Chapter 5

Output and analysis of trust data

This chapter provides guidance for data producers, media and researchers on how to deal with trust data once they have been collected. The chapter discusses the planning of statistical releases of trust data for a range of target audiences and highlights practical examples of various reporting styles. Advice is provided on the interpretation of results and analysis of microdata, including their challenges.

5.1. Introduction

After having discussed validity and measurement and the logistics of data collection in Chapters 2 to 4, the present chapter closes the circle by considering *how to use* trust data once they have been gathered. In other terms, what is best practice in releasing trust data? What key interpretative issues should be kept in mind when examining the data? What basic analytical techniques should (and should not) be applied?

The chapter has three main sections. Section 5.2 starts with the reporting of trust data and highlights the different types of end users of statistical releases, whose needs and capacities data analysts at national statistical offices have to be aware of. This section focuses on the type and depth of information that is most appropriate for different audiences and, to showcase best practice, provides examples of how national statistical offices (NSOs) and other data producers have released their results.

The chapter then turns to the interpretation and analysis of trust data. Section 5.3 discusses how to evaluate differences in trust between observations (e.g. groups of people or countries, and over time) and how to determine what to consider as *big* rather than *small* differences.

Section 5.4 discusses the empirical analysis of trust data – considering trust both as a driver of other outcomes of interest (e.g. economic growth, subjective well-being, health-seeking behaviour) and as a valuable outcome in its own right, for which we want to identify drivers (e.g. socio-demographic factors, life experiences, experiences with government). Basic analytic methods that are appropriate for selected cases are also discussed.

Section 5.2, on reporting, will be of most direct interest to statistical analysts working for producers of large-scale data, such as NSOs, as it concerns the kinds of outputs and analyses that they are most likely to report for different types of audiences. Section 5.3., dealing with interpretation, and Section 5.4, on analysis, are relevant to the broader research and policy audience concerned with trust.

Generally, this chapter's messages apply equally to measures of trust in both other people and institutions. Therefore, unless otherwise specified, the term trust as used in this chapter refers to both types of trust.

5.2. Reporting trust data

Ideally, NSOs would regularly collect and release high-quality trust data from large and representative population samples. As for other outcome measures, key users of trust data include policy makers, public service providers, civil society organisations, researchers and the wider public – all of whom may have an interest in whether, where and when conditions in society are improving or deteriorating. Given the heterogeneity of these audiences in terms of their prior knowledge and quantitative literacy, data producers should ensure that their releases mean something to the general public as well as to specialists (New Economics Foundation, 2009).

Statistical analysts in NSOs therefore have an important communication role, as they are the ones responsible for tailoring and presenting statistical information in the most appropriate way for these multiple audiences. Statistical releases can be thought of in terms of having different layers, or pages (see Table 5.1 for an illustration). The ***front page*** will mainly provide the general public as well as policy makers with a quick overview of the data. This front page should concisely yet precisely report levels of a single headline measure, or at most a small set of key measures, for strategic communication purposes and in order to avoid overloading laypeople with information. The core trust measures based on the prototype modules recommended by these Guidelines (see Annex 2) are ideally suited to feature on such a front page. Statistical staff should further include a brief commentary for the media, reiterating the importance and usefulness of measuring trust and describing the indicators used, the questions they are based upon, any chosen reporting thresholds, as well as the sample size and response rates. The importance of such commentary cannot be overstated: it immensely increases the likelihood that statistical releases are taken up by media and, given that news outlets often reprint the statistical analyst's commentary directly, they will also ensure that key features of the data are accurately reported. This is all the more important since very often the general public and policy makers will not read the statistical release itself, but are more likely to rely on a variety of media reports, including newspaper articles or social media posts.

Table 5.1. **Suggested structure of statistical releases of trust data**

	Purpose	Target audience	Specific content	General content
Front page	Give concise and precise overview of current trust levels	• Policy makers interested in a quick overview- General public • Most media outlets	• Headline measure levels • Commentary for media	All pages should include full descriptions of the question wording, answering scales, the sample size and any chosen thresholds
Second page	Provide more detailed information (e.g. trends over time, distributions across different population groups)	• Government analysts • Media outlets interested in a deeper analysis of trust	• Additional trust measure levels • Breakdown by socio-economic groups (including standard deviations) • Previous trust measure levels	
Third page	Facilitate access to and use of microdata	• Researchers and analysts wanting to conduct further analysis	• Microdataset • Questionnaire	

StatLink http://dx.doi.org/10.1787/888933584241

The ***second and third pages*** of statistical releases cover trust data in greater detail. The second page usually targets policy analysts, journalists and researchers who want to analyse trust beyond singular "quick and dirty" headlines. This analysis might look at trends in trust and its distribution, i.e. how trust has changed over time, and whether trust varies between groups (from different demographic groups to regions within a country up to the international level). This page hence expands the front-page content by reporting on relevant trust items other than the core headline measures (i.e. the prototype question Modules B-F described in Annex 2). Data can be further broken down by different population groups (e.g. by age, gender, employment status, income quartile, marital status, region of residence, and other factors identified as relevant for the specific country context), along with information on sample size, standard deviations, and the results of any significance tests for group differences that have been conducted. If available, trust data for earlier points in time should also be included for comparison purposes. In case the question wording, response scale or survey vehicle have changed over time, this should be pointed out.

The third page of data releases is designed for researchers and policy analysts who want to conduct deeper analysis with the trust data. In addition to directions on where and how to

access the trust microdata (ideally microdata should always be made publicly available), this page features a full description of the dataset and a copy of the full questionnaire, so that researchers can assess the order in which trust items were placed within the wider survey and take joint distributions with other relevant variables into account.

In general, audiences will find data releases easier to digest if there is some degree of consistency between the way the latest trust levels are reported on the first page of statistical releases and the way more detailed trends are described on the second page. Different ways of presenting and visualising trust data in practice are illustrated in Table 5.1.

The front page of statistical releases: Reporting levels of trust

The level of a trust measure essentially answers the question of whether the amount of trust in the population of interest is high or low. There are three main approaches to report that level. First, the frequency of responses can be described by category, i.e. the proportion of the population who select each response category of the respective response scale. The other approaches summarise the data for the purpose of easier presentation: this can be done either in relation to one or more threshold (e.g. the proportion of the population falling above or below the chosen threshold) or via a measure of central tendency (e.g. the mean, the median or the mode).[1]

The section below outlines the advantages and disadvantages of each of these three approaches, with visual examples of relevant statistical releases by either NSOs or other data producers. While data producers could refer to this for general advice on what type of reporting is more or less useful for trust measures, the final choice on presentation will also depend on the exact purpose of the release at hand.

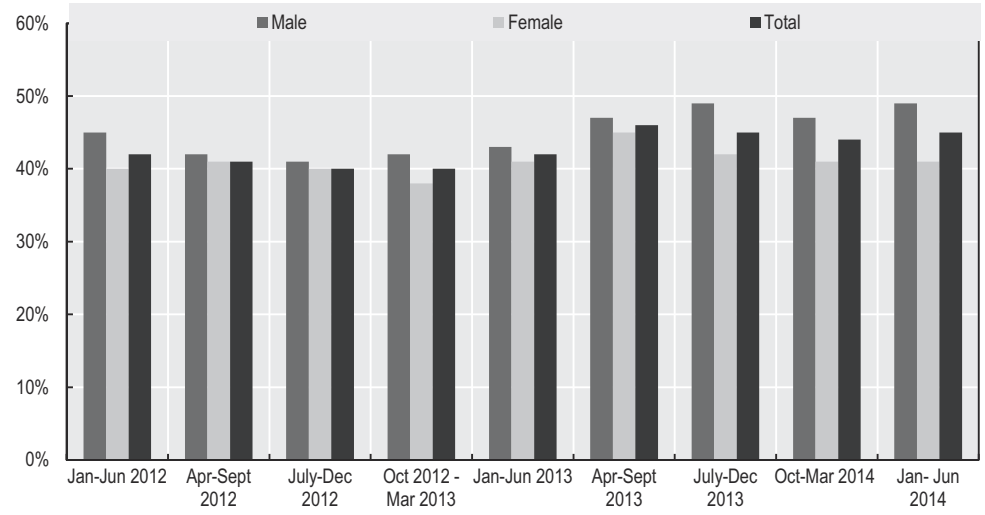
Reporting results by frequency in each category

At first glance, indicating the proportion of respondents selecting each response category seems appealing: as no categories are merged and the entire distribution is described, no information is lost, and the data producer makes no (arbitrary) decisions about how to collapse and present data. Nevertheless, there are good reasons why this type of presentation might not be appropriate for trust data. Unless the number of categories is limited, presenting the whole distribution of responses for each measure is likely to be overwhelming: non-technical audiences will, without some guidance about the *story* told by the data, not be able to directly compare and evaluate such distributions. Furthermore, while this approach works reasonably well for dichotomous trust measures (where reporting the share of responses having chosen to either trust or distrust is equivalent to reporting each response category), Chapter 3 of the Guidelines has provided arguments as to why a 0-10 response scale is preferable over a dichotomous one. Since a 0-10 response scale is relatively long, reporting the share of respondents for each of the 11 points on the scale is not recommended.

Reporting results by proportions above or below thresholds

Reporting on the proportion of responses falling above or below one or a set of thresholds offers a way around the problems of managing a large number of scale responses and facilitating respondent comprehension of the story behind the data. For example, responses can be either reported as the percentage of respondents falling above or below a certain cut-off point or collapsed into *high*, *medium* and *low* categories. Figures 5.1 and 5.2 below provide examples of each approach.

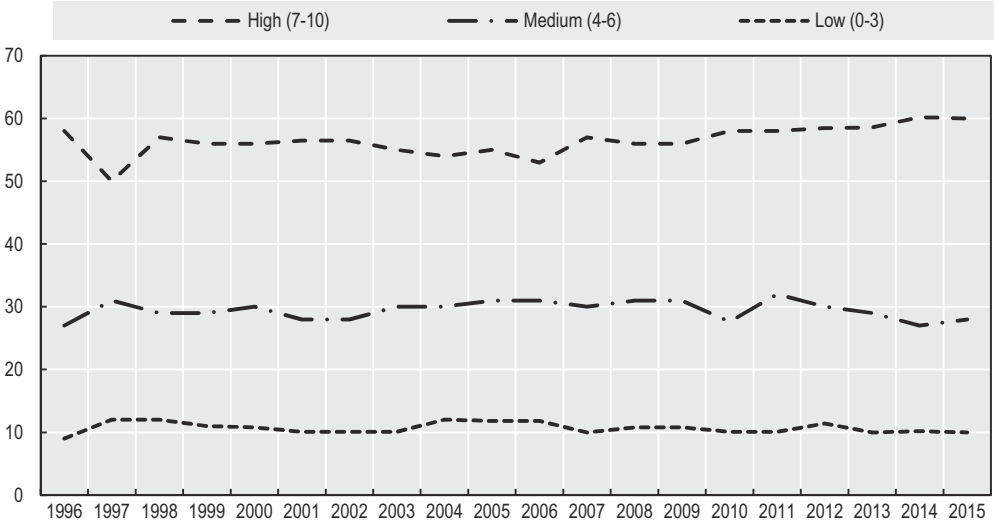
Figure 5.1. **Trust in public service by gender in New Zealand**



Note: Response options range from 1 ("Do not trust them at all") to 5 ("Trust them completely") to the question: "Overall, to what extent do you trust the public service?" The data show the percentage of respondents who gave a 4 or 5 out of 5.
Source: New Zealand State Services Commission (2014), Kiwis Count Survey (database), *www.ssc.govt.nz/kiwis-count-datasets*.

StatLink http://dx.doi.org/10.1787/888933583880

Figure 5.2. **Interpersonal trust in Sweden, 1996-2015**



Note: Response options range from 0 ("It is not possible to trust people in general") to 10 ("It is possible to trust people in general") to the question: "According to your view, to what extent is it possible to trust people in general?"
Source: SOM Institute (2015), The National SOM Survey (database), *http://som.gu.se/som_institute/-surveys/national-som*.

StatLink http://dx.doi.org/10.1787/888933583899

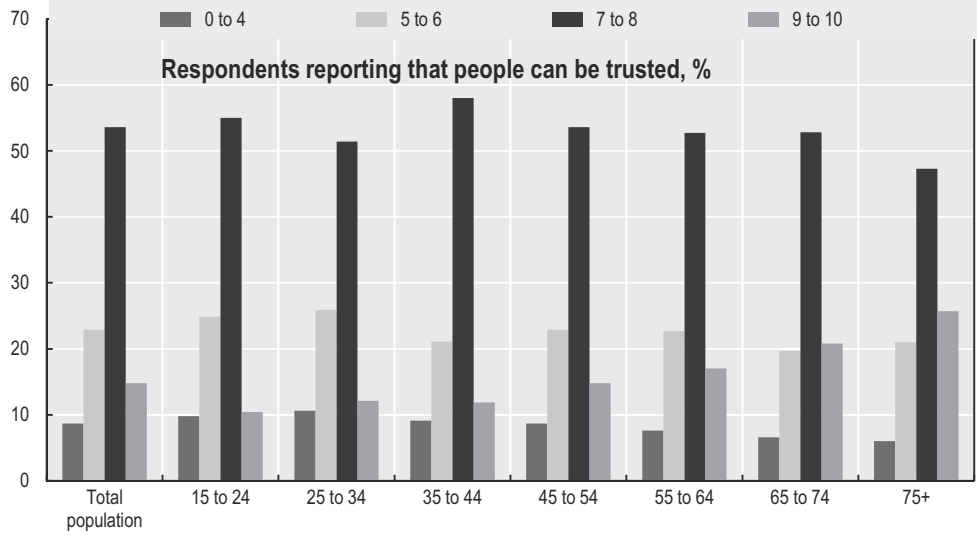
Figure 5.1 shows the proportion of the New Zealand population reporting trust in public services from 2012 until 2014, disaggregated by gender. The threshold for trust in this case was selected to be "4 and above" on a 5-point scale (where 1 is *do not trust them at all* and 5 is *trust them completely*) when answering the question: *Overall, to what extent do you trust the public service?* Figure 5.2 provides an example of collapsing data into several threshold-based categories. The data (from the SOM institute, an independent survey research organisation at the University of Gothenburg in Sweden) refer to responses to the question, "according to

your view, to what extent is it possible to trust people in general?", based on a response scale from 0 ("not possible") to 10 ("possible"). The SOM researchers coded respondents selecting the response categories 0-3 as having *low* trust in people in general, those ticking categories 4-6 as having *medium* trust, and those ticking 7-10 as having *high* trust. The suggested output for the question modules presented in Annex 2 also introduces the possibility of reporting the share of responses falling *below* the threshold of 4 on a 0-10 scale, hence focusing on the part of the population that does not trust other people or institutions.

Both figures make the benefits of threshold-based measures salient – the threshold provides an anchor and reference frame for interpretation, and thus eases the digestion of the results. However, threshold-based measures are far from perfect. Any summary measure, by definition, can be misleading, as it does not provide information on the distribution of data. Collapsing categories means losing some of the data richness and can lead to a narrow policy focus on shifting people from just below to just above a threshold. This is particularly problematic if only one threshold at the upper end of the response scale is used (as, for example, in Figure 5.1), as it may be important, for policy makers in particular, to recognise the existence of people at both extremes of the trust spectrum. In addition, reporting estimates based on thresholds runs the risk of presenting two very similar distributions as very different, when the only differences are around the threshold, or vice versa (OECD, 2013). Lastly, the use of threshold-based measures risks arbitrarily assigning individuals to high, medium and low trust categories without much evidence of what these categories mean in practice, and without testing them for real-world validity (Blanton and Jaccard, 2006).

One compromise between managing scale length by collapsing response scales and not assigning arbitrary and potentially misleading categories is to avoid labelling the categories as high or low, but to neutrally refer to their place on the scale, for example "0-4" or "5-6" (see Figure 5.3 for an example). However, even with this approach, it is unclear which criteria govern the selection of the scale ranges to be bundled together. Overall, given the challenges

Figure 5.3. **Interpersonal trust by age groups in New Zealand, 2014**



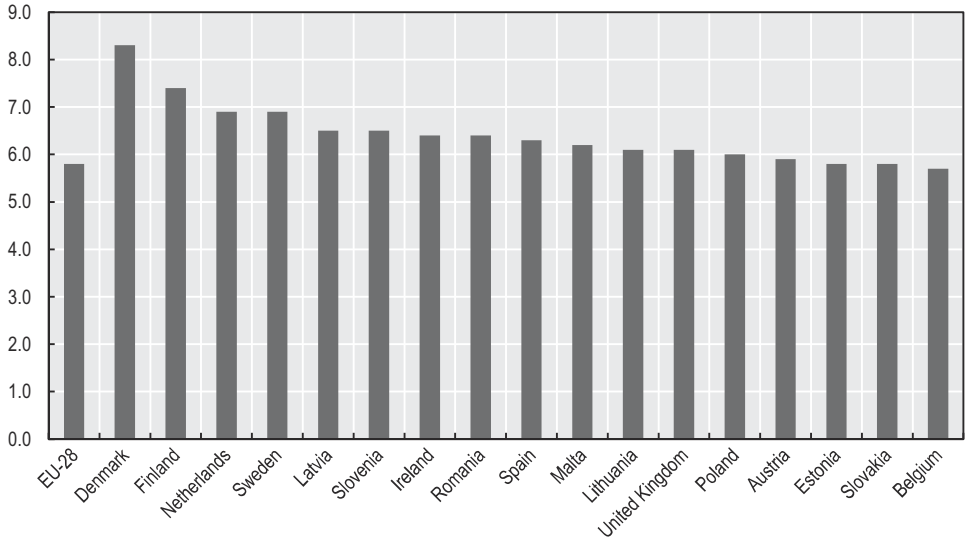
Note: Based on a scale where 0 is "not trusted at all" and 10 is "trusted completely" as answer to the question: "In general how much do you trust most people in New Zealand?"
Source: Stats New Zealand (2014), New Zealand General Social Survey (database), *www.stats.govt.nz/nzgss2014*.
StatLink http://dx.doi.org/10.1787/888933583918

associated with setting suitable thresholds, statistical analysts should avoid adopting these types of summary measures uncritically. At the very least, information about which threshold was chosen should always be presented alongside such data.

Reporting results by central tendency measures

The third approach for presenting levels of trust involves summarising the central tendency of a distribution in a single and easily understandable number. While the median and the mode may lack sensitivity to changes over time or to differences between groups due to the limited number of scale categories, the mean is a more useful summary statistic. Figure 5.4 shows the mean average value (on a 0-10 response scale) to EU-SILC's generalised trust question, whereas Figure 5.5 does the same for the measure on trust in public institutions from the same survey. Figure 5.5 also highlights another good practice when reporting data on trust in institutions, i.e. separately showing results for various institutions (in this case, the political system, the police and the legal system) rather than merging them into some sort of summary index. Since the mean, as all other measures of central tendency, consists merely of a single number that takes up little space in data releases, this leaves more room for comparisons between different institutions in the same figure or table.

Figure 5.4. **Interpersonal trust in European countries, 2013**



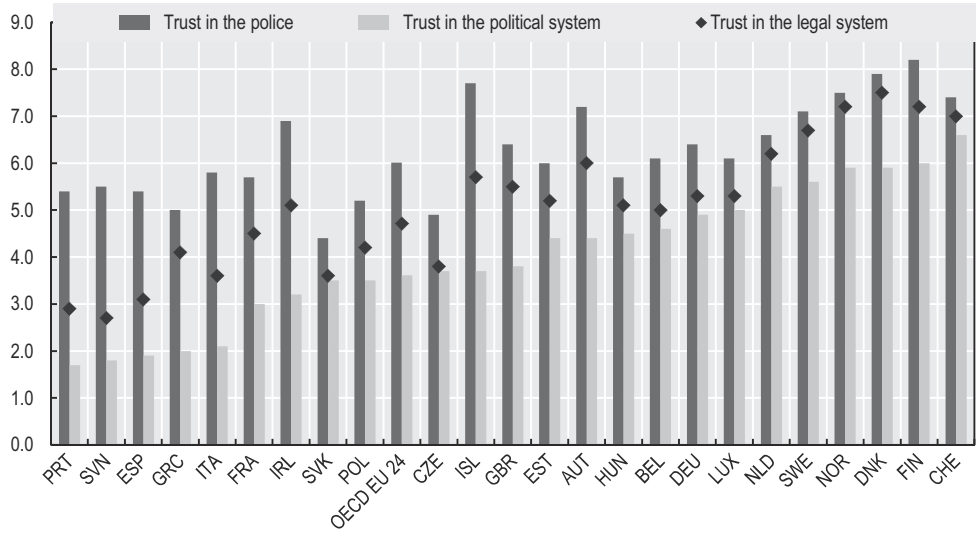
Note: Response options ranged from 0 ("You do not trust any other person") to 10 ("Most people can be trusted") to the question: "Some people say that you can trust most people. Others think you cannot be too careful in dealing with other people. Do you think most people can be trusted?" The OECD EU average is the population-weighted average of the values included in the chart.

Source: Eurostat (2015), European Union Statistics on Income and Living Conditions (EU SILC) (database), *http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=ilc_pw03&lang=en*.

StatLink *http://dx.doi.org/10.1787/888933583937*

Of course, the mean has disadvantages as well: in particular, it requires treating data as cardinal even when they are, in fact, ordinal.[2] However, several studies suggest that this does not generally lead to biased results (Diener and Tov, 2012). More importantly, though, outliers can strongly affect the mean value, and the mean does not provide information about the data's underlying distribution. Therefore, the mean should be complemented with

Figure 5.5. **Trust in public institutions in European countries, 2013**



Note: Response options range from 0 ("No trust at all") to 10 ("Complete trust") to the question: "How much do you trust: The political system in [country]/ The legal system in [country]/ The police in [country]?" The OECD EU average is the population-weighted average of the values included in the chart.
Source: Eurostat (2015), European Union Statistics on Income and Living Conditions (EU SILC) (database), *http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=ilc_pw03&lang=en*.

StatLink *http://dx.doi.org/10.1787/888933583956*

information about the distribution of data. One obvious choice is the standard deviation, but this is quite abstract and not easy to communicate to a non-technical audience. Other measures of dispersion, such as interquartile ranges, can thus be considered as alternative, depending on the statistical literacy of the release's target population.

The second page of statistical releases: Reporting change over time and differences between groups

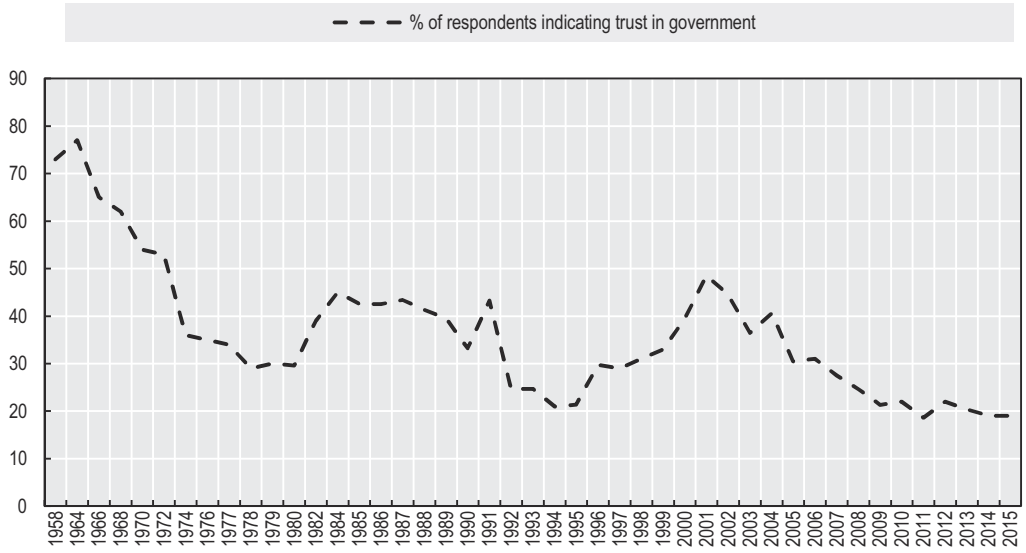
The report of trust data on the second page of statistical releases should generally cover two elements: changes over time, and differences between either societal groups or countries. Both of these elements are essential to provide some kind of "external reference point" to the headline trust measure of a country. How does the current level of trust compare to the score recorded one, two, three, four, five years ago? How does it compare to the trust score of other countries at similar levels of development? Do all groups within society have a similar score, or are there big discrepancies? More comprehensive reporting of trust measures should, ideally, answer all these questions.

In fact, all these questions point to the core mission of national statistics. If policy makers, civil society and researchers want to understand how to increase trust or to prevent its decay, the characteristics of the groups at both extremes of the trust scale need to be closely examined and better understood. Breaking down national statistics (e.g. by age, gender, education, place of residence, ethnicity, religion, occupation, socio-economic conditions, employment status, health status, etc.) and contrasting these measures of group performance against the overall population can enhance that understanding. Comparisons over time and between groups can also guide the quest for the potential drivers of trust.

Presenting measures of changes over time and of differences between groups requires reporting tools that are similar to those used for presenting central tendency measures. For example, changes over time can be monitored by tracking changes in mean trust through

time series (Figure 5.6) or by calculating changes in the mean score between various points in time (Figure 5.7).

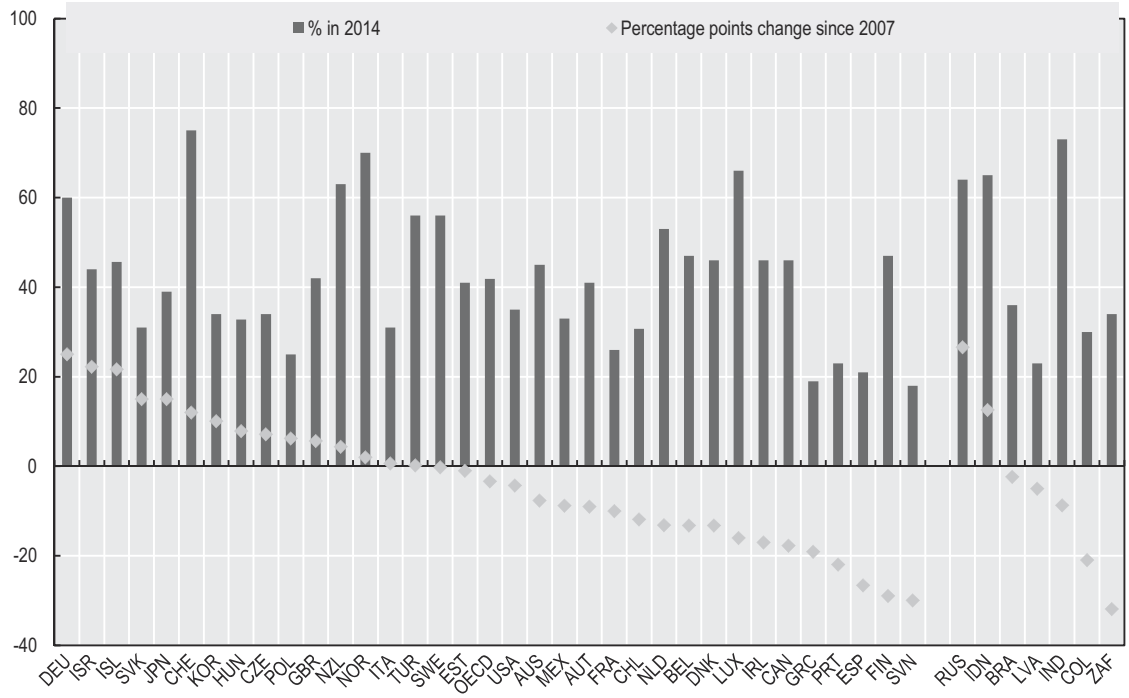
Figure 5.6. **Trust in government in the United States, 1958-2015**



Source: Pew Research Center (2016), *Beyond Distrust: How Americans View their Government, www.people-press.org/2015/11/23/1-trust-in-government-1958-2015*.

StatLink http://dx.doi.org/10.1787/888933583975

Figure 5.7. **Trust in national government in 2014 and changes since 2007, OECD countries**



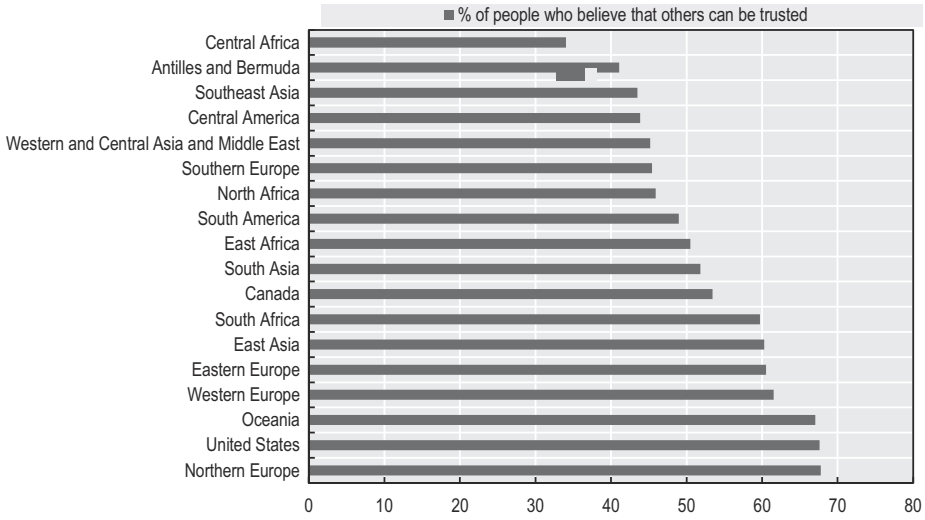
Note: Data refer to the percentage of people who answered *yes* to the question: "Do you have confidence in national government?" Countries are ranked in descending order according to the percentage point change between 2007 and 2014. Data for Chile, Hungary and Iceland refer to 2013 rather than 2014. Information on data for Israel: *http://dx.doi.org/10.1787/888932315602*.

Source: Gallup (2014), Gallup World Poll (database), *www.gallup.com/services/170945/world-poll.aspx*.

StatLink http://dx.doi.org/10.1787/888933583994

Group differences can be examined by presenting these 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 for several groups of interest (Figure 5.8). Table 5.2 provides an example of how to include estimates of statistical inference testing.

Figure 5.8. **Interpersonal trust by region of birth in Canada, 2013**



Source: Statistics Canada (2013), *Trends in Social Capital in Canada, www.statcan.gc.ca/pub/89-652-x/89-652-x2015002-eng.htm.*

StatLink http://dx.doi.org/10.1787/888933584013

Table 5.2. **Interpersonal trust by population group in Canada in selected years**

People who reported most people can be trusted[1]

	2003	2008	2013
		percentage	
Total	55	48†	54†
Men (ref.)	57	48†	55†
Women	54*	47†	52*†
Age group			
15 to 24 years	52*	47†	48*†
25 to 34 years	51*	45†	52
35 to 44 years (ref.)	56	47†	53†
45 to 54 years	61*	50*†	55†
55 to 64 years	58	49†	57*
65 to 74 years	53	47†	55
75 years and older	54	48†	54
Province			
Newfoundland and Labrador	64*	49†	56†
Prince Edward Island	69*	51†	63
Nova Scotia	62	53†	59
New Brunswick	56*	47†	51*†
Quebec	35*	32*†	36*
Ontario (ref.)	60	51†	57†
Manitoba	64*	52†	58†
Saskatchewan	67*	54†	60†
Alberta	63*	55*†	60*
British Columbia	65*	57*†	63

1. "Generally speaking, would you say that most people can be trusted or that you cannot be too careful with people?"
† Significantly different from 2003 ($p < 0.05$).
* Significantly different from the reference category ($p < 0.05$).
Source: Statistics Canada (2013), *Trends in Social Capital in Canada, www.statcan.gc.ca/pub/89-652-x/89-652-x2015002-eng.htm.*

StatLink http://dx.doi.org/10.1787/888933584260

Depending on the target audience's interest, other relevant changes in trust (e.g. changes in the overall distribution, or differences in rates of change between different societal groups) can additionally be reported.

Since both sample size and standard errors are essential when comparing two or more observations, this information should be reported alongside group means (OECD, 2013). This information is easier to interpret when presented graphically, e.g. though box plots or error bars.

5.3. Interpreting trust data

This section focuses on how to interpret trust data after they have been released. Apart from NSO staff, this information is relevant for professional communities (i.e. media, think tanks) concerned with communicating the key messages and trends of fresh data releases, as well as for researchers and analysts working directly with microdata. Most analyses of trust data will naturally be concerned with examining *differences* between observations. Two essential questions to answer for these types of analyses will be whether these differences are actually meaningful in practical, real-life terms and to what extent any differences are due to measurement artefacts and errors. Both issues are addressed below.

Essentially, any assessment of the meaningfulness of differences between observations requires an understanding of what size of difference is likely to be encountered, and what difference can be categorised as *small* or *big*. These evaluations, by their nature, will be relative – for example, in comparison to values from other groups, countries or previous time points. A necessary precondition for making such relative comparisons is the availability of all relevant comparison points – in other words, knowledge about the full data universe of trust measures.

At present, knowledge about the trust data universe, its properties and boundaries is incomplete – which is in fact one of the *raisons d'être* for these Guidelines. Nevertheless, Tables 5.3 to 5.6 represent a first attempt to summarise what is currently known about differences between observations for trust measures. For measures of interpersonal trust and trust in institutions (the police, judicial system and government), each table details the size of the differences for three types of measure (i.e. a 0-10 scale, a threshold measure of 7 and above for 0-10 scales, and a binary measure) across the two dimensions of comparison (cross-sectional or between observations, and over time or within observations).

Within the cross-section dimension, differences are further broken down by differences between groups (age groups, groups with different educational attainments and different labour force status were selected as examples, as these are common socio-economic and demographic distinctions) and differences between countries. The size of the difference is presented relative to the respective sample mean – for example, the first cell in the first row of Table 5.3 indicates that, compared to the overall 2013 EU-SILC mean score on a 0-10 scale, the respondents aged 16 to 24 reported a level of interpersonal trust 0.2 scale points higher. The same column, for the threshold measure, shows that, compared to the total sample percentage of the 2015 United Kingdom ONS National Opinion Survey, 16.7% fewer respondents aged 16 to 44 chose an interpersonal trust score of 7 or above on a 0-10 scale (ONS, 2016). For differences between countries, the range of differences from the respective sample mean is shown; for instance, in the case of 0-10 scale trust measures, the score of the country with the lowest interpersonal trust (Bulgaria) was 1.6 points below the EU-SILC 2013 sample mean, while that of the country with the highest score (Denmark) was 2.5 points above the sample mean, resulting in an overall range of 4.1 points.

Table 5.3. **Magnitude of differences for measures of interpersonal trust**

	Cross-section					Over time					
	Between population groups			Between countries		Within individuals			Within countries		
	Difference in mean scores (population group – total sample)		Source	Difference in mean scores (country – total sample)	Source	Difference in mean scores		Source	Difference in mean scores		Source
0-10 scale	By age By employment status By level of education	16-24 years: 0.2 25-49 years: 0 50+ years: 0.06 In employment: 0.1 Unemployed: -0.5 Tetriary education: 0.5 Upper sec. education: -0.1 Lower sec. education -.0.2	EU-SILC (2013)	Low end of distribution: -1.6 (Bulgaria) High end of distribution: 2.5 (Denmark)	EU-SILC (2013)	n/a		n/a	2004-06 2004-10 2004-14	Minimum: 0.02 (Switzerland) Sample average: |0.2| Maximum: -0.56 (Ireland) Minimum: -0.012 (Finland) Sample average: |0.25| Maximum: 0.78 (Poland) Minimum: -0.01 (Norway) Sample average: |0.23| Maximum: -0.75 (Ireland)	European Social Survey
Threshold measure: 0-10 scale	Difference in % of respondents that selected 7-10 (population group – total sample)		Source	Difference in % of respondents that selected 7-10 (country – total sample)	Source	% of respondents that switched 7-10 threshold category		Source	Difference in % of respondents that selected 7-10		Source
	By age By employment status By level of education	16-44 years:-16.7% 45+ years: 6.5% In employment: 3.6% Unemployed: -0.7 % With degree/equivalent: 12.1% Below degree level: 1.5% No qualification: -4.6%	ONS National Opinion's Survey (2015)	Low end of distribution: -21.2% (Poland) High end of distribution 31.7% (Denmark)	European Social Survey (2014)	n/a		n/a	2004-06 2004-10 2004-14	Minimum: -0.05% (Ukraine) Sample average: |3.47%| Maximum: -10.65% (Ireland) Minimum: -0.07% (Spain) Sample average: |4%| Maximum: -13.54% (Ireland) Minimum: -0.22% (Norway) Sample average: |3.71%| Maximum: -11.86% (Ireland)	European Social Survey
Threshold measure: binary scale	Difference in % of respondents that selected "trust" (population group – total sample)		Source	Difference in % of respondents that selected "trust" (country – total sample)	Source	% of respondents that switched binary category		Source	Difference in % of respondents that selected "trust"		Source
	By age By employment status By level of education	16-44 years: -0.3% 45+ years: 0.2% In employment: 2.1% Unemployment: -3.5% With degree/equivalent: 19% Below degree level: -3.7% No qualification: -12.7%	ONS National Opinion's Survey (2015)	Low end of distribution: -21.1% (Phillipines) High end of distribution: 43.5% (Netherlands)	World Values Survey (Wave 6, 2010-14)	1998-2000 2003-05 2007-08 1998-2006	23.69% 24.07% 22.77% 52.59%	British Household Panel Study	Wave 5 (2005-08) – Wave 6 (2010-14)	Minimum: -0.29% (Malaysia) Sample average: |5.64%| Maximum: 22.94% (Netherlands)	World Values Survey

Source: OECD calculations based on the OECD Trust Database.

StatLink http://dx.doi.org/10.1787/888933584279

Table 5.4. **Magnitude of differences for measures of trust in the police**

	Cross-section					Over time		
	Between population groups			Between countries		Within countries		
	Difference in mean scores (population group – total sample)		Source	Difference in mean scores (country – total sample)	Source	Difference in mean scores		Source
0-10 scale	By age	16-24 years: -0.2 25-49 years: -0.3 50+ years: 0.17	EU-SILC (2013)	Low end of distribution: -2.4 (Bulgaria) High end of distribution: 2.2 (Finland)	EU-SILC (2013)	2004-06	Minimum: -0.007 (France) Sample average: |0.19| Maximum: -0.69 (Ukraine)	European Social Survey
	By employment status	In employment: -0.03 Unemployed: -0.9				2004-10	Minimum: 0.07 (Norway) Sample average: |0.36| Maximum: -1.45 (Greece)	
	By level of education	Tetriary education: 0.3 Upper sec. education: -0.1 Lower sec. education: -0.2				2004-14	Minimum: -0.06 (Finland) Sample average: |0.43| Maximum: 1.52 (Czech Republic)	
Threshold measure: 0-10 scale	Difference in % of respondents that selected 7-10 (population group – total sample)		Source	Difference in % of respondents that selected 7-10 (country – total sample)	Source	Difference in % of respondents that selected 7-10		Source
	By age	n/a		Low end of distribution: -25.79% (Poland) High end of distribution: 26.38% (Finland)	European Social Survey (2014)	2004-06	Minimum: -0.16% (United Kingdom) Sample average: |3.19%| Maximum: -8.22% (Ireland)	European Social Survey
	By employment status	n/a				2004-10	Minimum: -0.09% (Hungary) Sample average: |5.6%| Maximum: -19.77% (Greece)	
	By level of education	n/a				2004-14	Minimum: -0.65% (Finland) Sample average: |7.53%| Maximum: 22.13% (Czech Republic)	
Threshold measure: binary scale	Difference in % of respondents that selected "trust" (population group – total sample)		Source	Difference in % of respondents that selected "trust" (country – total sample)	Source	Difference in % of respondents that selected "trust"		Source
	By age	n/a		Low end of distribution: -43.74% (Venezuela) High end of distribution: 28.26% (Niger)	Gallup World Poll (2015)	2006-08	Minimum: 0% (Cameroon, Japan, Tanzania, Latvia, Canada) Sample average: |6.11%| Maximum: 25% (Chad)	Gallup World Poll
	By employment status	n/a				2006-12	Minimum: 0% (Vietnam, Hungary, Bolivia, Germany) Sample average: |7.16%| Maximum: 32% (Nepal)	
	By level of education	n/a				2006-15	Minimum: 0% (Paraguay, Serbia, Canada, Kosovo, Belgium, Singapore, Botswana) Sample average: |8.96%| Maximum: 29% (Nepal)	

Source: OECD calculations based on the OECD Trust Database.

StatLink http://dx.doi.org/10.1787/888933584298

Table 5.5. **Magnitude of differences for measures of trust in the judicial system**

	Cross-section					Over time		
	Between population groups			Between countries		Within countries		
0-10 scale	Difference in mean scores (population group – total sample)		Source	Difference in mean scores (country – total sample)	Source	Difference in mean scores		Source
	By age	16-24 years: 0.3 25-49 years: 0.1 50+ years: 0	EU-SILC (2013)	Low end of distribution: -1.9 (Slovenia) High end of distribution: 2.9 (Denmark)	EU-SILC (2013)	2004-06	Minimum: -0.019 (Hungary) Sample average: |0.28| Maximum: -1.25 (Ukraine)	European Social Survey
	By employment status	In employment: 0.17 Unemployed: -1				2004-10	Minimum: 0.01 (Finland) Sample average: |0.46| Maximum: -1.63 (Greece)	
	By level of education	Tetriary education: 0.7 Upper sec. education: 0.1 Lower sec. education -0.6				2004-14	Minimum: 0.08 (Ireland) Sample average: |0.43| Maximum: 0.97 (Czech Republic)	
Threshold measure: 0-10 scale	Difference in % of respondents that selected 7-10 (population group – total sample)		Source	Difference in % of respondents that selected 7-10 (country – total sample)	Source	Difference in % of respondents that selected 7-10		Source
	By age	n/a		Low end of distribution: -29.53% (Poland) High end of distribution: 31.92% (Denmark)	European Social Survey (2014)	2004-06	Minimum: 0.03% (Hungary) Sample average: |3.31%| Maximum: -10.19% (Ukraine)	European Social Survey
	By employment status	n/a				2004-10	Minimum: -0.64% (Slovakia) Sample average: |6.09%| Maximum: -18.95% (Greece)	
	By level of education	n/a				2004-14	Minimum: -0.07% (Ireland) Sample average: |6.9%| Maximum: 16.36% (Norway)	
Threshold measure: binary scale	Difference in % of respondents that selected "trust" (population group – total sample)		Source	Difference in % of respondents that selected "trust" (country – total sample)	Source	Difference in % of respondents that selected "trust"		Source
	By age	n/a		Low end of distribution: -44.17% (Ukraine) High end of distribution: 37.83% (Singapore)	Gallup World Poll (2015)	2006-08	Minimum: 0% (Kyrgyzstan, Ukraine, Japan, Mexico) Sample average: |6.33%| Maximum: 32% (Poland)	Gallup World Poll
	By employment status	n/a				2006-12	Minimum: 0% (Canada, Zambia, Montenegro, Denmark, Greece) Sample average: |7.56%| Maximum: -29% (Uganda)	
	By level of education	n/a				2006-15	Minimum: 0% (Guatemala) Sample average: |9.68%| Maximum: -39% (Cyprus)[3, 4]	

Source: OECD calculations based on the OECD Trust Database.

StatLink http://dx.doi.org/10.1787/888933584317

Table 5.6. **Magnitude of differences for measures of trust in government**

	Cross-section					Over time		
	Between population groups			Between countries		Within countries		
	Difference in mean scores (population group – total sample)		Source	Difference in mean scores (country – total sample)	Source	Difference in mean scores		Source
0-10 scale	By age	16-24 years: 0.5 25-49 years: -0.1 50+ years: 0.06	EU-SILC (2013)	Low end of distribution: -1.8 (Portugal) High end of distribution: 3.1 (Switzerland)	EU-SILC (2013)	2004-06	Minimum: -0.007 (Germany) Sample average: |0.28| Maximum: -1.44 (Ukraine)	European Social Survey
	By employment status	In employment: 0.17 Unemployed: -1.1				2004-10	Minimum: -0.05 (Portugal) Sample average: |0.61| Maximum: -2.29 (Greece)	
	By level of education	Tetriary education: 0.7 Upper sec. education: 0.2 Lower sec. education: -0.6				2004-14	Minimum: -0.1 (Belgium) Sample average: |0.52| Maximum: -1.19 (Slovenia)	
Threshold measure: 0-10 scale	Difference in % of respondents that selected 7-10 (population group – total sample)		Source	Difference in % of respondents that selected 7-10 (country – total sample)	Source	Difference in % of respondents that selected 7-10		Source
	By age	n/a		Low end of distribution: -12.86% (Slovenia) High end of distribution: 13.35 (Norway)	European Social Survey (2014)	2004-06	Minimum: 0.19% (Hungary) Sample average: |1.8%| Maximum: -7.31% (Ukraine)	European Social Survey
	By employment status	n/a				2004-10	Minimum: -0.44% (France) Sample average: |4.88%| Maximum: -10.87% (Greece)	
	By level of education	n/a				2004-14	Minimum: 0.59% (Belgium) Sample average: |5.31%| Maximum: 14.99% (Norway)	
Threshold measure: binary scale	Difference in % of respondents that selected "trust" (population group – total sample)		Source	Difference in % of respondents that selected "trust" (country – total sample)	Source	Difference in % of respondents that selected "trust"		Source
	By age	n/a		Low end of distribution: -37.71% (Ukraine) High end of distribution: 45.29% (Singapore)	Gallup World Poll (2015)	2006-08	Minimum: 0% (Peru, Singapore, Philippines, Turkey, Denmark) Sample average: |10.84%| Maximum: 39% (Ecuador)	Gallup World Poll
	By employment status	n/a				2006-12	Minimum: 0% (Vietnam, South Korea, Kazakhstan, Israel, Panama) Sample average: |12.97%| Maximum: 52% (Ecuador)	
	By level of education	n/a				2006-15	Minimum: 0% (Japan, Singapore) Sample average: |13.51%| Maximum: -46% (Cyprus)[5]	

Source: OECD calculations based on the OECD Trust Database.

StatLink http://dx.doi.org/10.1787/888933584336

For changes over time, the tables present, if data are available, differences within individuals and within countries for various time ranges. For instance, in Table 5.3, longitudinal data from the British Household Panel Study from 1998-2008 show the share of respondents who switched between *no trust* and *trust* categories on a binary interpersonal trust measure. In the case of within-country differences, the minimum, sample average and maximum differences for various points in time are shown.

These tables show clearly that our current picture of the trust data universe is far from complete. The lack of data is particularly salient for individual-level panel data – only Table 5.3, the one focusing on interpersonal trust, features results, and only for a dichotomous measure. Furthermore, the underlying data in all tables varies greatly in quality. While EU-SILC data represent official statistical measures, especially for longitudinal data, non-official surveys such as the European Social Survey, the Gallup World Poll and the World Values Survey had to be relied upon. These surveys often suffer from small sample sizes and varying degrees of quality control. In some cases, national rather than cross-country data were used (i.e. the ONS National Opinion Survey and the British Household Panel Study). Furthermore, the tables are not significantly expanding the data universe beyond the geographic boundaries of Europe, with the exception of the Gallup World Poll and the World Values Survey. Lastly, not all measures in the table are based on the same question wording – for instance, the European Social Survey asks about trust in politicians, rather than in government. Thus, while these tables can serve as a useful reference point for people interested in situating effect sizes in a comparative context, firm conclusions should not be drawn until better, high frequency, global trust data are collected.

Furthermore, although the tables focus on mean differences in trust, inequality of trust within groups, across countries and over time are additional aspects to consider. Although standard errors of trust estimates are currently often not examined or reported in policy publications, they can provide relevant information about the state of trust and point to differences in the distributions between observations. A similar exercise featuring differences in trust distributions should thus be repeated in the future.

Keeping these limitations in mind, some preliminary trends are nevertheless detectable:

- Differences in interpersonal trust tend to be smaller than differences in institutional trust across the board. For example, the within-country average difference in 0-10 scale measures of trust over the 2004-14 period is 0.23 scale points for interpersonal trust, 0.43 scale points for trust in the police, 0.46 scale points for trust in the judicial system and 0.52 scale points for trust in government.
- When examining cross-sectional differences between societal groups, the employment status and educational attainment groups display larger differences than age groups.
- Individual-level interpersonal trust, at least on a binary measure, seems to be much more volatile over time than aggregate measures of the same construct would suggest. British Household Panel Study data exhibit a high level of instability in individual-level reports of trust, with nearly half of all sample members changing their trust response at least once over the 1998-2006 period of observation.[6]
- The choice of measures can make a difference: for similar points in time, the 0-10 scale measure, the "7 and above" threshold measure and the binary measure do not always result in the same countries being at the bottom and top ends of the difference distribution. For instance, focusing on the 2004-06 period of observation using European Social Survey data on trust in the police, France displays the smallest difference of -0.007 scale points and

Ukraine the largest difference of -0.69 scale points for a 0-10 scale measure, whereas the United Kingdom and Ireland take these places for the threshold measure, with differences in respondents falling over the threshold of -0.16% and -8.22%, respectively.

- Differences between countries are larger than differences within countries over time for all measures and types of trust considered.
- Effect sizes seem to grow larger as the sample considered increases in geographical coverage beyond Western nations. This suggests that as many countries as possible should be considered in order to fully comprehend the trust data universe and the volatility of trust measures in different contexts.

In conclusion, what constitutes a small or big difference in trust depends on the nature of the difference considered. It is important to consider the full range of potential drivers when interpreting results, including the influences stemming from measurement error described in the next section. The key message here is that knowledge on the overall data universe of trust will remain limited until higher quality and more frequent data become available. The current lack thereof significantly impacts the interpretations and conclusions that can be drawn about the magnitude of effect sizes.

Alternative influences on effect sizes and the possible impact of culture

The large differences between trust measures can be influenced by a variety of factors. These include the limits imposed by the response scale, issues of reverse causality and the possible impact of "culture" on trust estimates. These factors are briefly described below, together with an evaluation of their severity.

Box 5.1. **Demand effects and question framing**

Apart from differences between groups or time points, trust items may be subject to another type of difference. Namely, stakeholders might also be interested in assessing whether interpersonal trust depends on the type of group to be trusted (people in general, people you know personally, friends, etc.) and whether different public institutions (the national government, public agencies providing different services, the judiciary, state and local governments) are trusted to varying degrees. However, one should be careful not to over-interpret the size of differences between targets of trust when respondents are asked to provide their ratings for a long list of differently framed questions. Indeed, demand effects (changes in behaviour by respondents due to cues about what constitutes appropriate behaviour by the survey or experimental design) might lead to the assumption that slightly different information is wanted for each item. In other words, being asked repeatedly about similar – yet related constructs – might lead respondents to adjust their answers, thus inflating the magnitude of differences between trust in a range of social groups or institutions. The survey design should always be reviewed to assess whether it could have encouraged demand effects, including the order in which different types of people and institutions are listed, and whether any adjustments are needed. For instance, as outlined in the prototype module instructions in Annex 2, in cases where assessing trust in different institutions is important, the order in which questions are presented should be randomised across respondents.

One practical challenge that can influence the size of differences for trust items is the design of response scales. The specific features of the response scale influence which responses are theoretically possible and how responses are distributed. Since trust data are

collected with bounded scales with a limited number of response categories, the average trust measure can never exceed the top response category (unlike some measures where the scale is unbounded, such as income). Nevertheless, despite this theoretical limit, most countries are currently far from the top of the scale: for example, while, in the 2013 EU-SILC data, Denmark has the highest score (8.3 out of 10) for interpersonal trust, the overall sample mean lies at 5.8. In the case of institutional trust, the top scores in the EU-SILC 2013 data do not exceed 8 points. Looking at global trust measures, in some countries (e.g. Singapore, Kazakhstan, Vietnam) between 81% and 91% of respondents in the Gallup World Poll state that they trust government based on a dichotomous question item – however, these findings may reflect the small sample size and the influence of political considerations on respondents (this underscores the importance of managing sensitivity concerns through sound survey design and mode, as described in Chapter 3). Overall, apart from a few outliers, there seems to be quite a lot of "upwards space" left until countries hit the maximum ceiling allowed by the respective response scales.

The possibility of two-way causality between trust and its determinants may also influence and limit the size of the differences likely to be observed. Two-way causality occurs where there are reciprocal and causal relationships between two variables, running in both directions. For example, some researchers have argued that membership in voluntary associations increases interpersonal trust, while others have asserted that one's level of trust in others also drives the likelihood of volunteering (Stolle, 1998; Uslaner, 2000). Two-way causality points to the importance of selecting an appropriate time frame when examining changes in trust, based on what is known about the variables in question and the possible causal pathways through which they might take effect. Although longer time-frames might be required to detect significant changes in interpersonal trust data, measures of trust in institutions in particular might be relatively bumpy over short time periods. For example, Chanley (2002) found that trust in the US federal government experienced a sharp rise after 11 September 2001, a phenomenon Deaton (2012) refers to as short-term "cognitive bubbles" in reaction to significant national events. Further research will be needed to identify which events have a short-term impact and which have a longer-term impact on trust levels. For example, levels of institutional trust in Europe, particularly in politicians and the government, have so far not bumped back to pre-financial crisis levels. In general, time-series data on trust should be examined over longer time periods to identify meaningful changes associated with persistent societal and economic shifts.

Cultural differences in how respondents understand, process and reply to subjective questions, and the frames of reference they rely on when doing so, may also drive difference in international comparisons of trust. Especially in the case of interpersonal trust, it may seem puzzling that countries at similar levels of economic development report quite different mean levels of trust. Figure 5.4 above portrayed the mean distribution of interpersonal trust across the EU-24 countries: average country scores range all the way from 5 (France) at the bottom to 8.3 (Denmark) at the top end of the distribution. Countries with relatively high levels of GDP, notably Germany and France, are located near the low end of the distribution. In fact, the low ranking of France (within Europe) holds not only for trust in other people, but also for other self-reported measures such as subjective well-being and trust in the market. Senik (2011) attributes this "French unhappiness puzzle" to unique cultural factors and mental attitudes of French people.

Chapter 3 already described the evidence around cultural response styles and the methodological steps in survey design that can be taken to reduce the risk that scales and

questions might be understood differently by respondents. This section focuses on how to interpret and deal with cultural biases, once the data have been collected. In the following, possible sources of cross-country differences in average trust levels are highlighted and methods to "correct" data for cultural bias are briefly introduced – while also raising the question of whether such corrections should be conducted at all. Before attributing differences in average trust between countries at similar levels of economic development to "cultural bias", it is important to remember that these differences may also have multiple sources. A useful distinction can be made between "cultural impact", i.e. genuine sources of variance between cultures, and "cultural bias", i.e. inter-cultural differences stemming from measurement artefacts (Van de Vijver and Poortinga, 1997; Exton et al, 2015). The potential sources of cross-country variance described below all carry different implications for the validity of between-country comparisons of trust data, as well as for any actions that one might take to mitigate the impact of cultural bias.

Sources of cultural impact

On the one hand, there are country-specific differences for which one would not necessarily want to "correct". For example, one source of international variance in trust scores, in addition to economic variables, involves cross-country differences in the *social and political context* and other life circumstances of residents, all of which might impact upon trust. These drivers may or may not be related to culture: they include social relationships, the homogeneity of society in terms of income and ethnic diversity, religion, unemployment, the rule of law, crime, corruption, and the type and quality of institutional arrangements (Alesina and Ferrara, 2000; Algan and Cahuc, 2013; Guiso et al., 2006; Stolle et al., 2008; Rothstein and Uslaner, 2005; Jordahl, 2007).

The socio-demographic structure of the particular sample in each country may also contribute to differences in reported trust: trust is shaped by individual background characteristics, such as a respondent's age, gender, income, education and employment status, and features of the community where he or she lives. For example, there is a robust relationship between both interpersonal and institutional trust and educational status (Stolle et al., 2008; Helliwell and Wang, 2010; Carl and Billari, 2014). Theoretical reasons that have been put forward for this include the fact that education – and the higher level of income associated with it – empowers people, enabling them to make their own choices and to accomplish their goals and expectations (Hudson, 2006). An advanced education is also likely to be linked with a better understanding of how public institutions function.[7] Moreover, people who are better off financially and more educated are likely to enjoy more opportunities and channels to take part in society (e.g. through volunteering and political participation), which is conducive to developing and maintaining larger and more diverse social networks (Helliwell and Putnam, 2007; OECD, 2015).[8] Unemployment is another socio-economic factor that has been associated with the degree of an individual's trust both in other people and in institutions, with the unemployed exhibiting lower trust compared to their fellow citizens (Chabanet, 2007; Bârgăoanu et al., 2015). It has also been found that interpersonal trust increases slightly with age (Stolle et al., 2008; Putnam, 2000; Tokuda et al., 2008; Li and Fung, 2012; Clark and Eisenstein, 2013). While the exact reasons why older people report higher trust in others are under debate, one potential explanation for age-related increases in trust is that older adults are more motivated to give back to others, therefore believing them to be good and trustworthy in return (Poulin and Haase, 2015). Generally, it is very important to examine each sample, including across countries, to understand how its composition might have driven effect sizes.

There may also be differences between countries in how people *feel* about trusting other people or institutions. These differences in feelings are influenced by many features, such as an individual's reference group, past experiences, and the historical roots of the country both of origin and of current residence (which can set collective *frames of reference*). "Frame-of-reference effects" refer to differences in the way respondents answer survey questions based on their own life experiences as well as on their knowledge about the experiences of others, both within and outside their comparison group (Sen, 2002; Ubel et al., 2005). These experiences set the frame of reference, relative to which a respondent's own current circumstances and feelings are evaluated. The frame of reference may contribute to appraisal styles that influence the connection between objective life circumstances and subjective feelings – for example, the degree of optimism or pessimism that individuals feel about the actions of other people. Frames of reference produce real differences in levels of trust, rather than simply differences in how people report those feelings. They thus do not bring into question the validity of trust measures per se. Evidence from other self-reported measures, such as subjective well-being, suggests that while framing effects may influence the size of differences between groups and countries, they are not sufficiently large to prevent the detection of the impact of life circumstances (OECD, 2013). This source of cross-country variance can reflect cultural impact rather than bias and might even add to the predictive validity of the overall trust measure (i.e. in its association with real-life trusting behaviour towards other people and institutions).

Sources of cultural bias

On the other hand, *linguistic differences* and *cultural response styles* are likely to add cultural bias to the data, reducing their overall validity and predictive ability. Linguistic differences are likely to play a key role when trust constructs are not perfectly translatable across languages – for instance, many languages do not make the conceptual distinction in English between *trust* and *confidence*. The translatability of constructs is also likely to be more challenging when response scales are verbal rather than numerical (Veenhoven, 2008). Cultural response styles, described in Chapter 3, refer to group differences in scale use or differences in how individuals *report* their feelings. For example, a "modesty" or moderate-responding bias might have a downward influence on self-reports, without having a negative impact on private feelings of trust. Similarly, tendencies towards *extreme responding* (i.e. using scale end points) or more socially desirable responding could imply differences in modes of cultural expression, rather than substantive differences in the trust actually experienced. Both linguistic differences and cultural response styles represent sources of bias that should be minimised at source through survey design (Chapter 3) or translation (Chapter 4) or by adjusting the data *ex post* to correct for the bias.

Methods for examining and removing cultural bias include the use of objective outcomes as counterfactuals in the analysis (here, experimental measures of trust might be useful in the future), fixed-effects models to control for country and regional characteristics, or vignettes to measure the different ways in which individuals and/or cultures may interpret or benchmark the same survey question (OECD, 2013). Another technique for investigating the effects of culture that is gaining prominence among scholars is the use of migrant data, i.e. the comparison of response styles of natives and migrant respondents within the same country for particular outcomes of interest (Senik, 2011; Exton et al., 2015).

However, none of these approaches has yet convincingly distinguished between a substantive cultural impact and cultural bias. The relatively small number of countries

sampled in the existing research also makes it difficult to extrapolate results more widely – meaning that there is little that can be said even about the expected magnitude of cultural effects, particularly at a global level. Access to further high-quality data on trust from large and nationally representative samples will help to shed light on the issue of what proportion of average-level cross-country differences can be attributed to cultural biases. This information will help to determine whether the benefits of data adjustments outweigh the risk of removing the influence that all unmeasured country differences (including the influence of a country's policy environment, social networks and a wide range of valid cultural differences) have on how trust assessments are formed and maintained. A further practical limitation in using vignettes and migrant data to correct country averages of trust is that the impact of culture in the data cannot be quantified in absolute terms – rather, it is always *relative* to other countries in the sample. This provides a further challenge if the goal is to adjust national-level data to provide *culture-free* estimates; it implies that only a large and representative global sample could be used as a basis for such adjustments. Given these limitations and the current state of evidence, these Guidelines recommend against using methods to correct national trust data for cultural influences for the time being.

5.4. Analysing trust data

Moving beyond descriptive presentations of trust metrics is of interest to both the general public and policy makers for several reasons. First, interpersonal trust and trust in institutions are valuable outcomes in their own right. Especially when identifying vulnerable groups and international benchmarking are core elements of monitoring trust, a better understanding of what causes, maintains or destroys trust can help explain some of the observed differences between countries and groups. Analysing the drivers of trust can then support both the identification of those areas where appropriate policies could raise trust as well as the informed appraisal of various policies that might have unintended impacts on trust.

Second, trust also impacts on a multitude of other outcomes of interest, including economic growth, subjective well-being, health status, crime levels and the willingness to participate in the democratic process. This makes it imperative to better understand how to improve trust and to unpack the exact pathways of how trust drives people's well-being.

Different methods of analysis are described below, before looking at selected examples of how both interpersonal and institutional trust has been analysed in practice. The section closes with some words of caution about the analytical challenges that might arise when working with trust data.

Data requirements and study design

Better understanding the drivers and impacts of trust involves a process of identifying the variables that have causal relationships with trust or on which trust has a causal effect and examining the mechanisms through which these effects take place.

Before presenting a range of basic analytical methods applicable to trust data, a few comments on general data requirements are warranted. Preferably, surveys of trust are already designed with some idea of the final analysis in mind and include a wide range of covariates to draw upon, including a number of standard demographic and control variables and measures of potential drivers of trust. Chapter 4 elaborates further on which data to collect alongside trust during the measurement stage. Since any analysis of drivers requires access to micro-level data, the trust datasets gathered by NSOs should ideally be

anonymised and made publicly accessible to government analysts, researchers and organisations that have an interest in informing policy and public debate.

The nature of the study design with which the data was acquired is a key determinant of the degree to which researchers are able to make inferences about the causality of relationships between variables. The "gold standard" for establishing causality in social research is the so-called experimental design, or randomised control trials (RCTs). RCTs, well known for their use in trials of new medications, involve the random allocation of individuals to control groups (not receiving an intervention) and treatment groups (receiving intervention A, B, C, etc.). The impact of the intervention is then established by comparing the outcomes of the treatment group with those of the control group, effectively presenting a "counterfactual", or "what would have happened to the treatment group in the absence of the intervention". The use of RCTs in policy research, especially with regard to trust-relevant interventions, is rare and expensive. Furthermore, they are often open to ethical criticism (whenever it can be difficult to justify why treatment should be withheld from one group); also, it is not always clear whether the findings established in a small-sample RCT will hold when an intervention is scaled up or implemented in another or broader context.

A second-best option is quasi-experimental study design. One form of a quasi-experimental pre-test/post-test design is regression discontinuity design, in which the causal effects of interventions are elicited by setting a cut-off or threshold above or below which an intervention is assigned. Regression discontinuity designs are appropriate when randomisation is not feasible, but still require that the intervention itself is under the control of the researchers. Another form of a quasi-experiment is the "natural experiment", in which individuals exposed to the experimental and control conditions are determined by nature or by other factors outside the control of the researchers, but where the process governing the exposure arguably resembles random assignment. However, natural experiments are difficult to find (as they tend to happen by chance), and many times not all baseline data of interest were collected or are available via administrative sources. In a way, international comparisons between countries, in which a particular intervention was implemented in one country but not in others, are a particular case of a natural experiment. However, it is generally very difficult to infer causality from international comparisons, given the variety of uncontrolled differences between countries in terms of both sample characteristics and other variables of interest. This is especially the case when relying on cross-sectional (rather than longitudinal) data. Ideally, researchers would work with panel data (longitudinal surveys collecting repeated measures for the same person over time), as this offers an opportunity to explore whether a change in a given determinant is associated with a subsequent change in a person's reported trust. Data quality, and the generalisability of findings, are further enhanced when panel data stem from large and representative samples such as those obtained by NSOs.

However, since large-scale comparable panel data are rarely available, most studies concerned with the impacts and drivers of trust must rely on cross-sectional datasets. Strictly speaking, such analyses are concerned with identifying covariates rather than causality. Nevertheless, even when direct causal inferences cannot be made with cross-sectional data, evidence from other sources about the direction of causality can be used to enrich the interpretation of the results.

Methods of analysis

Apart from the nature of the research question, the most appropriate method for the analysis of trust depends largely on the type of data collected and the method of collection.

The simplest test for the strength of a relationship or association between two variables is a bivariate correlation. The Pearson or product-moment coefficient can be calculated when the data are assumed to be normally distributed and the expected relationship between them is linear; Spearman's Rank and other non-parametric tests are available for ordinal data and non-linear relationships. Partial correlation enables examining the relationship between two variables while removing the effect of other variables. Correlations indicate the possible existence of a predictive relationship between two variables, but they do not imply causation.

For a more thorough examination of the impacts and drivers of trust in cross-sectional, international and longitudinal studies, regression analysis is widely used. Regression is a correlation-based statistical technique that examines how well a set of explanatory or independent variables can predict a given dependent variable, i.e. the chosen trust measure. Regression allows assessing the impact of several independent variables simultaneously in one model, and can be used even when explanatory variables are correlated with one another. However, the "best" regression solution (in terms of variance explained per independent variable) is produced when each independent variable is strongly correlated with the outcome variable but uncorrelated with other variables, whether these other variables are included or excluded from the model. Where curvilinear relationships are expected, as in the case of the U-shaped relationship between age and trust, squared values are typically used in regression models. For income, which is expected to have an asymptotic relationship with many outcome variables, values are often transformed into logs.

A range of regression models are available, depending on the nature of the independent variable: linear regression models are recommended for continuous variables, while ordinal and dichotomous outcomes are usually analysed with Probit or Logit models. Since trust measures are not continuous, a Probit or Logit model is the most appropriate in studies where trust is the outcome of interest. However, the output of linear regression models is generally easier to interpret, and in many cases there are few differences between estimates based on linear regression and Probit/Logit models (Diener and Tov, 2012). Therefore, it is recommended that both types of analysis are run when dealing with trust data, and that linear regression estimates are reported when the results do not differ.

More advanced methods of analysis, not described in detail here, include multilevel structural equation modelling, which allows for combining micro-level and macro-level information, and propensity score matching, which estimates the effect of an intervention by accounting for the covariates that predict receiving the treatment.

How to assess whether the results are significant

Associations between variables can be analysed by examining their correlation coefficients (denoted as r). These range from -1 to +1, with -1 implying a perfect negative linear association and +1 a perfect positive linear association. The square of the coefficient (or r^2) denotes the share of the variation in one variable that is related to the variation in the other. Thus, an r^2 of 0.36 (i.e. r of 0.60) means that 36% of the variance in the dependent variable is explained by the variance in the independent variable. The statistical significance of a correlation coefficient indicates the likelihood that the coefficient would be found in a sample by chance when no significant association actually exists between the variables.

In regression-based analyses, the overall model "fit" of the observed data is described in terms of the proportion of variance in the dependent variable that is explained by the variance in the independent variables (the overall multiple-correlation coefficient, or R^2).

Statistical significance is used to indicate whether the overall model provides a better-than-chance prediction of the dependent variable. In order to further understand how each independent variable contributes to the prediction of the dependent variable, one examines the set of regression coefficients for the independent variables. In linear regression, the size (and sign) of the coefficient for each independent variable indicates how much the dependent variable is expected to increase (if positive) or decrease (if negative) when the independent variable increases by one unit, while holding all the other independent variables constant.

The analysis of trust data in practice

The results from selected econometric analyses that have been carried out with trust data are described below to illustrate the different ways in which trust data have been modelled as well as the effect sizes that have typically been produced so far. The description distinguishes between studies that have considered trust as an outcome variable (i.e. where the analysis is concerned with establishing the determinants of trust) and those that have viewed trust as an input variable (i.e. where the analysis aims at quantifying the impact of trust on other outcomes of interest). In this context, the use of variables at different units of analysis, namely at the individual or community/country-level, is highlighted as well.

Trust as an outcome

Anderson and Tverdova (2003) provide an example of a study that is concerned with trust in institutions (in their case, trust in civil servants, expressed on a 1-5 scale) as an outcome. The authors combine individual-level trust data from the International Social Survey Program (ISSP) with information on respondents' political allegiance (the ISSP survey also includes an item about which party the individual voted for in the last election) as well as with Transparency International's country-level Corruption Perception Index (CPI) for 16 countries at different maturity stages of democracy. Controlling for various system and individual-level factors, including current macroeconomic performance, economic development, level of democracy, political interest, electoral participation, socio-economic status and standard demographic variables, the authors use multilevel structural equation-modelling to establish the impact of corruption and political allegiance on trust. Their findings indicate that individuals in countries with higher levels of corruption have less trust in civil servants, and that respondents who support the majority political party are significantly more trusting of civil servants than those in the minority. For example, a typical respondent in a country where corruption is absent scores 4.26 on the 5-point scale measuring respondents' trust in civil servants. In contrast, a respondent in a country in the most corrupt category scores 2.76, while those in a country in the mid-range of the corruption scale (=5) rate civil servants 3.33. Looking at the effect of political allegiance in the majority or minority party on trust in civil servants, those supporting the majority party score 3.98 in the average country, while those favouring the minority party score 3.83.

An example of a study that looks at the determinants of interpersonal trust is the research by Alesina and La Ferrara (2000) on US communities. Using micro-level trust data from the US General Social Survey (GSS) from 1974-94 as dependent variable, the authors rely on multiple Probit regression analysis (since the trust item is measured on a binary scale) to test for the impact of a rich set of independent variables: these include individual characteristics and experiences also featured in the GSS (education, income, marital status,

age, gender, religious affiliation, traumatic experiences such as divorce, disease, accidents and financial misfortune) as well as community characteristics drawn from administrative data (income inequality, racial and ethnic homogeneity, crime level). The authors find that the factors reducing trust most significantly are: a recent history of traumatic experiences; belonging to a group that historically feels discriminated against, such as minorities (Blacks in particular); and, to a lesser extent, being a woman, being economically unsuccessful in terms of income and education, or living in a racially mixed community and/or in one with a high degree of income disparity.[9] For example, in econometric terms, an increase in the Gini index by one standard deviation decreases the likelihood of trust by 2.5 percentage points. Interestingly, when the authors repeat their analysis with trust in a range of public institutions as dependent variable, they find no effect of community racial and income heterogeneity. These findings suggest that these factors influence interpersonal interactions but not trust in institutions.

Trust as a driver of other outcomes

Interpersonal trust has been analysed intensively not only as an outcome, but also as a determinant of other variables of interest. In particular, interpersonal trust has been linked in cross-country studies to income per capita and economic growth (Putnam et al., 1993; Knaack and Keefer, 1997; Ahn and Hemmings, 2000; Temple, 2000).

For example, Algan and Cahuc (2013), using a sample of 106 countries[10] over the years 1981 to 2008, regress the log of income per capita on average general social trust, including controls for education, ethnic fractionalisation and population size. Their results indicate that increasing trust by one standard deviation leads to a rise in income per capita of 0.18, or 2% of the sample mean. Even when introducing additional control for institutional history and quality, interpersonal trust remains significant at the 5% level. The authors replicate these findings using data from 800 regions around the world, relying on weighted regressions using the number of individuals polled per region. In this case, however, the positive and significant correlation between trust and GDP per capita disappears once country fixed effects are introduced into the model, suggesting that cross-country heterogeneity rather than within-country variation in interpersonal trust and income per capita drives the result. Apart from income per capita, the authors also examine the relationship between trust and economic growth. Regressing average annual growth between 1990 and 2009 on average trust between 1981 and 1990 for a 52-country sample as featured in the World Values Survey, and controlling for the country's investment level, their findings yield a correlation coefficient of 0.48 between trust and growth, significant at the 5% level. The overall R^2 of the model of Yann and Cahuc is 0.706, i.e. explaining 70% of the variance in economic growth.

Apart from economic outcomes, interpersonal trust has also been found to predict a range of social outcomes, such as different dimensions of health status and health-related behaviour (Lochner et al., 2003; Lindström, 2005; Brown et al., 2006; Poortinga, 2006; Petrou and Kupek, 2008), crime rates (Buonanno et al., 2009) and subjective well-being (Helliwell and Wang, 2010). Boarini et al. (2012) explore the determinants of subjective well-being across OECD countries and conclude that country-average trust predicts individual life satisfaction at the 5% significance level. Drawing on Gallup World Poll data, the authors' linear regression, controlling for demographics, socio-economic status and other variables related to well-being, estimates that the impact on life satisfaction of a one-unit change in aggregate average interpersonal trust is equivalent to multiplying current income by 1.23.

Generally, studies examining either the causes or the impacts of trust illustrate that the choice of the unit of analysis – i.e. whether to consider variables at the individual or country level – can make an important difference to the results. Most often, it is community and country characteristics that influence trust levels, while aggregate trust levels have been shown to affect individual-level outcomes, such as life satisfaction and health status. This is, in some way, good news for policy makers, since it suggests that policy interventions at the aggregate level (local, regional, national) are the most relevant.

Challenges to interpreting coefficients

A few words of caution about some analytical challenges in econometric exercises are warranted. While these challenges are not unique to studies examining the drivers or impacts of trust, they should be kept in mind when interpreting results, particularly if these are used to inform policy.

First, normally the size of regression estimates can be interpreted as the effect on the dependent variable of a one-unit change in the independent variable. However, regression coefficients can be affected if the independent variables in the equation are strongly correlated amongst themselves. The literature distinguishes between mediation, confounding and suppression effects resulting from the inclusion of additional independent variables (Dolan et al., 2007). Furthermore, if several measures of the same driver are included in the model, their intra-correlations can crowd one another out, to the limit case where an otherwise relevant driver can fail to reach significance and hence be overlooked (Boarini et al., 2012 refer to this scenario as "over-identification"). Therefore, the decision of which and how many variables to include in the analysis should always be informed by a clear theoretical structure and by an understanding of the hypothetical causal pathways between the different factors.

Second, estimates will be affected by endogeneity problems when the variable of interest is correlated with the model error term, an issue that is often referred to as the "omitted variable problem". In this case, variables that are omitted, i.e. not included in the model, but which are causally related to both outcome and predictor variables, can make it seem as if there is a significant statistical relationship between the latter two, although none exists in reality. In a hypothetical example, a relationship between interpersonal trust and community-level ethnic diversity might be caused by community-level income inequality, which it will be important to include as a measure in the model as well.

Third, the ability to understand the direction of causality between trust and other variables is strongest when experimental, quasi-experimental or longitudinal panel data are available. More often than not, this will not be the case, and analyses have to rely on cross-sectional data. In this case, the results need to be interpreted alongside evidence about the causal direction from other sources. One method often used to overcome issues of reverse causality in regression analyses is to include instrumental variables. An instrumental variable is one that has a direct association with the independent variable in question (e.g. trust), but not with the outcome of interest (e.g. GDP growth). Although it can be difficult to identify appropriate instrumental variables, several researchers have exploited the inherited trust of US immigrants as an instrument for trust in their origin country in cross-country regressions of GDP growth (Sangier, 2010; Algan and Cahuc, 2010).

Very importantly, the possibility of shared method variance should always be kept in mind when interpreting the results of analysis of trust data. Shared method variance refers

to variance that is attributed to the measurement method, rather than the constructs of interest. In the case of trust, the main concern is that, when drivers are also measured through self-reported data, self-report biases (including social desirability biases, response styles, cultural bias, etc.) can inflate the estimated impact of self-reported drivers relative to those measured through other means (e.g. objective observations). Questions that have very similar response formats (e.g. 0-10 scales) are especially likely to have correlated errors.

When comparing the effects and impacts of trust, particularly in cross-sectional data, it is therefore important to consider how each of those variables was measured. Whenever possible, the use of longitudinal data is encouraged, as individual fixed effects can be controlled for in these analysis formats. Alternatively, non-self-reported measures of constructs of interest should substitute self-reported items if data are available – for example, behavioural measures of trust collected through behavioural games.

5.5. Conclusion

This chapter has addressed different ways of working with trust data *after* it has been collected, ranging from reporting interpretation all the way to analysis. It has outlined best practice where available, raised awareness of challenges to interpretation and analysis, where relevant, and pointed to gaps in current knowledge where they exist. In doing so, the chapter aims to serve as a useful reference handbook for data analysts, researchers and journalists interested in outputting, reporting on and analysing trust data. The key messages of the chapter can be summarised as follows:

- 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. The front page of statistical releases should provide a quick overview for the general public and policy makers and focus on reporting levels of a single headline measure alongside a brief media commentary. In contrast, the second and third pages of statistical releases should cover trust data in greater detail (e.g. trends in and distributions of trust, different types of trust) for those who want to dig deeper. The third page of data releases is designed for researchers and policy analysts who want to carry out analyses themselves and require access to the trust microdata and the survey instruments.
- 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 thresholds, 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.
- 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 the chapter provides an initial attempt to document the magnitude of differences (between population subgroups, 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 and the possible impact of culture on trust estimates. This chapter stressed the point that, since it is very difficult in practice to distinguish between cultural impact and cultural bias, and the methods available to remove cultural bias face significant limitations, it is generally not recommended to correct trust data for cultural influences.
- 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 chapter has presented a range of analytical methods applicable to trust data, providing examples of studies that have relied upon trust as either an input or output variable. These studies illustrate that the choice of the unit of analysis – i.e. whether to consider variables at the individual or country level – matters, that community characteristics often influence trust levels, and that aggregate trust levels influence individual-level well-being outcomes.
- Lastly, the chapter has highlighted the importance of keeping common econometric challenges in mind when working with trust data. These include omitted variable bias, over-identification, reverse causality and shared method variance.

Notes

1. The mean, median, and mode are the most common measures of central tendency. The mean refers to the sum of all measurements divided by the number of observations in the dataset; the median is the middle value that separates the upper 50% from the bottom 50% of the dataset; and the mode is the most frequent value in the dataset. The median and the mode are the only measures of central tendency that can be used for ordinal data, whereas the mode is the only central tendency measure applicable to categorical data.
2. Ordinal data are those measured on scales where the intervals between scale points are not assumed to be equal, but there is an underlying sequence or rank order. For example, we assume that a 5 is lower than a 6 and a 6 is lower than a 7, but we do not assume that the distance between 5 and 6 is equivalent to the distance between 6 and 7. Linear regression relies on continuous variables, where cardinality is assumed, i.e. the size of the number on a scale is expected to have a direct linear relationship with the amount of the variable in question. Tabachnick and Fidell (2001), however, note that, in the social sciences, it is common practice to treat ordinal variables as continuous, particularly where the number of categories is large – e.g. seven or more – and the data meet other assumptions of the analysis.
3. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".
4. Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.
5. See Notes 3 and 4.
6. The importance of considering changes in individuals' levels of trust is highlighted by Bilson et al. (2017). The authors, using panel data from the Household, Income and Labour Dynamics in Australia (HILDA) survey from 2005-2014, investigate the effect of individual income on interpersonal trust, demonstrating the importance of accounting for individual-level fixed effects: the income coefficient switches from positive and statistically significant, in a pooled regression, to negative and

statistically significant, in a fixed-effects panel model. These findings underscore the need for individual-level panel data for all trust measures.

7. However, some researchers have also found that the education gradient in trust, especially in the US context, has generally diminished over time (see Dalton, 2005).

8. It should be noted that there is some debate about the direction of causality between educational attainment and trust, with some authors asserting that education impacts trust, and others claiming that trust causes differences in education (see Bjørnskov, 2006, for a discussion).

9. Putnam (2007) reports similar effects of community contextual variables on individual levels of interpersonal trust in the United States, including significant negative impacts of poverty rates, non-violent crimes, an index of ethnic homogeneity, and population density at the census tract level. Bjørnskov (2006), for an international sample, and Helliwell and Wang (2010), focusing on Canada, also report similar findings. Nevertheless, these results should not be generalised lightly: they need to be interpreted with the historical context of each country in mind, as well as while recognising that no consensus currently exists on how to define and measure ethnicity and ethnic fragmentation.

10. The authors use binary measures of trust drawn from the World Values Survey (1981-2008), the European Values Survey (1981-2008) and the Afrobarometer (2005).

References

Ahn, S. and P. Hemmings (2000), "Policy influences on economic growth in OECD countries: An evaluation of the evidence", *OECD Economics Department Working Papers*, No. 246, OECD Publishing, Paris, *http://dx.doi.org/10.1787/581718238837*.

Alesina, A. and E.L. Ferrara (2000), "The determinants of trust", *National Bureau of Economic Research*, No. w762, *www.nber.org/papers/w7621*.

Algan, Y. and P. Cahuc (2013), "Trust, growth and well-being: New evidence and policy implications", *IZA Discussion Paper,* No. 7464.

Algan, Y. and P. Cahuc (2010), "Inherited trust and growth", *The American Economic Review,* Vol. 100(5), pp. 2060-2092.

Anderson, C.J. and Y.V. Tverdova (2003), "Corruption, political allegiances, and attitudes toward government in contemporary democracies", *American Journal of Political Science*, Vol. 47, pp. 91-109.

Bârgăoanu, A., L. Radu and D. Varela (2015), *United by or Against Euroscepticism? An Assessment of Public Attitudes towards Europe in the Context of the Crisis*, Cambridge Scholars Publishing, Cambridge.

Bilson, J.R., M. Jetter and I. Kristoffersen (2017), "Gender differences in the link between income and trust levels: Evidence from longitudinal data", *IZA DP,* No. 10585.

Bjørnskov, C. (2006), "Determinants of generalized trust: A cross-country comparison", *Public Choice*, Vol. 130, pp. 1-21.

Blanton, H. and J. Jaccard (2006), "Arbitrary metrics in psychology", *American Psychologist*, Vol. 61, pp. 27-41.

Boarini, R. et al. (2012), "What makes for a better life? The determinants of subjective well-being in OECD countries – Evidence from the Gallup World Poll", *OECD Statistics Working Papers*, No. 2012/03, OECD Publishing, Paris, *http://dx.doi.org/10.1787/5k9b9ltjm937-en*.

Brown, T.T. et al. (2006), "The empirical relationship between community social capital and the demand for cigarettes", *Health Economics*, Vol. 15 (11), pp. 1159-1172.

Buonanno, P., D. Montolio and P. Vanin (2009), "Does social capital reduce crime?", *Journal of Law & Economics*, University of Chicago Press, Vol. 52(1), pp. 145-170.

Carl, N. and F.C. Billari (2014), "Generalized trust and intelligence in the United States", *PLoS ONE*, Vol. 9(3).

Chabanet, D. (2007), "Chômage et exclusion sociale: L'échec européen", *Politique Européenne*, Vol. 2, pp. 157-187.

Chanley, V.A. (2002), "Trust in government in the aftermath of 9/11: Determinants and consequences", *Political Psychology*, Vol. 23, pp. 469-483.

Clark, A.K. and M.A. Eisenstein (2013), "Interpersonal trust: An age-period-cohort analysis revisited", *Social Science Research*, Vol. 42, pp. 361-375.

Dalton, R.J. (2005), "The social transformation of trust in government", *International Review of Sociology*, Vol. 15(1), pp. 133-154.

Deaton, A. (2012), "The financial crisis and the well-being of Americans; 2011 OEP Hicks Lecture", *Oxford Economic Papers*, No. 64, pp. 1-26.

Diener, E. and W. Tov (2012), "National accounts of well-being", *Handbook of Social Indicators and Quality of Life Research*, pp. 137-157, K.C. Land, A.C. Michalos and M.J. Sirgy (eds.), Dordrecht: Springer.

Dolan, P., T. Peasgood and M. White (2007), "Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being", *Journal of Economic Psychology*, Vol. 29, pp. 94-122.

Eurostat (2015), *European Union Statistics on Income and Living Conditions (EU SILC)*, *http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=ilc_pw03&lang=en* (accessed 23 March 2017).

Exton, C., C. Smith and D. Vandendriessche (2015), "Comparing happiness across the world: Does culture matter?", *OECD Statistics Working Papers*, No. 2015/04, OECD Publishing, Paris, *http://dx.doi.org/10.1787/5jrqppzd9bs2-en*.

Gallup (2014), *Gallup World Poll, www.gallup.com/services/170945/world-poll.aspx* (accessed 06 February 2017).

Guiso, L., P. Sapienza and L. Zingales (2006), "Does culture affect economic outcomes?", *The Journal of Economic Perspectives*, Vol. 20(2), pp. 23-48.

Helliwell, J.F. and R.D. Putnam (2007), "Education and social capital", *Eastern Economic Journal*, Vol. 63(1), pp. 1-19.

Helliwell, J.F. and S. Wang (2010), "Trust and well-being", *National Bureau of Economic Research*, Working Paper No. 15911.

Hudson, J. (2006), "Institutional trust and subjective well-being across the EU", *Kyklos International Review for Social Sciences,* Vol. 59(1), pp. 43-62.

Jordahl, H. (2007), "Inequality and trust" *Working Paper Series, No.* 715, Research Institute of Industrial Economics.

Knack, S. and P. Keefer (1997), "Does social capital have an economic payoff? A cross-country investigation", *Quarterly Journal of Economics*, Vol. 112 (4), pp. 1252-88

Li, T. and H.H. Fung (2012), "Age differences in trust: an investigation across 38 countries", *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*.

Lindström, M. (2005), "Social capital, the miniaturization of community and high alcohol consumption: A population-based study", *Alcohol and Alcoholism*, Vol. 40 (6), pp. 556-562.

Lochner, K.A. et al. (2003), "Social capital and neighborhood mortality rates in Chicago", *Social Science & Medicine*, Vol. 56, pp. 1797-1805.

New Economics Foundation (2009), contributing authors: J. Michaelson, S. Abdallah, N. Steuer, S. Thompson and N. Marks, "National accounts of well-being: Bringing real wealth onto the balance sheet", *New Economics Foundation*, January, London, *www.nationalaccountsofwellbeing.org/learn/download-report.html*.

New Zealand State Services Commission (2014), *Kiwis Count Survey* (database), *www.ssc.govt.nz/kiwis-count-datasets* (accessed 20 April 2017).

OECD (2015), Education at a Glance 2015: OECD Indicators, OECD Publishing, Paris, *http://dx.doi.org/10.1787/eag-2015-en*.

OECD (2013), *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/9789264191655-en*.

ONS (2016), "Statistics on trust for methodological testing from the opinion's survey, October 2015 to May 2016", released 10 November, Office for National Statistics UK, ONS, Newport, *www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/adhocs/006326statisticsontrustformethodologicaltestingfromtheopinionssurveyoct2015tomay2016*.

Pew Research Center (2016), *Beyond Distrust: How Americans View their Government*, *www.people-press.org/2015/11/23/1-trust-in-government-1958-2015/* (accessed 18 February 2017).

Petrou, S. and E. Kupek (2008), "Social capital and its relationship with measures of health status: Evidence from the Health Survey for England 2003", *Health Econ.*, Vol. 17, pp. 127-143.

Poortinga, W. (2006), "Social relations or social capital? Individual and community health effects of bonding social capital", *Social Science & Medicine*, Vol. 63(1), pp. 255-270.

Poulin, M.J. and C.M. Haase (2015), "Growing to trust", *Social Psychological and Personality Science,* Vol. 6(6), pp. 614-621.

Putnam, R.D. (2007), "E Pluribus Unum: Diversity and community in the Twenty-First Century – the Johan Skytte Prize Lecture", *Scandinavian Political Studies*, Vol. 30, pp. 137-74.

Putnam, R.D. (2000), *Bowling Alone: The Collapse and Revival of American Community*, Simon & Schuster (eds.), New York.

Putnam, R.D., R. Leonardi and R.Y. Nanetti (1993), *Making Democracy Work*, Princeton, NJ, Princeton University Press.

Rothstein, B. and E.M. Uslaner (2005), "All for all: Equality, corruption, and social trust", *World Politics, Vol.* 58, pp. 41-72.

Sangnier, M. (2010), "Does trust favor macroeconomic stability?", *PSE Working Papers,* No. 2009-40.

Sen, A. (2002), "Health: Perception versus observation", *British Medical Journal*, Vol. 324, pp. 860-861.

Senik, C. (2011), "The French unhappiness puzzle: The cultural dimension of happiness", *Paris School of Economics Working Papers*, No. 2011-34, Paris-Jourdan Sciences Économiques.

SOM Institute (2015), *The National SOM Survey* (database), *http://som.gu.se/som_institute/-surveys/national-som* (accessed 05 March 2017).

Statistics Canada (2013), *Trends in Social Capital in Canada, www.statcan.gc.ca/pub/89-652-x/89-652-x2015002-eng.htm* (accessed 10 March 2017).

Stats New Zealand (2014), *New Zealand General Social Survey*, *www.stats.govt.nz/nzgss2014* (accessed 20 April 2017).

Stolle, D. (1998), "Bowling together, bowling alone: The development of generalized trust in voluntary associations", *Political Psychology*, Vol. 19(3), pp. 497-525.

Stolle, D., S. Soroka and R. Johnston (2008), "When does diversity erode trust? Neighborhood diversity, interpersonal trust and the mediating effect of social interactions", *Political Studies*, Vol. 56, pp. 57-75.

Tabachnick, B.G. and L.S. Fidell (2001), *Using Multivariate Statistics*, Fourth edition, Needham Heights, MA: Allyn and Bacon.

Temple, J. (2000), "Growth effects of education and social capital in the OECD countries", *OECD Economic Studies*, Vol. 2001/2, OECD Publishing, Paris, *http://dx.doi.org/10.1787/eco_studies-v2001-art11-en*.

Tokuda Y. et al. (2008), "Interpersonal trust and quality-of-life: A cross-sectional study in Japan", *PLoS ONE, Vol.* 3(12).

Ubel, P.A. et al. (2005), "What is perfect health to an 85-year-old? Evidence for scale recalibration in subjective health ratings", *Medical Care*, Vol. 43(10), pp. 1054-1057.

Uslaner, E.M. (2000), "Producing and consuming trust", *Political Science Quarterly*, Vol. 115, pp. 569-590.

van de Vijver, F.J.R. and Y.H. Poortinga (1997), "Towards an integrated analysis of bias in cross-cultural assessment", *European Journal of Psychological Assessment*, Vol. 13(1), pp. 29-37.

Veenhoven, R. (2008), "The international scale interval study: Improving the comparability of responses to survey questions about happiness", *Quality of Life and the Millennium Challenge: Advances in Quality-of-Life Studies, Theory and Research, Social Indicators Research Series*, V. Moller and D. Huschka (eds.), Springer, Vol. 35, pp. 45-58.