	Sexual interference by partner or ex-partner without consent as an adult
29085	Adverse life events	Sexual intercourse by partner or ex-partner without consent as an adult
29086	Adverse life events	Experienced a violent of sexual assault

Further, sensitive phenotypes may also be present in summary and record-level health linkage data systems (including but not limited to the codes ICD-10, READ, and SNOMED). These are listed below and should not be publicly displayed.

Showcase Category ID	Resource ID	Description	Coding systems	Subject
2000	141140	Inpatient summary fields and record-level diagnosis/procedure data	ICD-9, ICD-10, OPCS-3, and OPCS-4	Clinical or procedure codes indicating abuse, assault, or violence
		Inpatient summary fields and record-level data related to admission source or discharge destination	Various (see resource)	Codes indicating possible criminal offences or incarceration
100093	115559	Death register summary fields and record-level data on cause of death	ICD-10	Clinical codes indicating abuse, assault, or violence
3000	591	Primary care record-level clinical coded data	READ/CTV-3	Codes indicating abuse, assault, violence, exploitation, or harassment
				Codes indicating possible criminal offences or incarceration
				Codes related to people other than the participant (e.g. family members)

For more information about the specific Showcase fields and/or record table columns of relevance, please refer to Showcase [Resource 596](#) or the resources attached to the categories below.

Guidance on reporting small numbers

Any publicly available summary field should incorporate a minimum count of 5 where, on the basis of a single characteristic or combination of characteristics, there may be an increase in the potential re-identification risk. UK Biobank is not asking researchers to suppress or perturb the data, rather to use a minimum number with a particular field. Common strategies for disclosure control include:

- Collapsing categories to reduce the sparsity of the data;
- Aggregating the data over a greater period, or a larger geographical area;
- Rounding to a specific base to avoid very small numbers (no less than 5 – please note that any reported totals should use the rounded values to avoid the possibility of reverse-engineering the raw counts);
- Suppressing very small numbers (please note that in this case and reported totals should be derived excluding the suppressed counts, to avoid the possibility of reverse-engineering).

Any questions or queries please ask UK Biobank's Access Team (access@ukiobank.ac.uk).