Overview and recommendations

The main points and recommendations from the Guidelines are summarised in the following. These are organised in four sections, each reflecting the content of the substantive chapters of the Guidelines. The section corresponding to Chapter 4 (measuring trust) is outlined in more detail as this chapter provides the most detailed and prescriptive recommendations relating to the collection of trust data.

1. Concept and validity

The relevance of measures of trust is not in doubt. Measures of interpersonal trust – particularly generalised trust – are of fundamental importance to assessing the well-being of societies, to measuring social capital, and to understanding the drivers of other social and economic outcomes. This is reflected both in the large and growing literature on the drivers and consequences of generalised trust and in the wide range of national and international initiatives that include generalised trust as an outcome of interest, ranging from national initiatives to measure well-being such as those of the United Kingdom and Israel through to the UN Sustainable Development Goals. Similarly, there is a high level of interest in institutional trust, which is reflected both in measurement initiatives and in a large and varied academic literature. In addition to being important for measuring well-being, understanding institutional trust is essential to understanding government effectiveness and the functioning of democratic systems of government.

The accuracy of trust measures is less clear. In general, while evidence on the validity of measures of generalised trust is strong, there is relatively little evidence focusing on the validity of limited trust measures. Measures of generalised trust perform well in terms of face validity, construct validity and convergent validity. This holds whether the measures are assessed at the cross-country level or at the level of individual responses. While there is some question over the test-retest reliability of some measures of generalised trust at the individual level, country-level results are highly reliable across different data sources and over time.

As with all intangible concepts, measuring generalised trust raises a number of issues about respondents' interpretation of the question in front of them and about their subjective judgement, but these issues are not intractable. There is good evidence that, despite these issues, questions on generalised trust produce valid data, and extensive research is providing new insights into the remaining measurement issues. In fact, it is likely that more is understood about the validity of measures of generalised trust than is the case for many more traditional elements of official statistics. From this perspective, measures of generalised trust can be considered as fit for purpose and should be measured in official surveys where relevant.

The picture for measures of institutional trust is more mixed than is the case for measures of interpersonal trust. While institutional trust measures generally perform well in terms of construct validity, the situation is less clear with respect to face validity and convergent validity. There is thus reason to believe that such measures might be biased in some circumstances, and for a number of key aspects of validity there is simply no evidence one way or the other. For this reason, despite the policy relevance of measures of institutional trust, it is less easy to state that they are fit for purpose within the context of official statistics. This does not mean that such measures have no place in the official statistical system, however. Rather, it suggests that such measures should be considered more experimental and should be implemented in contexts where their experimental status is clear; this is particularly important for national statistical offices.

On the other hand, the relevance of such measures suggests that they should be a high priority for further research, both in the academic community and within national statistical agencies. Many of the key methodological questions regarding the accuracy of measures of institutional trust will require the sample size and response rates that only national statistical offices are able to deliver.

2. Methodological considerations in the measurement of trust

This section examines the impact of different sources of bias and measurement error on trust questions. While trust measures are more sensitive to response biases than more objective measures (such as educational attainment or life expectancy), these biases are also likely to occur in other self-reported measures that are already being collected by NSOs. Although it is essential to be aware of these biases and of the most appropriate question and survey design strategies to mitigate them, the existence of measurement error *per se* is not an argument against gathering data on trust. No matter which approach to question design is adopted by data collectors, standardisation is critical to ensure meaningful comparison over time and between groups and countries.

Question wording

The evidence on question wording (especially that drawn from split sample experiments) shows that good question wording matters for results.

- Question wording should avoid referring to concepts other than trust and be specific and precise to the situation of interest.
- For interpersonal trust, a neutral question wording is recommended: data collectors should refrain from referring to "caution in dealing with other people", as this wording can prime more vulnerable groups to report lower trust.
- For institutional trust, specifying what institutions are expected to do can make a difference to respondents in some cases.
- Overall, question wording should be precise enough to be understood by respondents, without getting into subtle nuances that might also pose problems for translatability across countries. If the concepts that different questions try to capture are too narrowly related, respondents might have difficulty differentiating between them (e.g. trust vs. confidence).

Response formats

The way response options are presented can have a significant impact on the distribution of responses.

- For trust items, a numerical 0-10 scale with verbal scale anchors is recommended, as it allows for a high degree of variance in responses, increases overall data quality and facilitates translatability across languages.
- The response order should be presented consistently (i.e. 0-10 instead of 10-0) in order to minimise mental switching between positive and negative normative outcomes.
- Verbal descriptions of the scale anchors should represent absolute responses (e.g. *completely/ not at all*) to minimise acquiescence bias and socially desirable responding and to allow for the full spectrum of possible responses.

Survey context

Trust measures should be considered within the broader survey context in which they are placed. As with the standardisation of wording and response formats, consistency of order within question modules across surveys and over time is essential to guarantee the quality and comparability of trust measures.

- Since order effects occur most often when two or more questions deal with the same or closely related issues, trust items should either be separated within the survey as much as possible or buffered by intervening text.

- Whenever lists of trust items are used, a survey should move from a broad to a narrow level of specificity within a group of questions, e.g. by placing items about generalised trust before questions about limited trust.

- Generally, trust questions should not be asked immediately after items that are likely to elicit strong emotional responses or that refer to experiences with other people or institutions.

- Questionnaire designers should equally reflect on the potential effect that trust questions themselves can have on subsequent items, in particular those dealing with similar content.

- To minimise the impact of holidays, seasons and elections, data collection is recommended to be spread throughout the year or at least over multiple weeks.

Survey mode

Evidence suggests that trust questions can be sensitive, triggering respondents to answer in socially desirable way or be unwilling to answer at all. This is especially true for measures of trust in institutions.

- Self-administered surveys, compared to interviewer-led ones, perform better in terms of minimising social desirability.

- In all survey modes, sensitivity-related response biases can be reduced by decreasing the respondent's concerns about data protection (e.g. via confidentiality assurances) or by controlling the survey situation (e.g. not having enumerators give out information about their own social identity).

- If face-to-face interviews are the only option, the use of innovative interviewing methods such as the sealed envelope or unmatched count technique could be explored.

Response styles and cultural context

Cross-cultural response styles are very difficult to verify externally against a common standard or actual behaviour. Even where the existence of response styles has been established, they do not necessarily harm overall data quality.

- If data producers want to mitigate the possibility of response style bias, they should, rather than relying on *ex post* statistical adjustment techniques, focus on designing the questionnaire so that items are as simple, easy to interpret and minimally burdensome as possible.

- The overall survey design (including its length and how it is introduced) needs to pay particular attention to respondent burden, motivation and fatigue in order to maximise data quality.

● Question formats that are more prone to response biases should be avoided: for example, *agree/disagree* and to a lesser degree *yes/no* response formats are more likely to prompt acquiescence.

Issues for further research

Further research is needed on both institutional and interpersonal trust, but especially on the former, for which there is very little methodological evidence available.

● First, with regard to question wording for institutional trust, experimental testing should be used to establish which specifications (e.g. *to act in the national interest*, or *to improve the life of someone like me*, or *to do what is right*) matter the most for which institutions. Ideally, these experiments should be carried out across more than just one country.

● Second, while it has been suggested that the use of a mixture of positive and negative statements can mitigate both *yay* and *nay* saying, this approach needs to be further tested to rule out the risk of confusing respondents when the same scale end presents something positive in one item and something negative in a following one.

● Third, with regard to order effects, it is not yet clear in which cases these occur for trust questions. More targeted methodological research is needed to discover what type of question or what type of context triggers which effect in order to further inform survey design. While there is some evidence that transitional text between questions can act as a buffer to mitigate order effects, various text versions should be tested for their impact on trust questions.

● Fourth, more research that validates response styles from different cultures against external references, such as actual trusting behaviour in real life or experimental games, would enrich the current body of cross-cultural trust research.

3. Measuring trust

The aim of this section is to outline best practice in the measurement of trust. This covers decisions about what to measure and the best approaches to measurement. Issues of sample design, survey design, data processing and questionnaire design are covered here.

Planning for the measurement of trust

● Decisions about what to measure should always be grounded in a clear understanding of user needs. Important questions to consider include: i) What are the policy questions?; ii) Is the trust content being proposed appropriate to respond to the policy questions?; iii) Does the measure proposed allow monitoring changes over time or comparing population groups?; iv) What population groups are of greatest interest to the user?; v) Does the user's interest lie in comparing outcomes of different groups or in understanding the relationship between different aspects of trust?; vi) Is the user's primary interest in generalised trust, limited trust or institutional trust? If the focus is on the latter, which institutions are of primary interest?; vii) What are the frequency requirements of the users to monitor changes over time?; and viii) What within-country comparisons are required, such as geographic level?

● It is imperative to consider not only how best to measure trust *per se*, but also what other measures should be collected alongside measures of trust for analytical purposes. These should include: i) Age; ii) Gender; iii) Marital status; iv) Household type; v) Presence of

children; vi) Household size; vii) Geographic information; and viii) Migration status/ Country of birth/Year of arrival.

- In addition to the demographic measures identified above, which can be considered essential, a number of additional variables may also be useful: i) Language spoken at home; ii) Living in urban/rural areas; iii) Income; iv) Wealth; v) Employment status; vi) Educational attainment; vii) Health status; viii) Social contact and networks; ix) Civic engagement and governance; x) Personal security and victimisation; xi) Subjective well-being; xii) Ethnic identification; and xiii) Religion.

Survey and sample design

Sampling

- Responses to questions on trust are inherently personal, and consequently the unit of measure must be the individual. This implies that the sampling frame must produce a representative sample of individuals or households as if all individuals are personally interviewed.

- In general, measures of trust would be collected for the entire adult population (aged 15 and older).

Frequency of data collection and duration of enumeration

- It is not possible to provide specific guidelines for how frequently measures of trust should be collected that cover every contingency, since the range of possible data uses is large and the frequency at which data are needed will vary depending on the intended use and on the type of measure in question.

- For the purposes of monitoring well-being and for assessing trends in social capital, an annual time series should be regarded as the minimum in terms of frequency of enumeration.

Duration of enumeration

- The duration of the enumeration period (i.e. the period of time over which information is collected) is important for measures of trust. Unlike measures of educational attainment or marital status, for which it does not usually matter at what point during the year the data are collected, the precise timing of the collection period might have an impact on measured trust.

- Ideally, enumeration of trust data would take place over a full year and would include all days of the week, including holidays. This would ensure that measures of trust provide an accurate picture for the whole year. Where a year-long enumeration period is not possible, enumeration should, as far as is possible, be spread proportionately over all days of the week.

Sample size

- Large samples are highly desirable in any survey: they reduce the standard error of estimates and allow both a more precise measure of trust as well as a greater degree of freedom when producing cross-tabulations and analysis of results for population sub-groups. With measures of trust, sample size is particularly important because of the relatively small changes in trust associated with many areas of analytical interest.

Survey mode

- In terms of data quality, Computer-Assisted Self-Interviewing/Computer-Assisted Personal Interviewing (CASI/CAPI) with show cards should be considered best practice for collecting trust data. The presence of an interviewer allows for a strong rapport to be built with the respondent, while show cards help with data quality.

- The confidentiality provided by CASI sections to the interviewing should help address respondent reluctance to provide accurate answers to potentially sensitive questions.

- Where other modes are used, it is important that data producers collect information to enable the impact of mode effects to be estimated. National statistical agencies, in particular, should consider experimentally testing the impact of the survey mode on responses to the core measures of trust and publish the results along with any results from CATI or CASI surveys.

Survey vehicle

- Where trust, governance or social capital are the key area of interest, it may be appropriate to build a special module focused specifically on trust. This is especially the case where the use of trust data focuses on measuring social capital or on evaluating governance.

- As trust measures are of analytical interest in a broad range of different contexts, a limited range of trust questions can usefully be included in a wide range of surveys.

Question placement

- Important trust questions should be included in the core section of the survey. Although it is not possible to place trust questions at the start of every survey, the effect of bias due to context effects can be limited if trust questions are included in a fixed portion of the survey questionnaire. While this does not eliminate bias, it will not affect analysis of differences in levels across population groups or over time.

- Trust questions should not be placed immediately after questions that are likely to prime respondents with regard to trust, or that respondents might use as a heuristic for determining their response to the trust question. This includes questions on social contact, victimisation, political beliefs, risk or insecurity. The best questions to precede trust questions are demographic questions.

- Transition questions should be used to refocus respondent attention. However, it is important to consider the risk that transition questions might introduce their own context effects. For example, drawing attention to a respondent's personal life may lead them to focus on personal relationships rather than on strangers when answering subsequent questions about interpersonal trust. Development of effective transition questions is a priority for future work.

- Introductory text should be used to distinguish between question topics. Well-worded text preceding each question or topic can serve as a buffer between measures of trust and sensitive questions. Further cognitive testing or experimental analysis of the impact of different types of introductory text would be of high value.

Question order

- In terms of ordering question modules, evidence suggests that moving from the general to the specific is the best approach. For interpersonal trust, a question on generalised trust should be placed ahead of more specific questions relating to limited trust. Also,

questions on interpersonal trust should be asked before questions on institutional trust, which tend to be more specific and are also likely to have a significant priming effect.

- Questions on trust in institutions should proceed from better-known institutions to more obscure ones. Where comparisons between levels of trust in different institutions are important, the order in which questions are presented should be randomised for each respondent. If this is not possible for the whole sample, pilot testing should involve the randomisation of question order so that the size of any bias in measured trust is known.

Translation

- Initial translation should ideally be carried out by at least two independent translators who have the destination language as their mother tongue and who are fluent in the source language. Translators should be informed about the goal of the study and be familiar with the background, origin and technical details of the source questionnaire, as well as with the nature of the target population. As with any survey design, cognitive interviewing and field testing should be undertaken, with the results reviewed before the full survey goes into the field.

Questionnaire design

- These Guidelines provide five prototype question modules for the measurement of trust. Module A contains a set of **core measures**, which include a single primary measure of generalised interpersonal trust that is intended to form the baseline for international comparisons, and is the highest priority for inclusion in any attempt to measure trust.

- Module A contains those questions for which evidence of validity and policy relevance is the strongest and which are most apt to achieve some degree of international harmonisation. Unlike all of the other question modules included in these Guidelines, the core module is intended to be used without significant amendment and in full.

- Modules B to E are focused on different approaches to measuring trust. These modules are not intended to be used in their entirety or unaltered, but provide a resource for national statistical agencies that are developing their own questionnaires.

Survey implementation

- Interviewer training is crucial to the quality of responses in any survey. To manage risks around respondent attitudes to questions on trust, interviewers should be well-briefed, not just on what concepts the questions are trying to measure, but also on how the information collected will be used.

- Evidence suggests that measures of trust are relatively non-problematic for respondents to answer. Item-specific non-response rates for interpersonal measures of trust are similar to those for marital status, education and labour market status, and much lower than for those on household income. Non-response rates for questions on institutional trust are somewhat higher, but still lower than is the case of income. This suggests that, in general, trust questions are not perceived as problematic by respondents.

- Normal data-cleaning procedures include looking for obvious errors such as transposed numbers, duplicate records, loss of records, incomplete responses, out-of-range responses or failure to follow correct skip patterns. Some issues are of particular relevance to trust data. In particular, where a module comprising several questions with the same scale is used, data cleaning should also involve checking for response sets (see Chapter 3).

4. Output and analysis of trust data

The output and analysis of measures of trust is inherently complex and difficult to reduce to a succinct series of recommendations. This section draws together general advice on the output of trust data along with some information on interpreting results and on issues involved in undertaking analysis of trust data.

Reporting trust data

Analysts tasked with reporting trust data have an important communications role to play and should take into account the intended target audience and its needs. There are several ways to output trust data, each with unique pros and cons.

- Trust levels can be presented by reporting frequencies in each category, the proportions above or below a given threshold, or central tendency measures (i.e. *mean*, *median*, *mode*).
- Some rules of thumb for best reporting practice include refraining from arbitrary labels for thresholds (e.g. *high*, *low*) and complementing mean levels with information about the distribution of data, such as the standard deviation.
- Changes over time can be monitored by tracking changes in mean trust through time series or by calculating changes in the mean score over various points in time.
- Group differences can be examined by presenting group differences over time, relative to a given threshold, or by showing the (absolute or percentage) differences in the proportion of respondents who have selected a specific answer.
- Both sample size and standard errors should be reported alongside group means.

Interpreting trust data

Essential questions for the interpretation of trust data deal with what should be considered a *small* or a *big* difference between observations in real-life terms, and the extent to which observed differences are influenced by measurement artefacts and errors. While these Guidelines provide an initial attempt to document the magnitude of differences (between population groups, between countries, over time) encountered up to now with existing data, many gaps remain, and knowledge on the overall data universe of trust will remain limited until higher quality and more frequent data become available. The magnitude of differences between and within observations (over time), can be influenced by a variety of factors that should be factored into any interpretation exercise. These include:

- the limits imposed by the response scale
- issues of reverse causality
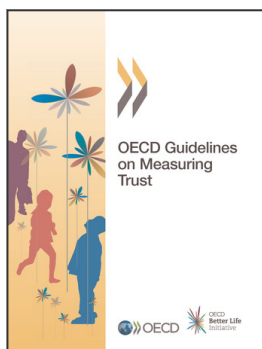- the possible impact of culture on trust estimates.

Analysis of trust data

Sound analysis of trust data requires access to data from which causal inferences can be made and that the relevant covariates, including standard demographic and control variables, are ideally collected in the same survey. The choice of the unit of analysis – i.e. whether to consider variables at the individual or country level – matters. In addition, community characteristics often influence trust levels, and aggregate trust levels influence individual-level well-being outcomes.

Keeping common econometric challenges in mind when working with trust data is very important. These include omitted variable bias, over-identification, reverse causality and shared method variance.

5. Next steps

These Guidelines do not aim at providing the final word on the measurement of trust or at developing a formal international standard for measuring trust. While formal statistical standards are an important part of official statistics, it is appropriate to produce them only when the measure in question is well understood and when there is a well-developed body of statistical activities pursued by NSOs, conditions that are not currently met in the case of trust. These Guidelines aim, rather, to bridge the gap between, on the one hand, the current situation of scattered measurement in official statistics and the more widespread – but still inconsistent measurement elsewhere – and a formal standard, on the other. By encouraging NSOs to collect trust data more systematically, i.e. by including a core set of internationally comparable questions in their surveys, the Guidelines aim to support the development of the evidence base that might in the future underpin decisions about developing an international statistical standard for measuring trust.