Measuring caste discrimination in Britain – a feasibility study

Nicholas Howat, Alice Ryley, Joel Williams, Rosie Mcleod, Amy Busby (TNS BMRB)

Hilary Metcalf (NIESR)

This research was commissioned by the Government Equalities Office (GEO). The findings and recommendations are those of the authors and do not represent the views of the GEO or government policy. While the GEO has made every effort to ensure the information in this document is accurate, the GEO does not guarantee the accuracy, completeness or usefulness of that information.

**Contents**

Executive Summary 5

Ethical considerations 5

Measuring caste discrimination 5

Survey design options 6

Conclusions 7

1. Introduction 9

1.1. Report structure 9

2. Method 10

2.1. Stakeholder interviews 10

2.2. Interviews with target respondents 10

2.3. Methodological review 11

2.4. Limitations of this research 11

3. Ethical considerations 12

3.1. Protection of individuals from undue harm 12

3.1.1. Respondents upset from identifying own caste 12

3.1.2. Concern for revelation of caste leading to anxiety 13

3.1.3. Victims of caste-based discrimination may feel distress at reporting this 14

3.2. Social benefits and harm 14

3.2.1. Social benefits 14

3.2.2. Increasing caste awareness/legitimacy 14

3.2.3. Targeting populations 15

3.3. Institutionalising caste 16

4. Measuring caste discrimination 17

4.2.1. Ability to recognise discrimination 20

4.2.2. Identification of relevant caste discrimination 20

4.2.3. Willingness to report caste discrimination in a survey 21

4.2.4. Summary 22

4.3.1. The interviewers 23

4.3.2. The interview mode 24

4.3.3. Language 24

4.3.4. Easing into caste discrimination questions 24

5. Survey design options 25

5.1.1. General sample options 25

5.2. Bespoke national survey 26

5.2.1. Assumptions and survey design decisions 26

5.2.2. Sample design 27

5.2.3. Survey development 34

5.2.4. Cost 35

5.3. Existing survey samples 35

5.4. Survey challenges 36

5.4.1. Response rates 36

5.4.2. Identifying a target population 36

5.4.3. Timing 37

6. Summary and conclusions 38

6.1. Summary 38

6.2. Conclusion 38

Appendix 1 – Interviews with target respondents 40

Appendix 2 –Target respondent topic guide 42

Appendix 3 – Key Stakeholder topic guide 45

# Executive Summary

**Background:**

Provisions within the Enterprise and Regulatory Reform (ERR) Act 2013 made it a duty to make caste an aspect of race under the Equality Act 2010. The provisions include the possibility of a review after five years. At present there is very **little evidence as to the existence or otherwise of caste discrimination in Britain** that may be captured by discrimination law. As a first step towards addressing this, the Government Equalities Office (GEO) commissioned this research to look at the **feasibility** of conducting a national survey to quantify the extent of caste discrimination. In particular this feasibility study looked at the ethical considerations in conducting a national survey, the extent to which caste discrimination could be reliably identified in a survey and how much a national survey could cost.

**Method:**

The research involved interviews with key stakeholders, academics and target respondents.

**Key findings:**

## Ethical considerations

* *Preventing undue harm*- Concerns from stakeholder groups and target respondents that people could potentially be *upset* at having to identify their caste as part of a survey. However, a survey of caste discrimination could just measure experiences of discrimination without specifically identifying the respondents’ caste. Furthermore, social researchers in Britain could deploy tested measures to minimise any harm, including specific training for interviewers surveying sensitive subjects.
* *Avoiding the legitimisation of caste*- Concern were raised both by some stakeholders and target respondents that a survey of caste discrimination would by its existence *lend legitimacy* to the concept of caste. These concerns were informed by a belief that caste was dying out, or was not relevant in a British context. Other stakeholders disagreed.
* *Sensitively targeting respondents* - Concern was raised that any survey of caste discrimination which targeted respondents solely from a specific ethnic background could make people from that background *feel that they were being singled out* and *treated differently*. This would need to be addressed if a national survey were to be conducted among a specific sub-group of the population with a detailed rationale given for any specific targeting.

## Measuring caste discrimination

* *Understanding caste* - The interviews with respondents, who we would expect to be targeted in a caste discrimination survey, indicated that regardless of whether they identified with the caste system or not, there was a strong understanding about caste and how it worked across Hindu and Sikh respondents. Respondents were often able to identify their own caste or at least how they would be perceived by those who did observe castes if they did not. Misunderstanding of the concept of caste would not be a barrier to a national survey being conducted.
* *Gaining reliable responses* - One challenge identified by respondents who believed caste discrimination existed was that victims may be reluctant to report their experiences. It was felt that this could be seen as de-facto identification of their caste and also that they may not wish to discuss potentially painful experiences. There was a general consensus that introductory questions which could cover other forms of discrimination should be asked before addressing the concept of caste and associated discrimination, both to ease respondents into sensitive topics and to gain an understanding of the challenges surrounding honesty.

## Survey design options

* *Random sampling*- For any national survey to be considered robust a random sampling method would need to be used as it will have the most credibility among various stakeholders. Non-random sampling methods usually cost less to implement but the accuracy of the resulting estimates rests upon assumptions that may be untenable – or at least may be regarded as untenable.
* *Face-to-face interviewing* - For a survey on caste discrimination, given the topic is sensitive, using a self-completion mode would be preferable. However, it is likely that some of the respondents may not read English well and require translation or the presence of an interpreter. The objective of inclusivity would therefore favour an interview mode. A secondary advantage of the interview mode is that self-completion sections can be inserted into the data collection process to handle the most sensitive areas, at least for those with sufficient language skills. As such we would recommend in person interviewing.
* *Combining a tailored sample with household screening*- If the national survey is to target particular subgroups, and throughout the feasibility study we assumed that it would focus on respondents from a South Asian background, household screening would be required to identify eligible respondents. However household screening for sub-populations can be particularly resource intensive and increase costs significantly. In order to minimise this it is possible to tailor a sample design so that it is disproportionately focussed on areas with a high prevalence of the sub-group of interest, although this runs the risk of under-representing respondents who live in areas where there are not as many people of the target background.
* *Placing an additional module on an existing survey* -One alternative to a standalone national survey would be to place specific items on caste discrimination on an existing national survey. This would significantly reduce the costs of any research but a number of conditions would need to be met for this to be viable. Any existing survey will need to include:
	+ a sufficient number of respondents from a South Asian background
	+ questions that have some relevance to the caste discrimination.
* Possible options include the *Understanding Society* survey or the *Labour Force Survey* which has a large focus on employment and items around employment discrimination would therefore not necessarily be problematic. Understanding Society would seem to be the most appropriate survey that items on caste discrimination could be placed.

## Conclusions:

1. There appears to be no significant ethical or methodological barriers that could not be overcome to allow a survey measuring the extent of caste discrimination to be carried out.
2. Respondent distress can be kept to a minimum if the survey is restricted to measuring caste discrimination rather than attempting to identify respondents’ caste.
3. Detailed contextual information should be provided on why the items are being asked. Interviews with respondents suggested that respondents displayed an understanding of the concept of caste even if they were vague on the details – particularly among second and third generation respondents.
4. A standalone survey would require a significant amount of financial resource to conduct with a robust sample that could provide reliable estimates on the extent of caste discrimination and change over time. Alternatively, placing items measuring caste discrimination on an existing national survey (eg. Understanding Society) could cause problems in terms of practicalities and timings. Financial and practical considerations would have to be addressed when making a decision as to whether a survey measuring caste discrimination should be conducted.

# **Introduction**

The extent and nature of caste discrimination in Britain is something that has been contested since the prospect of a new Equality Act was announced under the last government. Representations were made to include caste under the Act but these were rejected on the grounds that there was no strong evidence of caste discrimination existing in Britain. Instead when the Equality Act was passed in 2010 section 9(5)a allowed ministerial discretion to enact legislation on caste as an aspect of the protected characteristic of race. A subsequent amendment to the Enterprise and Regulatory Reform (ERR) Act 2013 changed the discretionary power making it a duty to make caste an aspect of race for the purposes of the Equality Act 2010 (Dhanda et al 2014). This amendment included within it the possibility of review after five years to allow the government to repeal or amend the legislation.

As a first step to improving the evidence base, the Government Equalities Office (GEO) committed to conducting a feasibility study in to the possibility of measuring caste discrimination in Britain. The feasibility study was intended to look at the methodological challenges that any research study would face, the ethical considerations that needed to be addressed and the potential cost of any research. A consortium of TNS BMRB and the National Institute of Economic and Social Research (NIESR) was commissioned to conduct this feasibility study using their experience of in conducting similar research studies.

## Report structure

The report is structured thematically incorporating content from the different study strands in each section. **Section 2** presents on overview of the method; **Section 3** of the report looks at the ethical considerations around any potential study including the possibility of any research causing harm.

**Section 4** looks at issues relating to how potential respondents may define caste and concerns around how willing respondents may be to discuss these issues. The final section (**Section 5**) looks at the possible survey design options for any research including conducting a stand-alone survey and including items on other existing surveys.

#  **Method**

This section sets out the methodological approach to the feasibility study. The research included three main strands:

* interviews with representatives of stakeholder groups and academics who have studied discrimination or issues relating to caste identity,
* interviews with respondents to test how questions about caste would be understood
* a methodological review of potential data collection strategies and sample design options.

In Britain, concerns around caste discrimination have tended to revolve around communities from a South Asian background. As a result of this, the feasibility study focussed on a survey covering respondents from a South Asian background as this was the most likely form subsequent research would take.

## Stakeholder interviews

The interviews took place with two key types of stakeholders: stakeholder organisations and academics. Stakeholder organisations provided divergent views on caste discrimination legislation and academics made suggestions regarding practical and ethical issues of conducting such a study. As the nature and scale of caste discrimination is contested, stakeholders with a range of views were sought.

Stakeholders from organisations representing people with roots in South Asia were interviewed following research suggesting caste awareness in Britain is concentrated among this population (Metcalf and Rolfe 2010, Dhanda et al 2014a and 2014b). They were asked for their views on: the ways and willingness of South Asian people to talk about caste, associated discrimination, and best practice in doing this; who should participate in a survey which aimed to measure caste discrimination, including likely response rate and reliability of responses; risks and effects of such a survey. A total of 14 organisations were contacted for the research and 7 agreed to be interviewed. The organisations were selected to cover a range of views on caste discrimination legislation and to represent different religious beliefs (including Hindu, Sikh and secular organisations). Interviews were conducted face-to-face and over the telephone with representatives from the groups in September 2014. Discussions were recorded where participants gave consent, and the data was analysed by comparing and synthesising responses.

In parallel to the engagement with stakeholders, interviews were also conducted with key academics to better understand the ethical and practical issues that might be involved with any research in to the extent of caste discrimination. These interviews covered the benefits of conducting this research as well as the harm that could felt by individuals the perceived risks to the wider community. They also looked at steps that could be taken to reduce any potential harm and maximise the effectiveness of the research.

## Interviews with target respondents

In order to better understand some of the practical issues around measuring caste discrimination, qualitative face-to-face interviews were conducted with respondents from a South Asian background. The interviews were conducted with a mixture of first, second and third generation respondents from Indian, Pakistani and Bangladeshi backgrounds. Respondents were also sampled to include a mixture of ages and socio-economic backgrounds. In total, 25 interviews were conducted in October 2014 at venues in Birmingham and London. A full breakdown of the sample is included in Appendix 1. As the research was intended to assess whether conducting a national study of caste discrimination was feasible, rather than exploring its nature, it was felt that a total of 25 interviews would be sufficient.

The interviews covered respondents’ understanding of caste and the terminology they understood, willingness to discuss caste and discrimination and the most appropriate mode for conducting a survey on these issues. Views on the extent of caste discrimination in Britain today as it pertains to employment and the provision of services were also explored.

The report contains a number of direct quotes from target respondents along with their broad demographic details.

## Methodological review

The methodological review looked at the potential strengths and weaknesses of different data collection strategies that could be employed in a survey to measure caste discrimination in Britain. This was done by looking at existing national surveys (such as USoc) and how these could potentially be drawn upon to support a survey to measure caste discrimination in Britain. Census data was also used to explore the feasibility of accessing a representative target population (i.e. people with a South Asian heritage) to form a stand-alone survey. This review included an assessment of the potential biases that could be introduced by different recruitment strategies, the impact that the mode of data collection could have on response and sample design options which maximise coverage in the most resource efficient manner. Finally the review also looked at the potential costs of conducting a stand-alone survey and the issues that placing items on an existing national survey would raise.

## Limitations of this research

This review set out to examine the feasibility of conducting a national survey to measure the extent of caste discrimination. It did not seek to produce evidence as to whether caste discrimination exists or how widespread it might be. Although interviews were conducted with a number of potential target respondents, they were not sampled in a systematic random manner which would mean that their experiences could be taken to be representative. In addition to this, while the study did look at the extent to which caste can be investigated as a concept, it did not seek to finalise question wording and, if a study were to be conducted in the future, further development work would be necessary.

#  **Ethical considerations**

In consideration of ethical issues associated with measuring caste discrimination, we were guided by the RESPECT guidelines, drawn up as a Code of Practice for the European Commission[[1]](#footnote-2). This has three main principles:

1. upholding scientific standards;
2. compliance with the law; and
3. avoidance of social and personal harm.

Of particular relevance to a survey of caste discrimination in Britain is the third principle (avoidance of social and personal harm) and, specifically,

* 1. protecting research participants from “undue intrusion, distress, indignity, physical discomfort, personal embarrassment or psychological or other harm”; and
	2. ensuring that the “results should benefit society, either directly or by improving human knowledge and understanding” and that researchers “should aim to avoid or minimise social harm to groups and individuals”, i.e. whether any possible harm to groups or wider society would be outweighed by benefits.

The following considers these two issues in turn. Our assessment draws on previous research into caste and caste discrimination (particularly, Metcalf and Rolfe, 2010; Dhanda et al*.*, 2014a and 2014b) and interviews with key stakeholders and academics.

## Protection of individuals from undue harm

### *Respondents upset from identifying own caste*

One potential source of harm from a study of caste discrimination is that the act of asking about caste itself causes upset for some respondents. This could arise where respondents view their caste as shameful and may also arise for those who reject the notion of caste. This was something that was brought up by interviews with some of the target respondents:

*“I suppose if you’re a lower caste you don't want to sort of highlight it, would you? You just want to get on with it.” (Female, 2nd generation Indian, Hindu respondent in the 46-55 age group)*

*“I think it is derogatory and it’s of a time, my parents left Indian 50 years ago, totally irrelevant, you know the caste, what you’re ticking should reflect what you are now, not your history, you know your history could reflect … you could have had ten different jobs, why pick the one that’s the worst? Do you know what I mean? It doesn’t make any sense. And if you do put that down, that’s almost, I think discriminatory itself because caste and social class is technically illegal isn’t it? I think it’s ethically illegal so I don’t think that should be recorded at all, yes.” (Male, 2nd generation Indian, Hindu respondent in the 36-45 age group)*

The first point to make is that a survey of caste discrimination **would not necessarily have to ask what caste respondents are** or are perceived to be by others. The extent of, and changes in, perceived discrimination can be identified without this information.

Target respondents who were concerned about the negative impact from asking individuals their caste felt that simply asking about experiences of discrimination based on caste would not have the same problems.

*“Discrimination yes, I would definitely ask, because discrimination by itself is almost empowering someone to say; do you know what, look, whatever your individual characteristic, caste, religion, whatever, you are given the opportunity to have a voice and say something about it, so yes definitely, I’d ask about discrimination.” (Male, 2nd generation Indian, Hindu respondent in the 36-45 age group)*

However, knowing an individual’s caste would provide more information on the nature of discrimination (e.g. amongst whom it is more prevalent) and so may help develop approaches to address it. Some of the academic experts consulted suggested that similarly useful information could be gathered by asking respondents about the caste that others might attribute to them or asking about their ‘caste/ background’ (e.g. that attributed to their ancestors), as neither requires respondents to see themselves as having a caste. Given the number of castes such a question would need to allow verbatim reporting.

Furthermore, we are not entirely convinced that this approach would cause much less undue distress. We are also sceptical whether it would do more than identify that most perceived caste discrimination is against people of low caste and so not assist us to understand caste discrimination further. Of more use, and without the ethical difficulties for the individual, would be to ask their religion or that of the perpetrators of the caste discrimination. This would better add to our understanding of caste discrimination and so may help develop approaches to address it.

### *Concern for revelation of caste leading to anxiety*

Whilst confidentiality and anonymity is an ethical requirement of social surveys, the possible damage which an individual might suffer through their (attributed) caste being made known to others with whom they associate is great. Concern over the survey leading to their caste (or attributed caste) being revealed to associates could cause significant anxiety for respondents therefore representing undue harm. To address this any study would need to reinforce confidentiality measures. It will be important that respondents are made aware of the seriousness with which confidentiality has been taken. Providing appropriate reassurance of this to respondents should reduce possible distress and concerns.

It should be noted that providing reassurances about confidentiality and the security of data is something that a number of social research studies face and procedures have proven to be effective. These studies have covered sensitive subjects including racial and religious discrimination and experiences of crime.

In the case of caste, the extent to which a respondent feels distressed by identifying their caste/jati/biradari (or that attributed to them) may be affected by the interviewer. Ideally, respondents and interviewers would be caste/jati/biradari matched. However, this is impossible (as the respondents’ caste/jati/biradari (or that attributed to them) is unknown).

To minimise any distress caused by respondent/interviewer interaction, interviewers should be fully briefed on the sensitivity of caste/jati/biradari identity, on how to ensure that respondents feels comfortable to discuss caste/jati/biradari with them and to avoid any possible indication that the interviewer themselves might be making any assumptions about the interviewee’s caste.

If respondents are not going to be asked about their caste/jati/biradari (or that which might be attributed to them), to reduce distress, it would be useful that they are reassured that this is the case at an early stage in the interview.

Although caste/jati/biradari discrimination need not be restricted to discrimination against people of low (attributed) caste/jati/biradari, those of low caste may believe that reporting perceived caste/jati/biradari discrimination against themselves indicates their low caste/jati/biradari status. Therefore, the same sensitivities may apply to this as to asking about (attributed) caste/jati/biradari and so careful reassurance would also be required to minimise distress.

### *Victims of caste-based discrimination may feel distress at reporting this*

A survey of caste discrimination cannot be guaranteed not to cause distress to some respondents, but it should seek to minimise this. Other than through reluctance to provide information on one’s (attributed) caste, the degree of distress and measures to address this would be similar to that for other victim surveys. These include surveys that collect information on very personal and upsetting forms of victimisations including domestic abuse and child maltreatment.

Debriefing procedures would need to be drawn up in the light of the level of sensitivity (both to caste and to revealing discrimination) so those upset have recourse to assistance. Models for sensitive debriefing exist for other sensitive surveys. For example, in surveys which cover topics of a sensitive nature respondents are provided with contact details for relevant services (such as helplines) as a precautionary measure in case the interview elicited anything that they wanted to discuss further. This practice is done on the Crime Survey for England and Wales (CSEW) and works well.

## Social benefits and harm

### *Social benefits*

The social benefits of a survey on caste discrimination are similar to those for race (and other groups covered under the Equality Act 2010): having more information on the extent and nature of caste discrimination will better inform policy to tackle such discrimination. The survey, depending how it is constructed, could identify not only the **extent** of perceived (ie. self-reported) caste discrimination, but also **where** (e.g. location), **amongst which groups** (e.g. by religion) and **how** (e.g. employment, goods and services) this is most likely to occur. If the respondents’ caste (or attributed caste) were identified in the survey, this would also lead to a better understanding of discrimination against different castes throughout the caste hierarchy.

In addition to this, a survey would provide a quantitative measure of the extent of caste discrimination within Britain. This could then be repeated at later intervals to identify the extent to which such discrimination is increasing or decreasing over time. This in turn could inform the level of resource which the government needs to deploy to address this issue and whether it is proving to be effective.

### *Increasing caste awareness/legitimacy*

A number of issues have been raised by stakeholders and target respondents as being potentially harmful at a wider social level in respect of conducting a study to measure caste discrimination. These hinged on it raising awareness of caste and this being damaging. In particular, there was concern that, by increasing caste identification, research would lead to divisions within some ethnic minority communities and, possibly, to an increase in caste discrimination.

These arguments were also informed by a belief that caste awareness was dying out and should be allowed to so do. One respondent from the target population described how she first came across the concept of caste when meeting new people at university. When she then raised it with her parents they told her that they had chosen not to engage with what they saw as an ‘outdated concept’. This view was echoed by a number of respondents in the research who said that children and young people shouldn’t have the subject of caste raised in a survey which they would then go home to ask their parents about.

*“If you asked somebody- If you ask me, I would be so angry, because caste, I don't think that even exists.” (Male, 1st generation Indian, Hindu respondent in 55 and over age group)*

It should be said that the assumption that caste awareness is declining is not universally accepted. A number of respondents made quite forceful arguments that the concept still played a part in Britain today.

*“I think you should ask the question [about caste discrimination], definitely it’s there. People are influenced by Caste. I think the majority of us are hypocrites by saying Caste doesn’t matter and when you ask them direct you know, this is what I do. If your son or daughter was to bring a lower Caste what would your reaction be and then they really have to be quite honest as well.” (Female, 1st generation Indian, Hindu respondent in the 46-55 age group)*

Ultimately, the concern about the possibility of a survey of caste discrimination increasing its awareness, or granting it legitimacy, has to be balanced against the need to measure the extent of the problem. The lack of any reliable data on the extent of caste discrimination does of course mean judgements of this kind are very difficult to make. If it was felt that if the need to establish prevalence measures was more pressing then steps could be taken to ensure that the survey items themselves were asked in a sensitive form which did not assume that respondents had a caste simply because of their background and allowed them to reject the concept as not being applicable to them.

### *Targeting populations*

There was strong opposition among anti-legislation stakeholder organisations and some target respondents, about the potential focus of the research on people from a south Asian background (Indian/Pakistani/Bangladeshi) and the exclusion of other ethnic groups. In particular, any focus on people from a South Asian background was seen to confer blame, condemnation and insult to the ‘Hindu community’ by implicating the population in discriminatory practice. Asking questions about something that people think is ‘left behind’ is expected to cause annoyance, but also harm. Some representatives from stakeholder groups expect that the majority of respondents would consider this ‘imposition’ a ‘retrograde step’ in the progress towards ending caste awareness.

However, there is concern among pro-legislation groups that this is an excuse being used by the anti-legislation groups and that this could then hide discrimination which is taking place (the Anti Caste Discrimination Alliance refer to it as a ‘Hidden Apartheid’[[2]](#footnote-3)).

There was a suggestion that ‘descent’ would be a preferable term on the basis that it encompasses caste and can be distinguished from race, but would speak to a wider range of ethnic origins. It was felt that this could include other populations from Africa and the Middle East, such as Somali, Yemeni and Afghani and other ‘Arabs’ believed to use descent-based social indicators as a basis for discrimination in some cases. Pro-legislation stakeholders also supported the idea that the survey should ask about discrimination based on ‘descent’ because they incorrectly believe that this reflects the wording of the Equality Act 2010, and that people from Japanese and West African communities could/should be included.

Some target respondents also mentioned some concerns about focussing on respondents from a South Asian background. This was particularly the case for 2nd and 3rd generation respondents who felt that they were British and so felt that being targeted for these measures would serve to marginalise them and treat them as being somehow “different” from the rest of society.

*“If you are going to just ask South Asian [people], that means you are ‘outcasting’ them from the rest of the society, because why were you just getting asked? Automatically this question will be raised” (Female, 2nd generation Indian, Hindu, age 26-35)*

The concept of ‘Britishness’ was brought up with regard to asking about caste. In particular, a male respondent in the 55 and over age category who had been recruited as a Hindu, when interviewed said that he rejected any identity that he may have had when he lived in India (e.g. caste or religion) and that to try and apply caste to him in the UK by asking about it insulted the efforts he had made to assimilate as a British man. It was “otherising” him to ask about his caste (or associated discrimination) as if he practiced different systems to a white British man, even if it was in the context of trying to prevent discrimination. By “otherising” he meant that to distinguish between him and a white British man would create an unequal ‘them’ and ‘us’ division which supported discrimination against him as a non-white British man. This is because he would belong to the inferior ‘them’ category, or ‘the other’.

However, target respondents felt that this issue could be allayed somewhat if the survey provided contextual information explaining why these items were being asked. This contextual information would need to explain the purpose of the survey in reference to the legislation and the need for the government to assess the extent of discrimination in Britain.

## Institutionalising caste

The risk of ‘institutionalising’ caste through surveys was identified as a risk by those stakeholders both for and against caste legislation. The concern was that it could recreate and normalise terms and practices that are unwanted – and that among some stakeholders are considered irrelevant. Concern among those opposed to legislation was that a survey discussing caste discrimination was ‘hypothesis-driven’ survey, i.e. suggesting discrimination may exist. To this extent, the exercise itself could create a new rift between people by ‘reawakening’ the concept in their minds.

The concern about ‘institutionalising’ caste was also something that was mentioned by some target respondents. In particular there was a concern that caste should not become a standard measure that was recorded on forms alongside other demographic information that is routinely collected such as ethnicity, gender and religion.

*“Yes I think it’s quite, it’s really old-fashioned, people have moved on from that, they need to move on you know. Because it’s almost like saying, if you record something, you’re asking for it, then it exists” (Male, 2nd generation Indian, Hindu respondent in the 36-45 age group)*

This latter concern in particular is one that could be relevant in relation to the actual legislation, rather than a survey to measure caste discrimination itself, so would not necessarily preclude conducting research on the issue.

#  **Measuring caste discrimination**

This section examines some central issues affecting the feasibility of conducting a survey of caste discrimination relating to target respondents (i.e. people with a South Asian heritage): their understanding of ‘caste’; their identification of what caste discrimination might be; and, data collection approaches.

Throughout, we are concerned about the feasibility of measuring caste discrimination as defined under the Equality Act, i.e. using the definition of caste relevant to the Equality Act and, in respect of discrimination, limiting consideration to those activities covered by the Act (see Section 4.1 and 4.2 below for definitions).

The chapter is based on NIESR’s previous research into caste in Britain, recent EHRC research reports (Dhanda *et al.* 2014a, 2014b) and interviews with target respondents, stakeholders and academics.

* 1. **Understanding of ‘caste’**

Regardless of whether they identified with the caste system or not, there was a strong understanding across Hindu and Sikh respondents about caste and how it worked. Respondents were often able to identify their own caste or at least how they would be perceived by those who did observe castes even if they themselves did not. How confidently respondents were able to explain about caste varied according to age and migrant generation status (and therefore their likely direct experience of caste); older and first generation respondents tended to know more.

In order to conduct a survey of caste discrimination, it is important that questions in the survey result in respondents using a consistent definition of ‘caste’ and that this is the definition relevant to the Equality Act. The definition of caste given in the Explanatory Notes to the Equality Act 2010 is as follows:

*The term “caste” denotes a hereditary, endogamous (marrying within the group) community associated with a traditional occupation and ranked accordingly on a perceived scale of ritual purity. It is generally (but not exclusively) associated with South Asia, particularly India, and its diaspora. It can encompass the four classes (varnas[[3]](#footnote-4)) of Hindu tradition (the Brahmin, Kshatriya, Vaishya and Shudra communities); the thousands of regional Hindu, Sikh, Christian, Muslim or other religious groups known as jatis[[4]](#footnote-5); and groups amongst South Asian Muslims called biradaris[[5]](#footnote-6). Some jatis regarded as below the varna hierarchy (once termed “untouchable”) are known as Dalit. (Equality Act 2010: Explanatory notes Section 9: Race paragraph 49, p15).*

As is clear from the definition, the term ‘caste’ covers a number of different social systems, but with key common elements (hereditary, endogamous and hierarchical encompassing ritual purity). The terminology used varies by group (caste, jati and biradari, as detailed in the note, as well as other terms, e.g. clan). Thus a key question is whether, amongst survey respondents, there could be a common understanding of ‘caste’ as defined under the Act.

Whilst there are differing views on the extent to which South Asians in Britain continue to practice a caste/jati/biradari system, it is apparent from the literature and from our interviews that the term ‘caste’ continues to have meaning to many people of South Asian heritage (note that, in exploring the meaning of caste interviewers used the terms ‘jati’ and ‘biraderi’ as well as caste).

On the whole, younger second or third generation respondents seemed to know less about caste than older respondents of any generation, and when they did, this was through school or seeking knowledge due to interest. Some had heard or learnt about caste from parents but this was often something that was not passed down to children living in the UK, even from first generation parents for whom caste was very real.

*“In all the time I have known my parents, my dad only spoke about it recently, like the last two years, about our history and the caste system and stuff like that, so, but he wasn’t being specific about the caste system, so I think that would be the difficulty maybe, yes you wouldn’t be able to just naturally go into the conversation about the caste system, there would have to be some kind of lead up to it maybe… The young ones really, if they don’t actually even speak their parent’s language or whatever, they probably won’t have a clue at all [about caste]” (Male, 2nd generation Indian, Hindu respondent in the 36-45 age group).*

For some respondents caste was seen as outdated and lacking in relevance in the UK. This was mirrored by the stakeholders we interviewed who suggested that young people in particular would be expected to have low awareness of the concept of caste, and difficulty in discussing it. Even if presented with a list of terms and names (which would be rejected by others on ethical grounds), young people may not recognise and would not identify with the terms.

There were a range of explanations as to what caste means. This tended to vary with respondents’ religion. In the interviews, regardless of whether they identified with the caste system or not, many Hindu and Sikh respondents had a good understanding of the concept and how it worked. The caste system in Hinduism was described as a hereditary social system through which hierarchies are defined predominantly by surname, occupation and education, and wealth. Caste in Sikhism was described similarly. However, not all Hindu and Sikh respondents could either identify their own caste or how they would be perceived by those who did observe caste.

Some Pakistani and Bangladeshi respondents had knowledge of caste but this was on a much lesser scale than for Hindu and Sikh respondents. Pakistani and Bangladeshi respondents identified a sort of caste system similar to that of Hindus and Sikhs, although the definition was a lot looser. It tended to focus more on different beliefs and practices within religion (for example, distinctions between Sunni and Shia Muslims), and reference was made again to occupation. Whereas in Hinduism and Sikhism practices associated with caste felt very alive for some people, the concept did not seem to be as sensitive among Pakistani and Bangladeshi respondents. It appeared more to indicate subtle differences, which although could cause conflict, did not always prescribe innate hierarchy.

Reference was made to castes within Islam, but even then some Muslim respondents predominantly had knowledge of the Hindu caste system, albeit as an abstract concept. There was also some confusion in differentiating between the meaning of caste and ethnicity (or race or religion), with some respondents needing to be brought back to caste several times when trying to recount experiences of discrimination.

There was also some indication that understanding of caste might vary with age and migrant status, with older people and migrants having a greater understanding. For others, the picture may be more mixed. Certainly some stakeholders believed young people, in particular, to have low awareness of the concept of caste and that they would have difficulty discussing it. They suggested that, even if presented with a list of terms and names, young people may not recognise and would not identify with the terms. However, others suggested that, although some young people saw caste as outdated and lacking in relevance in the UK, they continued to have a theoretical understanding of caste. Moreover, there were indications of caste continuing to have practical relevance, particularly in respect of choice of spouse, marriage partner.

A further issue in understanding caste is the lack of a common terminology, both for the system and for specific castes. Interviewees agreed there was not a single, or even few words, that everyone would understand and relate to. However most suggested ‘jati’ and ‘biraderi’ were the key terms used.

There are two important conclusions about how people understand caste and how this understanding affects the ability to undertake a survey on caste discrimination in the UK.

1. The concept of caste appears to be widely understood. Although some may not understand the concept, identifying this would contribute to our knowledge of caste discrimination and, over time, would provide an indicator of change in caste’s significance. (However, note that an individual’s lack of understanding of the concept does not mean they cannot be discriminated against by others, who attribute a caste to that person.)
2. In summary, the interviews with target respondents seemed to indicate that the concept of caste was understood and that therefore questions about caste-based discrimination could be asked in a survey. However, ‘caste’ may be interpreted in a number of ways. This means it would be very important to develop a widely understood interpretation for use in the survey. As part of this, development work would also be needed to refine terminology for ‘caste’ and to ensure that the full range of terms are presented to respondents. One approach is to list caste groups. However, this presents two problems. Firstly, there are thousands of jati (and a shortened list of those relevant to Britain does not exist) and, secondly, some low caste names are seen as offensive, particularly to those from low caste who have rejected the caste system. Therefore, it will be important to develop a comprehensive, but simple, definition, which may need to be supplemented by comprehensive terminology for the systems (e.g. jati, biraderi). It could also be tested whether use of a few example caste names would be helpful.
	1. **Identifying caste discrimination in the British context**

Three issues arise in relation to identifying caste discrimination in a British context. Firstly, whether people can identify discrimination against themselves; secondly, whether discrimination of relevance to the Equality Act can be identified; and thirdly, whether victims of perceived caste discriminationwill report it in a survey.

### *Ability to recognise discrimination*

The first issue is general to discrimination and not specific to caste. Without close examination of the claimed discrimination (for example, as may be provided in a tribunal), it may not possible for an individual to establish whether they have been discriminated against or not: their perception may be erroneous. They may erroneously believe they have suffered discrimination or, conversely, they may not have recognised the discrimination they have suffered. This is a generally recognised problem, but has not prevented the use of self-report to provide information on discrimination (notably the national Citizenship Survey). Given the acceptance that this problem does not negate the usefulness of such surveys, this issue was not explored further within this study. (Note that whilst the level of discrimination may be reported with error, repeat surveys may give a valuable picture change.)

### *Identification of relevant caste discrimination*

In relation to the identification of caste discrimination relevant to the Act, although the existence of caste discrimination in Britain relevant to the Equality Act is disputed, previous research has identified discrimination which would fall under the remit of the Act, i.e. discrimination in respect of employment and the provision of goods and services (see, for example, Metcalf and Rolfe, 2010). Research has also identified discrimination and caste segregation outside the remit of the Act. Of particular prominence has been marriage within caste (unsurprisingly, given endogamy, i.e. marriage within caste, is part of the definition of caste) and, to those claiming discrimination, religious practices and access to places of worship. It is important that a survey can identify caste discrimination relevant to the Act (whether or not it collects information on other types of discrimination or segregation).

Within the interviews, the issue of employment discrimination was raised. However, as in previous research, the issue of caste affecting social relations, including marriage, was prominent, a factor which is outside the Equality Act. According to the representatives from stakeholder organisations we spoke to, though, discriminatory practices surrounding marriage and the caste system are problematic.

What respondents considered to be discriminatory behaviour based on caste largely predominated in marriage practices which were enforced to maintain and protect the status of families. This was reported as applying to Muslims, Sikhs and Hindus. For example, a female 3rd generation Sikh respondent in the 18-25 age group began by saying that caste and associated discrimination was less prevalent in India than it used to be, and nearly non-existent in the UK now. However she then went on to say that where marriage was concerned she would be expected by her father to marry a man within the same caste. Suggesting that caste and associated discrimination is also an issue that exists outside of Indian Hindu and Sikh communities, a female Bangladeshi 3rd generation respondent in the 25-36 age groupsaid “*In our culture casting still exists. In fact, my family is very relaxed with casting, but I know there are other families out there who casting is a big thing*”. She added that whilst she was personally against the use of caste:

*“…my close family, very own close families, they believe in casting and they still continue to believe and carry it out. But it just means that the caste is also connected to behaviour, the way you are, manners and it is connected to a lot of things. And you might be, occupation wise, you might be very good. You could be a professional, but yes, you will be still looked down upon, so much that some people don’t even sit next to them. Just because you are from a lower caste, you won’t even sit next to the person or speak to the person, or anything the person touches, you can’t touch that. It’s like that”. (Female, 3rd generation Bangladeshi, Muslim respondent in the 25-36 age group).*

Discrimination based on caste reportedly occurred in employment, and a male Sikh respondent in the 36-45 age category strongly believed that this kind of discrimination was still very much alive in the UK today. He said that particular castes may cluster in an organisation due to preferential treatment from managers towards people they perceived to be of their own caste. This makes it difficult for some people to find employment in certain sectors. A male, 1st generation Hindu in the 55 and over category made reference to the employment tribunal case of the Begraj couple in 2013 for unfair dismissal on the grounds of caste[[6]](#footnote-7). He made reference to his understanding of certain elements of the case which he felt highlighted the difficulties of proving caste discrimination. Also the changing nature of caste in the UK, due to the fact that the couple involved identified as Dalits (or ‘untouchables’) but were working in a law firm; an employment status that would never had occurred in the traditional caste system.

Other forms of perceived caste discrimination that were highlighted by respondents as existing in the UK relate to the setting apart of different castes for the purposes of social functions. A male, 2nd generation Hindu respondent in the 36-45 age group explained that organisations existed whose purpose it is to represent certain castes within Hindu communities in Britain, and all families that belong to each organisation come from a particular caste. These groups organise charity events, day trips, holidays and collectively celebrate Hindu festivals for people from specific Hindu castes, and excluded those from others.

A further form of perceived caste discrimination highlighted by respondents was caste-based ‘jokes’. A male 2nd generation Hindu respondent in the 18-25 age group said “*sometimes you hear jokes and stuff made about it [caste]… it’s kind of like dumb jokes*”. He furthered this to say “*I wouldn’t say they’re malicious, they’re jokes based on the caste system*”. In previous research, the issue of physical, as well as verbal, harassment has also been raised.

These findings suggest that while people may recognise what unlawful discrimination is if they experience it, caste discrimination may need defining because people would have different understandings of what is actually unlawful. For example, some may mistakenly assume, unless clarified, that prejudice linked to marriage and other social behaviours constitutes relevant unlawful discrimination, which are not forms of discrimination that fall under the Equalities Act 2010.

### *Willingness to report caste discrimination in a survey*

Given the accepted use of self-report surveys on discrimination, here we limit our discussion on willingness to report perceived discrimination to issues specific to caste. It is assumed that most caste discrimination in Britain is against people of low caste (or attributed low caste). If this is the case, it may affect the extent to which caste discrimination would be reported in a survey because the respondent may believe reporting caste discrimination would identify them as low caste (whatever their caste). This, indirect, identifier of caste may raise the same issues as asking respondents about their caste, discussed in the chapter on ethics. Unlike identification of a respondents’ caste, obviously reported caste discrimination is central to the survey.

It will therefore be important that respondents are given strong reassurance about confidentiality, data security and that the survey is not interested in identifying their caste. Respondents would need to be provided with ample information on how their details would be used, where their responses would be stored or published, and information about anonymity and confidentiality with explicit justification for why the information was needed and important. It may also be useful to point out that caste discrimination takes place against people of any caste.

### Summary

In summary, it is clear that questions on a survey about caste discrimination will need to be specific in relation to what constitutes discrimination as defined under legislation. In particular, it will need to ensure that respondents’ experiences of social prejudice because of caste, which do not impact directly on employment or service provision, are not incorrectly identified as constituting discrimination.

However, while some respondents felt that asking questions to identify caste would not be problematic, there were also those interviewed who would said that they would take offence at such a question being asked in a survey. Given that identifying the respondent’s caste is not critical to measuring the extent of discrimination, we would recommend that any national survey be restricted to asking about discrimination rather than attempting to establish respondents’ caste.

Initially, it was thought by respondents that younger respondents would be more willing to talk openly about caste, and there would be no problem with openly asking about this. However, the interviews showed that their responses may be affected by concerns about the security of their data. To overcome this, respondents would need to be provided with ample information on how their details would be used, where their responses would be stored or published, and information about anonymity and confidentiality with explicit justification for why the information was needed and important. Depth of knowledge or understanding about caste may be weaker with younger respondents too.

It was thought by respondents that caste may resonate more with older generations, based on their previous experiences, and that they may have more to say on the issues of caste discrimination. However for respondents who understood caste as a tangible concept, issues surrounding pride and shame may prevent them from openly discussing or responding to survey questions on this issue. This would particularly be the case if they identified as being from a low caste and felt they were being asked to reveal this. Stakeholders further supported this view, suggesting that it would be ‘unsettling’ to ask whether someone considered themselves of a certain caste, or was considered to be by other people. “Even if you ask how others see you, it is incredibly loaded” (Respondent from stakeholder organisation).

It was also suggested by respondents that even when you had a respondent who was willing to discuss caste and discrimination, problems would still arise as you could expect to find that someone of a high caste may not actually report discrimination either because they do not experience it or they impose it on other people. Similarly, another issue was identified by respondents and stakeholders which related to the requirement for interpreters and being able to access those who could not speak English, who would more likely be prevalent amongst lower castes.

*“Some sections of society will promote their caste because it’s of a higher caste. If you’re lower then you’re not going to do that.” (Male, 2nd generation Indian, Hindu respondent in the 36-45 age group)*

There was a general consensus that introductory questions which could cover other forms of discrimination – such as age, ethnicity, gender and disability – should be asked before addressing the concept of caste and associated discrimination, both to ease respondents into sensitive topics and to gain an understanding of the challenges surrounding honesty.

* 1. **Data collection approaches**

### *The interviewers*

As discussed above, the survey must ensure the use of a consistent interpretation of caste, of relevant discrimination and that the interviewer handles the interview with a high degree of sensitivity. Although questionnaire design and information provided to respondents would seek to ensure the first two aspects, it will be essential that the interviewers a have a good understanding of caste, as questions will arise and also because the interview will require sensitive handling.

Some respondents felt that when interviews were to be conducted in person the interviewer and respondent should be matched and, specifically be of the same gender, community and caste. Whilst there was some attraction to matching by community, this could cause problems if the interviewer and respondent were of different castes, particularly as people of the same community would be expected to be better able to recognise each other’s caste and, for low caste respondents the fear of having their caste more widely revealed may be greater. Therefore some interviewees felt that it would be preferable if face-to-face interviewers were from a completely different ethnic background, as respondents would be less likely to feel any embarrassment or shame talking about these issues.

As might be expected, views on the background of interviewers tended to be held more strongly by people who had been directly affected by issues relating to caste. For those who reported that they found the issue of caste comfortable to discuss, they did not feel that there would be any particular type of interviewer that would be more or less appropriate to conduct research on the subject. Because of the degree of sensitivity, we would suggest that, subject to language requirements (see below) that not only should matching based on community not take place, but that it should be avoided. The stakeholders we spoke to felt that it was important that interviewers were provided with knowledge on issues relating to caste and caste discrimination. Ensuring that interviewers were fully briefed on the subject should mean that they will be equipped to apply the necessary sensitivity. As such, any survey would probably require full day in person briefings rather than briefing by video or paper.

### *The interview mode*

Regarding respondent concerns about confidentiality and anonymity raised in sections 3.1.2 and 4.2.3, using different modes of interview may increase honestly or confidence in sharing experiences of caste discrimination, such as answering questions in a self-completion or over the telephone. However, some of the target respondents felt that there was value in building a rapport with an interviewer and therefore it would be important to ask any questions face to face.

### *Language*

As has been discussed earlier in this section, understanding of caste and its practical relevance may vary with migration status. Some migrants will have difficulty completing a survey in English. In order to ensure the survey covers all types of the relevant population (migrants and British-born) (i.e. is not biased), respondents should have the option of being interviewed in their preferred language. Moreover, recruitment into the sample should take language facility into account to prevent any bias. Although some concerns were expressed by interviewees that respondents should not be interviewed by someone from the same community as them, or who was able to recognise caste, in practice it would be difficult to screen interviewers and respondents to ensure that this was the case. The only real option would be to explore the possibility of translators who would definitely have a different ethnic background to the target respondents (e.g. White British who fluently spoke the respondent’s preferred language). However, this would significantly increase costs.

### *Easing into caste discrimination questions*

There was a general consensus from the stakeholder groups, literature review and from prior experience conducting work on this topic that introductory questions which could cover other forms of discrimination should be asked before addressing the concept of caste and associated discrimination, both to ease respondents into sensitive topics and to gain an understanding of the challenges surrounding honesty.

#  **Survey design options**

This section looks at the survey design issues and potential options for conducting a survey to measure caste discrimination in Britain and monitor change over time. It includes detailed recommendations about how to carry out a new tailored survey of the South Asian population and an outline of the possible options for placing items on existing national surveys. Parts of this section look in detail at issues relating to survey design and are therefore necessarily technical in nature.

We have assumed that random sampling methods would be required here, given the need to establish robust estimates of the prevalence and incidence of caste discrimination within the target population. Non-random sampling methods usually cost less to implement but the accuracy of the resulting estimates rests upon assumptions that may be untenable in particular cases. Random sampling methods are best when the estimates must have broad credibility across various stakeholders, especially where they hold opposing viewpoints.

### *General sample options*

As described in section 2, when conducting the feasibility study we have assumed that the target population would be GB residents who identify with a South Asian ethnic background. It could be decided that the population of interest for any survey is actually significantly wider or narrower. If that were to be the case some of the design issues and attendant costs would change.

Qualitative research suggests that the nature and incidence of caste discrimination may vary across South Asian sub-groups, with differences related to religion and culture. Religion (e.g. Muslim, Sikh, Christian, Hindu, Buddhist) and family’s pre-migration background (e.g. landowners, peasants or merchants; location) may influence caste discrimination. These too are associated with family background locations (e.g. Indians, Bangladeshis, Pakistanis and Sri Lankans; the Punjab, Gujarat, Kerala). Where the issue of interest in a survey may differ in incidence across sub-groups, to get a more reliable estimate of overall incidence, it is useful to structure the sample (i.e. set sample sizes for each subgroup) to ensure the survey captures the views of smaller sections of the population. Without structuring, the results can be swamped by those of the majority group(s). It is necessary to have data on the incidence of the structuring criteria across the whole population in order to be able to reweight survey data back to the population.

For caste discrimination, the key structuring criteria are inter-related e.g. almost all Pakistanis in Britain are Muslim (although some South Asian Muslims are Indian) and nearly all Sikhs are from the Punjab. The Census provides data on South Asian family country of origin and also religion. Structuring could be conducted using either or both of these variables, depending on the aims of the survey. We have structured by family background nationality (Indian, Pakistani etc.). This illustrates the approach and provides an example of costs. Should a survey be commissioned, further consideration should be given to whether structuring should be conducted by family background nationality, religion or combining religion and nationality. Key considerations include the aims of the survey in providing understanding of the nature of caste discrimination and ways to tackle it and the effect on different communities. Costs would differ from those illustrated below, although the main influence on costs would be the number of structuring characteristics, rather than the change in criteria.

## Bespoke national survey

### *Assumptions and survey design decisions*

For any random sample survey, the primary design decision is what mode to use for recruiting the sample and what mode to use for data collection. The mode for recruitment and the mode of data collection do not have to be the same but, in practice, it is unusual to separate the two.

#### Sample recruitment mode

Because there is no existing comprehensive list of households containing individuals with South Asian ethnic heritage, a general sample of households must be screened to find the eligible households. Screening may be carried out by interviewers who can then immediately collect data from identified eligible sampled individuals. However, this is an expensive method if the eligibility rate is low (across Great Britain, it is estimated that only 4% of households contain anyone eligible for this survey).

An alternative is to use a letter-based screening mechanism in which sampled households are invited to complete a short questionnaire - either online or on paper - which will establish their eligibility for the survey. The response rate to this method is likely to be considerably lower than the response rate to an in-person (ie. face to face) screening method but it may be the only practical method if the eligibility rate is very low. It can also be combined with in-person screening of non-responding households if the eligibility rate is high enough to make this approach feasible. Costs may be reduced if part of the sample is screened using the letter-based method even if the remainder must be screened in-person. For a survey on caste discrimination, we recommend using in-person screening only. This is because the anticipated eligibility rate in targeted areas (see discussion on the sample design in section 5.2.1below) is sufficient for it to be economical and we would expect fewer screening errors due to respondents not correctly identifying themselves than would be obtained from a self-completion method (for which there is no in-person validation).

#### Data collection mode

For most sensitive topics a self-completion questionnaire method would be expected to provide the most accurate data. When an interviewer asks the questions instead, some respondents choose (either consciously or subconsciously) to present themselves a step or two removed from reality. In other words, the social nature of the interview has an impact on the responses given. Some respondents will report what they take to be generally desirable opinions or behaviours; others will disguise a lack of knowledge or opinion by selecting the most plausible positive response rather than admitting to having no answer to give. This is called ‘social desirability bias’. These phenomena can lead to measurement error.

On the other hand, interviewers can help respondents understand what is being asked of them, can clarify responses, and can query unexpected responses. Interviewers can therefore reduce measurement error as well as introduce it. In practice, the ideal data collection mode is likely to vary from topic to topic and from item to item within a questionnaire.

For a survey on caste discrimination, given the topic is sensitive, as demonstrated in the earlier sections of this report, and given the diversity of understanding of the concepts involved, using a self-completion mode may be preferable. However, it is possible that some of the respondents may not read English well and require translation or the presence of an interpreter. The objective of inclusivity would therefore favour an interview mode. A secondary advantage of the interview mode is that self-completion sections can be inserted into the data collection process to handle the most sensitive areas, at least for those with sufficient language skills.

On balance, an interview mode rather than self-completion seems appropriate for data collection.

#### Data collection within the household

For some surveys, data is collected from all individuals in the sampled household, but for this survey we recommend that data is collected from just one adult individual sampled at random within a household. Given the sensitivity of the topic, a multi-individual data collection model, in which all the people in a household are interviewed, is likely to lead to data ‘contamination’ as those responding first may influence both the participation decision of others in the household *and* the way they respond. As such interviewing one person per household is likely to lead to more reliable data.

### Sample design

#### Population distribution

The 2011 Census provides a count of individuals classified by ethnic group. The question used in the Census is not perfect for our needs and three and a half years have now passed since the data was collected. Nevertheless, it should provide a more accurate count of individuals with South Asian ethnic heritage[[7]](#footnote-8) than any other source. The total in 2011 was 3,220,942.

Census counts are available down to a granular geographic level. The smallest level is the Census Output Area, containing an average of 120 households. However, the smaller the area, the more quickly Census data becomes out of date. A more robust set of counts is available for Lower Level Census Super Output Areas (LSOAs), each of which contains approximately 600 households. There are 34,553 LSOAs in England and Wales[[8]](#footnote-9). This is the database we have used to develop the sample design that follows.

For the purposes of a new survey, we have defined five sub-classes within the larger population of individuals with South Asian ethnic heritage. The precise Census codes for each sub-class are provided in Table 4.1. Arguments could be made to shift Census codes from one sub-class to another but we have placed any code that is not clearly part of one of the four named sub-classes – Indian, Pakistani, Bangladeshi and Sri Lankan – into the ‘Other South Asian’ sub-class. For example, the Tamil ‘homeland’ is split between India and Sri Lanka, and the Punjabi ‘homeland’ is split between India and Pakistan. These codes are therefore placed into the ‘Other South Asian’ sub-class.

Table 4.1: Five sub-classes within the South Asian group

|  |  |  |
| --- | --- | --- |
| **Sub-class** | **Census categories** | **Census total individuals** |
| Indian | Asian/Asian British: Indian or British Indian | 1,412,958 (43.9%) |
| Pakistani | Asian/Asian British: Pakistani or British Pakistani | 1,124,511 (34.9%) |
| Bangladeshi | Asian/Asian British: Bangladeshi, British Bangladeshi | 447,201 (13.9%) |
| Sri Lankan | Mixed/multiple ethnic group: Sri Lankan; Asian/Asian British: Sinhalese; Asian/Asian British: Sri Lankan; Other ethnic group: Sri Lankan | 150,240 (4.7%) |
| Other South Asians | Mixed/multiple ethnic group: East African Asian; Mixed/multiple ethnic group: Kashmiri; Mixed/multiple ethnic group: Tamil; Asian/Asian British: British Asian; Asian/Asian British: East African Asian; Asian/Asian British: Kashmiri; Asian/Asian British: Punjabi; Asian/Asian British: Tamil; Other ethnic group: Kashmiri; Other ethnic group: Punjabi; Other ethnic group: Tamil  | 86,032 (2.7%) |
| **All South Asians** |  | **3,220,942** |

Source: 2011 Census

These sub-classes are not distributed evenly across England and Wales. They tend to be **clustered** into particular parts of major English cities (London, Birmingham, Leicester, Bradford, Manchester, Coventry and Leeds in particular) plus a small number of key towns (Luton, Slough, Blackburn, Oldham etc.). Within this the individual sub-classes are also not distributed evenly with certain areas having concentrations from different ethnic backgrounds. This clustering means that although individuals of South Asian ethnic heritage represented only 5.7% of the population of England and Wales in 2011, a sample design which targets those areas with higher concentrations of potential respondents can ensure a much higher survey eligibility rate with only a small sacrifice in the coverage of each sub-class. The different distributions of each sub-class also means that it is possible to target an approximate sample size for each sub-class without knowing the ethnic composition of specific households.

#### Sample frame

All random sample surveys in England and Wales utilise the Postcode Address File (PAF) as a sampling frame because there is no easily accessible database of individuals or households in the UK. A sample of addresses is drawn from the PAF and, at each address, one household is sampled (should there be more than one at the address).

We can estimate the number of addresses in each LSOA that will be eligible for the survey, so long as we know:

1. the average number of individuals per household per Census sub-class, and
2. the average number of households per address.

Census information for (i) is not available at the detailed level we require so we have used household size information from the 2013-14 Crime Survey of England & Wales as a proxy. The CSEW is a very high quality survey with a large annual sample size (35,000 households) and is therefore a suitable source to use. [[9]](#footnote-10)

For (ii), we know from other surveys that approximately 90% of PAF addresses labelled as ‘residential’ contain at least one household[[10]](#footnote-11) and that only 3% of these are multi-household addresses. An estimate of 0.9 households per sampled address is therefore a fair assumption to make. This will vary between areas but the extent of this variation is known only approximately so may be ignored for now.

Based on our assumptions, Table 4.2 uses this data to estimate that a total of c. 950,000 households would be eligible for the survey, from a total of c. 24 million households and c. 26-27 million addresses. There is an implicit assumption here that all households are mono-ethnic. In reality, households will comprise a mixture of with eligible and non-eligible individuals, which has the effect of increasing the total number of eligible households, which in turn means that *the number of eligible households estimated in Table 4.2 should be regarded as a lower bound estimate*.

Table 4.2 Estimated number of survey eligible households

|  |  |  |  |
| --- | --- | --- | --- |
| **Sub-class** | **Average household size (Crime Survey of England and Wales 2013-14)**  | **Census total individuals** | **Estimated number of eligible households (% of eligible households)** |
| Indian | 3.02 | 1,412,958 | 467,867 (49%) |
| Pakistani | 3.94 | 1,124,511 | 285,409 (30%) |
| Bangladeshi | 3.52 | 447,201 | 127,046 (13%) |
| Sri Lankan | 3.29\* | 150,240 | 45,666 (5%) |
| Other South Asians | 3.29\* | 86,032 | 26,150 (3%) |
| All eligible individuals | 3.38 | 3,220,942 | 952,137 (4.0% of all households; 3.6% of all addresses) |
| Non-eligible individuals | 2.29 | 3,220,942 | 23,080,773 |
| **All individuals** | **2.33** | **56,075,912** | **24,032,910** |

\*Mean for all non-Indian/Pakistani/Bangladeshi South Asians used due to small sample sizes in the CSEW

Sources: CSEW 2013-14 & 2011 census

#### Sample design

If only c. 4% of addresses contain survey eligible individuals (Table 4.2), a representative sample design – in which every address has an equal probability of selection - will be rather expensive, whatever mode is used for data collection. However, by making use of the uneven geographic distribution of each sub-class, we can raise this to a more practical level of 15-20% so long as we accept that a minority of each sub-class will not be covered by the survey. Not covering parts of a population in a survey sample design can lead to biased estimates to the extent that any variable of interest, in this instance this would be experiences of caste discrimination, is correlated with the local survey eligibility rate. It seems reasonable to expect that the absolute level of caste discrimination *will* vary with the density of the relevant ethnic community so bias is a risk. Nevertheless, for reasons of practicality and cost, we would not advocate surveying *any* areas where fewer than 5% of households would be eligible. Non-coverage bias is a risk that must be taken (but not ignored).

There are two ways of raising the survey eligibility rate and they can be used in combination:

1. Addresses may be stratified on the basis of their (LSOA level) ethnic composition and sampling fractions set for each stratum. The sampling fraction may vary between strata if desired (see below). One stratum should include all LSOAs in which the estimated survey eligibility rate falls below a minimum threshold (e.g. 5%). In this stratum the sampling fraction should be set at zero, excluding it from the survey.
2. Sampled addresses in survey-eligible strata may be randomly allocated to different eligibility criteria (for example only households with respondents from an Indian background would be eligible for the survey). For the majority of addresses, all sub-classes would be eligible but at a minority of addresses, only the less numerous sub-classes would be eligible.

Although these methods may be used in combination, a satisfactory result *may* be possible just method (i). The use of method (ii) inevitably leads to a reduction in the survey eligibility rate as fewer households will contain eligible respondents and a corresponding increase in the cost per completed questionnaire. However, it should be employed when method (i) is insufficient by itself to generate the required number of interviews for particular sub-groups. In this instance the sub-groups which would be most likely to require additional numbers would be the Sri Lankan and ‘’other-South Asian’ groups. If achieving specific numbers for these groups was a priority then method (ii) would need to be employed and the survey costs would increase significantly.

In the illustration that follows, we have assumed that in-person methods will be employed to screen addresses, sample one individual at each eligible address, and to collect the data from that person. As argued above, this seems the only plausible method given the nature of the topic and the target population.

In-person data collection methods enforce a two-stage ‘cluster’ sample design in which a sample of small areas (LSOAs in this case) is drawn at random from each stratum and then a sample of addresses is drawn from within each small area. Cluster samples are statistically less valuable than un-clustered samples[[11]](#footnote-12) but data collection costs are much lower due to the reduced travel time between addresses. Unless the sample of addresses itself is enormous, a cluster sample will typically yield a higher effective sample size than an un-clustered sample of equal cost[[12]](#footnote-13).

For our illustrative design, we have set a minimum household eligibility rate of 5%. This reduces the total number of LSOAs from 34,553 to 6,163 and covers 77% of the target population (individuals of South Asian ethnic heritage). Coverage varies between sub-classes; from 71% of the Sri Lankan sub-class to 85% of the Pakistani sub-class. The estimated proportion of households that would be eligible for the survey is estimated at 17%. This could be raised but only at the expense of reduced population coverage. For example, if the minimum household eligibility rate was raised to 7.5% (instead of 5%), the average survey eligibility rate would rise from 17% to 22% but the population coverage would go down from 77% to 69%. In our view, the 5% minimum is a reasonably compromise between cost and population coverage.

The 6,163 LSOAs have then been allocated to one of ten strata, each maximally different from the others with regards to the estimated number of households in each ethnic sub-class. A ‘*k*-means’ clustering algorithm was used for this purpose. Stratification on this basis ensures minimal sample variance in terms of ethnic composition, a desirable feature if the credibility of the survey rests upon the representativeness of the sample.

Each stratum *may* be given a different sampling fraction to skew the sub-class distribution in the sample of addresses so that the less numerous sub-classes are represented more robustly than they would be under a proportionate design. However, disproportionate sampling of this type requires compensatory weights and these can reduce the statistical value of the whole sample. The implications need to be well understood before taking this step. In this case, although there is an argument for reducing the Indian sub-class sample size and increasing the Bangladeshi sub-class sample size, varying the stratum level sampling fractions appears to be a statistically inefficient way of doing so and we do not recommend it.

Table 4.3 outlines what we might expect under both a proportionate design. It also provides an indication of the confidence intervals around proportion estimates *if the overall sample size was around 2,000* and for variables with no significant clustering effect. Clustering effects differ between variables but are usually present to some degree. Therefore, the *quoted confidence intervals should be treated as the minimum than can be expected[[13]](#footnote-14)*.

Table 4.3 Expected sample shares and confidence intervals for an overall sample size of 2,000

|  |  |  |
| --- | --- | --- |
| **Sub-class** | **Percentage of overall sample** | **Confidence interval[[14]](#footnote-15)** |
| Indian | 46.3% | +/-3.2pp |
| Pakistani | 33.0% | +/-3.8pp |
| Bangladeshi | 13.4% | +/-6.0pp |
| Sri Lankan | 4.4% | +/-10.4pp |
| Other South Asians | 2.8% | +/-13.1pp |
| **All eligible individuals** | **100.0%** | **+/-2.2pp** |
| *Survey eligibility rate* | *17.4%* |

Source: 2011 census

Larger sample sizes would decrease confidence intervals while smaller sample sizes would increase them. However, the impact of sample size on the width of a confidence interval is not linear. The ‘return on investment’ of an increase in sample size is smaller the larger the baseline sample size[[15]](#footnote-16).

The confidence intervals in Table 4.3 look reasonable for the three most numerous groups but the table also makes it clear that only a very large overall sample – or a smaller but highly skewed sample - would provide robust *specific* evidence about the Sri Lankan and Other South Asian sub-classes. It is our view that the cost of meeting this objective would be too great and that an overall sample size of around 2,000 – providing precise population-level evidence and reasonably precise evidence for the three most numerous sub-classes - would be appropriate to the importance of the topic.

#### Measuring change over time

After completing an initial ‘baseline’ survey, a second survey would be required if change over time in the extent of caste discrimination were to be measured, as per the review clause in the legislation. Any change that is observed may not be entirely due to the legislation, it may reflect greater awareness of the harm done by caste discrimination or changing attitudes towards caste but it is plausible to believe that it would be a contributory factor.

The most obvious form the second survey could take would be a repeat of the first survey in its design but there are two alternatives that would lead to greater precision in our estimates of change.

The first option is to re-interview as many respondents from the first survey as possible, topping up with a fresh sample to achieve a target sample size. This option will lead to the most precise estimates of change because the ‘paired samples’ component is not affected by any random differences from the first survey sample (such as arise when fresh samples are drawn, even if the sampling protocol is identical). The impact is diluted if a ‘top up’ fresh sample is required but it still retains a higher level of precision for change estimates than an entirely new sample. The difficulty for a designer is knowing *how* much extra precision is retained. There is a risk in re-interviewing respondents that the first interview acts as an intervention, affecting either the respondent’s opinions or behaviours, or affecting their *reporting* of otherwise unchanged opinions or behaviours. As with the potential benefit of greater precision, it is impossible to quantify this risk in advance. Generally speaking, such ‘conditioning’ effects are mild but there are exceptions. To some extent these can be assessed *post hoc* by comparing estimates drawn from the repeat sample from those drawn from the concurrent ‘top up’ sample and an adjustment can be made to the estimate to reflect the findings[[16]](#footnote-17), if necessary.

A second option is to draw a fresh sample of addresses from within the same set of LSOAs sampled for the first survey. Technically, this is a ‘paired sample’ as well, only of neighbourhoods not of individuals. The correlation between responses given at the first and second surveys will be weaker for this reason but might be large enough to increase precision for some measures. This design also minimises the risk of conditioning effects discussed in relation to the re-interview design. One drawback is that the costs are higher compared to the re-interview design described above. Interviewers must screen as many addresses per interview as they did for the first survey, instead of being able to re-interview a number of already sampled cases.

As already noted, the problem for the designer is that the potential gains in precision from either of these designs are (i) unknown, and (ii) likely to differ from measure to measure. Consequently, it makes sense to set a second survey sample size that assumes there is little or no gain in precision for estimates of change. From this start-point, we can take the confidence intervals quoted in table 4.3 and multiply them by 1.4 to obtain the confidence interval for an estimate of change. So, if the 95% confidence interval for an estimate from the first survey is +/-2.2 percentage points, then the 95% confidence interval for an estimate of change between the first and second surveys is +/-2.2\*1.4 = +/-3.1 percentage points. The second survey sample size may be selected to obtain a target confidence interval. Indeed, both the first and second survey sample sizes may be decided in tandem if funding is available for both from the start.

An alternative - much used in experimental research - is to set a sample size so there is a minimum probability that a real change of *x*% between time points 1 and 2 will be reflected by a ‘statistically significant’ difference between the first and second survey estimates. Traditionally, that minimum probability is set at 80% but other values could be selected. Returning to table X.3, if we *double* the quoted confidence interval, we obtain the smallest real change that will – with at least 80% probability - be reflected by a ‘statistically significant’ difference between the first and second survey estimates *if the sample size of 2,000 is retained for the second survey*. So, if the quoted confidence interval is +/-2.2 percentage points from a first survey sample size of 2,000, then a second sample size of 2,000 would mean that real changes of 4.4 percentage points or greater would be reliably ‘detected’. As before, the second survey sample size can be adjusted up or down to obtain sufficient ‘power’ to detect a particular magnitude of change – and this may be done for the first and second surveys in tandem if the funding is in place.

#### Assignment structure

A reasonable target number of LSOAs to sample is around 400 to generate a national estimate of caste discrimination, yielding an average of five interviews each. It is possible to sample fewer LSOAs and carry out more interviews in each one but we should not underestimate the work required to screen households for eligibility, which is additional to the interviewing work. In LSOAs with relatively low survey eligibility rates the *majority* of the work will be screening, while in LSOAs with relatively high survey eligibility rates, the majority of the work will be interviewing. When refining the sample design, care needs to be taken to ensure that the workload in each LSOA is neither too great nor too small for a single interviewer[[17]](#footnote-18).

To assist with this, the primary stratification of LSOAs by ethnic composition may be complemented by a second layer of stratification, this time determined by overall survey eligibility rate. In ‘high eligibility’ sub-strata, the sampling probability of each LSOA may be increased while the conditional sampling probability of each address is *de*creased. Conversely, in ‘low eligibility’ sub-strata, the sampling probability of each LSOA may be decreased while the conditional sampling probability of each address is increased. The *overall* address sampling probability (taking account of both sample stages) remains the same across all sub-strata but this approach ensures that the variation in workload across assignments is reduced to a manageable level.

### *Survey development*

Before a national survey of caste discrimination could be conducted there would need to be an extensive programme of survey development. At present, there are no major surveys that attempt to measure caste discrimination and therefore questions would need to be developed specifically for this survey. This is particularly important given the potential sensitivities around the subject matter and the need to be precise in identifying caste discrimination. In particular, the survey will need to be able to differentiate discrimination that would fall under the Act from discrimination that may occur in the personal sphere (See Section 4.2.2). It will also be necessary to ensure that the questions are designed so that respondents understand that they should only be reporting caste discrimination rather than any discrimination based on ethnicity or religion.

We would recommend that ‘cognitive testing’ is used to develop specific questions on caste discrimination for the survey. In cognitive testing the focus is on how respondents are understand and are interpreting the questions which they are being asked and what lies behind their answers rather than whether or not they can complete a questionnaire. This added focus means that questions which on the surface can seem to work because respondents are able to give answers can be more thoroughly evaluated and underlying problems identified.

After the questionnaire has been cognitively tested, we would recommend that a pilot of 100 interviews is conducted to test all aspects of the survey (such as interviewer briefings; respondent contact procedures; the wording, routing and length of the questionnaire; general testing to ensure that the questions and survey overall would work in a real life setting). This would further test questionnaire content but also allow for other survey processes such as the screening procedures to be evaluated. The pilot could also provide a guide to possible levels of response, although this would necessarily only be indicative as the confidence intervals even with a pilot of 100 interviews would be relatively large (+/- 10pp).

### *Cost*

Tailored national surveys, with the scale and design described above, are resource intensive and as a result can be relatively expensive. As part of the feasibility study we have calculated that a single cross sectional survey as described above would cost £870,000 excluding VAT if conducted in 2015. This would cover substantial set up costs as well as the large-scale screening exercise to identify the target respondents in the absence of a comprehensive list of this specific population. In putting together these costs we have assumed an interview length of 30 minutes and that a post screening response rate of 60% could be achieved which is what was achieved on the 2010 Ethnic Minority British Election Study (EMBES). We have also assumed that respondents would be paid an incentive of £20 to maximise response in the same manner as the 2010 EMBES.

The cost for a repeat cross sectional survey in 2018, in line with the five year review clause of the ERR 2013, would be slightly lower at £850,000. If a panel design were adopted for the second survey with respondents from the first wave interviewed again alongside top-up sample the data collection cost would be reduced to £530,000. The reason that this approach would provide cost savings is that there would be no need to repeat the screening exercise to the same extent as the first wave, which contributes greatly to the cost.

While the sample design discussions above relate to England and Wales, these costs would be for a GB-wide survey including Scotland.

## Existing survey samples

An alternative to a standalone national survey would be for items on caste discrimination to be placed on an existing national survey. The main benefit of this approach is that the data collection costs would be greatly reduced as the administration costs could be shared across the main survey and the additional items needed to record caste discrimination. There are, however, a number of conditions that would need to be met before a set of items on caste discrimination could be included on an existing survey.

The first condition is that any survey would need to include a sufficient number of respondents from a South Asian background to provide a robust sample size in order to draw reliable conclusions (as described above). There are only a limited number of surveys that would fit this criteria criterion: the Labour Force Survey (LFS), the Crime Survey for England and Wales (CSEW) and Understanding Society (USoc). The second condition is that the items on caste discrimination should have some relevance to the other items included in the survey. The modular approach of USoc means that items on caste discrimination would fit in to the interview structure without being jarring for the respondent and therefore would not appear too far out of context with the rest of the interview. It would also be possible to filter on ethnicity so that items were only asked of respondents from a South Asian background. The LFS has a large focus on employment and items around employment discrimination would therefore not necessarily be problematic, however ethnicity is asked towards the end of the survey so items may have to be asked of the full sample. Items on caste discrimination would appear most problematic in the context of the CSEW. There are now obvious places from which such questions could flow from and again items could not be filtered on respondent ethnicity as this asked at the end of the survey.

Given the factors outlined above the most appropriate survey items on caste discrimination could be placed on would be USoc. The particular challenges in placing items on USoc would be securing the permission of the scientific leadership team of the study and timing. This latter point could be particularly problematic as the earliest that data from an initial survey would be available would be November 2018 and from a follow up wave in November 2019.

## Survey challenges

### *Response rates*

One of the biggest challenges a survey of this nature faces will be ensuring that response rates are sufficiently high so that all stakeholders can have confidence that the results are an accurate representation of the level of caste discrimination in Britain. As stated above we believe that a post screening response rate of 60% is achievable based on our experiences of EMBES. However if the response rate were to be significantly lower then the risk that the survey is not an accurate reflection of the level of discrimination increases. Indeed it should be noted that even with a response rate of 60% there is a significant risk of non-response bias with those who consider themselves to be victims of caste discrimination being either significantly over or under represented

### *Identifying a target population*

As stated earlier, this feasibility study has assumed that any national survey to identify caste discrimination would target an eligible population of South Asian respondents. However this would mean that caste discrimination by people from other ethnic background would not be recorded. There has been some indication that caste discrimination may be present in Somali communities for example. However if the survey target population needed to be expanded and minimum numbers of interviews be conducted from populations with a relatively low presence in Great Britain the costs would increase.

In addition to this, some stakeholders may argue that just focussing on respondents from a South Asian background casts a negative light on those people, treating them as somehow different and liable to commit this kind of discriminatory practices. These sentiments were also mentioned by some target respondents. One way to address this would be to widen the target population so that includes respondents from all ethnic backgrounds. However, the rationale for targeting respondents from a South Asian background arises from their representation in campaigns against the label of caste and associated discrimination. Therefore the voices of respondents from a South Asian background would seem like the most appropriate to hear in understanding how caste and associated discrimination may exist in Britain. To widen the target population beyond this would necessitate a significant increase in the overall number of interviews. This would again increase costs when it is likely that there may be no advantage in doing so.

Whilst there is inevitably a level of uncertainty surrounding such a unique survey and sample, the basic sampling approach is established and we are confident that any potential challenges could be mitigated with sufficient piloting and testing to ensure that the presentation of the survey worked well, and to equip interviewers with ample information to deal with subject-specific queries and cultural issues.

### *Timing*

The **survey development** stage for a national survey of caste discrimination would realistically be conducted over the course of six months. The **fieldwork period** would need to be relatively substantial to ensure that the maximum response rate is achieved. We would recommend a fieldwork period of at least 4 months. Further time will be required to quality assure the final dataset.

#  **Summary and conclusions**

## Summary

Box 1 summarises our proposed approach to quantifying the extent of caste discrimination in Britain.

**Box 1: Summary of approach to a survey of caste discrimination**

**Survey design**

* The most appropriate approach would be a bespoke national survey. The addition of a module to an existing survey (such as Understanding Society) would not provide the most robust findings and would result in significant delay to data.

**Sample**

* The target population would be residents in Great Britain who identify with a South Asian ethnic background, for which the sample design would provide 77% coverage of.
* The survey would require a random probability sample, stratified to ensure a reasonable number of the target populations are reached. It would use Census data in aggregate with the Postal Address File to sample respondents initially.
* The database used for the sample would use Lower Level Census Super Output Areas (LSOAs) which contain around 600 households each. The number of eligible households in each LSOA can be estimated using the Crime Survey for England and Wales as a proxy (c.950,000 households in total for the survey).
* Five sub-classes of target respondents within the wider population have been defined and are unevenly distributed across England and Wales. The sample design should target those areas to ensure the highest possible survey eligibility rate, which although runs the risk of under-representing respondents in areas where not as many people from the target background live, would minimise costs significantly.
* In-person screening is then recommended to find eligible households. Whilst in-person screening would be a costly option, it would produce fewer screening errors and the anticipated eligibility rate for targeted areas would make it more economical.

**Data collection**

* Traditional interviewer led face to face survey. Whilst this is the most expensive method, it is the most robust in mitigating challenges such as language barriers and illiteracy, which may arise in self-completion modes. Self-completion modes can potentially instead be used as a supplement to cover more sensitive issues where possible.
* One interview per household with a randomly selected person would avoid data contamination and provide the most reliable results.

**Measuring change over time**

* Following completion of an initial baseline survey, change in levels of caste discrimination over time could be measured by a second survey which would be in line with the review clause of the Equality Act 2010 legislation. The most cost effective way to do this would be to re-interview people from the original sample.

**Challenges**

* Due to the unique nature of the study, substantial development work is required (such as cognitive testing of questions and piloting) which will increase costs.
* Considerable time spent on the set-up of the survey would mean that results would not be ready for about a year.
* Gaining sufficiently high response rates to ensure credible results that accurately represent levels of caste discrimination will be a challenge. A post-screening response rate of 60% is achievable, although there is a risk of non-response bias for those who consider themselves at risk of caste discrimination being over or under represented.
* The assumed target population is of a contested nature, but inclusion of respondents beyond this (such as those with a relatively low presence in Great Britain) would necessitate significant increase in the overall number of interviews, and therefore cost.

The major ethical considerations identified in the feasibility study concerned the potential impact that revealing their caste could have on respondents. In particular there was a concern that some respondents from a lower caste may become upset at having to reveal this fact and also there may be some level of distress among respondents who rejected the very notion of caste. However, identifying the respondent’s caste would not necessarily be required for a survey to be able to measure caste discrimination. While this information could prove useful in better understanding the nature of that discrimination it would be required to establish its prevalence. There was also some concern that people who perceive themselves to be victims of caste discrimination may feel distress at reporting this. This is an issue that many social research surveys measuring victimisation face and there are protocols that have been established over the years that allow respondents to be directed towards appropriate support.

Other concerns that were identified in the feasibility study focussed on the perception that research on caste discrimination would be setting people from a South Asian background apart from the rest of British society. This was particularly a concern for second and third generation respondents who do not view caste as being at all relevant to their lives today. To address this any survey about caste discrimination would need to provide contextual information which would explain the government’s rationale for researching the issue.

The concept of caste was well understood among respondents of an Indian background, even among those who rejected the concept. The concept was grasped less immediately by respondents from a Pakistani or Bangladeshi background but interviews showed that it could be introduced in such a way that they could recognise it, if required. There were some concerns that respondents would not be open in discussing their experiences of caste discrimination which would affect the response rate and nature of response, but it was felt that if appropriate introductory questions were asked to ease respondents in to this potentially sensitive subject honest answers would be given.

The requirement to provide a reliable measure of caste discrimination in line with ONS standards means that random sampling methods are the only appropriate way of conducting this research. If a standalone survey were to be conducted it would be necessary to conduct a large-scale screening exercise to identify respondents from a South Asian background as there is no comprehensive list identifying such households. This approach would necessarily require a significant amount of resource although a longitudinal panel could be utilised to reduce this somewhat.

Overall, we would recommend a target sample size of 2,000 individuals to provide a reliable estimate. An alternative approach could be to place items on an existing national survey, such as Understanding Society, but this would face challenges relating to timings even if it did require significantly less resource.

## Conclusion

There are no significant ethical or methodological barriers that could not be overcome to allow a survey measuring the extent of caste discrimination to be carried out. Respondent distress can be kept to a minimum if the survey is restricted to measuring caste discrimination rather than attempting to identify caste and if detailed contextual information is provided as to why the items are being asked. In addition to this, interviews with respondents suggested that respondents displayed an understanding of the concept of caste even if they were vague on the details – particularly among second and third generation respondents.

There are however, practical constraints that would need to be taken in to consideration. A standalone survey would require a significant amount of resource to conduct with a robust sample that could provide reliable estimates on the extent of caste discrimination and change over time. There is the possibility of placing items measuring caste discrimination on an existing national survey of which Understanding Society would seem to be the most appropriate. However this could cause problems in terms of timings with results not available for two waves of findings until the end of 2019. These practical considerations are what will have to be addressed when making a decision as to whether a survey measuring caste discrimination should be conducted.

# **APPENDICES**

# **Appendix 1 – Interviews with target respondents**

Recruitment for the target respondent interviews was carried out by TNS BMRB’s specialist qualitative recruiters. Recruiters were briefed by researchers about the study and provided with screeners for the recruitment process. When being recruited, possible respondents were provided with information about the purpose of the study and what would be required from them. They were also asked whether caste, jati or biradari played any role in their or their family’s life. This was to ensure that we only recruited people who would be able to understand and share their views, and potentially experience of, caste.

In order to ensure respondents would be willing to take part and to travel a short distance to the interview venues, they were offered a cash incentive. This was £50 for respondents in London, and £35 for respondents in Birmingham. Respondents in Birmingham were also offered an additional £15 to cover travel expenses if they were bringing an interpreter with them.

Recruitment was focused on gaining a broad range of characteristics within respondents who would understand, or had experience of, caste and caste discrimination. These respondents were identified as most likely being from an Indian Hindu or Indian Sikh background, so respondents from these ethnic backgrounds were prioritised. A limited number of respondents from other South Asian backgrounds were also included, as, as mentioned earlier in this report there is evidence to suggest that an associated caste system may exist in the South Asian community outside of Indian Hinduism and Sikhism and so it was important to capture this. Loose quotas were placed on the characteristics of gender, age, and social grade, as seen below:

|  |  |
| --- | --- |
| **Broad categories** | **Interviews achieved** |
| Gender | MalesFemales | 1213 |
| Ethnic group | IndianIndian HinduIndian SikhIndian MuslimBangladeshiPakistani | 294145 |
| Age group | 18 – 25 years26 – 35 years36 – 45 years46 – 55 years 55+ years | 57643 |
| Social grade | ABC1C2DE | -310543 |
| **Total** |  | **25** |

In total 25 interviews were carried out. Respondents were recruited from London and Birmingham and interviewed in central locations in these two cities across the course of one week. Interviews took place face to face and were conducted by researchers from TNS BMRB. Interviews lasted up to 45 minutes, and interpreters were used where necessary.

The interviews covered a range of topics and used questions about demographics and general experiences of discrimination to build rapport with respondents before introducing the subject of caste. The topic guides can be found Appendices 2 and 3, but the overall areas covered were:

* Introductory questions about general discrimination
* Understanding of terms associated with descent other than ethnicity
* Role of caste/descent in society nowadays
* Experiences of the concept of caste being used
* Comfort in answering questions about caste
* Willingness to answer questions about caste
* Barriers to honest answers about experiences of caste discrimination
* Ethical considerations in conducting a survey about caste and associated discrimination
* Other issues not covered in the interview

# **Appendix 2 –Target respondent topic guide**

* Introduction (5mins)
* Introduce self, TNS BMRB independent research agency working on behalf of the Government Equalities Office (GEO)
* Explain aims of the discussion: *to find out views about discrimination to inform how a future survey may measure this.*
* Explain GEO’s reason for commissioning the research: *to explore methodological and technical issues (including possible ethical ones) associated with reliably and accurately measuring forms of discrimination over time.*
* The purpose of a future survey*: the survey would be focused on understanding how much/little discrimination occurs, and changes in this level over time.*
* Explain recordings (only available to research team)
* Anonymity and confidentiality (all answers will be kept confidential, and you will not be identifiable by the information you give)
* Length approximately 30 minutes
* Right to withdraw
* Introductory questions about discrimination (5 mins)
* Ever experienced crime, discrimination or prejudice based on:
	+ Age
	+ Nationality
	+ Ethnicity
	+ Religion
	+ Gender
	+ Sexual orientation
	+ Disability
		- For each this could include in the workplace or education (bullying, promotion, task allocation), provision of services etc
* Experiences of crime, discrimination or prejudice against other people (as above)
* Understanding concepts (5-10mins)
* Understanding of the terms associated with descent other than ethnicity (prompt jati, biradari if needed)
	+ What does it mean (how is it referred to, discussed, in what settings, by whom/to whom, specific terms, what feelings does it evoke)
	+ For those who recognise jati, biradari (or other identified similar concepts) –
		- Talk through the different labels (how many are there, are they all recognised by everyone?)
		- How else is this kind of descent referred to?
		- How do they personally identify their descent in terms other than ethnicity?
* What role does descent play in society nowadays
	+ How do the concepts exist in the UK nowadays?
	+ Does discrimination based on this exist in the UK? (How, who, conditions, examples)
* Experiences of concept being used
	+ Experiences of concept being used (when, how, to whom)
	+ Do respondents themselves use it? How?
	+ Ever experienced discrimination on basis of descent themselves?
* Comfort in answering questions (5-10 mins)
* How do you think people would feel being asked questions about descent (prompt jati, biradari if needed) Probe types of respondents - age, gender, nationality, religion
	+ Reasons for views
* Willingness to talk about descent and associated discrimination (probe how this may differ according to types of interviewers - age, gender, nationality, religion)
	+ reasons for views - Why/why not
	+ How might we make them more comfortable, more likely to respond openly and accurately (Probe types of settings/modes: group, online, face to face, self-completion, location etc)
* Will people feel able to answer questions honestly and directly? Why? (for example, it could explore the following subject areas):
	+ *Structural reasons*
	+ *awareness– whether attribute to themselves*
	+ *whether feel that others attribute to them*
	+ *whether think about what another person’s is*
	+ *aware of discrimination*
	+ *own experience of discrimination*
		- For each: how would we make them comfortable/ able to be accurate
* Should we expect people to exaggerate discrimination, and what they have experienced; or conversely, to understate it? Why?
* Ethical issues (5 mins)
* Do they see any ethical issues in surveying people about this type of discrimination (elaborate on ethical issues if needed – e.g. consent, respondent harm, risk, confidentiality etc)
	+ How might people respond?
	+ What might the effect(s) be of discussing this discrimination be (Probe – on individuals/communities/groups/society)
* Other issues (5 mins)
* Any other challenges/issues in conducting a survey on descent discrimination
	+ Are there ways to ensure an accurate survey of discrimination, e.g. preparing people in their community to respond
	+ Other comments to inform the research which explores whether we can measure discrimination

Thanks and close

# **Appendix 3 – Key Stakeholder topic guide**

* Introduction (5mins)
* Introduce self, TNS BMRB independent research agency
* Explain aims of the discussion: *how to ask people about ‘caste’ without causing them any discomfort or harm; and what specific language and vocabulary to use, given ‘caste’ is described in different ways by different individuals*
* Explain GEO’s reason for commissioning the research: to explore methodological and technical issues (including possible ethical ones) and the costs associated with reliably and accurately measuring caste discrimination over time, using the terms ‘caste discrimination’ to refer to the type of discrimination which would be covered by the Equality Act 2010.
* The purpose of a survey*: the survey would be focused on aiming to understand how much caste discrimination occurs, and changes in this level over time.*
* Set ground rules for the discussion
* Explain recordings and anonymity of feedback
* Length 90 minutes
* Ways of talking about caste (15mins)
* How does it get referred to and discussed [in what conditions and by whom?]
* What would people recognise and respond to as terms
	+ generic (e.g. ‘low caste’, ‘high caste; Dalit; untouchable’)
	+ specific (clan; biraderi)
* Willingness to talk about caste and caste discrimination, both generally and in a survey context
	+ How might we make people more comfortable
	+ Differences by caste / religion?
* Response rate and reliability issues (25 mins)
* Who should participate in the survey – what is the appropriate sample population
	+ Reasons for views
* Would people participate in the survey
	+ Reasons for views - Why/why not
	+ How might we make them more comfortable, more likely to respond openly and accurately
* What are people likely to consider to be ‘discrimination’?
* Should we expect people to recognise discrimination if they experience it?
	+ Why/ why not
	+ Spontaneous and probe:
		- In voluntary work, education (predominantly bullying of pupils by other pupils), demeaning behaviour, violence
		- In relation to work (bullying and harassment, recruitment, promotion, task allocation, dismissal) and provision of services
* How might we encourage and support people to be as accurate as possible regarding experience of caste discrimination?
* Should we expect people to report discrimination in a survey, if they experienced it
	+ Why/ why not
	+ How might we encourage them to do so
	+ Why/ why not
* Will people feel able to answer questions honestly and directly. Spontaneous views, then offer examples of types of question
* Explain: for example, it could explore the following subject areas
	+ *awareness of caste – whether attribute caste to themselves*
	+ *whether feel that others attribute caste to them*
	+ *whether think about what another person’s caste is*
	+ *aware of discrimination based on caste*
	+ *own experience of caste discrimination*
		- For each: how would we make them comfortable/ able to be accurate
		- Should we expect people to exaggerate discrimination, and what they have experienced; or conversely, to understate it
* Ethical issues (30 mins)
* Do they see any ethical issues in surveying people about caste discrimination
	+ What might the effect(s) be of discussing caste discrimination
	+ What are the risks of discussing caste discrimination -spontaneous discussion at first, then prompt for following issues.
		- it will increase awareness of caste (and it is dying out)
		- it will increase caste discrimination
		- causes discomfort/humiliation
		- How to minimise these ethical problems
* Other issues (15 mins)
* Any other major challenges/issues in conducting a survey on caste discrimination
	+ Are there way they could help to ensure an accurate survey of caste discrimination, e.g. preparing people in their community to respond
	+ Other comments to inform the feasibility study

Thank and close

1. see http://the-sra.org.uk/research-ethics/ethics-guidelines/ [↑](#footnote-ref-2)
2. See *Hidden Apartheid – Voice of the Community: Caste and Caste Discrimination in the UK,* Anti Caste Discrimination Alliance (http://acdauk.org.uk/pdf/Hidden%20Apartheid%20-%20Voice%20of%20the%20Community%20-%20Exec%20Summary.pdf) [↑](#footnote-ref-3)
3. *Varna* is a Hindu religious concept. Its importance and role within the Hindu tradition has varied over time. [↑](#footnote-ref-4)
4. Also zat (Ballard, 1994). *Jati* can be defined as ‘occupational castes’, of which there are about 6,000 (Chahal, undated). In the Indian sub-continent, many people worked in the occupations linked to their *jati* (e.g. carpenter, shoemaker). *Jati* vary across the sub-continent, so, for example, the *jati* for carpenters in Gujarat is not the same as in Tamil Nadu. *Jati* should not be confused with *Jatt*: the first (pronounced with a long a) refers to occupational castes, the second (pronounced with a short a) is a specific occupational caste, agricultural workers from the Punjab. [↑](#footnote-ref-5)
5. More commonly spelt biraderi (e.g. Ballard, 1994); also spelt biradri (Chahal, undated). *Biraderi* (or brotherhood) is found amongst Muslims from the Indian sub-continent. [↑](#footnote-ref-6)
6. A former practice manager and solicitor at the same firm claimed that they had been discriminated against as a result of their marriage, as he was perceived to be of a lower caste than her. This was the first case brought to an employment tribunal under caste discrimination legislation. [↑](#footnote-ref-7)
7. We have used the term ‘South Asian ethnic heritage’ to cover (i) those who associate themselves strongly with a particular South Asian country (or ethnic group whose ‘homeland’ is in the South Asian area), and (ii) those with a much weaker self-association with a South Asian nation or ethnic group but who nevertheless recognise that their family background is South Asian. Both groups should be covered by any survey on the topic of caste-based discrimination. [↑](#footnote-ref-8)
8. While the work for the feasibility study was based on LSOAs in England and Wales this could be expanded to include the equivalent Scottish geography of data zones. Including Scottish data zones would not alter the conclusions or costings detailed in this feasibility study but any final design would need to include them in order for the survey to be fully representative of GB as a whole. [↑](#footnote-ref-9)
9. More details about this survey can be found on the website for the Office of National Statistics (ONS) at: <http://www.ons.gov.uk/ons/guide-method/method-quality/specific/crime-statistics-methodology/guide-to-finding-crime-statistics/crime-survey-for-england-and-wales--csew-/index.html> [↑](#footnote-ref-10)
10. The residential PAF includes addresses of second homes, properties that were residential but are now vacant (or even demolished), and properties that have not yet been built or offered to market. There are also a small number of misclassified addresses that are actually business/organisation addresses or that contain communal living quarters that are usually excluded (for practical reasons) from surveys of individuals. [↑](#footnote-ref-11)
11. At least for population estimation. Cluster samples are better if the purpose is to study between-area variations in experience. [↑](#footnote-ref-12)
12. The ‘effective’ sample size is used when estimating the precision of estimates and takes the sample clustering into account. [↑](#footnote-ref-13)
13. They may be smaller than this for proportion estimates that are very low or very high. [↑](#footnote-ref-14)
14. 95CI (proportion estimates in the 40-60% region) ***if n=2,000*** and accounting for effects of disproportionate sampling but not effects of sample clustering [↑](#footnote-ref-15)
15. For information, the confidence intervals in table 4.3 should be multiplied by √(n2/2000) where n2 is an alternative sample size. [↑](#footnote-ref-16)
16. It is worth noting that, if the second survey is two years after the first, the re-interview portion of the sample will contain no 16-17 year olds or immigrants since the first survey. Samples from these groups will be obtained from the ‘top up’ sample only and will need to be weighted upwards to reflect their lower overall sampling probability. [↑](#footnote-ref-17)
17. Sample clustering effects also tend to be smaller if the number of interviews per area is low so there is a statistical advantage to setting limited workloads as well as a practical advantage. [↑](#footnote-ref-18)