

## **RSS RESPONSE TO THE RACE DISPARITY UNIT'S CONSULTATION ON STANDARDS FOR ETHNICITY DATA**

30 August 2022

### **1. Introduction**

This is the RSS's response to the [Race Disparity Unit's consultation](#) on its draft [Standards for Ethnicity Data](#). These are proposed standards that would apply to the public sector for collecting, analysing and reporting ethnicity data.

We are broadly very supportive of the Standards and think that they could be a real help for public sector analysts in improving our understanding of disparities between ethnic groups. In this document we make a number of recommendations for further strengthening the Standards. The main changes that we recommend are:

- The inclusion of a clear process for involving professional statisticians on technical and challenging areas and highlighting the areas where users of the Standards might especially want to consider consulting a statistician.
- The current draft recommends aggregating to five ethnic groups (white, black, Asian, mixed, other) where there is not enough data to use the full harmonised list of ethnicities. We argue that the Standards should instead recommend aggregating to seven groups in this situation: white British, black, Asian, mixed, white other, other, not specified). First, given that white other now represents around ten percent of the population we think it is worth separating the white category into white British and white other – especially as, in most cases, white British is the appropriate category to use for comparison. Second, the most appropriate way to count people who do not wish to identify themselves as a particular ethnicity is to record this – and adding a not specified category would allow this.
- Greater emphasis on the importance and relevance of collecting data on religion – as some groups (eg Sikhs and Jews) can be missed without collecting this information.

### **2. Should the government publish the current draft of the Standards for Ethnicity Data?**

The RSS welcomes the draft Standards for Ethnicity Data – we think that the current draft is good but that it can and should be strengthened further before it is published.

First, it is important to recognise that the Race Disparity Audit in 2017 was an excellent piece of work that successfully gathered together much of what we knew about ethnicity. The [Ethnicity Facts & Figures website](#), which this work led to, is a really valuable resource. While – as we outline in this section – there is more that can be done, it is important to first recognise the good work that has been done in this area to date.

As a general point, it is critical that the voices of ethnic minority groups are taken into account when decisions are made about what questions about ethnicity to research, how the data should be collected and how the results should be presented. This is touched on in the draft, but we would suggest that whenever the Standards state that "stakeholders or users" of the data need to be consulted, it should be made explicit that this includes consultation with the communities affected by the topics considered.

Below are some specific changes that we would like to see made to the draft standards before they are published.

#### Who the standards are for (1.3)

This section is unclear on the extent to which the intent is for the private sector to follow the public sector in adopting these standards – and particularly the harmonised categories for ethnicities. While the Standards are not directly written for private sector organisations, it should be recognised that these organisations are likely to be

influenced by them. It would be helpful for the document to include some reflection on how government would like private sector organisations to engage with the Standards.

There are differences in the purpose that public and private sector organisations collect and use data for, and we do not think it is straightforwardly beneficial for private sector to adhere to public sector categories for ethnicity in all situations. In some circumstances – eg, if the anticipation is that the data being collected will be made public in form – it would be useful to use the public sector categorisations to enable comparisons. However, if the purpose is for solely internal analysis private sector organisations could be encouraged to enable people to say how they perceive their ethnicity. We also note that comparisons between organisations are more important for the public sector than for the private sector and care should be taken to ensure that the public sector desire for commonality does not prevent the private sector from undertaking their own approaches to collecting and analysing ethnicity data. This would be potentially valuable to the public sector as it may provide a contrasting bottom-up approach to categorisation, rather than the top-down approach for the public sector. Given the acknowledgement in section 1.4 that ethnicity is self-defined and subjective, we would like the Standards to encourage both a top-down approach (from the public sector) and bottom-up up approach (from the private sector).

There is also a difference between large and small organisations. For large organisations, allowing each person to self-identify may make the data impossible to reasonably analyse, and it would be better for the organisation to follow these standards. For some, eg very specialised organisations, there may be a compelling reason to consider a more granular or individual view of ethnicity that reflects salient ethnic differences for that organisation that are not included in the standard. It might be helpful for the Standards to also encourage private sector organisations, like public sector, to consider why they are collecting the data, and if those reasons are best served by an approach that permits benchmarking or by an approach that permits self-definition or a different choice of categories.

### Data Collection (2.1)

This section starts by setting out some key principles – before data is collected it is important to think about how it will be used and to collect feedback from users. However, we would question whether the section on using GSS harmonised standards might potentially undermine that message and lead to confusion. Harmonisation should not be a goal in and of itself – it should only be sought when the goal is to compare across populations. If that is not the objective, then seeking harmonisation might mean that suboptimal data (from the perspective of meeting user needs) is collected.

The linked-to RDU blog on harmonisation states that the key is to ensure that categories can be mapped to harmonised categories rather than harmonising outputs. This approach seems right, however it could result in users having to put a large number of categories in their survey to enable mapping. One way to avoid this issue would be to encourage two questions on ethnicity – ie, ask people both for their ethnicity using harmonised categories followed by a second question using their preferred categories. This would also allow an analysis of how much the preferred categories overlap with harmonised. We support the general approach set out in the section on self-reporting – though we are concerned that there are some mixed messages here and an explicit reference to this two-step approach could be helpful.

We would like to see stronger guidance about collecting data on religion. The current wording is "You should include national identity and religion questions to improve the acceptability of the ethnicity question". However there are significant groups in Britain (for example Sikhs and Jews), who cannot be identified within the current ethnicity classification. Only by collecting both ethnicity and religion can they be identified and statistics compiled for them. We would welcome a stronger guidance on this.

In the section on the representativeness of ethnic groups, we would like to see a reference to how people using the guide could get help from an appropriate professional statistician if they need support.

There are a group of people who may not choose to answer ethnicity questions – this section would benefit from including some guidance on how these people should be treated.

Finally, a small point, this is a long and important section that people are likely to want to refer to – this would be made easier with subsection numbering.

### Data Analysis (2.2)

In the discussion of weighting, it should be made clear that this can only be done if external data of good quality is available so that it is possible to determine the weights. Typically the Census will provide this, but that won't necessarily be possible in all cases. This is another area where it might be helpful to include some information on how users of the Standards can receive assistance from a professional statistician.

In the section on harmonised categories, the big five aggregated groups (white, black, Asian, mixed, other) are presented as though it is the default. It is important to question this grouping and, at least, provide an explanation of why these five groups should be used in the absence of reliable data for the full harmonised set of classifications. This is especially important given the point that is, correctly, made in the section about comparators: in many cases the appropriate comparator will be white British and not white as whole. Indeed the grouping of white other (taking in Gypsy, Roma and Irish Traveller, white European etc) is now around ten percent of the population. This is a large enough grouping that we think it is worth splitting out white British and white other into two groups, which would make it easier to use white British as a comparator in a wider range of circumstances. In our comments on data collection, we suggest that the Standards consider how to handle those who do not wish to state an ethnicity. Our view is that it is important to record the number of people choosing not to say, and that this should be always be included in analysis. So, we would suggest – rather than the five groups identified – that seven groups (white British, black, Asian, mixed, white other, other, not specified) be used when the full harmonised classifications are not suitable.

The point made about the geographic clustering of some ethnic groups leading to the potential for counterintuitive results is important. This is especially important given that 40% of ethnic minority people in the UK live in London whereas (based on the 2011 Census data) only around 9% of the white population do (and 7% of the white British population do). Given some of the ways in which the capital is an outlier in the UK – eg, around salaries – this is a vitally important aspect of data analysis to be aware of. We would suggest adding a more full explanation and some additional examples to help emphasise this point.

The section on how to handle confounders – whether differences between ethnic groups should be measured by analysis of raw data or after adjustment – is also important and sets out the principles well. This is another area where we think it helpful to set out a process for receiving advice from a professional statistician.

### Data Reporting (2.3)

The Standards correctly emphasises the importance of ethnicity data and states that reporting it should be a priority. We would suggest adding a related point on reporting data on religion. There are sources, such as the Crime Survey of England and Wales, where religion is collected, but not always analysed and reported, due to relatively small sample numbers. In this case pooling together three years of data can enable further reliable insights to be produced, showing, for example, differences of experience of crime and fear of crime among Sikhs and Jews. This type of analysis should be carried out routinely, not just under special requests. Under the ethnicity classification, Sikhs are unidentifiable within the much larger Indian category, and Jews within the white category.

The only other change we would recommend in the section on data reporting relates to the section on measures of reliability. This section is very statistical – if the people using this guide are not statisticians this is another area where we think it would be worthwhile recommending that users consider consulting a professional statistician. Sometimes advice is needed and this is a good opportunity to point people to it.

We should also recognise that there is a tension between a person's ethnic identity and society's perception of their identity – differences in certain social measures between different ethnicities are not due to the effect of race, but to the effect of the experience of racism. This is important to consider when reporting outcome differences by ethnicity – we would recommend that statistical summaries of observational data should always include cautions about misinterpretation of correlation as causality. It might also be useful when reporting to think about what “dark data” may have been missed and highlight this if possible. Finally, in terms of protecting confidentiality, it may be worthwhile for the Standards to suggest the use of differential privacy methods when making data available, or even perhaps in survey design when the topic is sensitive.

### Trustworthiness – data collection, reporting and analysis (3.1)

This section is important in explaining why the actions recommended in the previous section are important from the perspective of demonstrating trustworthiness and that is welcome. In particular the points about GDPR are essential.

We mention in our comments on section 2.1 that it is important to set out how we should treat people who choose not to declare their ethnicity. How this is treated has bearing on the trustworthiness of the data collection and analysis and so we think that this should also be discussed here.

The section on regular review is important. We would recommend strengthening this section further by noting that the Inclusive Data Task Force's report states that classifications used should be reviewed regularly (every five years) to reflect societal changes.

### **3. Monitoring the use and impact of the Standards for Ethnicity Data?**

The Standards propose a role for the Office for Statistics Regulation (OSR) in helping to monitor the use and impact of the Standards. We support the idea of involving OSR in this way. We are, in fact, keen that the OSR understands its remit to be broader than this: the private sector can contribute to this debate and we would like to see the OSR flagging good practice in the private sector so that it can be incorporated into practice in the public sector.

### **4. The development of similar standards for other groups, eg disabled people**

We would strongly support the development of similar standards for other groups and would be happy to help however we can. It is important that, if these standards are developed, the affected groups are fully engaged in the development of data standards.