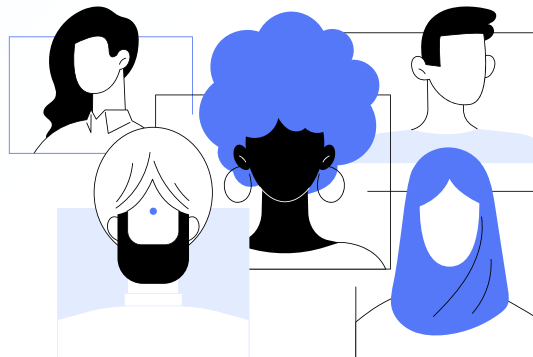




Assessment of data and **Risk of Bias** when using data Ethnicity and Migration

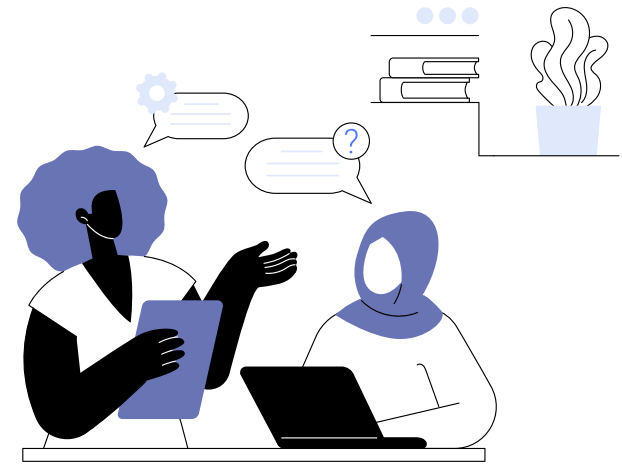


Guiding questions

To be used in consultation with detailed explanatory document.



PART 1



Questions to ask regarding data collection

1.1

Questions to ask regarding data collection

What was the method of gathering data – was it collected by an organisation or an individual?



In the case of survey data:

- Were the data collected face-to-face, online, or via an automated system?
- In the case of survey data: was the survey self-complete or interviewer-completed?
- Was an interviewer present even in the event of self-complete?

In the case of administrative data:

- Was this collected in a face-to-face setting (such as via a GP) or was it an online submission (such as registration for specialist services)?
- Was the information provided via a self-complete form or via face to face discussion?
- Was any demographic data related to ethnicity specified by the interviewee, or the interviewer?
- Were labels or categories assigned with regard to ethnicity or immigration status? Is information provided about the method of assigning categories?

1.2

Who is collecting the data?

What are the characteristics of the people collecting data?

How do they relate to the population being studied?



In the event of face-to-face collection, consider:

- Were the data collectors reflective of the population?
- Consider the potential effect of the relationship between the participant (or individual) and the data collector.
- In what ways were data collectors proximate to or representative of the population?
- Consider the most relevant characteristics in the context of the population, cultural norms, and the subject matter. These might include race, ethnicity, immigration or immigrant status, age, gender, sexual orientation, disability, accent, communication style.

- This can be applied to both survey and administrative data-collection settings.
- Did the data collectors consult with relevant groups in the design of the questions or the format?
- Consultation prior to the design of data collection is best practice to account for any norms of information disclosure relevant to particular minoritized populations. This applies to both survey and administrative data.

How did people feel about disclosing their data?



- **Was it made clear which institution or organisation was collecting the data?**

If yes:

- **Was the institutional affiliation clearly disclosed to the participants?**

Consider the impact of the organisation on the data collected. For example, people in insecure status might have less trust of state-affiliated or state-based institutions. Certain specialist services might carry cultural connotations that affect particular minoritized groups (for example violence against women, mental health).

- **To what extent does the collecting institution have power over the lives of the data providers?**
- **To what extent might the data providers perceive that the collecting institution holds power over them?**
- **Is there any record of exploitation between institution/organisation and population?**

This might be relevant in the case of administrative data such as in the context of police data. It may also be relevant in the case of institution-affiliated surveys such as Home Office.

- **Have any mitigating steps been taken at the point of data collection to address this power disparity?**

Data collectors should have documented any actions taken to address and neutralise a power disparity. If no mitigating steps were taken, this should be noted and addressed in data analysis.

- **In the case of specialist services data, are there any normative or cultural factors that will bias the data?**

For example, in the data field, consider if there are any known cultural biases to being referred to, or seeking out, specialist services. (See companion document for examples).

If potential biasing effects have been mitigated, what were the mitigation strategies?



Assess any reported strategies designed to reduce data bias. Ask:

- **Did strategies aim to neutralise bias or reduce bias?**
- **Were there potential new problems created by these strategies?**

For example, if a 'safe space' for disclosure is adopted to avoid people potentially being overheard in the home, this might create travel costs or time costs, thus creating new points of bias that might affect particular groups more than others.

1.3 How are the questions asked (culture, norms)?

What cultural norms might affect the likelihood of responding to a survey?

What cultural norms might have an impact on the disclosure of information in specific administrative settings?

Or in response to particular questions?



- **Are a subset of people being excluded because they are not reachable for social or cultural reasons?**

For example, are some populations less likely to be home/ less likely to have an individual phone number/ are less likely than other groups to visit a GP or to seek out specialist services?

- **Are people less likely to respond because their days are structured in a certain way?**

*For example, do they have after-school caring responsibilities, are they shift workers, and so on?
Are particular minoritized groups known to be over/under-represented in these areas?*

- **Are there cultural taboos or discomforts that might apply to question content?**

For example, questions about relationships might be affected by culturally specific gender norms. Questions about mental health might be interpreted in various ways. This could be applicable in administrative settings (such as GP) or in surveys.

- **Was there cross-cultural construct validity?**

Is the interpretation of the meaning and the content of the question suitably similar across languages and cultures to allow for confidence that the questions will be similarly understood by different people? Has this been considered and accounted for by the data gatherer?

If no - assess the topics, questions, language, syntax and semantics of questions asked (both in survey and in administrative settings).

1.4 How are the questions asked (practical, technological)?

Are there potential exclusions based on the means of collecting data?

Was writing ability and literacy accommodated in survey and administrative data?

What IT accessibility and ability was required at the point of data collection?



- **Was a particular confidence/comfort with IT needed to provide data?**

For example, surveys completed online? Data from referrals to specialist services completed online or via an app?

- **Are some minoritized groups likely to be under-represented because of technology accessibility and confidence/comfort with use?**

Consider this question in the context of intersectional characteristics; for example, where age and minoritized identity intersect is there any difference that should be accounted for?

Was particular hardware or software necessary to complete a survey?

Consider whether this is relevant to the dataset. For example, if the data is all online referrals for a specialist service, whilst people might refer via phone and can access the service this way, is the dataset making those people invisible? Are the two groups mutually representative?

- **Was language ability accounted for at all relevant stages?**
- **Where translation was provided, did this produce new points of exclusion?**

For example, accommodating the most common languages might make for a significant difference in the quality of responses between different groups of second language speakers, obscuring some while making others more visible.

- **In administrative data, what information is given about language and translation at the point of collection?**

For example, if working with GP records, do you have data on languages accommodated from each GP surgery whose records are included in the dataset?

- **Was writing ability and literacy accommodated in the data?**

For example, consider if this compromises self-complete surveys or self-referrals for specialist services datasets? Consider if this intersects with particular minoritized identities.

1.5

What do you know about the sampling approach?

What unconscious exclusions might be embedded in the sampling?



For survey data:

- **Who is included in the sampling strategy?**

For example, are only people living in households sampled? How does this effect the representativeness of the sample? Are certain populations more influenced by the sampling approach than others?

- **Were specific sample strategies used to reach hard to reach populations?**

For example, were ethnically diverse areas oversampled to boost ethnic minority groups in the sample?

- **Do population boosts create new bias?**

For example, would a boost of a highly ethnically diverse area mean that ethnic minorities in predominantly white areas are inadvertently missed without accounting for the idea that they might have different experiences?

1.6

What was the environment in which data was gathered?

Were surveys done in personal space (homes / doorsteps?)

What degree of confidentiality was assured?



- **If dealing with sensitive subjects, was the room 'safe'?**
- **Who was present in the room?**

In homes the possibility of being overheard may compromise the disclosure of some types of information. Minoritized populations may be more likely to have another person present during their interview which may affect decisions to disclose, or create pressure to keep issues within a family or cultural community. If using an interpreter, the additional presence and the relationship between the interpreter and the participant might affect disclosure.

- **In on-site settings, was the location and room a comfortable environment?**

Disclosure might be affected by the nature, comfort and privacy of the setting.

- **If information was gathered over the phone, were 'yes', 'no' questions asked to acquire information about who was present and the extent to which the respondent felt comfortable disclosing information?**
- **What was the wording of the mandatory reporting statement (if applicable)?**

In contexts where disclosure is preceded by a mandatory reporting statement the wording of the statement might be impactful. For example, if the statement refers to generalised illegality rather than the specific contexts in which mandatory reporting is required, this might deter participation of people in insecure immigration status.

PART 2



Questions to ask before and while interpreting data.

2.1

What assumptions and stereotypes might be affecting you as the person / people interpreting the data?

Reflect on your conscious and unconscious biases. Do you subscribe – knowingly or not – to any stereotypes?

What implicit or explicit biases might be contained within your analytical or conceptual categories?

Who is leading the research team conducting the analysis?



- **What assumptions generated your hypotheses?**

Biases might be revealed in your response to what the data shows you.

For example, these might apply to your internally held ideas of what victimhood, agency, bodies, resistance, perpetrators etc., look like. This includes both positive and negative assumptions. All should be included.

- **Is the data showing what you expect?**

Reflect on your expectations and on your reactions to un/expected results in relation to assumptions and potential unconscious biases.

- **What is driving your analytical or conceptual categories?**

For example, are categories formed by what is available? What theory suggests? Path dependency? Past studies? Might biases be contained in these formations that are being reproduced in your categories?

- **Is the research team reflective?**

- **Is there any hierarchy in the organisation of roles that might impact results?**

For example, extant hierarchy might impact the ability of people to comment critically on the analysis or analytical process. Even if the team is representative, if it is unequally organised this can still have an impact and should be considered.

2.2

Has your academic discipline or field of research internalised any hierarchies?

Is it possible your analysis will reproduce entrenched disciplinary hierarchies?

In your discipline are some things given uncritical value?



- **What concepts have you adopted uncritically?**

Reflect on your conceptual categories. Are they reproducing hierarchies and are those hierarchies arbitrary? For example, the nation state is accepted as a unit of organising the world, with citizenship as the indicator of belonging. However, if you value the experiences of citizens over the experiences of non-citizens are you arbitrarily reproducing a hierarchy?

- **Are there conceptual or analytical categories you have missed because of disciplinary norms?**

To mitigate the risk of reproducing hierarchies and assumptions held in discipline:

- **Is the research team multi-disciplinary?**
- **Are you getting any input or critical assessment of your analysis from outside of your academic field (either from academics in other fields, or from interested parties with relevant expertise)?**

2.3

What is your positionality in relation to the data?

What is your own relationship with the data or with the topic under investigation?

What is your relationship with the data providers?



- **Do you have a personal connection to the topic studied?**
- **Might any personal connection, whether in the present or past, be affecting your analysis?**

For example, if you are studying an ethnic group with which you have a personal experience of conflict (e.g. a Greek Cypriot academic studying a Turkish diaspora) reflect on whether this might unconsciously affect your analysis of the data?

- **Do you have (strong) feelings or opinions about the organisation who provided the data?**

For example, do you hold a relevant political bias that will affect how you interpret the data provided?

- **Will your analysis be affected by your knowledge regarding the goals of the data collection?**

The goals of data collection may differ from your analytical objectives. Consider if this will impact your analysis.

2.4

What are the potential effects of manipulating data?

Do you need to combine or change categories in order to use your data effectively?



- **Are you subscribing to hierarchies or groups that are imposed by an institution or authority?**
- **Are these groupings accepted by the population from whom the data was collected? Or are other categorization more appropriate?**

For example, you may combine Black African, Black Caribbean, and Mixed heritage as Black British while this is not always an accepted categorization by these groups. Alternatively, it may be more salient to combine Black Caribbean with Latin American groups depending on the exact research question.

It is important to note that detailed categorization of for instance ethnicity, may make intergroup differences not statistically significant as the group sizes may become too small. This may result in overlooking ethnic differences. Hence, careful testing of different categorizations is strongly advised to investigate which groups should and should not be combined.

2.5

Who has the authority to produce meaning on the subject matter?

Will the data tell you everything you need to know about the population to produce an informed analysis?



Reflect on who has the authority to produce meaning on the subject matter.

- **What gaps might there be between the data, your knowledge of the population, and the lived experience of the population?**
- **What strategies have you adopted to address these gaps?**
- **Will you present your hypotheses / research questions / expectations / analyses to members of the population studied for consultation?**

2.6

How can you retain the authenticity of the voice of participants?

Has your interpretation adequately reflected the voice of the population?



Reflect on how you can remain true to the voice of participants even while interpreting this through your own positionality

- **Is there cross-cultural construct validity between your interpretation and the understanding of research participants?**
- **What might be lost in the interpretation?**
- **Does this loss have an impact on the meaning produced?**

PART 3



Questions to ask regarding reporting of analysis

3.1

Could your reporting of the analysis be misinterpreted?

Could your report of your findings be misinterpreted in a way that creates harm?



Consider the wording used.

- **Might your reporting be adopted in a way that confirms social prejudices?**

This may apply to what you choose to report or what you choose to leave unreported.

- **Have you taken steps to ensure mitigation?**

This might involve recognising and debunking social prejudices and stating explicitly the reasons why your results do not confirm them

3.2

Does your reporting increase risk or the likelihood of harm to minoritized populations or marginalised groups?

Is there any way your report could put marginalised groups at risk?

Might your report produce a backlash against minoritized groups?



- **Could your results further disadvantage already-disadvantaged groups?**

For example, cost benefit analyses of by-and-for specialist services that do not account for population size disparities.

- **Who will your findings reach or not reach?**
- **What are the implications of only reaching certain people?**

E.g., Findings that are limited to special interest groups may increase the burden on those groups to problem-solve. If a study of race or ethnicity is only seen by people who are politically disinterested in these issues, it may be dismissed.

- **What strategies can you use to maximise reach?**

3.3

Could your results be misused or politicized?

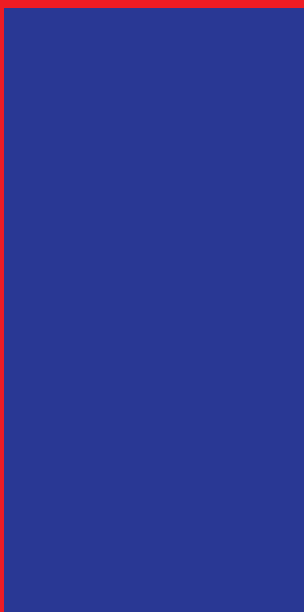
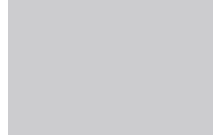
Have you considered mitigation strategies to prevent your results being misused?



Taking the steps indicated above to ensure against misinterpretation is the first necessary stage. However, recognise that both misinterpretation and wilful misuse can happen.

Consider:

- **Does your research prioritise the dignity, safety and well-being of participants, partners, researchers and any implicated minoritized groups?**
- **Do you clearly communicate any policy-relevant findings?**



The VISION research is supported by the UK Prevention Research Partnership (Violence, Health and Society; MR-VO49879/1), a Consortium funded by the British Heart Foundation, Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Health and Social Care Research and Development Division (Welsh Government), Medical Research Council, National Institute for Health and Care Research, Natural Environment Research Council, Public Health Agency (Northern Ireland), The Health Foundation, and Wellcome.

The views expressed are those of the researchers and not necessarily those of the UK Prevention Research Partnership or any other funder.

