GSR quota sampling guidance

What to consider when choosing between quota samples and probability-based designs

Gary Brown (ONS), Natalie Low (NAO), Ramona Franklyn (DfE), Emma Drever (DfT) Catherine Mottram (DfT)

November 2017

Why is this guidance needed?

Quota sampling appears to be used more often in government for policy-related research. This is likely to be a response to requirements for faster and faster evidence allied with continuing pressures on research programme budgets, and rising costs in relation to some of the random probability sample surveys used in the past. It is vital that those of us involved in commissioning and conducting research fully understand the benefits of well-designed quota sampling as well as the unavoidable shortcomings.

In this context, this guidance on quota sampling is both necessary and timely. The aim of the guidance is not to produce an academic treatise – if you want detailed theory, please refer to Patrick Sturgis's recent work[1] – but to provide practical advice.

Decisions about sampling methods are complex, hence this guidance cannot provide definitive instructions for each situation, or proffer a simple step-by-step approach. Instead it provides information on what to consider so that decisions on sampling methodology are fully informed.

Message from the authors

We hope that this guidance provides you with information that you need to consider when making decisions around which sample approach to adopt.

Such decisions are rarely straightforward and are likely now more complex than ever because of declining response rates, emerging evidence of a weak relationship between response rates and non-response bias and developments of new online alternatives to traditional face-to-face household random probability surveys.

These issues are all explained in the guidance along with some information on where quota samples may be more appropriately used and how they can be improved. All this information should help to ensure that decisions around sampling approaches are the right ones for the evidence needed.

[1] A dizzying number of publications, including the inquiry he chaired to investigate inaccuracies in the opinion polls during the 2015 General Election campaign

What is sampling and what are the key methods[2]?

Sampling is how to select 'sample units' (usually people, households or businesses – hereto referred to as respondents) to participate in research. Sampling is cheaper and less burdensome than surveying every respondent in the target population of interest (a census).

All sampling methods can be categorized as either probability or non-probability. The distinguishing factor being, respectively, whether the probability of a respondent being sampled is known or unknown. This guidance focuses mainly on a key non-probability method – quota sampling – but also provides information on probability sampling methods as a contrast.

Quota sampling (non-probability)

Quota sampling takes its name from the setting of 'quotas' of different types of respondent to survey.

> *For example, in a face-to-face survey an interviewer would be instructed to interview pre-specified numbers of men and women, or of people from different age groups.*

The aim of quota sampling is to control the composition of the final achieved sample 'by design'. The design may replicate the true composition of the population of interest, have equal numbers of different types of respondent, or over-sample a particular type of respondent.

> *For example, the quota may require equal numbers from different ethnic minority groups even though their population proportions differ.*

Through appropriate design, a quota sample will be more representative (in terms of the estimates it produces) of the population of interest than:

- a self-selecting sample – where any respondent can choose to take part, usually those who feel very strongly about a particular topic
- a convenience sample – where interviewers can sample any respondent they choose and don't have particular quotas to meet, usually those who are easiest to contact

Quota samples vary in terms of quality. Section 5 sets out practical tips for improving the quality of quota samples in key areas such as: the source of the sample; the quotas themselves; survey procedures; and technical adjustments such as weighting.

Probability sampling

Probability sampling requires a list of every respondent in the population – 'sampling frame' – from which respondents are selected with a known probability. See appendix 1 for the main methods.

> *For example, a sample of households selected from a list of all the residential addresses.*

The aim of probability sampling is to generalise, or make inferences, about the whole population sampled from, and be able to quantify the precision of this inference. The statistical theory which underpins this only applies if the sample is unbiased – i.e. there is no systematic difference from the 'true' population – and each respondent in the population has a known non-zero probability of selection.

> *For example, if a particular type of respondent has a 1 in 10 probability of selection, their responses can be assumed to represent the 9 un-sampled respondents of their type.*

[2] See Appendix 1 for a quick guide to sampling methods

What are the issues with different sampling methods?

Bias and non-response

The main statistical issue is bias – whether survey results are systematically different from the population 'truth'. Non-response errors arise (in both probability and quota samples) when different types of potential respondent are more or less likely to take part.

For example, a quota sample of men and women to assess prevalence of mental illness may suffer higher non-response from people suffering mental illness, and therefore falsely under report its prevalence.

The problem with quota samples is the sampling method itself. Until a quota is filled, the interviewer (or in the case of some self-complete online surveys, the respondents' themselves) determines which respondent to sample. Interviewers will naturally target the easiest respondents to contact, e.g. those walking by in the street where they are interviewing or those who have a greater desire to participate in the research, i.e. those with a high 'propensity' to respond. This means that they exclude respondents not easy to contact or with a low propensity to respond. Therefore, some sections of the population will never be sampled, and if these have different survey characteristics, the results will be biased. Even if an interviewer makes an equal effort to interview every contactable respondent, the respondents with a high propensity to respond – by definition – are more likely to participate, so the achieved sample would still be unbalanced and potentially biased[3].

> *There are particular concerns about bias in online panel surveys. In recent years, many organisations have set up large online panels of potential respondents. They then use quota sampling to conduct large-scale surveys with a sample of these web panel members, which can result in achieved sample sizes of thousands of respondents within short timescales for relatively little cost. Panel samples may be particularly prone to bias due to self-selection in joining the panel, exclusion of the off-line population and panellists becoming "experienced" respondents to repeated surveys.*

The problem with probability samples is participation – unlike quota sampling, interviewers can't determine who to interview and must make repeated efforts to contact their sample and can't substitute refusals with any person who fits the quota. Non-response does occur and evidence suggests that it is getting harder achieve high response rates (though random probability samples tend to have higher response rates than quota samples). Non- response may be due to the lack of an accurate and up-to-date sampling frame, or simply propensity to respond. Importantly, a low response rate does not automatically mean a sample is biased. Many studies have shown a weak relationship between response rates and non-response bias (for more information on this see Appendix 3). However, if the non-response is not random – i.e. it is associated with the intended outcome of the survey research – and it cannot be adjusted for, it may lead to bias. In practice, it remains difficult to measure bias for random samples.

[3] Many studies have shown certain parts of the population, such as those from more disadvantaged socio-economic backgrounds or young men are less likely to respond

Note that survey estimates are prone to other types of non-sampling error which might be more substantial, in particular measurement error, which can apply to both random and quota sampling approaches.

Time and money

Probability samples of households have traditionally been conducted face-to-face, with the sample selected from the Postcode Address File[4]. Even with a multi-stage sample, i.e. interviews only within selected clusters, travel costs can be very high. Combined with the need to make repeated efforts to interview a sampled respondent, this methodology can be very expensive and time-consuming.

Although quota samples are traditionally considered quicker and cheaper than probability samples, if due care and attention is given to training interviewers, randomly selecting data collection points – i.e. where an interviewer samples their quota, then the costs and time spent implementing a good quality quota sample will also be high.

A couple of things to consider

Researchers should beware quota samples using the term 'random' in their name – they are still non-probability!

- *Randomly selecting a location – such as a shopping centre – for quota sampling will improve the quality, but the resulting quota sample should still be treated as such. Compare this method to cluster sampling, where the sampling frame is split into clusters, some clusters are selected randomly and a probability sample is selected within these clusters, which does result in a probability sample*
- *Selecting a quota sample of people, and then randomly allocating them into control and treatment groups would measure whether the control and treatment group have different outcomes – but only for the selected quota sample. For the evaluation to be generalizable to the population, a probability-based approach would be needed for the initial sample selection*

Recently there have been moves to develop online probability approaches, which offer cheaper and quicker ways of conducting random probability sample surveys, including:

- *Address Based Online Surveys (ABOS), where a random sample of addresses is sent a postal letter containing a weblink to complete an online survey. This method has now been adopted for the Community Life Survey (more information can be found in this presentation by Kantar Public)*
- *Random sample online panel, with the participants recruited by a random sample household survey (more information can be found in this presentation by NatCen)*

[4] There is no complete list or sampling frame of telephone numbers for the general population, hence probability surveys of households are only appropriate if it's a survey of a sub-group for whom sampling frames with telephone contact details are available. Therefore telephone surveys for probability sample household surveys are rare, though they are more common for business surveys where sample frames with telephone contact details are more commonly available

When should quota sampling be used?

Quota sampling can be useful for a broad brush picture of attitudes, behaviours or circumstances:

- understanding the range of concerns facing respondents about potential policy decisions
- testing a communications intervention before wider roll-out
- testing questions[5] in the pilot stage of a survey
- getting preliminary information on an issue
- helping to scope a policy intervention

However, quota sampling should only be used **in government** if there are compelling reasons for **not** using a probability sampling approach – which should be the default choice for survey research.

> *In areas like quantitative commercial research and political opinion polling quota sampling is used as matter of course – though for the latter this is now a matter of some controversy after recent polling data led to incorrect predictions of a number of election results (e.g. the 2015 General Election, Brexit)!*

You should particularly try to avoid quota sampling in the following situations:

- **When you want to draw robust conclusions about the population as a whole and quantify your precision.** This includes analysing differences between sub-groups or changes over time.
    - Quota sampling only provides robust information about the responding sample which cannot be generalised to the wider population – unlike probability sampling, which is designed to provide information about the wider population with a certain level of confidence.
    - Quota sampling results will not be truly representative of the population even after a great deal of effort in the design of quotas and weighting to account for all expected differences between domains, and only pseudo-measures of precision are possible.
- **If your research is about informing important policies, government forecasts, important/controversial/high profile debate or is relied on as evidence for a select committee.** Quota samples are not suited for measuring key issues due to inherent problems with respondent bias – participants are more likely to be willing, easily accessible and interested in the survey topic.

[5] There would need to be compelling reasons even for this use. At the Office for National Statistics, the Questionnaire Design team in Methodology use probability sampling, not quota, for selecting samples for piloting questions via cognitive interviewing.

Quick guide to: What sampling method to use?

In practice, decisions about sampling methods come down to many different factors and often need to be taken on a case-by-case basis, ideally in consultation with a relevant statistician or researcher.

To help guide often complex decisions about which sampling methods to use, we have come up with **four golden questions** plus an **additional consideration**. These simplify many nuances that exist between probability and quota sampling, but are a useful quick guide.

FOUR GOLDEN QUESTIONS

1. **How are you going to use the results?**
   Will a survey result that may be biased or not representative of the target population impact policy being developed or the decision-making? If the answer is 'yes', then the preference would be a probability sample

2. **Do you need conclusions about the wider population?**
   Is it important to can draw precise conclusions, based on the sample, about the population as a whole? If the answer is 'yes', then the preference would be a probability sample

3. **Are there technical reasons preventing probability sampling?**
   While probability sampling has benefits, there may be technical reasons why it isn't suitable:
   - it requires a suitable sampling frame, and this may not be available
   - it may, for some reason, be very difficult to achieve good response. Where response rates are low, the potential for bias is higher: but bias is ultimately caused if the profile of non-responders is systematically different to respondents, leading to flawed survey results

4. **How much scope is there for collecting a good quality quota sample?**
   Generally, the more information known about a population and subject area, and the more focused the research question, the more potential there is for a good quality quota sample

FINAL SANITY CHECK

- **How much time and money exists for the survey?**
  The amount of time available for research, as well as the size of the budget, are constraints all researchers work within: can the sampling approach be delivered within the allocated budget and in the time available?

  ***NB** Do not automatically rule out probability approaches just because budgets are small and/or timescales are limited, and do remember that it is possible to conducting a random probability sample household survey via online methods which is both quicker and cheaper than a traditional face to face household survey (more information is earlier in the document)*

See appendix 4 for a more detailed delineation of these questions.

What is best practice for quota sampling?

Once you have decided to use a quota sample, it is important to ensure both high quality inputs (**design**, **sample sources** and **training**) and carefully considered outputs (**weighting**, **reporting**). Quota samples can vary considerably in their quality, and this section highlights how you can ensure the most robust findings possible.

Design

The population characteristics used as the basis of setting quotas are most often demographic – gender, age, ethnicity, socio-economic group, working status etc. However, you can also use characteristics relating to opinions, attitudes and behaviours – political ideology, whether cycles, whether goes to the cinema etc. If you don't have high quality data to inform the sample design, you may need to consider a probability sample.

Best practice in setting quotas includes the following.

- **Think carefully about the characteristics used to set quotas.** Consider which demographic, attitudinal and behavioural characteristics, which can be identified for the whole population, are most likely to influence the research. Balancing these characteristics will reduce design bias. Data may be available from previous studies about which characteristics could influence the research and there may be census, administrative or survey data on how these characteristics are distributed across the population. Interviewers also need to be able to record quota characteristics for the sample.
- **There is no ideal number of characteristics / quotas**. When it comes to setting quotas, quality is more important than quantity. In general, the more quota categories, the more representative the design will be of the population and hence the less biased. However, there are practical limits to the number of quotas – too many and they may become unachievable as uncommon combinations will become impossible to fulfil.
- **Consider the use of "interrelated" or "interlocking" quotas.** Independent quotas match distributions of characteristics to the population – but can lead to unintended correlations. Setting "interrelated" or "interlocking" quotas for combinations of characteristics is more expensive (larger sample and more difficult) but improves weighting.
  For example, an independent quota sample selected by age and sex could have all younger people being female and all older people being male.

Once you have decided characteristics and quotas, you can introduce some randomisation by, for example, selecting locations at random and/or selecting a systematic quota sample (for example, every tenth individual). For quotas selected from sampling frames which include contact details of respondents, you can specify in advance a minimum number of re-contacts to be made before replacement.

Sample sources

In some situations you may be able to use a good quality random survey sample as a starting point for your quota sample. This will help to reduce selection biases. Use multiple sample sources can also help to reduce selection bias.

Training

Quota samples are biased towards people who are willing, easily accessible and interested in the survey topic. Bias is reduced for interviewer administered surveys, if interviewers convert refusals,

rather than replace. This means that interviews should practice and use refusal conversion techniques and collect information on refusal rates to assess bias.

In general, interviewer recruitment and training procedures should be rigorous – and make efforts to avoid 'contamination' between sampling techniques. Interviewers are generally skilled in either quota sample interviews or random sample interviews, but often not in both, so check whether interviewers also work on random sample surveys.

Weighting

For good quality research based on quota samples, a fully specified weighting model should adjust the results for non-response. In order to do this, information on refusal rates should be augmented by demographics of non-responders (i.e. paradata).

In addition, thorough and comprehensive post-stratification weighting should be applied to adjust, or calibrate, the achieved sample to known population totals for all the variables that could influence the research outcomes. Otherwise, mismatches between demographic characteristics between the sample and population could lead to biased results. Calibration will not completely remove bias due to propensity to respond, but will account for whatever can be explained by the population characteristics that are included in the weighting. See appendix 5 for more details.

Reporting

The important messages to convey are:

- why quota sampling has been used
- how to interpret the results

There may be financial and other pressures to use quota sampling and other non-probability methods – this is fine and perfectly understandable. However, the potential issues with quota sampling need to be made clear in reporting – and not hidden.

NB There are countless examples that go against these best practice reporting guidelines.

- **What can be said about the population as a whole?** Unless the research participants represent non-participants – for example, share the same attitudes, beliefs, have the same characteristics, behaviours, experiences – then quota sample results cannot be generalised to the population. Unfortunately, quota sampling will always be biased towards respondents willing to take part in the research – which may be very different and think very differently about the research topic than non-participants. Therefore, quota sampling research should only refer to survey respondents. *Any generalisations to the whole population should be strongly caveated with statements such as "If the quota sample was representative of the whole population, this would mean…" .*
- **Are the results 'statistically significant?** Any differences observed in the data need to be interpreted with caution. Try to triangulate findings with other evidence (such as other quantitative sources or themes from earlier qualitative research or comparable evidence from other organisations or countries) or review against the theory to see if the results make sense. There are a number of issues with using formal significance tests on quota sample data – bias, lack of known sampling probability, unknown population. Therefore, you must report any statistical significance testing of quota sampling results with great care. *Any reference to statistical significance should be strongly caveated with statements such as "If the data were generated from a probability sample, the following results would be… ".* NB Many market research agencies use significance tests as a matter of course, even though they are, strictly, invalid.

- **Can confidence intervals be used?** Again, confidence intervals do not apply for quota samples – as they also rely on statistical theory and a known probability of selection. Therefore, you must report on confidence intervals for quota samples with great care. *Any reference to confidence intervals should be strongly caveated with statements such as "If the data were generated from a random probability sample, the following confidence intervals would apply. As the data were generated from a quota sample, confidence intervals are unknown. "*

See appendix 6 for some examples of good and bad reporting of quota sampling research.

It is also important to be as transparent as possible about survey procedures in any technical reports or appendices. Survey contractors should be able to provide information on[6]:

- **people's willingness to take part in the survey** - If lots of people are refusing to take part in the research, this may mean that participants are more likely to be 'unusual' in some unknown way. Contractors should be able to provide information on feasible expected, and actual, refusal rates and outline their approaches to reduce non-response and bias.
- **Whether survey participants are different from non-participants with respect to the research topic** – Contractors should be able to provide insight into how key characteristics are distributed, and whether certain groups are more/less willing to respond.
- **Whether the sample profile matches that of the population**- it is worth checking that the sample profile matches the population for characteristics, which have not for some reason been used to set quotas. Mismatches in demographic characteristics may be an indicator of other hidden imbalances or bias, which may not be corrected by weighting or other adjustment techniques.

[6] Wretman, J. (2010) Reflections on Probability vs. Non-Probability Sampling. In M. Carlson, H. Nyquist & M. Villani (eds.), *Official Statistics - Methodology and Applications in Honour of Daniel Thorburn*, pp. 29-35

APPENDIX 1: Quick guide to key sampling methods

Probability sampling is the method used in most government surveys

What is it?

- A sample of respondents randomly selected from a list covering every potential respondent in the population – called a 'sampling frame'

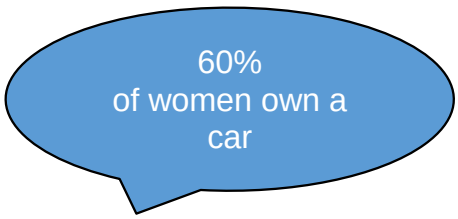
  For example, a list of all car owners

- Every respondent has a **known** and (usually) non-zero chance of being selected
- Respondents can be sampled proportionately or disproportionately – it is possible to over-sample some groups of particularly interest

What are the pros?

- Results/assertions can be generalized to the whole of the population of interest

  For example, women drivers



- Statistical tests can be used to test for accuracy of these generalizations
- Results are suitable for measuring differences between sub-groups OR changes over time

What are the cons?

- It *can* be expensive and time-consuming
- There can be challenges finding population lists or other sample frames
- It is getting harder over time to achieve good response rates – high non-response might indicate bias and makes findings subject to similar caveats as those from quota samples

Quota is a type of **non-probability** sampling method

What is it?

- Interviewers are given quotas of different types of respondents, usually aiming to reflect the population of interest

  For example, different quotas for men and women

- Interviews can be telephone, online or face-to-face – stopping commuters on their way to work
- Hybrid approaches – that have probability-based stages of selection, and also quota sampling stages, should still be treated as quota samples

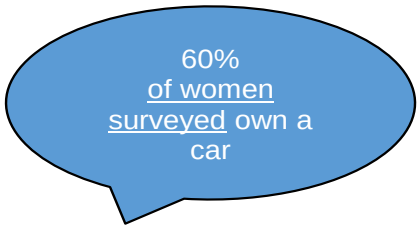
  For example, quota sampling at random locations

What are the pros?

- Surveys can be quicker, cheaper and/or easier as selected respondents are by definition 'easy to interview' which might not be true for randomly selected respondents from a list

What are the cons?

- There is an inherent issue of bias
- There is no theoretical basis – i.e. no justification – for generalizing results to the population



- There is no theoretical basis – i.e. no justification – for formal statistical tests of accuracy or significance
- Care is needed if results are used to measure differences between sub-groups OR changes over time

APPENDIX 2: Probability sampling methods

The main methods of selecting a probability sample are as follows:

- **simple random** – every respondent has an equal probability of selection, like picking names out of a hat
- **systematic** – respondents are selected from the sampling frame according to a fixed periodic interval (and a random start point), i.e. a random start between 1 and 10 and every 10th respondent thereafter (2, 12, 22 …) NB the sampling frame list is usually[7] ordered pre-selection by size, region, alphabetically, etc.
- **pps** – respondents are selected with a 'probability proportional to size' (pps), i.e. if a large sample of school children was required, larger schools would have a higher probability of selection than smaller
- **stratified** – the sampling frame list is split into a small number of non-overlapping respondent types ('strata') and a probability sample is selected within each group, ie this is equivalent to ordering the sampling frame for a systematic sample
- **clustered** – the sampling frame is split into clusters 'that occur naturally in the population', a sample of clusters is selected randomly and a sample (or census) of respondents within those cluster are interviewed, i.e. this is equivalent to schools being clusters of children for a pps sample NB clusters should be 'similar' and internally representative of the population
- **multi-stage** – a combination of the above, normally involving selecting a sample of clusters sample at the first stage, followed by another form of probability sampling within the selected clusters, i.e. also negates the requirement of a sampling frame for the whole population, as only needed for selected clusters

[7] The International Passenger Survey is an exception – a systematic sample is selected of passengers as they arrive at/depart from ports, airports, train terminals, where the only ordering is their arrival time

APPENDIX 3: Response rates and non-response bias

Response rates have traditionally been used as a proxy for how representative a survey is, and have generally been viewed as being inversely related to non-response bias. That is, a high response rate is seen as an indication that the survey has minimal non-response bias and vice versa. Response rates across most surveys have declined over the last 15 years, with more effort needed to achieve and maintain them, leading to rising fieldwork costs.

Recent studies[8] have explored the relationship between fieldwork efforts, response rates and non-response bias to assess whether increasing fieldwork efforts (and costs) is worth it. Such studies indicate that ***there is generally a weak relationship between response rates and non-response bias***. For example, Sturgis et al (2016) compared weighted survey estimates after different numbers of field calls (i.e. with different response rates) with the final estimates and found that higher response rates had only a marginal impact on estimates overall. However, there were occasional exceptions with some variables being particularly susceptible to non-response bias at lower response rates, such as volunteering rates.

There is therefore no easy answer to what response rate is 'good enough' for your survey. Bolling and Smith (2017)[9] outline 3 key questions to consider:

- **How vulnerable are your variables of interest to non-response bias?** Existing evidence should be used to try to identify these.
- **Will marginal increases in non-response bias compromise your conclusions?** Often trends are more important than point estimates, and non-response bias is likely to be constant over time.
- **Are there better ways of reducing non-response bias in your study?** For example, instead of trying to increase your overall response rate, target increased field efforts on selected non-respondents.

SOURCES:

Groves and Peytcheva (2008), *The impact of nonresponse rates on nonresponse bias, a meta-analysis*, Public Opinion Quarterly Vol.72 No.2, pp 167-189. Accessed at: http://askchisne.ucla.edu/chis/tac2015/Documents/SDSM%20TAC/Groves_Peytcheva%202008.pdf

Sturgis et al (2016), *Fieldwork effort, response rate, and the distribution of survey outcomes: a multi-level meta-analysis*. Accessed at: https://eprints.soton.ac.uk/398569/1/Paper_Draft_final_RR_April16_submitted.pdf

[8] For example, Groves and Peytcheva (2008) and Sturgis et al (2016).

[9] Declining response rates and their impact, presentation at the SRA on 29th June 2017, accessible at: http://the-sra.org.uk/wp-content/uploads/keith-bolling-and-patten-smith-declining-response-rates-and-their-impact.pdf

APPENDIX 4: Golden questions

<table>
<tr><th colspan="2">1. How are the results going to be used?</th></tr>
<tr><td>Use a quota sample if…✓

… aim is a broad brush picture of people's attitudes, behaviours or circumstances

• Sometimes, it's not vital to get an exact result… and it is not always necessary to be able to quantify accuracy

For example:
• testing questions in the piloting stage of a survey
• getting preliminary information on an issue
• helping to scope a policy intervention
• understanding the range of concerns people face about potential policy decisions
• testing a communications intervention before wider roll-out</td><td>Use a probability approach if…✓

… aim is precise estimates to inform important policies, government forecasts, to monitor and report trends or feed into wider public debate; or if the research is controversial or high profile

• Sometimes, it's important that results are representative of the population … and that accuracy can be quantified

For example:
• The Family Resources Survey collects data on Social Security benefits, assets, savings, housing costs and income – because it is based on a large probability sample and has good response, the results are representative of the UK and their accuracy can be quantified. Hence it is used to inform government policy and forecasts, as well as wider public debate
• Research used in Select Committees needs to use probability sampling to ensure results stand up to scrutiny: quota sampling is easier to discredit</td></tr>
<tr><th colspan="2">2. Are conclusions needed about the wider population?</th></tr>
<tr><td>Use a quota sample if…✓

... generalisation to the population is not needed, or indicative estimates for the wider population are acceptable with some

• Survey results do not always need to be generalised to the wider population

For example:
• research may be in early stages – ie scoping a policy or communications intervention – with light needed on issues re the intended approach before wider testing and roll-out
• for rare subgroups for which a sampling frame does not exist, probability sampling would not work but quota sampling could achieve sufficient numbers to analyse – and point to differences, without being able to test whether they are statistically significant at a point in time or in terms of change over time</td><td>Use a probability sample if…✓

… firm precise conclusions are needed about the population as a whole

• Most of the time, survey results need to be generalised to the population as a whole

For example:
• the National Travel Survey uses probability sampling and therefore results can be generalised to the whole UK population, ie "60% of women in the UK own a car"
• Understanding Society uses probability sampling, which enables small differences identified between subgroups to be tested for statistical significance
• the Labour Force Survey uses probability sampling, and so is able to accurately measure changes in UK unemployment rate changes over time</td></tr>
</table>

3. Are there technical reasons preventing probability sampling?	
Use a quota sample if…✓ … there is no sample frame (or contact details on the frame are unsuitable) or a probability sample would have non-response bias • If a sample frame is not available and cannot be built (due to time or resources) or the frame lacks contact information, a quota sample is a sensible alternative • If non-response bias would be an issue for a probability sample, a quota sample is a sensible alternative[10] *For example:* • *DfT recently conducted a quota survey of commuters as there was no available population list. Instead, commuters were intercepted at key locations, such as service stations (car drivers) and train stations (train passengers)* • *response rates are very low for random digit dialling, and bias can only be assessed by comparing the achieved sample to other sources of population information*	Use a probability approach if…✓ … you have a sample frame with suitable contact details • If there is a good sample frame, and can achieve a good response within budget, use a probability sampling approach *For example:* • *the Student Income and Expenditure Survey uses probability sampling because universities maintain databases of students. Hence, researchers are able to access emails, telephone numbers and postal addresses, facilitating an effective survey* • *with sufficient budget, a longer fieldwork period and more effective data collection modes – i.e. face-to-face surveys including an increased number of visits to each household – will ensure good response* • *information on non-responders from the sampling frame means potential bias can often be adjusted (i.e. weighted) for*
4. How much scope is there for collecting a good quality quota sample?	
Use a quota sample if…✓ … good knowledge of the population and subject being researched is available, or the research is focussed on a narrow policy area • A good quality quota sample requires good prior knowledge of population variables and sizes in order to set quotas *For example:* • *the DfT quota survey of commuters used a substantial body of prior knowledge about people's travel patterns (through the National Travel Survey)*	Use a probability approach if…✓ … only limited knowledge of the population and subject being researched is available, or the survey covers a broad range of unrelated areas • Without knowledge of population variables and sizes, appropriate quotas cannot be set so probability sampling should be used *For example:* • *the British National Survey of Sexual Attitudes and Lifestyles is probability as data on sexual behaviours is unavailable* • *the Community Life Survey covers networks*

[10] Cumming (2010): "Quota sampling is not an acceptable alternative to probability sampling with a reasonable response rate. It can however be an acceptable alternative if probability sampling is unlikely to produce a reasonable response rate"

• *a targeted quota sample to get a snapshot of attitudes towards a certain policy topic*	*and local action, volunteering, charitable giving, and well-being*
Final sanity check. How much time and money exists for the survey?	
First investigate all options around using a probability sampling approach within the time and budget…✓ … though quota sample surveys are usually cheaper and quicker, email and telephone surveys in particular can often be relatively cheap, even with probability-based methods • For some populations the difference in cost between a random probability approach and quota may be minimal, especially when serious efforts are adopted to improve the quality of the quota sample *For example:* • *NatCen has developed a probability-based online and telephone panel* • *The ONS omnibus survey uses probability sampling to achieve a random sample at low cost – researchers can pay to add questions (NB most omnibus surveys use random location quota sampling)* • *The Community Life Survey uses an address based online sample*	If after initial investigation a probability sample clearly cannot be used, a quota sample should not be the automatic first step…✓ … instead, researchers should: ○ discuss with policy colleagues what will be lost from **not** adopting a random probability approach ○ explore whether a high quality quota sample is possible • Discussions may result in commissioning a quota sample survey **or** changes to the research question and/or adjustments to the time and budget, resulting in commissioning a probability survey *For example:* • *if the problem is a lack of sample frame the first step could be building one, or even doing nothing at all since the evidence requirements cannot be met*

APPENDIX 5: More detail on post-stratification of quota samples

Simple post-stratification – to be avoided	***So how should post-stratification be used?***
Assume an achieved quota sample from a targeted rare population happens to be 100 male and 200 female, with a population of 400 of each gender. Simple post-stratification would assign a weight of 0.5 to the female responses. • Problem – a weight of 0.5 assumes that the samples are drawn randomly from the populations, and this is precisely what we know is not the case for quota samples • Solution – the first 100 sampled of each sex would be likely to have the same 'propensity to respond' and so would probably be similar in other characteristics, but the second 100 women sampled would be likely to have a lower propensity, so would not be similar, but weighting would equate them to the first 100 men. The unpalatable solution would be to only use the interviews from the first 100 women for population results. Rather than assume a quota sample is random, it is more realistic to describe it as the "lowest hanging fruit". What the quota sample actually delivers is the sub-population with the lowest probability of refusal.	To ensure data generated through a quota sample are as representative of the general population as possible, all relevant stratification domains should be used. For example, if collecting information known to be influenced by age, sex, ethnicity, education, and occupation – then the results should be post-stratified to known population totals for each of these variables: which implicitly relies on the quota sampling questionnaire collecting enough demographic information to properly post-stratify. • Knowing all key variables impacting on response could be through a pilot survey, previous studies or expert analysis. • Post-stratification requires a highly regarded independent data source. While this may be feasible in some studies, in others it may be less straightforward. It is also important thing that the survey data measures the characteristic in the same way as the benchmark data, and that there is sufficient sample size for each characteristic.

Case study: National Sexual Attitudes and Lifestyle Survey

Researchers used data from the National Sexual Attitudes and Lifestyles Survey (NatSal) to post-stratify a (web panel) quota sample survey on sexual behaviour. The aim was to reduce the potential bias of the quota sample survey.

The use of post-stratification techniques enabled researchers to take account of characteristics known to be related to the research topic, but which would have been difficult to set quotas for - for example, because they related to sexual attitudes, rather than straightforward demographics.

Despite post-stratification, the results of the quota sample survey and NatSal were significantly different on most key measures. The conclusion was that the quota sample survey must be biased to some degree – i.e. the quota sample survey participants must be different in some way to non-participants, despite researchers' best efforts to control findings.

While it is possible that the mode (web), nature of the sample (panel) and topic area (sexual attitudes) exacerbated these biases, it is useful to bear in mind that biases may be present when quota sample research is used (more information in this SRA pamphlet)

APPENDIX 6: Reporting do's and don'ts

<table>
<tr><th colspan="2">What can be said about the population as a whole?</th></tr>
<tr><td>Do say…✓

Results showed:
"60 per cent of women surveyed owned a car".
"50 per cent of young people in the sample had never voted".
"25 per cent of people surveyed thought that climate change is one of the biggest problems we face".
If the quota sample was representative of the whole population, the results would show:
"60 per cent of women own a car".
"50 per cent of young people have never voted".
"25 per cent of people think that climate change is one of the biggest problems we face".</td><td>Don't say ×

Don't use the present tense without clarification:
"60 per cent of women own a car".
"50 per cent of young people have never voted".
"25 per cent of people think that climate change is one of the biggest problems we face".</td></tr>
<tr><th colspan="2">Are the results 'statistically significant'?</th></tr>
<tr><td>Do say…✓
"In this report, we used a two-sample t-test to assess differences between groups, and highlight those where the p value<.05. However, due to the quota sampling methodology used, this is not an exact test of whether differences are statistically significant, i.e. indicative of real changes in the wider population[11].
"Of those surveyed, men were more likely than women to own a car (65% compared with 60%)".
"Survey results were that young people were less likely than older people to be satisfied with the service".
"Concern about climate change was higher amongst individuals surveyed in 2016, compared to those surveyed in 2010."
"Satisfaction levels with the service for young people responding to our survey were similar to those for older people."</td><td>Don't say…×
"In this report, we tested differences between groups using a two-sample t-test, and highlight as statistically significant those where the p value<.05."
"Men are significantly more likely than women to own a car (65% compared with 60%)"
"Young people are less likely than older people to be satisfied with the service".
"Concern about climate change has increased significantly between 2010 and 2016".
"Satisfaction levels with the service do not vary significantly with age."</td></tr>
</table>

[11] A significance test estimates whether differences are statistically significant but this is an estimation only

Can confidence intervals be used?	
Do say…✓ • "While it is difficult to quantify the accuracy of these results in the way that you can for random variability in probability sampling, this does not mean that errors/bias do not exist." • If the data were generated from a random probability sample, the following confidence intervals would apply. As the data were generated from a quota sample, confidence intervals are unknown. ”	Don’t say…× • “Survey findings are subject to a margin of error as they are based on a sample. Findings were statistically tested at the 5% significance level, and only differences which were statistically significant at that level are referred to in the text unless otherwise stated.” • "60 per cent of women own a car" (±3%)". • “We estimate that if all adults in the population had been asked, the proportion thinking that climate change is one of the biggest problems we face would probably be between 23% and 27%.”