

MEASURING WELLBEING INEQUALITY

WORKING PAPER ON THE SELECTION OF A HEADLINE INDICATOR

Written by: Annie Quick & Stephen Devlin

New Economics Foundation
www.neweconomics.org
info@neweconomics.org
+44 (0)20 7820 6300
@NEF

Registered charity number 1055254
© 2018 The New Economics Foundation

Contents

1	Introduction.....	5
2	Project methodology.....	7
2.1	Rapid literature search	7
2.2	Interviews.....	7
2.3	Surveys	8
2.4	New statistical analysis	8
2.5	A note on sampling.....	9
3	Criteria for an indicator on wellbeing inequality.....	10
3.1	Reflective of ethical and policy concerns	10
3.2	Methodological robustness.....	10
3.3	Other criteria.....	13
3.4	Conclusion	14
4	What should we measure?.....	16
4.1	Three ethical propositions	16
4.2	Understanding dispersion aversion	17
4.3	Understanding aversion to suffering and weighted universalism.....	22
4.4	Co-benefits	26
4.5	Conclusion	27
5	Review of indicators.....	29
5.1	Measures of dispersion	29
5.2	Alternatives to measures of dispersion.....	37
5.3	Data Analysis	41
6	Recommendation.....	42
7	Further research	45
8	References.....	46
	Appendix 1: Survey responses	50
	Appendix 2: Data analysis.....	63

EXECUTIVE SUMMARY

1. Background

Policy makers and academics are increasingly interested in wellbeing inequality. The vast majority of academic studies into wellbeing inequality currently use standard deviation of personal wellbeing measures such as self-reported life satisfaction or happiness. However, there has so far only been limited debate on whether standard deviation is the most appropriate measure, and its choice is rarely justified.

This working paper presents research commissioned by the Office for National Statistics (ONS) and carried out by the New Economics Foundation (NEF) in collaboration with the What Works Centre for Well-being. NEF was tasked with exploring the strengths and weaknesses of different measures of wellbeing inequality and to make a recommendation of a measure which could be reported by the ONS alongside mean wellbeing.

2. Project methodology

Measures were judged against a number of criteria:

- Reflective of public priorities - a good measure of what matters in relation to wellbeing inequality, for example inequality per se, or helping the worst off.
- Robust to methodological biases – robust to response biases, ordinality vs. cardinality and bounded scale effects.
- Easy to compute - simple to compute and analyse for non-specialists
- Easy to communicate - simple to explain and understand
- High predictive power – provides insight into other trends
- Sufficient variation – changes enough over time that it can feasibly be influenced by policy
- Low correlation with the mean - measures that provide information not already captured by the existing measure of mean wellbeing

We consulted with key stakeholders with an interest in wellbeing at a national level via in-depth interviews and a roundtable discussion. This was supplemented by an online survey where wider views were invited, including from non-experts.

3. Main findings

From these responses we identified three distinct reasons why people cared about wellbeing inequality. These were:

- 1) Dispersion aversion - that large differences in life experiences are inherently undesirable
- 2) Suffering aversion - placing a special status on individuals experiencing wellbeing below a particular threshold
- 3) Weighted universalism - valuing improvements in wellbeing for everyone, though with more value given to improvements in wellbeing for those at the bottom of the distribution

Although all three propositions were held to some degree by some participants, weighted universalism was more widely shared than the other positions. This suggests that an indicator should first and foremost reflect the wellbeing of the worst off, with diminishing weight given to those who are already doing well.

Almost all inequality measures we identified are, like standard deviation, measures of dispersion. While these indicators may have good uses for academic or other purposes, they failed our criteria for a headline indicator of being reflective of public and policy priorities. Of those that were available

There is a significant gap in the literature exploring alternatives to dispersion measures for wellbeing inequalities. We assessed three:

- The average of the bottom 40%
- The percentage below a threshold on the wellbeing scale
- Subjectively-weighted average

Based on our analysis, we propose the use of the percentage below a threshold as a headline indicator of wellbeing inequalities. For example, in the year ending September 2017, 4.46% of people – almost one in 20 – responded 4 or below when asked how satisfied they are with their life on a scale of 0 to 10. When reported alongside mean wellbeing, this indicator reflects the dominant values we found in our research, particularly concern with the wellbeing of the worst off, with diminishing weight given to those higher up the scale. It is easy to explain, compute and analyse.

4. Recommendation

This paper aims to open up discussion about appropriate indicators of wellbeing inequality. Based on analysis so far, we propose using a threshold for a single national indicator. Further work is needed to identify which threshold to use, although interviewees suggested 4, 5 or 6 on a scale of 0-10. Using a threshold of 4 would reduce the sample to those with lower wellbeing, while increasing the threshold to 5 or 6 might allow for a more preventative approach, focussing on interventions aimed at those who are struggling before they slip into very low wellbeing.

However, different indicators may be used for different purposes. We encourage researchers to reflect on which wellbeing inequality measure they choose and for a broader debate between key stakeholders on appropriate wellbeing inequality measures for different purposes.

1 INTRODUCTION

Policy makers and academics are increasingly interested in measures of wellbeing inequality. The vast majority of studies into wellbeing inequality currently use standard deviation of personal wellbeing measures such as self-reported life satisfaction or happiness. However, there is extremely little research on whether standard deviation is the most appropriate measure, and its choice is rarely justified.

The measures we use matter. While thousands of statistics are reported by the Office of National Statistics (ONS) every year, certain key indicators – for example Gross Domestic Product (GDP), or overall carbon emissions – are used as key headline indicators. In order to effectively guide policymaking, these measures should be robust to biases and a good reflection of the underlying phenomenon they are trying to measure, for example economic growth or contributions to man-made climate change.

These measures have a public as well as a technical role, helping non-specialists understand what progress is being made, informing political debate and enabling the electorate to hold decision-makers to account. For these purposes indicators must be meaningful and clearly understood. At a local level they may also be used by community actors to understand local need and guide action.

This research aimed to go back to first principles, asking two questions:

1. *What* should a wellbeing inequality indicator aim to measure?
2. *How* should this aspect of wellbeing inequality be measured?

Question 1 is a normative question, though it can be informed by empirical research. Question 2 is a practical question about how well indicators perform against different criteria such as methodological robustness and communicability.

These questions are also sequential. It is not possible to have a useful discussion about how wellbeing inequality should be measured without a clear understanding of what it is that the indicator should be measuring. While there has been some discussion in the literature on question 2 (Delhey & Kohler, 2011; Goff, Helliwell, & Mayraz, 2016; Veenhoven & Kalmijn, 2005), we have found no explicit discussion of question 1 as it relates to the measurement of wellbeing inequality.

In addressing these questions we have imposed some limitations on the scope.

Firstly, we have made the assumption that the distribution of personal wellbeing is of policy or public interest. This opinion is not universally held, including by some of the stakeholders we engaged with for this project. We heard concerns that it is natural that

wellbeing fluctuates throughout peoples' lives, and so the reduction of wellbeing inequality was an inappropriate or unrealistic goal for policy makers; that the use of wellbeing inequality in policy making would be a distraction from health inequalities, and the much more advanced evidence-base on how to reduce them; that a focus on wellbeing inequality is redundant as the best way to reduce wellbeing inequality is to improve average wellbeing.

These are important concerns that require discussion and may be informed by our research on what aspects of wellbeing inequality are of public and policy concern. However, while these issues have been discussed elsewhere by the authors (Quick, 2015), they are not the subject of this paper.

Secondly, in this paper we will refer to 'wellbeing' to mean subjective, personal wellbeing. Where a specific measure is needed for illustration we have used the ONS, eleven-point life satisfaction scale, but much of the normative discussions would apply to other key subjective measures such as happiness. Further research would be needed to explore how the issues of robustness would apply to a wider range of multi-dimensional wellbeing measures, or whether measures of ill-being such as stress or anxiety would require an alternative approach.

Thirdly, we will focus only on the inequality of wellbeing in itself – i.e. regardless of any other variable such as gender or income. These 'univariate' measures of inequality are used for income with measures such as the gini coefficient, or ratios. Wellbeing inequality can also be measured between groups, e.g. considering the difference in wellbeing according to ethnicity or education. These 'bivariate' inequality measures have received some attention in relation to wellbeing (Abdallah, Wheatley, & Quick, 2017; Harrison, Quick, & Abdallah, 2016) but are much more common in the health inequalities literature, reporting the gap in health outcomes between people living in different areas, or with different levels of education.

This project is the result of a collaboration. Work was carried out by the New Economics Foundation on behalf of the ONS with support and advice from the What Works Centre for Wellbeing. It is intended to inform further discussions about the adoption of an indicator on wellbeing inequality and be used in policy and research. While some aspects are intended for a technical audience, much of the discussion, in particular section 4 on normative aspects of wellbeing inequality, should be accessible to any interested audience.

2 PROJECT METHODOLOGY

Summary: To answer our research questions, we conducted a rapid literature search, interviews with academics, politicians and policy makers. To address the question of 'what' we should measure, we also ran an online survey. However, our samples for surveys and interviews were not representative, and this research should be seen as an initial stage in an ongoing programme of understanding wellbeing inequalities.

2.1 RAPID LITERATURE SEARCH

In order to understand the wider context, we conducted a rapid literature search of existing studies that considered the measurement of wellbeing inequalities. This included a targeted call for evidence amongst those in the field, reference chasing and a shallow online search. We included academic as well as grey literature.

As the New Economics Foundation and What Works Centre have worked on a number of projects involving inequalities in wellbeing over the past few years, information was also gathered from the following:

- Notes from an all-party-parliamentary group roundtable on inequalities in wellbeing from April 2016 (APPG for Wellbeing Economics, 2016)
- Notes from an expert roundtable as part of the 'Making Wellbeing Count' ESRC-funded project in collaboration with City University London and Cambridge University in January 2016 (Abdallah & Quick, 2016)
- Email conversations with a number of academics responding to the authors' published work
- Feedback from partners and advisors of the What Works Centre for Wellbeing and the Centre's Community Wellbeing Evidence Programme in relation to wellbeing inequality analysis being conducted as part of the programme.

2.2 INTERVIEWS

We conducted six interviews with policy makers, politicians and academics between February 2017 and April 2017. The primary purpose of these interviews was to gather views on question 1: What should we measure?

The interviews were qualitative and semi-structured and took place either over the phone or face-to-face. The interviewer walked the interviewee through a set of scenarios that had been developed for an online survey (see below) as prompts for discussion.

The second aim of the interviews was to gather technical expertise on specific aspects of measurement. For this aim, we contacted academics with specific expertise as issues emerged from the research. These conversations were undertaken mostly by email, though two telephone interviews were conducted.

In most cases this latter sample was also asked their views on ‘what to measure’ so in practice, the interviews overlapped.

2.3 SURVEYS

To complement the qualitative interviews we designed a short online survey to gain a greater breadth of responses on the same issues.

The survey asked respondents to indicate their agreement with statements about the importance of wellbeing inequalities relative to average wellbeing and the wellbeing of the least satisfied. It also asked respondents to make judgements about the desirable characteristics of the wellbeing distribution for a society overall, and about policies that affect people at different points on the wellbeing distribution. For example, one question asked respondents to choose between two wellbeing distributions where both have the same mean, but one has a higher standard deviation (and more people in both the most and least satisfied categories), and then to explain whether their choice was motivated by reducing inequality, reducing low levels of life satisfaction, or something else.

We administered the survey using the software Survey Monkey and disseminated it through contacts and social media in the ONS, the What Works Centre for Wellbeing and NEF. A total of 112 responses were received, including 20 from academics working on wellbeing issues and 14 policy professionals (seven worked directly on wellbeing, seven did not). The full questionnaire and results are in Appendix 1.

2.4 NEW STATISTICAL ANALYSIS

We used empirical data to test some of the indicators against our criteria (detailed below). We used the European Social Survey and the Annual Population Survey (UK), which provide both a country-level and local-level test. Section 6 and Appendix 2 provide details of the methodology and results.

2.5 A NOTE ON SAMPLING

The sampling strategy for the interviews was purposive, reaching out specifically to people whom we knew would have thoughts to share. The questionnaire was used to broaden this range of views and was not representative. As there has been almost no discussion of the normative aspects of the measurement of wellbeing inequalities so far, it was valuable to first collate the informed views of those already working in related issues. In order to get a wider perspective on the question 1, including from those with no knowledge of the area, it would be very valuable to do further work using other engagement techniques.

The survey was distributed through mailing lists and on forums known to the New Economics Foundation, What Works Centre and Office for National Statistics as well as family and friends of project staff. These channels are likely to have reached:

- Policy makers and academics who have worked on wellbeing inequalities and may have already been part of conversations about its measurement
- Policy makers, academics and community members who know about wellbeing measurement in general but may not have yet thought about wellbeing inequalities
- People with very little knowledge about either wellbeing or wellbeing inequalities

3 CRITERIA FOR AN INDICATOR ON WELLBEING INEQUALITY

Summary: In this section we explore different criteria against which to evaluate the indicators. We conclude that a good indicator of wellbeing inequality is reflective of public and political priorities; robust to methodological biases; easy to construct and analyse; can be communicated easily; is sensitive enough to reflect policy change; associated with other outcomes of interest; and adds additional information over and above the widely used measure of average wellbeing.

Criteria were initially drawn from literature on effective measures and indicators (Jeffrey & Michaelson, 2015; Whitby, 2011), but these were added to and developed with input from interviewees and project partners. In the next section we will move on to assess indicators against these criteria.

3.1 REFLECTIVE OF ETHICAL AND POLICY CONCERNS

First and foremost, the indicator must be a good measure of what matters in relation to wellbeing inequality (research question 1). By ‘inequality’ do we really mean variation, or are we just concerned about improving the wellbeing of the worst off? Are some kinds of inequality more or less amenable to policies or interventions?

3.2 METHODOLOGICAL ROBUSTNESS

The following issues were identified in relation to methodological robustness.

3.2.1 Ordinality vs. cardinality

Most measures of subjective wellbeing use an ordinal scale (the numbers indicate relative but not absolute differences), whereas many measures of dispersion, like standard deviation, are intended for cardinal scales (where the numbers indicate relative and absolute differences).

We therefore have to convert our subjective wellbeing results into cardinal responses. A key problem is that different transformations from ordinal to cardinal may result in different orderings of wellbeing distributions in terms of inequality (Dutta & Foster,

2013). This is especially acute for responses that are expressed in lexical categories (e.g. “very satisfied”) and, perhaps, less acute for responses that are expressed in numerical categories (e.g. the 0-10 life satisfaction scale) since it is possible that respondents do interpret the latter as a cardinal scale. However, even in the latter case we have to make an assumption about the interval between each response – the natural assumption may be that there is a distance of 1 unit between each point on the scale, but this is not necessarily reflective of how people answer the question – it implies that 0 and 10 are the true minimum and maximum scores and that, for example, a score of 8 is twice as good as a score of 4. If there is more difference between what people report as 1 and what people report as 2 than there is between 8 and 9, then this constant interval transformation is misapplied. This suggests that measures such as standard deviation are not theoretically consistent at ranking distributions and should not be used for subjective wellbeing measures.

3.2.2 Bounded scale effects

It is useful to distinguish between peoples’ actual wellbeing and the wellbeing score that people provide in surveys. We use wellbeing scores because we hope that they will be a useful quantification of peoples’ underlying wellbeing, but they are not the same thing: one is a number, the other is a state of being.

One characteristic of wellbeing scores is that they often on a scale of zero to ten. In order to measure wellbeing we impose limits. One could argue that actual wellbeing is also bounded, and that it cannot improve indefinitely. However, even if actual wellbeing does have limits, it may not be well reflected in the scale on which it is often measured. For people that report a score of 10 on the life satisfaction scale, is it possible for them to achieve higher wellbeing? For those that report 0, is it possible for them to be even less happy? If the answer to either of these questions is yes, there is the potential for responses to be artificially bunched at the top or bottom of the distribution. This poses a specific challenge to the study of inequalities in wellbeing.

First, it implies that a cardinal transformation of the ordinal responses is not quite accurate, which creates some concerns for the ranking ability of certain measures, including standard deviation, as discussed above.

Second, it means that distributions with a mean that is further from the centre of the distribution (either high or low) will have an artificially lower standard deviation. Quick (2015) describes a hypothetical distribution of wellbeing, and demonstrates that if everyone’s wellbeing improved by one point, the result would produce a lower standard deviation because those who had already scored themselves a 10 could not score

themselves any higher and so become bunched. Some measures, such as standard deviation, may be more closely associated to the mean, suggesting that they might be particularly affected by bounded scale effects. How might analyses overcome potential confounding from bounded scale effects? In particular, is controlling for mean wellbeing satisfactory, or does this depend on the specific distributions under analysis?

A related problem is whether to use ratio measures. A ratio measure describes how much more of something there is in relation to something else, for example the ratio of wellbeing between the top 80% and the bottom 20% describes how much more wellbeing the top 80% of people have compared to the bottom 20%. The use of ratio measures assumes that there does exist an absolute zero level of happiness and that this corresponds to zero on the life satisfaction scale. When you divide one number by another their absolute magnitude matters, unlike when you subtract one number from another (as when you calculate a range measure), in which case only the distance between them matters. In other words 3 divided by 2 is not the same as 9 divided by 8, whereas 3 minus 2 and 9 minus 8 are equivalent. The implication is that if we think that actual satisfaction levels go beyond the measured response scale then ratio measures may not be appropriate. Extreme response bias

One study has used data from the European Social Survey to explore different kinds of cultural biases in how participants respond to questions. While many kinds of bias did not seem to apply, the authors did find evidence of cultural differences in 'extreme response style' i.e. the extent to which respondents use the far ends of the scale. In this case, they did not find that this had an impact on the results for life satisfaction (Eurostat, 2012). However, if an inequality measure is particularly sensitive to the extremes of the wellbeing scale, it may be that this bias is more problematic than it is for studies of the mean because of the way in which inequality measures pay particular attention to extremes. The study only looked at cross-country comparisons, but it could be that there are cultural differences in extreme response style between demographic groups within countries. No further research was identified on this topic, so we were not able to assess indicators against this criteria, although if it does turn out to be a substantial issue, inequality measures that give particular weight to extreme responses may be less robust.

3.2.3 Unreliable results at the top of the distribution

During our analysis, we came across an irregularity in the life satisfaction scores at the top of the distribution. When plotting data from the British Household Panel Survey against well-known drivers of wellbeing such as income, health and employment, the

linear relationship between these indicators breaks down for the top score, which in the British Household Panel Survey is 7 (using data from 1991 to 2009). With every point increase in the life satisfaction scale respondents also had higher income, and were less likely to be long-term sick or disabled, or unemployed until the highest score, at which point the relationship reversed. So, people who reported the highest category of wellbeing were slightly more likely to be long-term disabled, unemployed and had lower income, than the second highest wellbeing score. This analysis has not been replicated on other datasets as far as we are aware, but it suggests either that the happiness of those at the top of the wellbeing scale is less connected to their objective life circumstances, or that response biases are operating. The latter could be some form of acquiescence bias, or be the result of survey fatigue, where reporting a ten is easier than thinking through the lower options.

3.3 OTHER CRITERIA

3.3.1 Easy to compute

Measures that are simple to compute or are pre-programmed into common statistical software are preferred over measures that are more complicated. Given the demand for wellbeing inequality measures that can be understood at a local level (Abdallah et al., 2017), a good wellbeing inequality indicator should be one that researchers and analysts working in local authorities or third sector organisations are able to compute themselves from raw data in a consistent way. There are two aspects to ease of computation;

1. how easy it is for non-specialists to compute the indicator itself and
2. how easy it is for technical experts to compute change over time.

3.3.2 Easy to communicate

Measures are preferred that are easy to explain and understand, and ideally can be explained in a sentence. Measures are also preferred that are concrete – i.e. can be understood in real terms. For example, Oxfam’s assertion that “the top 1% has as much wealth as the other 99% put together” is concrete in that it is meaningful on its own, while “the UK’s inequality-adjusted human development index is 0.84” is itself opaque; it only becomes meaningful when compared to other countries or periods of time.

3.3.3 Predictive power

An indicator is likely to be more useful to policy and political audiences if it tells us about other social or economic trends. One study for example suggested that one measure of wellbeing inequality (mean pair distance) was associated with the Brexit vote (Abdallah, 2016). A wellbeing inequality indicator that has high predictive power is more likely tell a useful and interesting story about peoples' lived experiences.

3.3.4 High variation and change over time

Measures are only useful if they change. Less sensitive indicators will respond in a slower or more muted way, making it hard to evaluate whether changes in policy or practice have had an impact.

3.3.5 Low correlation with the mean

Mean wellbeing is both widely reported and research suggests that it is strongly correlated with measures of wellbeing inequality (Goff et al., 2016; Ovaska & Takashima, 2010; Veenhoven, 2005). Interpretations differ as to whether this relationship is causal or possibly the result of bounded-scale effects, but either way, if a wellbeing inequality measure is to add value to the existing widely used measure of mean wellbeing, it should add as much information as possible. The lower its correlation with mean wellbeing, the more value it is likely to add.

3.4 CONCLUSION

A number of the methodological issues raised are under-researched. This is particularly the case in relation to extreme response bias and misleading responses at the top of the distribution. At this stage, we therefore have not included robustness to these biases as part of our criteria.

Our chosen criteria are therefore:

- Reflective of public and political priorities
- Robust to bounded-scale effects
- Robust to cardinal/ordinal scales
- Easy to construct
- Easy to analyse change
- Can be communicated in a sentence
- Can be understood in concrete terms

- High predictive power
- High variation
- Low correlation with the mean

A further criteria could be added on whether the measure can be used for different wellbeing measures, though this was not assessed in this project.

4 WHAT SHOULD WE MEASURE?

Summary: Participants had different reasons for caring about inequality in wellbeing, which can be divided into three ethical propositions. The first is a belief that policy should be focussed on those with very low wellbeing, to establish a threshold under which people should not fall. The second is the desire to reduce the gap between those with very high wellbeing and very low wellbeing, in the belief that such a gap may create social disruption or damaging social comparisons. The third is that we should be interested in improving everyone's wellbeing, but that our concern should be weighted to the worst off. This final proposition was the most dominant both in revealed preferences and through explicit expressions of peoples' views. However, the other two ethical propositions were also held by some participants.

4.1 THREE ETHICAL PROPOSITIONS

In this section we bring together findings from the interviews, stakeholder engagement, literature review and survey results to address the question underlying our first research question: *What* should be measured in relation to inequalities in wellbeing?

We identified three distinct ethical propositions that could motivate people to reduce wellbeing inequalities.

Dispersion aversion is the proposition is that large differences in life experiences are inherently undesirable. It is a pure aversion to dispersion in wellbeing outcomes. The corresponding policy approach is to reduce this gap, and this can be achieved either by reducing wellbeing at the top of the distribution or by improving the wellbeing at the bottom of the distribution.

Weighted universalism values improvements in wellbeing for everyone – this is what makes it ‘universal’. However, more value is given to improvements in wellbeing for those at the bottom of the distribution, compared to those whose lives are already going well. Under this proposition, we should not be concerned about dispersion per se. If policy can improve anyone’s wellbeing, then it should, but when there are trade-offs to be made, resources should be directed towards those at the bottom, with that weight diminishing up the scale.

Suffering aversion places special status on individuals experiencing wellbeing below a particular threshold (for example, a life satisfaction score of 3). The corresponding policy approach would be to focus on ameliorating suffering and to devote resources only to a group of particularly badly off individuals.

These ethical positions are distinct but not mutually exclusive. For example, it is consistent to have both an aversion to dispersion and a desire to ameliorate suffering. The positions are intentionally stylised and in practice, most people would hold at least two; however, the question of what we should measure boils down to the relative importance we give to these different ethical propositions.

A fourth ethical proposition is strict utilitarianism, in which the aim of policy should be to improve average overall wellbeing. As our research question was to identify a measure of wellbeing inequality to complement the existing use of the mean wellbeing, we did not explore this ethical proposition. As it happens, although people drew on utilitarian calculations in their decision-making, nobody we spoke to revealed a commitment to strict utilitarianism over and above the other positions.

4.2 UNDERSTANDING DISPERSION AVERTION

In economics and across the political spectrum, it is widely argued that policy should focus not only on reducing material poverty but also on reducing overall inequalities in income and wealth. This is not only because of the diminishing marginal returns of income, but also because it is argued that income inequality itself – independent of average incomes – has harmful social and health consequences (Pickett & Wilkinson, 2010). The explanations given for this relationship are often psycho-social, in particular that inequality breeds distrust and creates status competition which increases anxiety. It is also widely accepted that material poverty is relative to the living standards of the rest of the population. Indeed, in the UK, poverty is now measured in relative terms – defined as those living on 60% below the national median income.

The question for this project is whether we should be similarly concerned about inequality of wellbeing, as we are about inequality of income. Given that many would see wellbeing as an intrinsic outcome, while income is an instrumental output, are the same arguments valid?

Confusingly, the language of “reducing inequality” is commonly conflated with reducing poverty, low wellbeing, or inequity. Although these may have the effect of reducing inequality that is neither the object nor a necessary consequence. As explained by one interviewee:

“Sometimes the word inequality is also used to denote relative deprivation, for example when one deems a country more unequal when 80% of its inhabitants are unhappy (e.g. rating 4) and 20% are happy (e.g. rating 8) then in the reversed case of 80% happy and 20% unhappy”. (Kalmijn & Veenhoven, 2005, p. 14)

To resolve some of this confusion we can pose the levelling down thought experiment:
Would you prefer scenario A or scenario B?

Scenario	A	B
Person 1	£100	£50
Person 2	£20	£20

If equality is desirable per se then a change from Scenario A to Scenario B is a beneficial change in at least one respect, even though no one is individually better off (Parfit, 1997). In the case of income or wealth either scenario could be consistently justified as preferable depending on whether or not you believe equality is intrinsically valuable. However, in the case of wellbeing is this the case?

A similar question can be applied to health inequality. Health inequality measures are widely reported (almost always between population groups), but the policy solutions proposed are about how to increase the health of those groups who are struggling, rather than make very healthy people less healthy.

Many of the people we heard from felt intuitively that a wide dispersion was undesirable, with most people agreeing that equality is important even if there are fewer people with high wellbeing (Figure 1).

It is important for there to be a high degree of equality in people's wellbeing, even if that means there are fewer people at the top of the wellbeing scale.

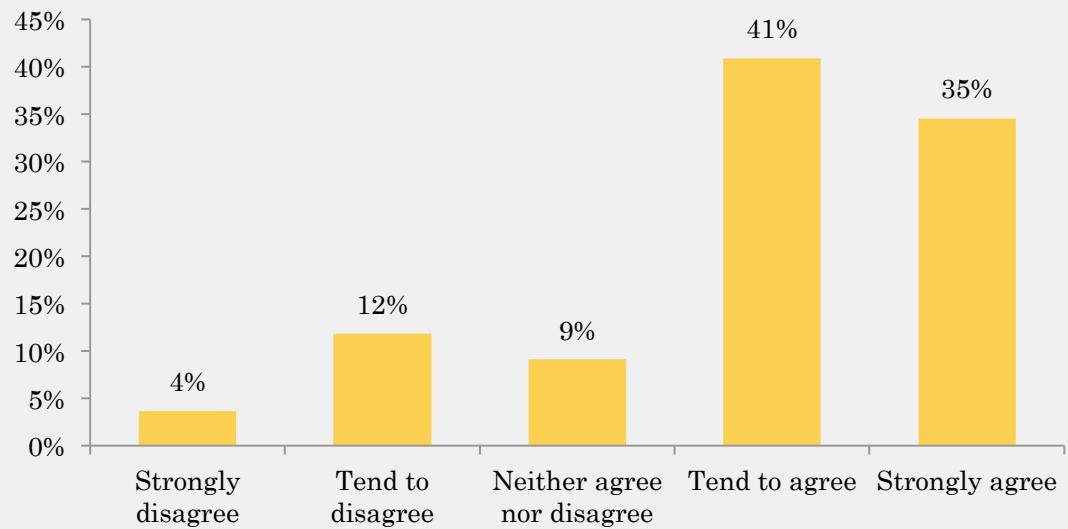


Figure 1 Responses to question 'Is it important for there to be a high degree of equality...'

Probably the most widely cited reason was that gaps in people's experience create other undesirable social outcomes.

"If there's too wide a gap between people who are happy with their life and people who aren't that's damaging for social cohesion, trust in society and in government and... for willingness to, for example, subsidise people through taxation. It leads to a fragmented society and a break down in the social contract".

Interviewee

"Distribution B [with higher dispersion] implies a greater level of inequality and more potential problematic impacts on everyone. You'd start to have less stability because of that level of inequality; you'll see far greater demand [for public resources] generated by low wellbeing - mental health, anti-social behaviour".

Interviewee

"Extremes produce unhappiness, alienation and discontent"

Survey respondent

However, this view was not universal, and when people were asked to distinguish between concern for inequality *per se*, compared to simply raising up the bottom, respondents were split (Figure 2).

Inequality of wellbeing doesn't matter as long as the wellbeing of the least satisfied people is increasing.

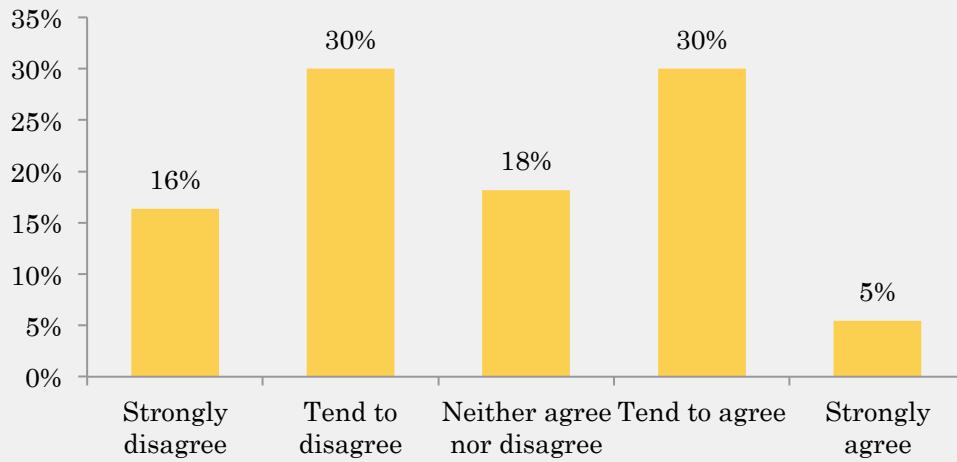


Figure 2 Responses to survey question 'Inequality of wellbeing doesn't matter...'

Some interviewees made an explicit comparison to income to highlight the difference to wellbeing.

"Income is a fixed pie - and you can redistribute that while preserving the mean. The idea that you would take happiness from someone and give it to another is absurd. So you always pull up the bottom. For happiness as a whole there is no keeping up with the Jones's effect".

Interviewee

"With income inequality there are ways in which high income actually causes negative impacts lower down, such as through conspicuous consumption. But with wellbeing, do really happy people annoy people who are unhappy? I don't think so".

Interviewee

We attempted to explore this question with survey respondents by adapting the thought experiment used for incomes described above, asking: if you had a wellbeing score of 6, would you prefer to live in a society where the average is 7, or where the average is 9? We called these scenarios A and B.

	Your wellbeing	Average wellbeing
Scenario A	6	9
Scenario B	6	7

Almost half answered A (Figure 3) suggesting that they personally wouldn't expect to be negatively affected by a lower relative position on the wellbeing scale. Peoples' explanations for this choice included the selfless desire for others to be happy, but also the self-interested perception that a society in which others were better off gave them more hope that they might see improvements in their own wellbeing.

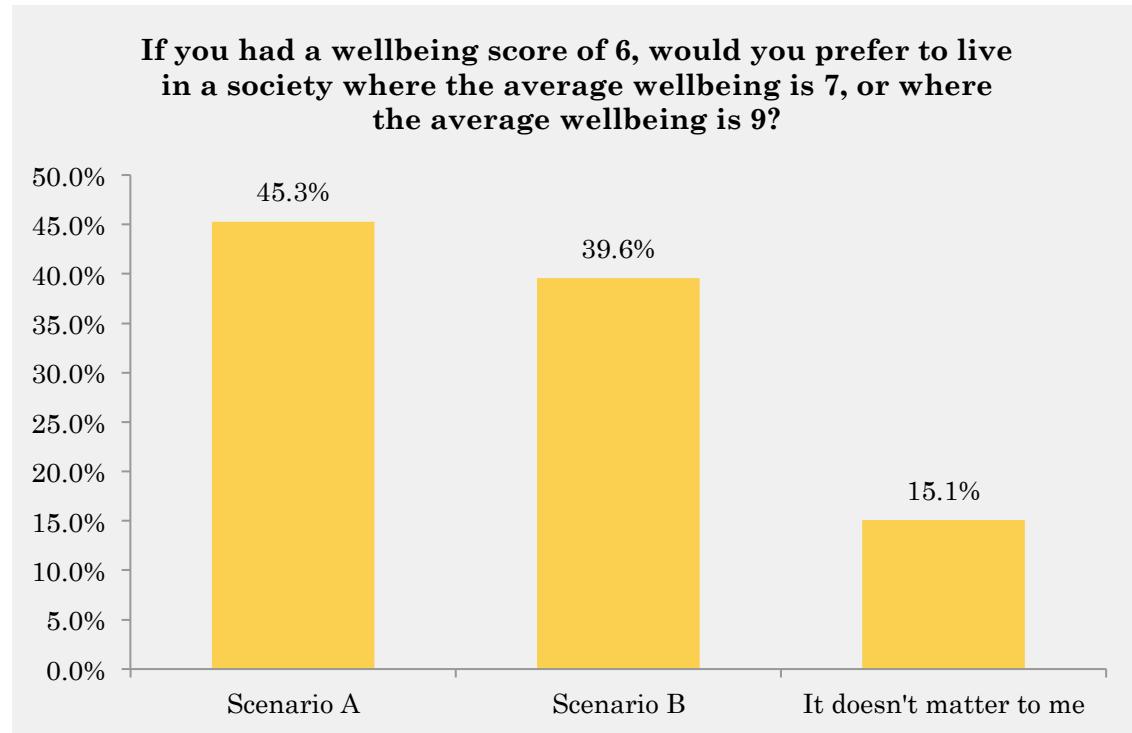


Figure 3 Responses to survey question 'if you had a wellbeing score...'

However, almost 40% said that they would prefer scenario B and the reasons they give do resonate with reasons people might give for an aversion to income inequality:

“I'd like to feel like I'm in the same boat as others.”

Survey respondent

“The rational choice should be B. I'd rather be nearer the average than looking at the wellbeing of everyone else who is better than mine... it might sound selfish but you can't understand personal wellbeing in isolation of your surroundings and your community.”

Interviewee

One could argue that the question is logically inconsistent. After all, given that wellbeing is a self-reported measure of how people feel their life is going, any negative impacts of low relative wellbeing status should be captured in the scores that people give. While that is theoretically true, it doesn't take into account predictions of future events, which featured quite strongly in peoples' decisions.

Two areas of research may provide some empirical evidence of these negative impacts of relative wellbeing status.

Firstly, there is a growing literature on the negative impacts of personal comparisons on social media, whereby people present a more positive picture of their lives on social media compared to how they feel in real life. Research suggest that these comparisons increase anxiety as people compare their lives to others' and feel inadequate as a result (Coyne, McDaniel, & Stockdale, 2016). Although some of these comparisons could focus on external conditions such as peoples' career, possessions or foreign holidays, a great deal of social media does consist of people expressing their emotions and how they feel about their life – not dissimilar to a subjective wellbeing question. Although we found no studies exploring whether material or experiential content was most responsible for social comparisons on social media, this could be an example of a 'keeping up with the Jones' effect for wellbeing.

Conversely, social contagion theory suggests the opposite – that wellbeing is transferred between people. Longitudinal analysis suggests that being around happy people makes people more likely to be happy themselves, and similar patterns have been found for health, health behaviours and illness (Fowler & Christakis, 2008; Smith & Christakis, 2008).

Aversion to dispersion was the most divisive of the ethical propositions. Unlike the other two ethical propositions, which most people held to some degree, some respondents expressed strong and well-supported arguments why in the case of wellbeing, reducing dispersion should not be an aim at all.

4.3 UNDERSTANDING AVERSION TO SUFFERING AND WEIGHTED UNIVERSALISM

There was a strong consensus that policy should prioritise the wellbeing of those who are struggling the most. This was reflected in the survey when asked to choose between three indicators. These indicators were chosen at the start of the project, when we aimed to pick a range of measures that would reflect different kinds of inequality. These were two measures that dispersion measures: standard deviation, and the ratio of wellbeing between the top 80% and the bottom 20%. We also chose one measure focussed on the bottom of the distribution: The average of the bottom 40%. This measure was preferred by more people than the two other dispersion measures (Figure 4).

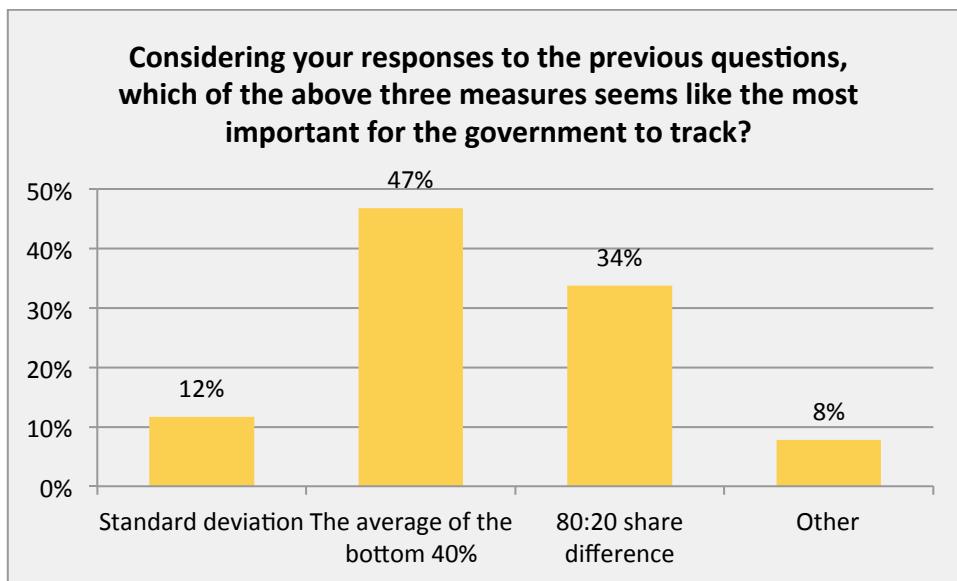


Figure 4 Survey responses on alternative measures of wellbeing inequality

Some interviewees referred to a threshold to aim for:

"We could establish a floor and maintaining that floor could be a useful departure point for economic and social policy. Once you've achieved that floor, you try to shift the floor up."

Interviewee

"I found myself counting what percentage of people were 5 or less in the two scenarios, or 6 or less."

Survey respondent

These comments suggest a particular aversion to suffering. In order for this ethical position to be dominant, one would need to be interested in improving the bottom of the distribution as a moral imperative regardless of what was going on at higher levels. Some survey respondents did suggest that improving the wellbeing of those really struggling could not be traded off against improvements higher up the distribution.

"There is a certain point of wellbeing below which people shouldn't fall, even if that slows down the process of improving average wellbeing."

Interviewee

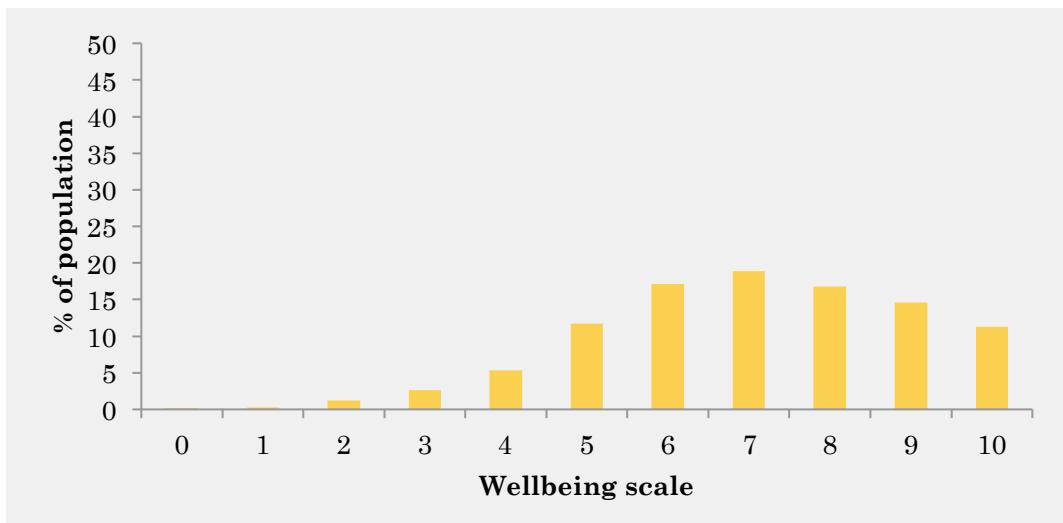
One civil servant expressed the view that their department was primarily concerned with improving the lives of the very worst off. They referred to the theory of the hierarchy of needs to argue that once people had obtained a certain level of wellbeing, and the autonomy that brings, people should be left to take responsibility to improve it.

The idea of providing a floor is common in arguments to guarantee a basic level of living standards. While this is possible for material goods such as income and housing, respondents recognised that there are some areas of personal wellbeing that are not – or should not – be influenced by policy. These include life events, for example, suffering a bereavement, and aspects of mental health that are hard to prevent or treat. One respondent expressed the view that achieving a society in which nobody fell below a certain point was unrealistic, and that we are all likely to fall to the bottom of the distribution at some point in our lives.

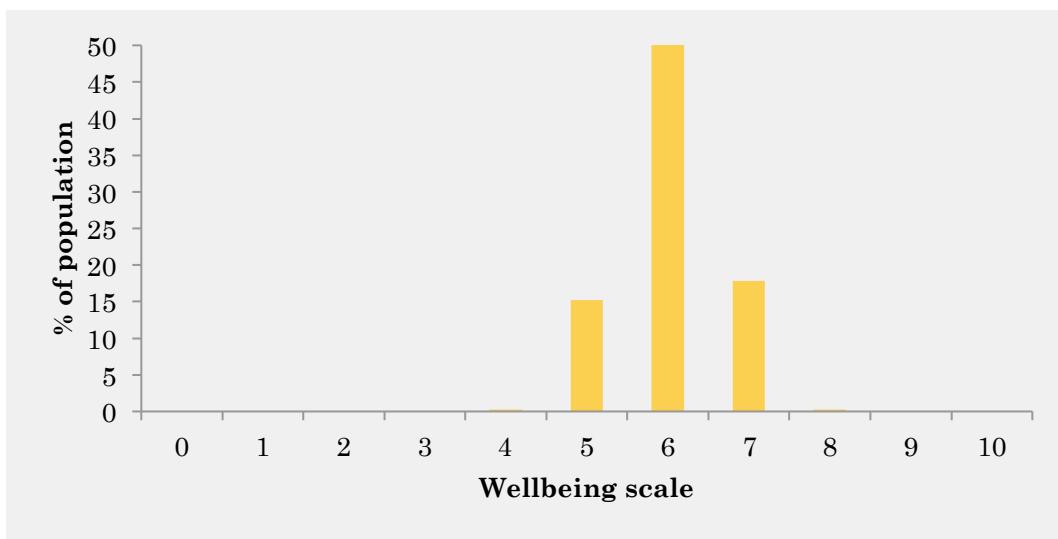
However, although there was strong agreement that those suffering most should be of greatest concern, it did not necessarily override concern for those higher up the scale.

We asked survey respondents to compare two distributions. In distribution C people are quite spread out, with some people having very low wellbeing but most people higher up the distribution. The mean is 7. In distribution D, people's wellbeing was concentrated around one part of the scale, with fewer people with very high or very low wellbeing. The mean wellbeing is 6.

Distribution C



Distribution D



If the dominant ethical proposition was aversion to suffering, the response should be D. However, a number of people discussed their decision-making in terms of applying mental weights to different parts of the distribution, suggesting they were trading-off improvements across the scale against each other.

“Relatively few people in lower categories in distribution C – indeed, far fewer than in the higher categories. While extra weight should be given to those lower down the scale, given the relative numbers in the low vs upper categories, it still was not enough to outweigh the benefit at the top end.”

Survey respondent

“Looking at the average implies that each person is weighted equally to the policymaker, which shouldn’t be the case - we should focus on those with lower wellbeing. So there is a case for using a measure like average of the bottom 40%, which is still about inequality.”

Interviewee

One interviewee argued that wellbeing should follow a comparable weighting process to those explored in the study of the elasticity of marginality with respect to income.

“This is really a question about the ‘marginal utility of utility’. I would like to see a function that is heavily weighted to the bottom, but even so if you can get someone from a 9 to a 10 without affecting anything else on the distribution, that has to be good.”

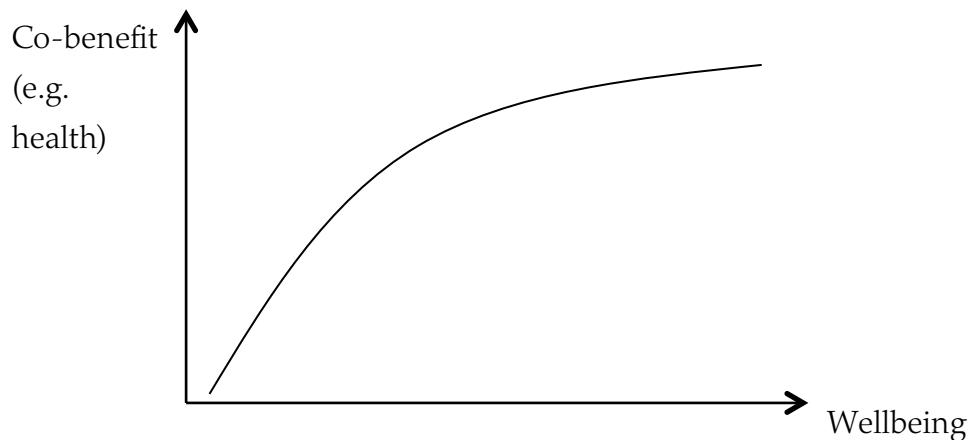
Taking a universal approach that gives particular weight to those at the bottom was also proposed by Michael Marmot in relation to health inequalities, coining the term ‘proportionate universalism’ (Marmot et al., 2010). Marmot argued:

"Focusing solely on the most disadvantaged will not reduce health inequalities sufficiently. To reduce the steepness of the social gradient in health, actions must be universal, but with a scale and intensity that is proportionate to the level of disadvantage. We call this proportionate universalism."

4.4 CO-BENEFITS

Although wellbeing is valued in itself, it can also be valued by policy makers because of the way in which it can improve other outcomes in a particular policy area, for example if people with higher wellbeing are less likely to become unemployed, engage in anti-social behaviour or become physically unwell. One additional motivation for having preferences about the nature of the wellbeing distribution is that the 'co-benefits' of wellbeing could be non-linear. In other words, improvements in some parts of the wellbeing distribution may produce more other benefits from the same change in wellbeing. This would be an instrumental motivation that would imply different weights along the distribution, in line with the weighted universalism ethical position.

If this is the case then it would suggest that public policy could achieve a greater "volume" of benefits by targeting interventions at particular parts of the distribution.



Binder and Coad (2010) find that the marginal effect of life satisfaction on health is greater for those in the lower end of the health distribution and Graham et al. (2004) find some evidence consistent with non-linearities in the effect of wellbeing on future income. Apart from these small glimpses, there has been very little research on this question.

Diener and Chan (2011) examine the effect of personal wellbeing on health (finding a positive relationship) and conclude that one of the priorities for future research is to examine whether there are non-linearities in this relationship.

4.5 CONCLUSION

Although all three ethical propositions were held to some degree by some people, our research suggests that weighted universalism was more widely shared than the other positions, and is theoretically coherent. This suggests that an indicator should first and foremost reflect the wellbeing of the worst off, with diminishing weight given to those higher up the ladder.

However, it is interesting to note the strength of dispersion aversion amongst some participants. The interviews and our own research have identified legitimate reasons to be concerned about wide gaps in wellbeing, including values-based reasons about fairness and social justice, as well as instrumental concerns such as social unrest.

Compared to the other two ethical propositions, however, aversion to dispersion suffers from a number of weaknesses as a strict ethical proposition.

Most importantly, nobody we spoke to said they would ever be in favour of decreasing the wellbeing of someone at the top of the distribution in order to reduce inequality, which would be the logical conclusion if dispersion aversion were the strongest ethical proposition.

For example, one interviewee held a strong aversion to dispersion, and preferred the 80:20 measure over the average of the bottom 40%. However, when pushed on this last point, they responded:

"Do I want anyone to be less happy? No."

Others suggested that although one might be theoretically averse to dispersion, in practical terms the only way this could be pursued would be to pull up the bottom, effectively making the first ethical proposition obsolete in policy terms.

"Income is a fixed pie - and you can redistribute that while preserving the mean. The idea that you would take happiness from someone and give it to another is absurd. So you always pull up the bottom."

Interviewee

Additionally, although many people expressed some aversion to dispersion, peoples' aversion often softened as they discussed it further. Given that these questions are usually considered with respect to income, it's possible that some respondents had not made the shift to thinking about wellbeing. Our initial survey testing and interviews demonstrated that people hadn't actually mistaken the questions to be about income

inequality. However, it may have taken some time for respondents to fully comprehend the different issues involved in relation to wellbeing.

5 REVIEW OF INDICATORS

Summary: Most inequality indicators are measures of dispersion, and therefore do not adequately meet our first criteria of being reflective of policy and public concerns. There is a gap in measures of inequality that give particular weight to those towards the bottom of the scale but, of those available, the 'percent below a threshold' was assessed to best meet our criteria based on existing research.

5.1 MEASURES OF DISPERSION

Most inequality measures are measures of dispersion. Almost all of the literature on wellbeing inequalities so far has used standard deviation (Fahey & Smyth, 2004; Goff et al., 2016; Ott, 2005; Ovaska & Takashima, 2010; Stevenson & Wolfers, 2008; Veenhoven, 2005). However, none of our interviewees (including those who had used the measure in their own analysis) argued in favour of standard deviation. Where other measures have been used or proposed, they are also measures of dispersion.

These measures do not reflect a weighted universalist position, and so fail our first criteria. However, given their widespread use, and their potential value for other purposes, we assess them against other criteria.

5.1.1 Standard deviation

Standard deviation is very widely understood in academic fields and is easy to construct and analyse in any statistical programme including Excel. Although its technical definition is not easy to communicate (the square root of the mean of the squares of the deviations from the mean), it can be understood in concrete terms, e.g. “68% of the population are within 1.2 points of the average on the life satisfaction scale,” in the case of a normal distribution.

Standard deviation poses a number of methodological concerns. The use of standard deviation implies a number of assumptions about personal wellbeing:

- that the response scale can be interpreted as cardinal, not simply ordinal
- that the response scale corresponds closely to a latent wellbeing variable that is itself bounded at 0 and 10

- that the intervals on the response scale map onto equal intervals on the latent variable's range

Standard deviation is not scale-invariant, which is a characteristic often cited as desirable for inequality measures. It can be made scale-invariant by standardising by the mean, giving the coefficient of variation. This suggests standard deviation is preferable to coefficient of variation (see below).

Goff, Helliwell, & Mayraz (2016) consider the extent to which standard deviation is susceptible to bounded scale effects. Their hypothesis is that inequalities in life satisfaction have a causal, negative relationship with average life satisfaction; i.e. that people are, on average, happier if they live in a country with more equal wellbeing. To test whether this relationship could be the result of a mechanical bounded scale effect, rather than a genuine relationship, they tested the relative associations between the mean and the standard deviations of life satisfaction on social trust – a variable known to be associated with inequality and average wellbeing. If the relationship between inequality in wellbeing and social trust were entirely confounded by mean wellbeing (as in figure b below) the relationship would lose significance when average wellbeing was introduced into the mix. Their regressions found that this was not the case, and standard deviation in life satisfaction remained significant even after adding average wellbeing into the mix. They concluded that a causal relationship, such as that shown in figure 5, is more likely.

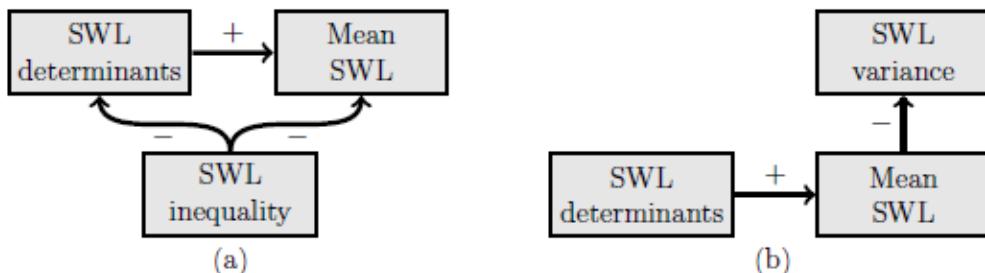


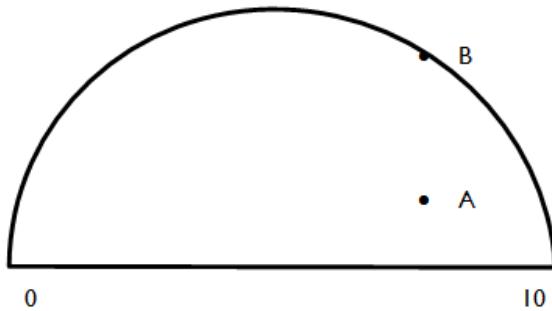
Figure 5 Models of causality between wellbeing inequality and social trust (Goff et. al., 2016)

This analysis provides some reassurance that the bounded scale of life satisfaction does not render standard deviation entirely defunct. However, it does not rule out the possibility that some bounded-scale effect is operating even if it does not eliminate significance in this case.

Two alternative measures have been proposed to overcome the bounded-scale effect; percent maximum standard deviation, and mean pair distance.

5.1.2 Percent maximum standard deviation

Delhey and Kohler (2011) note that the standard deviation has a different range of possible values depending on the mean. For very large or very small means the standard deviation can only take on a smaller range of possible values; for mid-ranging means the



standard deviation could be larger. For example, when the mean is 10 on a 10-point scale this means that everyone responded with the answer 10 and the standard deviation is zero. The possibility space for the standard deviation is, therefore a semi-circle, where the straight edge is the wellbeing scale and its amplitude is the maximum possible standard deviation for a given mean (Kalmijn & Veenhoven, 2014).

The concept of the percent maximum standard deviation (PMSD) is to standardise the standard deviation across this semi-circle. For example, the PMSD for distribution A would be the height of A divided by the height of B. The same standard deviation will yield a larger PMSD toward either end of the wellbeing scale, and a smaller one towards the centre.

Figure 6 Possibility space for standard deviation (Kalmijn & Veenhoven, 2014)

deviations for the 0 – 10 life satisfaction scale.

The figure below depicts the actual maximum possible standard



Figure 7 Actual possibility space for standard deviation for an 11-point wellbeing scale

Veenhoven (2012) objects that even though dispersion is dependent on the mean in this way this is in fact a real phenomenon that should not be corrected for. In other words, people that respond to a wellbeing survey with the maximum response are actually as satisfied with life as it is possible to be (or at least that we must accept their claim that this is the case) and therefore societies with high mean life satisfaction have genuinely lower wellbeing inequality.

A further critique is whether this transformation is substantial – Delhey and Kohler admit that in most cases the correction is very small. But they do also show that the transformed scores are less associated with the mean than simple standard deviation (Delhey & Kohler, 2011), suggesting there may be some statistical benefit to the correction.

An important disadvantage of this measure is that it is difficult to explain, cannot be understood in concrete terms and will be unfamiliar to researchers and technicians who may find it difficult to construct and analyse. It is virtually unused in existing literature aside from the debate between the proponents and opponents outlined above.

5.1.3 Mean pair distance (MPD)

The mean pair distance (MPD) is the average distance between two randomly selected individuals (pair). This is conceptually distinct to standard deviation which measures average distance from the mean. These differences are illustrated in figure

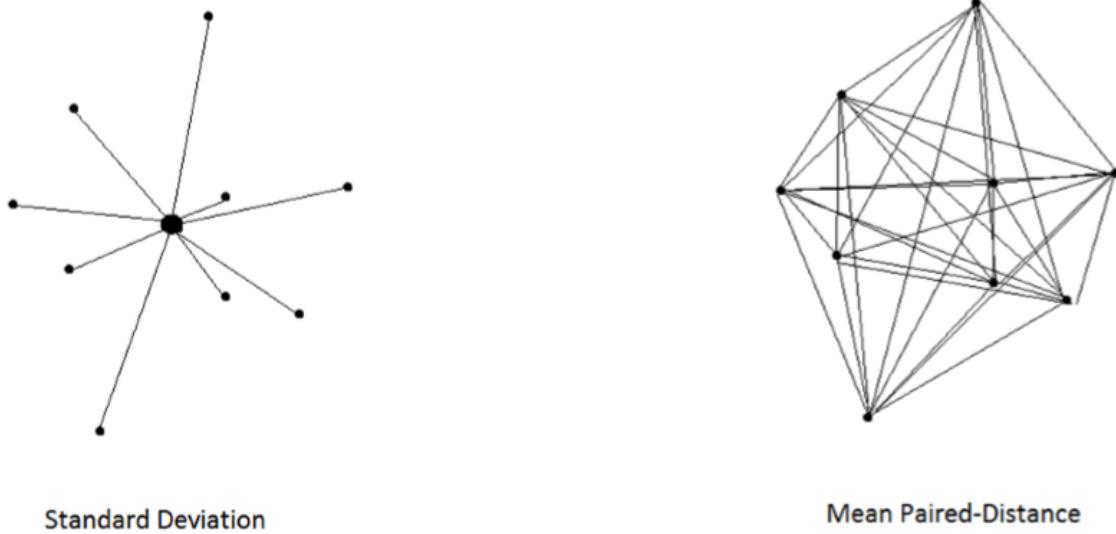


Figure 8 Illustration of models of standard deviation and mean pair distance

MPD is expressed in the same units as standard deviation (points on the wellbeing scale) but should always be higher than standard deviation. At the same time, similar types of distribution will tend to maximise the standard deviation and the MPD – a distribution with two peaks at either end of the scale. Therefore, standard deviation and MPD both indicate that a twin peak distribution has high inequality, but for different reasons.

The number of computations for MPD is much greater than standard deviation – for standard deviation there are N computations and for MPD there are N^2 , where N is the sample size. This isn't much of a problem given the computing power now available, but most statistical packages do not have MPD pre-programmed so this will not be a computation that most researchers will be able to immediately apply.

Although MPD is less widely used than standard deviation it does not seem inherently any more difficult to understand or communicate.

This measure is used in some applications in biology and geography, but is quite rare (unless it has another name that we have not identified), e.g. Daggitt et al. (2016). The only applications of this measure to wellbeing have been undertaken by the New Economics Foundation (Abdallah et al., 2017; Harrison et al., 2016).

5.1.4 Inequality-adjusted wellbeing / geometric mean

Veenhoven and Kalmijn (Kalmijn & Veenhoven, 2014; 2005) advocated combining central tendency and dispersion in one measure of inequality-adjusted happiness. This measure is depicted in the following diagram (the same space as the diagram in PMstandard deviation section above):

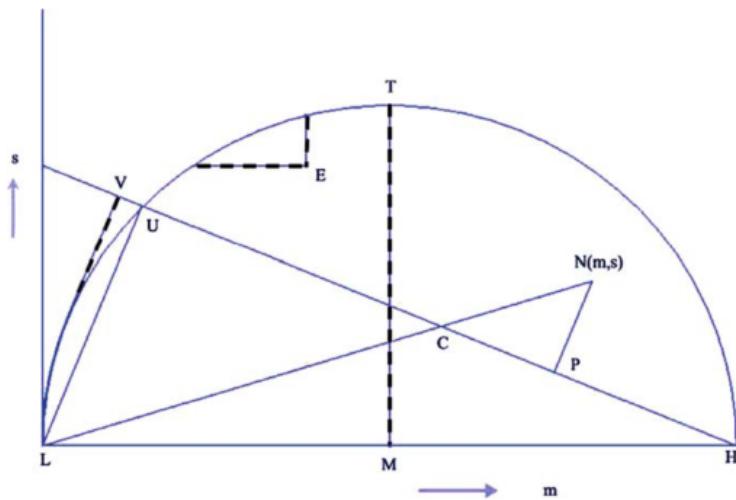


Figure 9 Possible combinations of mean and standard deviations (Kalmijn & Veenhoven)

Point U is assumed to be the worst possible outcome in that it has both low average wellbeing and high wellbeing dispersion. It is a compromise between the utilitarian's worst outcome (point L) and the egalitarian's (point T). To get the value of the index for any society its co-ordinates are orthogonally projected onto the line UH (e.g. point N becomes point P) – the final index value is the proportional distance along line UH . A society at point H would score 100%, while a society at point C would be closer to 50%. Real world scores vary from 20 to 74%.

This measure is an alternative to reporting both the mean and an additional measure of inequality and is strongly related to the mean. While this research aimed to produce a recommendation for an indicator to sit alongside the mean, a combined measure does overcome the concern about standard deviation being taken out of the context of the mean and vice versa.

This is somewhat more complex to calculate than many other measures due to the orthogonal projection calculation.

Another way of combining information on central tendency and dispersion in a single metric is the geometric mean. If all the values of a distribution are identical then the geometric mean is equal to the arithmetic mean; if values differ then the former is less

than the latter. The greater the inequality, the greater the disparity. This means that for two societies with the same average life satisfaction the one with the greater inequality will have a lower geometric mean – it therefore combines these two pieces of information.

A major disadvantage of this measure in the context of wellbeing is that it is incalculable if any of the values are zero (since the product of any series that contains any zeros is zero). In practice there is likely to be at least one zero response making this measure more difficult to use.

These measures are also difficult to communicate and understand in concrete terms.

5.1.5 Variance

Variance is equal to the square of the standard deviation. It is, therefore, an exponential transformation of that measure that will tend to proportionately exaggerate larger values of dispersion. Unlike standard deviation, variance is not expressed in the same units as the wellbeing scale (it is expressed in wellbeing points squared), which makes its interpretation less natural. We did not identify any reasons why variance would be preferable to standard deviation.

5.1.6 Coefficient of variation (relative standard deviation)

The coefficient of variation is calculated as the standard deviation divided by the mean. It is dimensionless, whereas standard deviation is measured in the same dimension as the wellbeing scale. Standard deviation is an absolute measure of dispersion, whereas the coefficient of variation is dispersion as a proportion of the mean. For example, if there are two distributions with a standard deviation of 2, one with a mean of 5 and the other with a mean of 7. Should these be considered equally unequal? If yes then standard deviation is a more appropriate measure, if not then coefficient of variation is better.

The coefficient of variation is strongly related to the mean since it is calculated as standard deviation over mean.

This measure is easy to compute and analyse, but requires a two-step explanation (both of standard deviation, and dividing it by the mean) and is difficult to understand in concrete terms.

5.1.7 Wellbeing Gini coefficient

The Gini coefficient is a very widely used measure of income inequality at a macro level. It is calculated as the ratio between two values: a) the area between a cumulative income share curve (Lorenz curve) and a 45° line and b) 0.5. It takes values between 0 (perfect equality) and 1 (a single individual earns all income).

There is a conceptual problem with the Gini index in the context of wellbeing, in that it measures the proportion of a quantity (e.g. total income) that accrues at different points along the population ranking (Kalmijn & Veenhoven, 2005). Can we really think of, for example, 50% of people having only 20% of the happiness? Happiness is not a quantity that can be transferred from one person to another so this feels instinctively wrong.

Kalmijn and Veenhoven (Kalmijn & Veenhoven, 2005) disqualify the Gini measure for the above reason – it requires the happiness scale to be treated at the ratio level (i.e. a 4 is twice as good as a 2), rather than at the interval level (i.e. a 4 is two units better than a 2).

Nonetheless Gadelman and Porzecanski (2013) use the Gini coefficient because that facilitates comparing happiness inequality with income inequality – something that is not as easy with standard deviation, for example. They show that income inequality is twice the level of happiness inequality.

Becchetti et al. (2010) also use the Gini to measure happiness inequality, in conjunction with the variance and the inter-decile range, though they don't assess its appropriateness.

Although the gini coefficient is widely used and so widely trusted, it is difficult to explain and will be more complex to calculate for those who are not used to using it in relation to income inequality.

5.1.8 S-domination (Dutta & Foster, 2013)/Lorenz dominance (Bellù & Liberati, 2007)

These approaches rank distributions based on the characteristics of their cumulative distribution functions. For example S-domination will rank a distribution higher than another if it has more mass above the median of the other distribution.

This avoids the need to assume that the response scale is cardinal, a key advantage for its proponents (Dutta & Foster, 2013). Therefore, these measures will consistently rank distributions even if one makes different assumptions about the “distances” between different points on the wellbeing scale.

The computations are complex in comparison to measures such as standard deviation – most researchers are unlikely to have existing knowledge of how to compute these measures. They are also difficult to communicate in comparison to other measures.

5.1.9 80/20 share ratio

The 80/20 ratio suffers from the same conceptual problem as the Gini coefficient in that it requires a construct along the lines of “the bottom 20% have only 5% of the happiness, while the top 20% have 40% of the happiness” (giving a 80/20 ratio of 8). This measure of wellbeing inequality is therefore theoretically incoherent.

5.1.10 80/20 difference

This measure would compute the mean wellbeing for the 20% of the population with the lowest wellbeing and the mean for the 20% with the highest wellbeing, and then take the difference. An alternative method would be to take the 20th and 80th percentile and calculate the difference between these two, but given the discrete nature of the wellbeing scale, this will always be a whole number and many different distributions would therefore generate the same value for this measure if calculated in that way.

This is strictly a relative measure – it is conceptually feasible for two distributions with the same 80/20 difference to have very different means – and is, therefore, associated with the dispersion aversion proposition.

The 80/20 difference is relatively easy to construct and analyse, and can be explained easily in concrete terms. For those interested in understanding dispersion, this indicator may be useful.

5.2 ALTERNATIVES TO MEASURES OF DISPERSION

There is a significant gap in the literature exploring alternatives to dispersion measures for wellbeing inequalities. The selection below have emerged during discussions and interviews, but there may be others which have not yet been thought of.

5.2.1 Average of the bottom 40%

This measure takes the 40% of the population with the lowest wellbeing scores and computes the mean for that group. This is not a measure of dispersion – the same value for this measure could result from distributions in which the top 60% of society have

very different levels of wellbeing. Equally, two distributions with the same level of dispersion but very different means would generate different values for this measure.

One issue that has been raised for this measure is that the discrete nature of the wellbeing scale means that the 40% threshold will almost always fall in the middle of a wellbeing score (e.g. 6). This would mean that the measure arbitrarily sorts some people with the same wellbeing score into the bottom 40% category and not others. This would be problematic if it were being used for individual-level analysis that seeks to explain or predict status in that category or use that status to explain other outcomes. However, it should not present a problem for aggregated population-level explanatory analyses, which is the intended use of inequality measures – after all, a single person cannot have an ‘inequality’ score. In the latter case individuals are sorted into the bottom 40% only for the construction of the indicator’s value – thereafter, their status in or out of that 40% is irrelevant. In this way, this measure differs from the common poverty measure of having income that is 60% below the median.

The only existing example of this measure being used is in a recent analysis of wellbeing inequalities at the local authority level in the UK (Abdallah et al., 2017).

The 40% threshold is relatively easy to construct (although not as easy as standard deviation) and to analyse. It is also easy to compute and can be understood in concrete terms.

A major disadvantage of this measure is that the 40% threshold is arbitrary and could be replaced by any other value – this question was raised by a number of interviewees and survey respondents. There could be an argument to make the threshold lower since such a measure would display more variation and may be less associated with the overall mean. Conversely, another respondent suggested using the bottom 50% on the basis that it is more inclusive and easy to conceptualise as the ‘bottom half’.

5.2.2 Percent below a threshold

This measure would mean identifying a threshold, for example, of ‘low wellbeing’ and simply calculating the percentage of people who fall below this threshold. On its own, it corresponds most closely to the ethical proposition concerning ameliorating misery, but when presented alongside mean wellbeing, it reflects a proportional universalist approach. Theoretically, this measure could conceivably fall to zero, and subsequently yield no further information about societal improvement. In practice, this situation is unlikely to ever occur as it is likely there will always be some number of people experiencing low wellbeing due to life events that aren’t amenable to policy. However, it

would still fail to yield any further information if it fell to the minimal level associated with irremovable misery. This may not be considered a deficiency if one strictly adheres to the ameliorating misery proposition – if the measure falls to zero then the problem is indeed solved.

One advantage of this measure is that it does not require an assumption of cardinality in the response scale, unlike most other measures. It is easy to calculate and of all the measures, it is possibly the easiest to communicate and conceptualise.

5.2.3 Percent outside modus

This measure computes the proportion of respondents that do not report the mode (most common) value.

One concern with this measure is that it doesn't necessarily distinguish between different severities of inequalities. For example, the distribution (1, 1, 2, 2, 8, 8, 8) has the same value for this statistic as the distribution (6, 6, 7, 7, 8, 8, 8). It is also insensitive to the absolute level of the mode. As with the percent below a threshold measure, this measure does not rely on an assumption of cardinality.

5.2.4 Entire distribution

One interviewee advocated reporting the entire wellbeing distribution in graphical form as an indicator, on the basis that this would be the most informative.

This is certainly useful information that should be published where feasible; however, it does not represent a measure that can be used as a single input for inferential statistics and can be quite difficult to interpret when asked to compare two distributions (for example between areas or over time). This indicator was therefore not included in the final list for consideration.

5.2.5 Subjectively-weighted average

It is accepted practice in government policy appraisal to apply different weights to impacts depending on whether they affect people who are higher or lower on the income spectrum. Intuitively, given that many people express more concern about individuals languishing at the bottom of the wellbeing scale, a similar method could be applied to wellbeing.

However, the rationale is somewhat distinct. In the case of income the reasoning for differential weights stems directly from the theory (and evidence) that marginal utility of

income decreases as income increases. As such, weighting is just an adjustment to bring monetary measures more in line with a utility measure. Wellbeing, on the other hand, already corresponds to a form of utility. Weighting in this case does not imply that notches on the wellbeing scale are intrinsically of unequal values (e.g. the difference between 1 and 2 being greater than the difference between 9 and 10), but rather that society as a whole would choose to prioritise particular people over others on the wellbeing spectrum.

Whereas the weights for adjusting income can be implicitly derived from existing evidence (HM Treasury, 2011) the weights for adjusting wellbeing would have to be established based on consultation with the public at large – a potentially resource intensive process.

This idea is similar to Veenhoven and Kalmijn's proposed approach (discussed above) of combining information on the mean and standard deviation in one measure, except that their method assigns an arbitrary level of dispersion aversion.

This kind of approach probably most closely approximates the proportionate universalism position as a standalone indicator, and so best meets our first criteria. However, it may be more difficult to construct for non-statisticians and it is difficult to communicate. Although the final number can be expressed in concrete terms (i.e. as a point on the wellbeing scale) its calculation is relatively opaque.

5.3 DATA ANALYSIS

Three of our criteria required further analysis to assess; variation, predictive power and correlation with the mean. We used two existing datasets to compare some of the main measures. We focused on four measures that were the most promising based on other criteria: standard deviation, mean pair distance, 80:20 difference and average of the bottom 40%.¹

We tested these measures in three ways:

- Which measures exhibit the most variation over time and between places?
- Which measures are most correlated with mean wellbeing?
- Which measures have greatest power in predicting other societal outcomes?

The datasets we used were the European Social Survey and the Annual Population Survey (UK), which provide both a between-country and within-country test.

Appendix 2 provides more details on these tests and the results.

For local authority data the measure that shows the most variation is the 80:20 difference while for country-level data the average of the bottom 40% (and average of bottom 20%) shows the most variation. Standard deviation shows the least variation in all cases.

The level of correlation between wellbeing inequality measures and mean wellbeing is relatively high, but highest for the average of the bottom 40%, while the others show fairly similar levels of correlation.

In terms of predictive power there were only very minor differences between the measures in how well they were able to predict other societal outcomes. This is unsurprising due to the high degree of correlation between the measures. It was therefore not included in the matrix below.

¹ This analysis was undertaken in parallel with other work, therefore indicators to assess were chosen at the outset, before a clear short-list had emerged.

6 RECOMMENDATION

Our primary objective was to recommend a single measure to be reported alongside mean wellbeing:

In addressing our first question of what should be measured, our research suggested that weighted universalism was more widely shared amongst those that we engaged with, and is theoretically coherent. Weighted universalism is not about dispersion per se, but about giving priority to improving the lives of the worst off.

For the purposes of recommending a single wellbeing inequality indicator to be reported by the ONS we therefore eliminated measures of un-weighted variance.

This is not to say that measures of variance are not useful for other purposes. For example, a number of interviewees suggested that inequalities in wellbeing may be a good predictor of future social unrest. This is a legitimate purpose for which measures of dispersion may be used, but is different to a headline statistic which should act as a focus for policy makers to improve.

This elimination left:

- The average of the bottom 40%
- The percentage below a threshold
- Subjectively-weighted average

Of these measures, the percent below a threshold was assessed to best meet our criteria.

The average of the bottom 40% and the subjectively weighted average are both meaningful and robust indicators that can provide useful information. Both indicators are sensitive to change across the bottom of the distribution, compared to a cruder cut-off that has to be chosen for a threshold.

However, this complexity comes at a cost. Both the subjectively weighted average and the average of the bottom 40% are conceptually somewhat difficult to explain and understand. One stakeholder suggested that an indicator should pass the 'Today Programme' test: it should be easy and simple enough to explain in a sentence on Radio 4's Today Programme. It was felt that only the percentage below a threshold passed this test and that, when presented alongside the mean wellbeing, it provides a good headline of changes in wellbeing across the distribution.

If a threshold measure is adopted this raises the question of what threshold should be used. This is a question that we were not able to fully explore within this project.

The ONS currently define 'low' wellbeing as those responding 0-4. It was suggested that if it was lifted to 5 or 6 then this might allow a more prevention based approach to the improvement of well-being.

Another consideration for where to set the threshold is sample size. If a low threshold was used – e.g. 4, a number of local authorities may not have adequate data to be able to report. Given the importance of local wellbeing inequalities, this would be a serious disadvantage. More work is needed to choose a threshold that can be reported across the UK.

Finally, the question of communication was repeatedly raised during our conversations. If public and political priorities in relation to wellbeing inequality area in fact focussed on bringing up the wellbeing of the worst off, is 'wellbeing inequality' the right terminology, or do we need to find alternative language to communicate this concept?

	Reflective of weighted universalism	Communicated easily	Can be understood concretely	Robust to cardinal/ordinal	Robust to bounded scale	Easy to construct	Easy to use in analysis	Variation	Correlation with mean
Standard deviation	Poor	Good	Good	Poor	Poor	Good	Good	Lower	Similar
Mean Pair Distance	Poor	Good	Good	Poor	Poor	Good	Poor	Lower	Similar
Subjectively-weighted average	Good	Poor	Medium	Poor	Medium†	Poor	Good	N/A	N/A
Average of the bottom 40%	Good	Good	Good	Poor	Medium††	Medium	Medium	Higher	Highest
80:20 difference	Poor	Good	Good	Poor	Poor	Medium	Poor	Higher	Similar
Variance	Poor	Good	Medium	Poor	Poor	Medium	Poor	N/A	N/A
% maximum standard deviation	Poor	Poor	Poor	Poor	Poor	Poor	Poor	N/A	N/A
Inequality-adjusted mean	Poor	Poor	Poor	Poor	Poor	Poor	Poor	N/A	N/A
Gini coefficient	Poor	Poor	Poor	Poor	Poor	Poor	Poor	N/A	N/A
Coefficient of variation	Poor	Poor	Poor	Poor	Poor	Medium	Poor	N/A	N/A
S-domination	Poor	Poor	Poor	Poor	Poor	Poor	Poor	N/A	N/A
Percent below threshold	Medium	Good	Good	Good	Good	Medium	Good	N/A	N/A
Percent outside modus	Poor	Good	Good	Good	Good	Medium	Good	N/A	N/A

Table 1 Overview of performance of measures against criteria

† assuming small weights are given to the top of the distribution

†† Robust to the top of the distribution but not to the botto

7 FURTHER RESEARCH

The measurement of wellbeing inequality has been woefully under-researched. In this paper we have mapped the issues that should be considered in choosing a measure of wellbeing inequality, and chosen a recommendation based on the research that we have found or undertaken.

However, our approach has been broad rather than deep, and a number of areas require further research. These include:

- Exploring normative priorities with a broader sample of political and policy stakeholders
- Examining non-linearities in the co-benefits of wellbeing
- Exploring alternative thresholds. This could include further engagement on peoples' priorities as well as some technical work to assess how different cut-offs perform in terms of variability, predictive power and correlation with the mean
- Explore the extent to which extreme response bias may be operating between UK populations
- Test the robustness of respondents reporting the top category of wellbeing

8 REFERENCES

- Abdallah, S. (2016). What wellbeing inequalities tell us about the EU Referendum result.
- Abdallah, S., & Quick, A. (2016). Inequalities in wellbeing. Background briefings and engagement from the Making Wellbeing Count for Policy project (p. 11). New Economics Foundation. Retrieved from http://b.3cdn.net/nefoundation/7831bc5dbdb172539e_64m6brzrn.pdf
- Abdallah, S., Wheatley, H., & Quick, A. (2017). *Measuring wellbeing inequality in Britain*.
- APPG for Wellbeing Economics. (2016). Roundtable on wellbeing, inequality and social deprivation: Discussion summary (p. 6). London: New Economics Foundation.
- Becchetti, L., Massari, R., & Naticchioni, P. (2010). Why has happiness inequality increased? Suggestions for promoting social cohesion. *ECINEQ (Society for the Study of Economic Inequality) Working Paper Series*.
- Bellù, L. G., & Liberati, P. (2007). *Policy Impacts on Inequality: Inequality and Axioms for its Measurement*.
- Binder, M., & Coad, A. (2010). Going Beyond Average Joe's Happiness: Using Quantile Regressions to Analyze the Full Subjective Well-Being Distribution. *Papers on Economics and Evolution*.
- Coyne, S. M., McDaniel, B. T., & Stockdale, L. A. (2016). "Do you dare to compare?" Associations between maternal social comparisons on social networking sites and parenting, mental health, and romantic relationship outcomes. *Computers in Human Behavior*, 70, 335–340. <https://doi.org/10.1016/j.chb.2016.12.081>
- Daggitt, M. L., Noulas, A., Shaw, B., & Mascolo, C. (2016). Tracking urban activity growth globally with big location data. *R.Soc. OpenSci.* <https://doi.org/10.1098/rsos.150688>
- Delhey, J., & Kohler, U. (2011). Is happiness inequality immune to income inequality? New evidence through instrument-effect-corrected standard deviations. *Social Science Research*, 40(3), 742–756.

- <https://doi.org/10.1016/j.ssresearch.2010.12.004>
- Diener, E., & Chan, M. Y. (2011). Happy People Live Longer: Subjective Well-Being Contributes to Health and Longevity. *Applied Psychology: Health and Well-Being*, 3(1), 1–43. <https://doi.org/10.1111/j.1758-0854.2010.01045.x>
- Dutta, I., & Foster, J. (2013). Inequality of happiness in the U.S.: 1972-2010. *Review of Income and Wealth*, 59(3), 393–415. <https://doi.org/10.1111/j.1475-4991.2012.00527.x>
- Eurostat. (2012). *Analysis, implementation and dissemination of well-being indicators*. London. <https://doi.org/10.13140/RG.2.2.28094.69440>
- Fahey, T., & Smyth, E. (2004). Do subjective indicators measure welfare? Evidence from 33 European societies. *European Societies*, EVS(1), 5–27. <https://doi.org/10.1080/1461669032000176297>
- Fowler, J. H., & Christakis, N. A. (2008). Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. *Bmj*, 337(dec04 2), a2338–a2338. <https://doi.org/10.1136/bmj.a2338>
- Gandelman, N., & Porzecanski, R. (2013). Happiness Inequality: How Much is Reasonable? *Social Indicators Research*, 110(1), 257–269. <https://doi.org/10.1007/s11205-011-9929-z>
- Goff, L., Helliwell, J. F., & Mayraz, G. (2016). The Welfare Costs of Well-being Inequality. *NBER Working Paper Series*, 27. <https://doi.org/10.1017/CBO9781107415324.004>
- Graham, C., Eggers, A., & Sukhtankar, S. (2004). Does happiness pay? An exploration based on panel data from Russia. *Journal of Economic Behavior and Organization*, 55(3), 319–342. <https://doi.org/10.1016/j.jebo.2003.09.002>
- Harrison, E., Quick, A., & Abdallah, S. (2016). Looking through the wellbeing kaleidoscope.
- HM Treasury. (2011). *The Green Book: Appraisal and Evaluation in Central Government*. <https://doi.org/http://greenbook.treasury.gov.uk/index.htm>
- Jeffrey, K., & Michaelson, J. (2015). *Five headline indicators of national success: A clearer picture of how the UK is performing*. London. Retrieved from

<http://www.ncbi.nlm.nih.gov/pubmed/15003161%5Cnhttp://cid.oxfordjournals.org/lookup/doi/10.1093/cid/cir991%5Cnhttp://www.scielo.cl/pdf/udecada/v15n26/art06.pdf%5Cnhttp://www.scopus.com/inward/record.url?eid=2-s2.0-84861150233&partnerID=tZOTx3y1>

Kalmijn, W., & Veenhoven, R. (2005). Measuring inequality of happiness in nations: In search for proper statistics. *Journal of Happiness Studies*, 6(4), 357–396. <https://doi.org/10.1007/s10902-005-8855-7>

Kalmijn, W., & Veenhoven, R. (2014). Index of Inequality-Adjusted Happiness (IAH) Improved: A Research Note. *Journal of Happiness Studies*, 15(6), 1259–1265. <https://doi.org/10.1007/s10902-013-9474-3>

Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? *Psychological Bulletin*, 131(6), 803–855. <https://doi.org/10.1037/0033-2909.131.6.803>

Marmot, M., Allen, J., Goldblatt, P., Boyce, T., McNeish, D., Grady, M., & Geddes, I. (2010). *Fair society, healthy lives (The Marmot Review): Strategic Review of Health Inequalities in English post-2010. The Marmot Review*. <https://doi.org/10.1016/j.puhe.2012.05.014>

Ott, J. (2005). Level and inequality of happiness in nations: Does greater happiness of a greater number imply greater inequality in happiness? *Journal of Happiness Studies*, 6(4), 397–420. <https://doi.org/10.1007/s10902-005-8856-6>

Ovaska, T., & Takashima, R. (2010). Does a Rising Tide Lift All the Boats? Explaining the National Inequality of Happiness. *Journal of Economic Issues*, 44(1), 205–224. <https://doi.org/10.2753/JEI0021-3624440110>

Parfit, D. (1997). Equality or Priority? *Ratio*, 10(3), 1467–9329.

Pickett, K., & Wilkinson, R. (2010). *The Spirit Level: Why Equality is Better for Everyone*. London: Penguin.

Quick, A. (2015). Inequalities in wellbeing: Challenges and opportunities for research and policy, 44(1055254).

Quick, A., & Abdallah, S. (n.d.). *Why personal wellbeing is good for the nation's health*.

- Smith, K. P., & Christakis, N. A. (2008). Social Networks and Health. *Annual Review of Sociology*, 34(1), 405–429.
<https://doi.org/10.1146/annurev.soc.34.040507.134601>
- Stevenson, B., & Wolfers, J. (2008). Happiness inequality in the United States. *NBER Working Papers*, 37(June 2008).
- Veenhoven, R. (2005). Return of inequality in modern society? Test by dispersion of life-satisfaction across time and nations. *Journal of Happiness Studies*, 6(4), 457–487. <https://doi.org/10.1007/s10902-005-8858-4>
- Veenhoven, R. (2012). The medicine is worse than the disease: Comment on Delhey and Kohler's proposal to measure inequality in happiness using "instrument-effect-corrected" standard deviations. *Social Science Research*, 41(1), 203–205. <https://doi.org/10.1016/j.ssresearch.2011.11.009>
- Veenhoven, R., & Kalmijn, W. (2005). *Inequality-adjusted happiness in nations
egalitarianism and utilitarianism married in a new index of societal performance*. *Journal of Happiness Studies* (Vol. 6). <https://doi.org/10.1007/s10902-005-8857-5>
- Whitby, A. (2011). *The BRAINPOoL Project: Beyond GDP – From Measurement to Politics and Policy (Final Report)*. Brainpool (project website). London. Retrieved from <http://www.brainpoolproject.eu/about-2/background/>

APPENDIX 1: SURVEY RESPONSES

The first questions asked respondents the extent to which they agreed or disagreed with a number of statements.

Question 1 (110 responses)

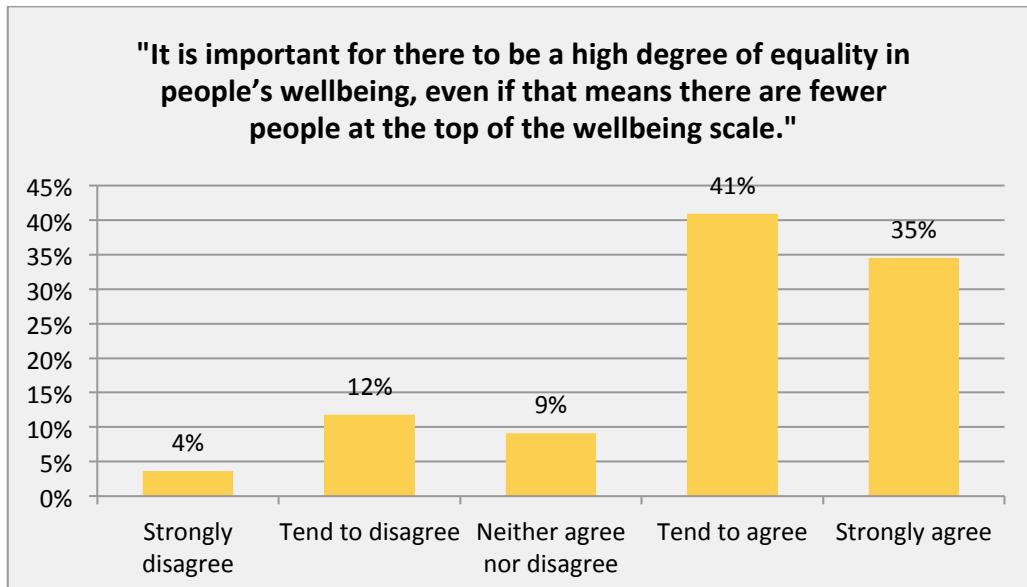


Figure 10 Survey response, Q.1

Question 2 (110 responses)

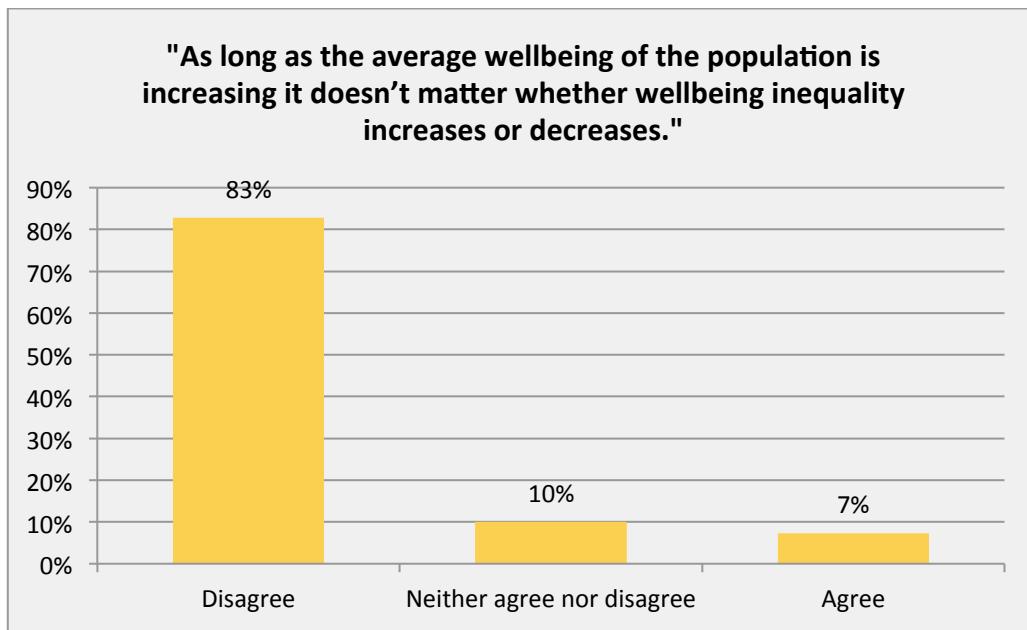


Figure 11 Survey response, Q.2

"Strongly (dis)agree" and "tend to (dis)agree" have been combined to protect anonymity.

Question 3 (110 responses)

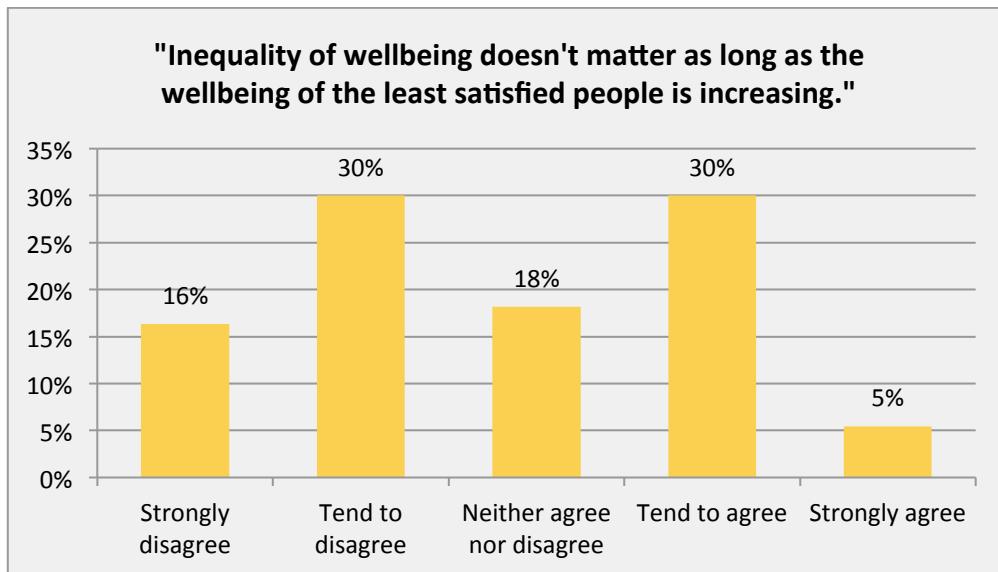


Figure 12 Survey response, Q.3

The next set of questions asked respondents to state their preference in a number of scenarios.

Question 4 (106 responses)

	Your wellbeing	Average wellbeing
Scenario A	6	9
Scenario B	6	7

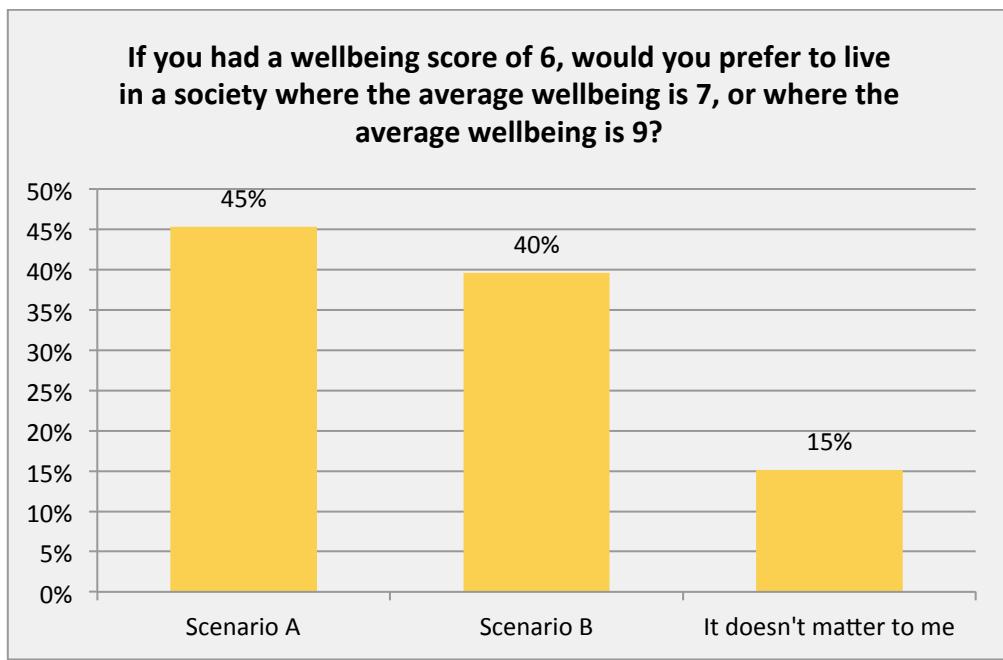


Figure 13 Survey response, Q.4

Question 5 (49 responses)

Question 5 gave respondents the opportunity to explain their answer to Question 4. The table below shows tallies of the answers grouped by theme, as well as the corresponding answers to Question 4.

Theme code	Theme	Q4 answer
1	There's more potential for my wellbeing to grow	Scenario A
		Scenario B
2	Greater overall wellbeing / higher average	Scenario A
3	I'd like to feel like I'm in the same boat as others (& not feeling so would negatively impact my wellbeing)	Scenario B
4	I'd rather be just below average than 3 points below	Scenario B
5	I would like others to be happy even if I wasn't	Scenario A
6	My wellbeing is identical in either case	It doesn't matter to me
7	Inequality	Scenario B

8	I'm assuming my score of 6 takes into account my perception of others wellbeing	It doesn't matter to me
---	---	-------------------------

Table 2 Survey responses Q.5 grouped by theme and shown in order of frequency

Question 6 (100 responses)

Respondents were asked to consider two distributions, A and B.

In Distribution A, most people score somewhere in the middle of the wellbeing scale, with no one experiencing either the very lowest or the very highest levels of wellbeing. The average life satisfaction of A is 5.

In Distribution B, people's scores are more spread out, with most people somewhere in the middle but many more people both at the high and low ends of the scale. The average life satisfaction of B is also 5.

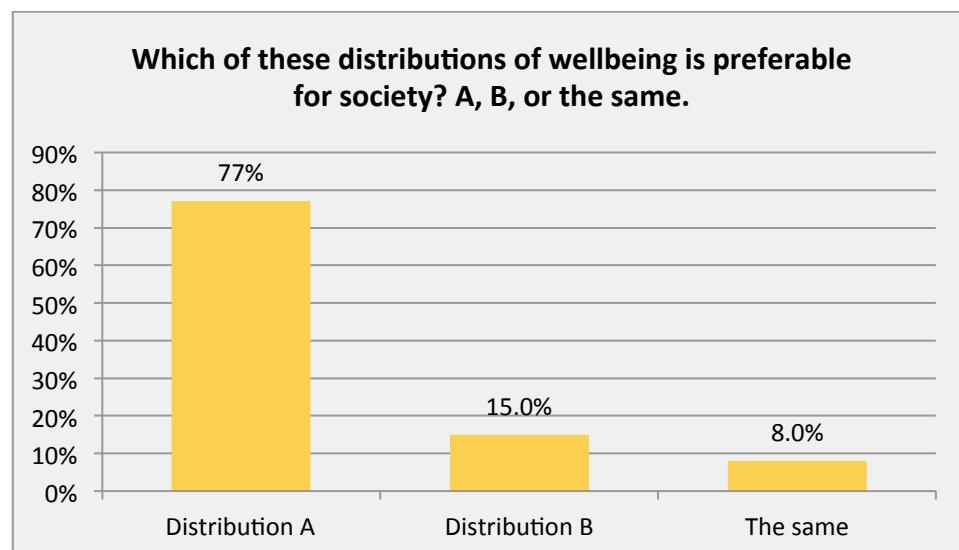


Figure 14 Survey responses Q.6

Question 7 (77 responses)

What is the most important reason why A is better than B?

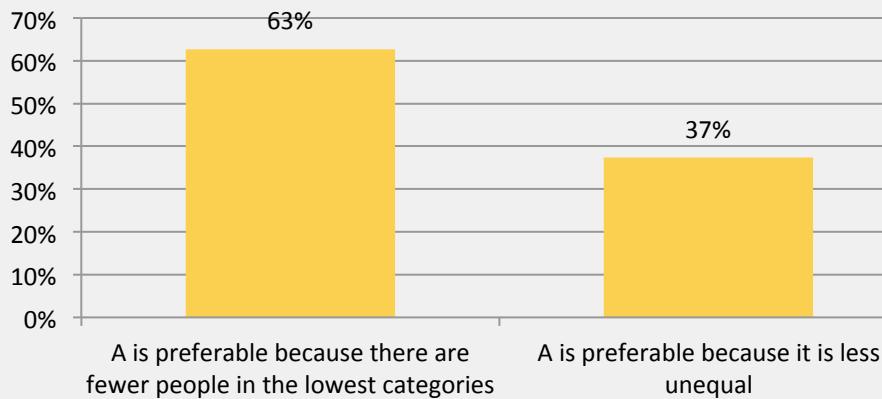


Figure 15 Survey responses, Q.7

Reasons for responding 'Other' included

- "Extremes produce unhappiness, alienation and discontent"
- "There are more people clustered around one category"

Question 8 (14 responses)

What is the most important reason why B is better than A?

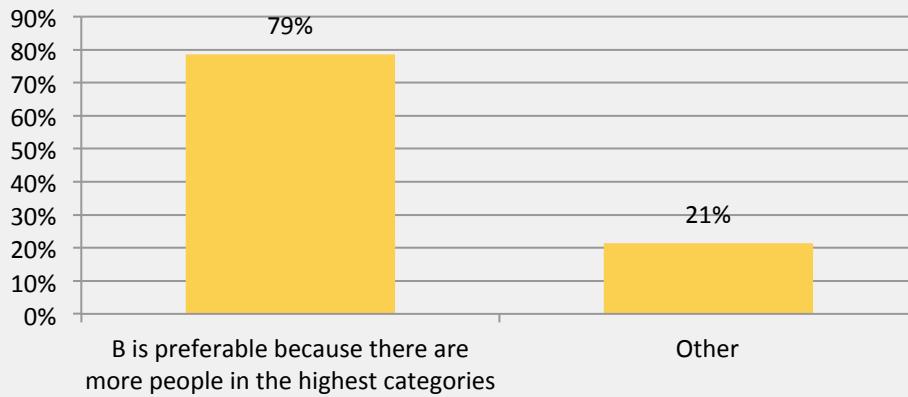


Figure 16 Survey responses Q.8

3 respondents answered 'Other' and gave the following reasons

- "Better spread"
- "The differences are less stark and whilst there are people with extremely low wellbeing it appears that there is potential here to move to higher levels of wellbeing."
- "B is preferable because the average distribution is higher"

Question 9

Question 9 asked respondents to consider two distributions, C and D.

In Distribution C, people are quite spread out, with some people having very low levels of wellbeing, but most people in the upper half of the scale. The average is 7.

In Distribution D, people are very concentrated around one part of the scale, with fewer people at the high and low parts of the scale. The average is 6.

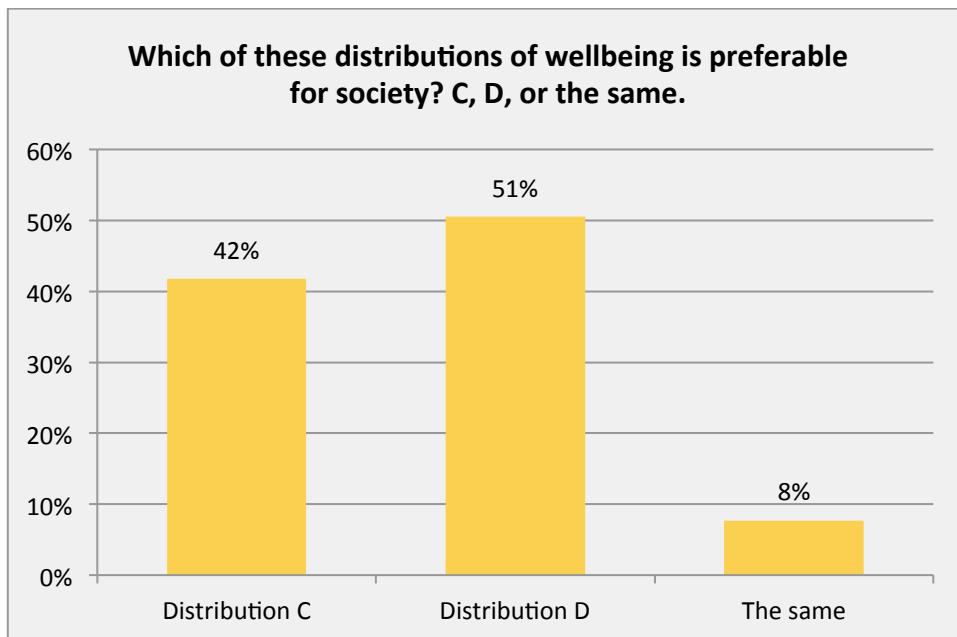


Figure 17 Survey responses Q.9

Question 10 (43 responses)

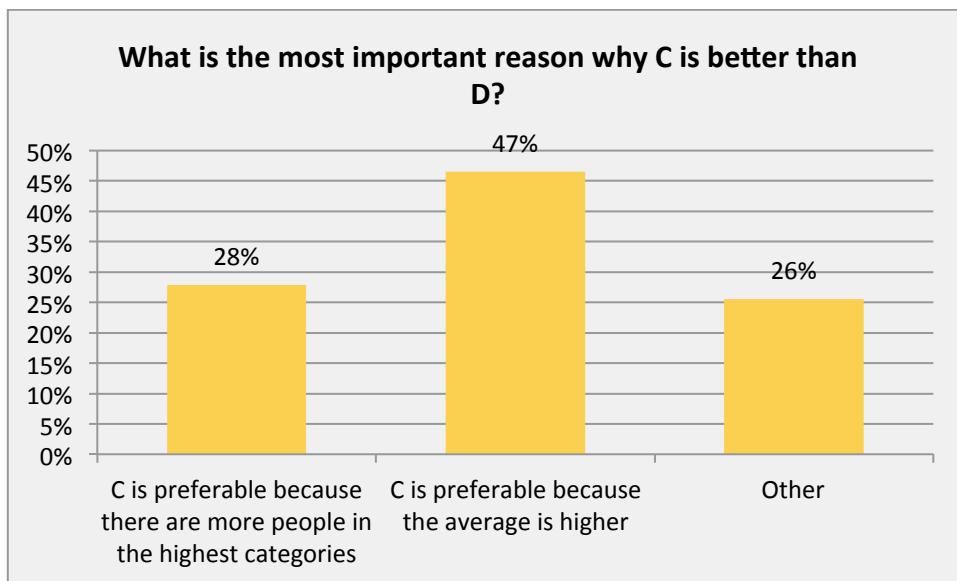


Figure 18 Survey responses Q.10

11 respondents answered 'Other', and gave the following reasons:

- *Both of these are very important reasons*
- *both*
- *Relatively few people in lower categories in distribution C - indeed, far fewer than in the higher categories. While extra weight should be given to those lower down the scale, given the relative numbers in the low vs upper categories in this case, it still was not enough to outweigh the benefits at the top end.*
- *C is preferable because the average is higher and a higher proportion have a wellbeing score over 7 than under.*
- *C is also more realistic*
- *No acceptable to have people at the very bottom so low totally unacceptable in 21st civilised society*
- *Few people very unhappy, so the inequality is acceptable*
- *It would be easier to move these people into the higher scores as they are already nearer to the higher end.*
- *Well, D is ridiculous: it also suggests that there is a dampening effect on well-being in this society that's reducing the number of people with congenitally high well-being scores*
- *I found myself counting what percentage of people were 5 or less in the two scenarios, or 6 or less.*
- *Some people get a chance at being in the highest categories*

Question 11 (46 responses)

What is the most important reason why D is better than C?

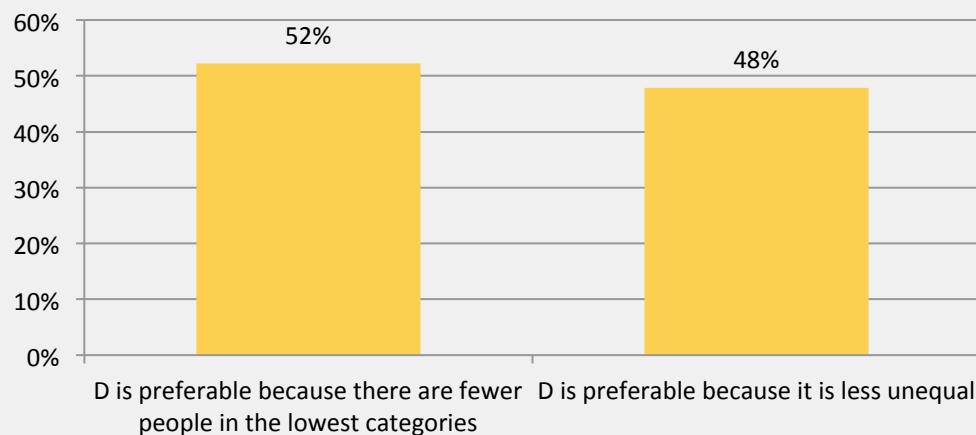


Figure 19 Survey responses Q.11

Question 12 (89 responses)

Consider the following distribution of people at different levels of wellbeing (0 – 10). The government can choose one of two policy options that cost the same amount:

Policy A will increase the wellbeing of one person from 1 to 2

Policy B will increase the wellbeing of one person from 5 to 6 and another from 6 to 7

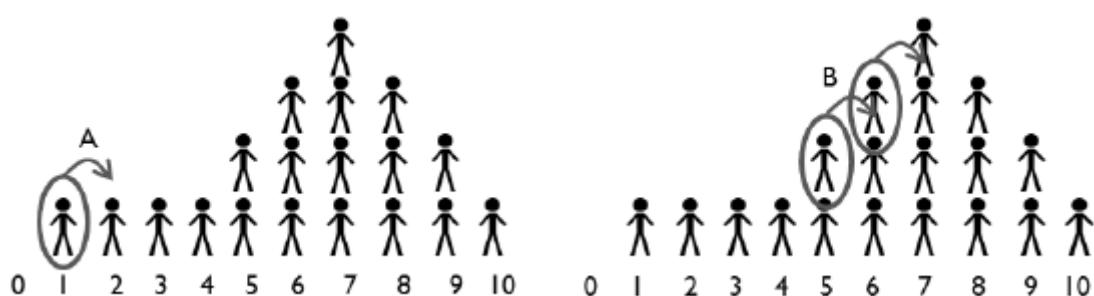


Figure 20 Illustration of two policy scenarios, A and B

Should the government choose policy A or policy B?

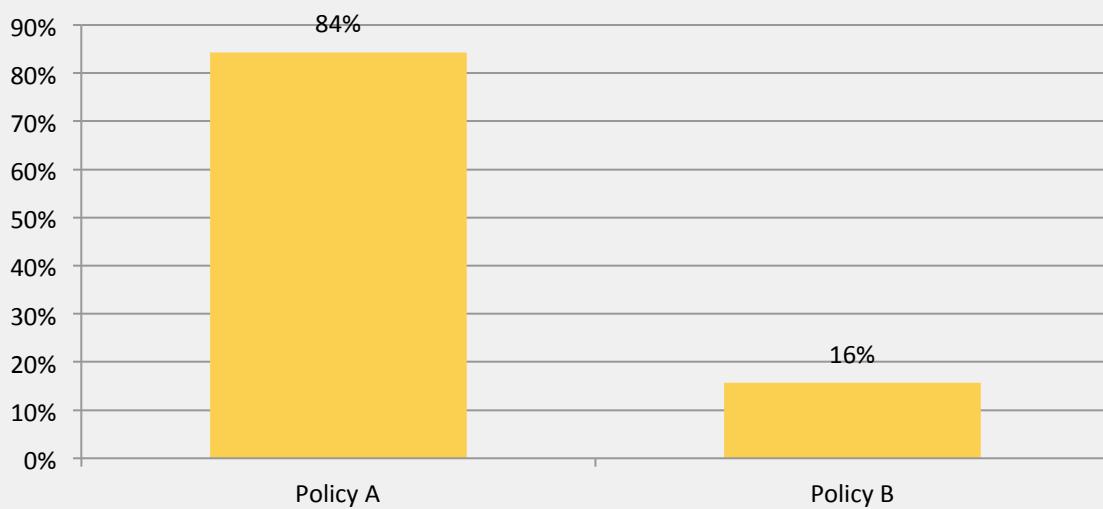


Figure 21 Survey response, Q.12

Question 13 (39 responses)

Respondents were asked to explain their answer to the previous question if they wanted. The table below shows tallies of the answers grouped by theme, as well as the corresponding answers to Question 12.

Theme code	Theme	Q12 answer
1	Addressing the worst off is more important than those who are already doing ok	Policy A
2	More peoples' wellbeing will be improved	Policy A Policy B
3	Equality	Policy A
4	Addressing the worst off will have biggest impact on society	Policy A
5	Reduction of extremes	Policy A

Table 3: Question 13 answers grouped by theme, in order of frequency

NOTES

- 4 of the 14 people who chose Policy B gave a reason in Question 13 (29%), compared to 35 of 74 who chose Policy A (47%)

- One interesting reason for choosing Policy A was “*The largest inequalities for other determinants are probably found amongst those in the lower categories*” – suggesting the respondent considers wellbeing inequality less important than other inequalities

Question 14 (77 responses)

This optional question asked for respondents' opinions on the three options for measuring how wellbeing is distributed across the population.

Standard deviation is a measure that calculates how far a random person is from the average level of wellbeing. The higher the standard deviation the more spread out the population is along the wellbeing scale.

The average of the bottom 40% is a measure that calculates the average wellbeing score for the 40% of people that have the lowest wellbeing. The higher this measure the fewer people there are with very low wellbeing.

80:20 share difference is a measure that calculates the difference between the wellbeing of the bottom 20% and the top 20%. The higher the 80:20 share difference the bigger the disparity in wellbeing between the highest and the lowest.

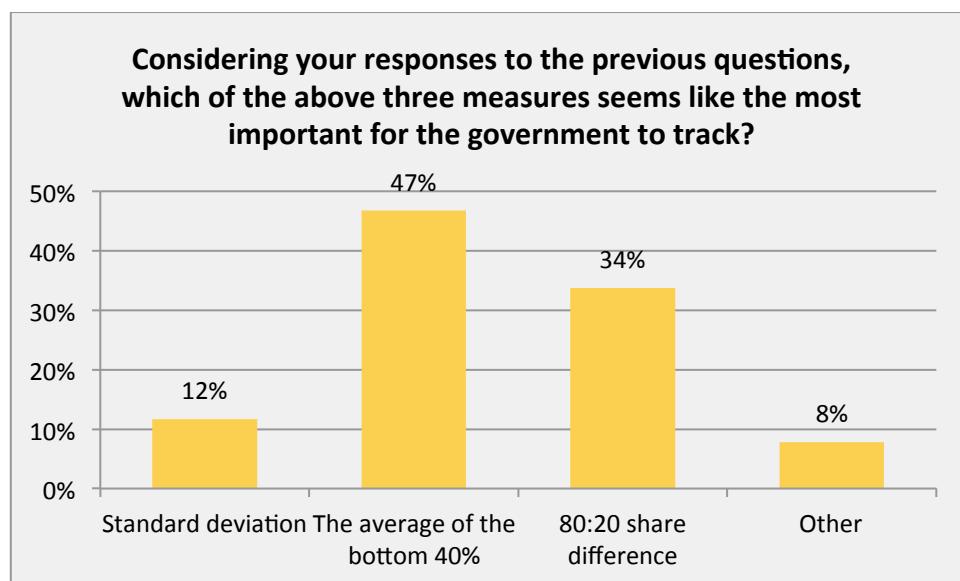


Figure 22 Survey responses Q.14

i. Question 15 (42 responses)

Respondents were asked why they thought their chosen measure in Question 14 was best. Table 3 below shows tallies of the answers grouped by theme, as well as the corresponding answers to Question 14.

Theme code	Theme	Q14 answer	Tally
1	Focusing on those at the bottom of the spectrum is the most important thing	Bottom 40%	15
2 & 3	Inequality = looking at the gap between the top and bottom 80:20 captures both absolute and also level of inequality	80:20 share difference	6
N/A			18

Table 4: Question 15 answers grouped by theme

Table 4 below lists pros and cons for each of the measures from respondents' comments in Question 15.

Measure	Pros	Cons
Standard deviation	<ul style="list-style-type: none"> Tells a fuller story by including all responses Measures relative equality Easy to understand and visualise 	<ul style="list-style-type: none"> Only measures spread
Average of the bottom 40%	<ul style="list-style-type: none"> It is understandable It caters a sufficiently large group to get political support Allows us to focus on those most in need, while also shifting the whole distribution 	<ul style="list-style-type: none"> Does not reflect wellbeing of the whole population Does not reflect overall level of inequality
80:20 share difference	<ul style="list-style-type: none"> Captures both absolute levels of wellbeing and inequality of wellbeing Gives information about relativity Absolute difference is more likely to provoke a political response than the other measures 	<ul style="list-style-type: none"> Ignores those in the middle Does not give enough information about distribution

Table 5: Pros and cons for measures in Question 14

Respondents who answered 'Other' to Question 14 provided the following comments:

Q14 'Other' answer	Q15 comments
The average of the bottom 20%	I find this hard to answer without knowing how well-being will be measured. I think for some elements of the measure current measures need to be compared with previous measure e.g. not just the level of good employment but whether or not that level is increasing or decreasing.
Neither	None. Better to look at tax payer liability shares
standard deviation, but with average + Proportions	None can be considered in isolation of the average and high/ low proportions. Standard deviation adds the most to average and proportions.
Mean pair distance	Just to keep it in the mix :)
See comments	I'm not sure a single measure is the right way to go. It will be too blunt. We should have three objectives. 1 - Improve the wellbeing of the whole population. 2. Reduce the inequality of wellbeing distribution. 3. Improve the wellbeing of the whole population to above 5. This requires three measures.
Track the average of the bottom 40 percent and the 80:20 share difference	

Table 6 'Other' responses to question 14

Question 16 (87 responses)

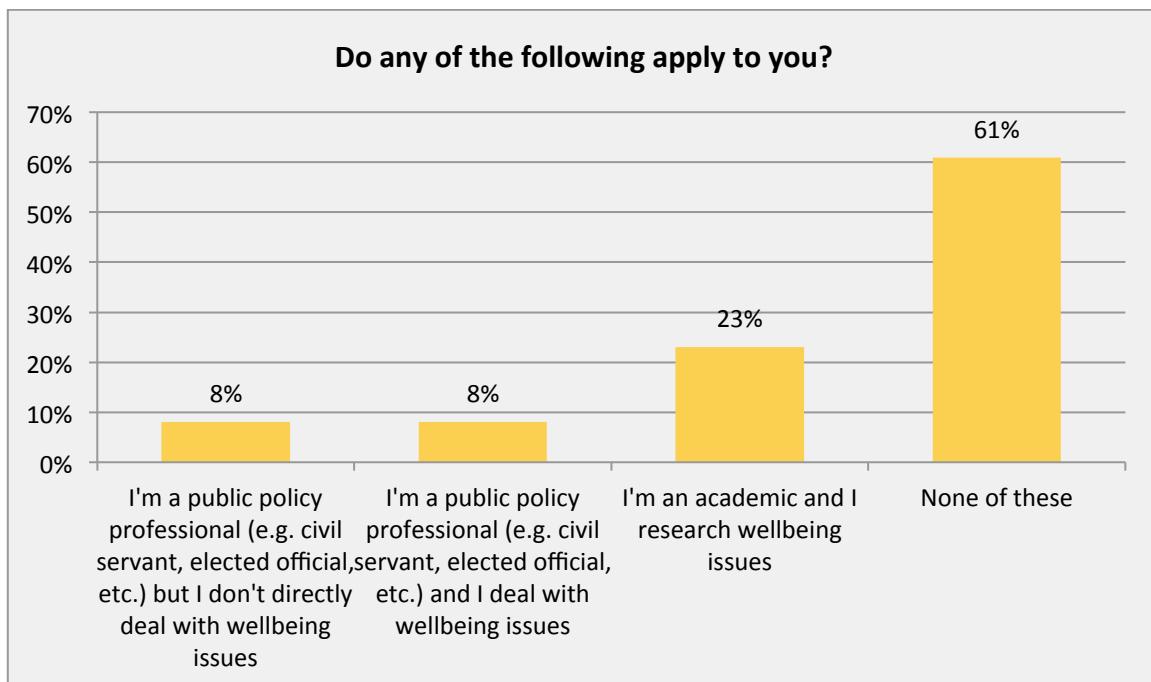


Figure 23 Survey responses Q.15

APPENDIX 2: DATA ANALYSIS

We used two datasets to explore three questions:

1. Which measures exhibit the most variation over time and between places?
2. Which measures are most correlated with mean wellbeing?
3. Which measures have greatest power in predicting other societal outcomes?

The datasets we used were the European Social Survey and the Annual Population Survey (UK).

We used six rounds of data from the European Social Survey (every two years starting in 2002). We merged data to form country-level observations covering 36 countries in an unbalanced panel (i.e. there is not data for each country in every year) with a total of 155 country-level observations.

We also used four years of data from the UK Annual Population Survey covering 2011 – 2014. We merged data to form local authority-level observations for 212 local authorities in an unbalanced panel (we do not have data for 2011 in some areas) with a total of 773 local authority-level observations.

These datasets provide the best opportunity to compute regional level wellbeing inequality measures alongside a range of other economic, social and environmental regional indicators.

1. Which measures exhibit the most variation over time and between places?

An inequality measure (or any indicator for that matter) is only useful if it changes. Less sensitive indicators will respond in a slower or more muted way, which can make it more difficult to distinguish changes that are driven by real effects, rather than just random variation, or noise.

We looked at the variation (as measured by standard deviation) in different measures in our panel datasets, splitting that into variation within regions over time and variation between regions.

For UK local authority-level data the 80:20 difference displays the greatest overall variation, followed by the average of the bottom 40%. The standard deviation shows the least variation. This is true for both life satisfaction and happiness. For all measures the variation between local authorities is greater than the variation over time – this is not terribly surprising since we have only 4 years of data and over 200 local authorities in the sample.

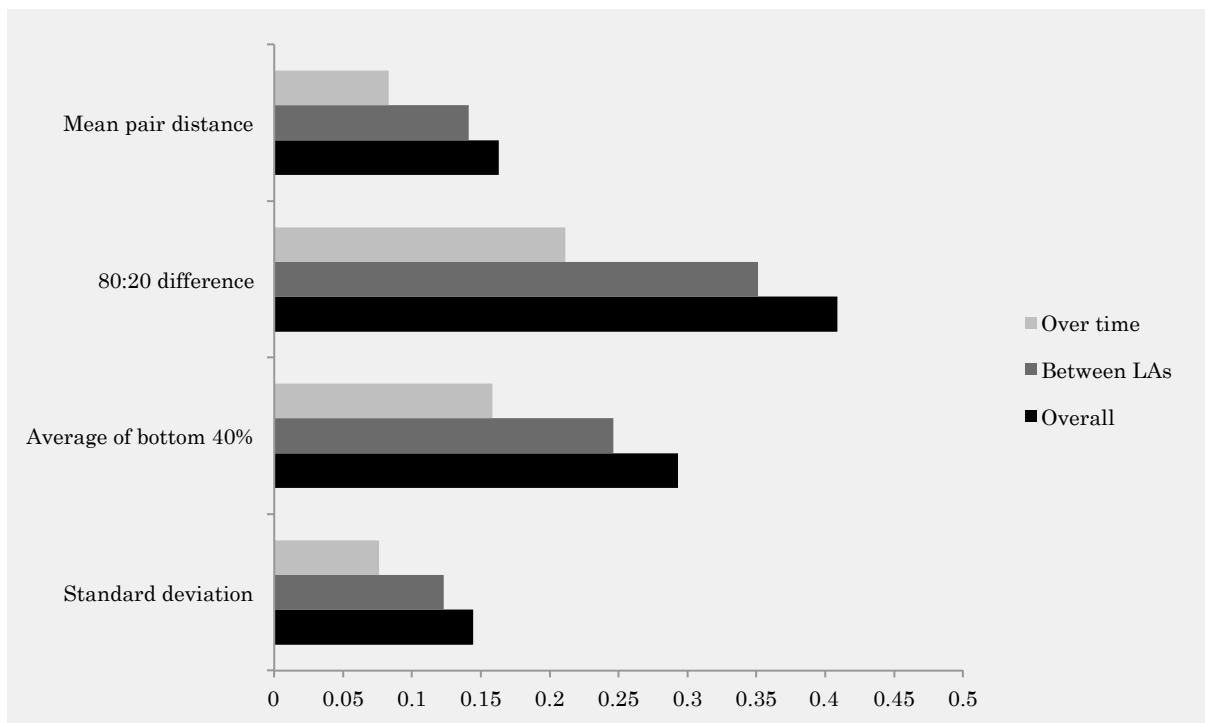


Figure 24. Variation (standard deviation) of different inequality measures for life satisfaction over time and between local authorities

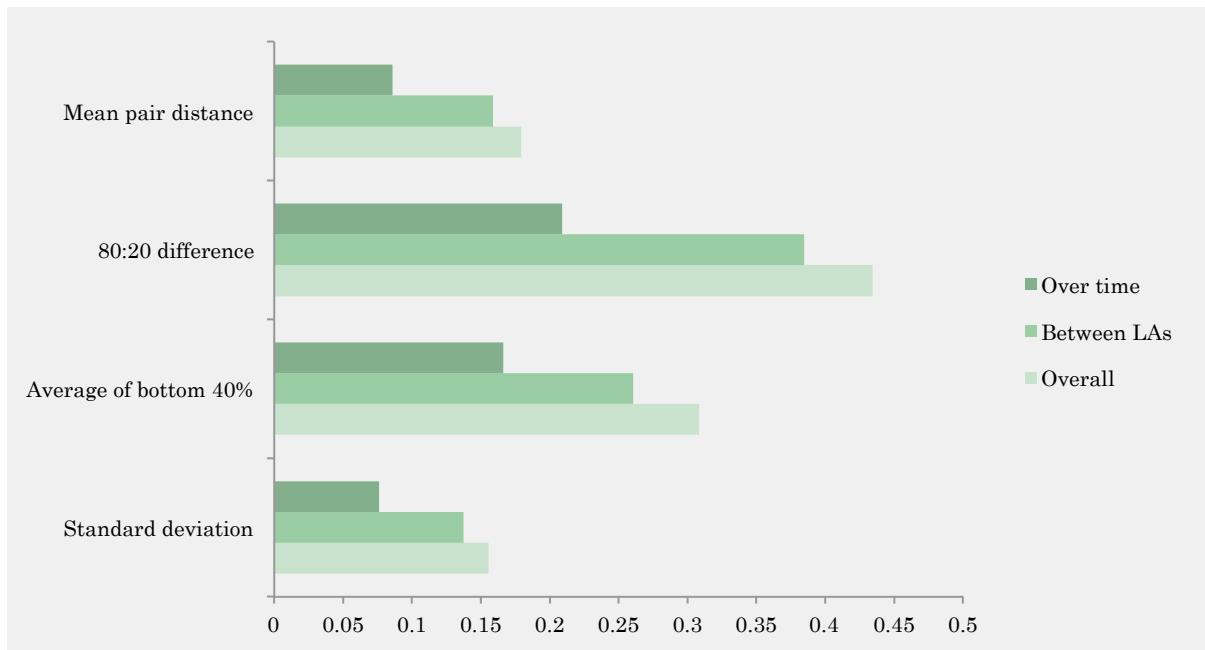


Figure 25. Variation (standard deviation) of different inequality measures for happiness over time and between local authorities

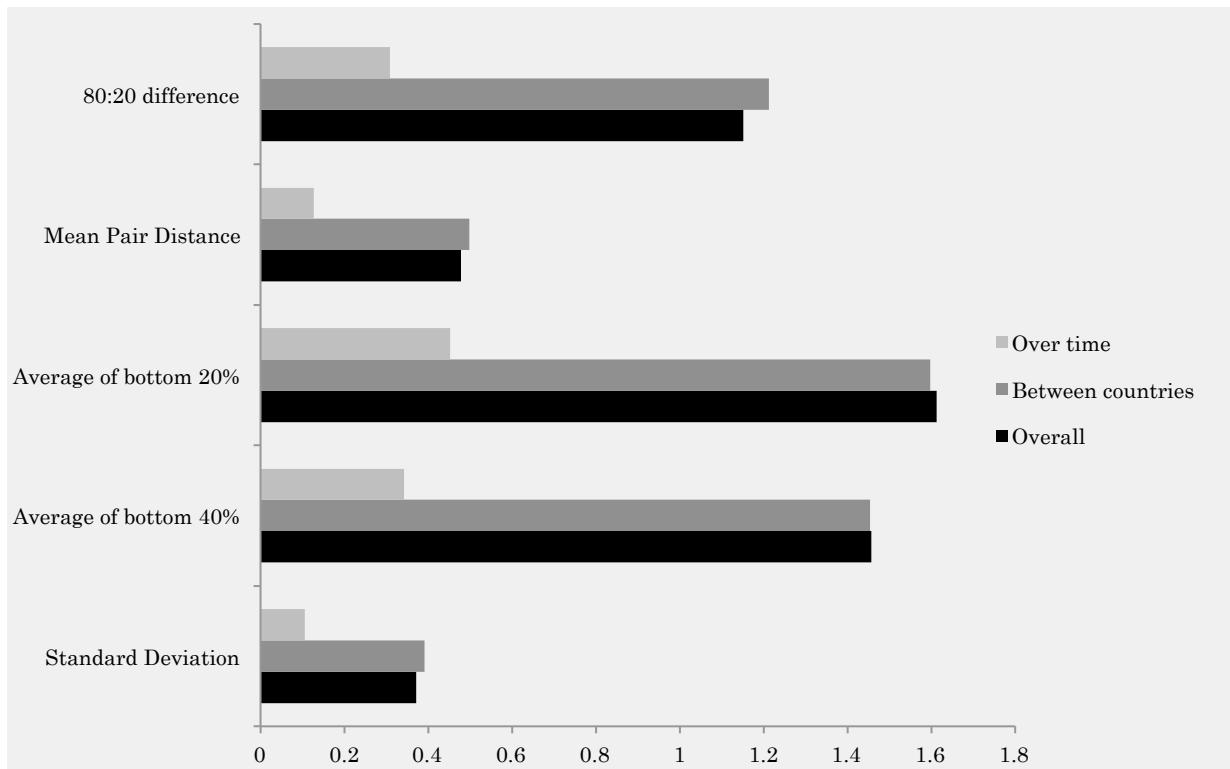


Figure 26. Variation (standard deviation) of different inequality measures for life satisfaction over time and between countries

2. Which measures are most correlated with mean wellbeing?

An inequality measure is most useful if it provides information that is not already provided by some other measure that we currently observe. If an inequality measure were 100% correlated with another variable then it would not be providing additional information that could not already be derived from this other variable. Mean wellbeing is both widely reported and likely to be correlated with most measures of inequality – therefore, we check the degree of correlation between different wellbeing inequality measures and the mean. The higher this correlation the less new information is obtained from the measure over and above the mean.

Figures 27 to 29 show correlations between inequality and the mean for life satisfaction and happiness at the local authority level and life satisfaction at the country level. The level of correlation between wellbeing inequality measures and mean wellbeing is relatively high, ranging from 0.60 to 0.97. The overall level of correlation is greater in the ESS country-level dataset than in the APS local authority-level dataset. The clearest pattern is that the correlation is highest for the average of the bottom 40%, while the others show fairly similar levels of correlation.

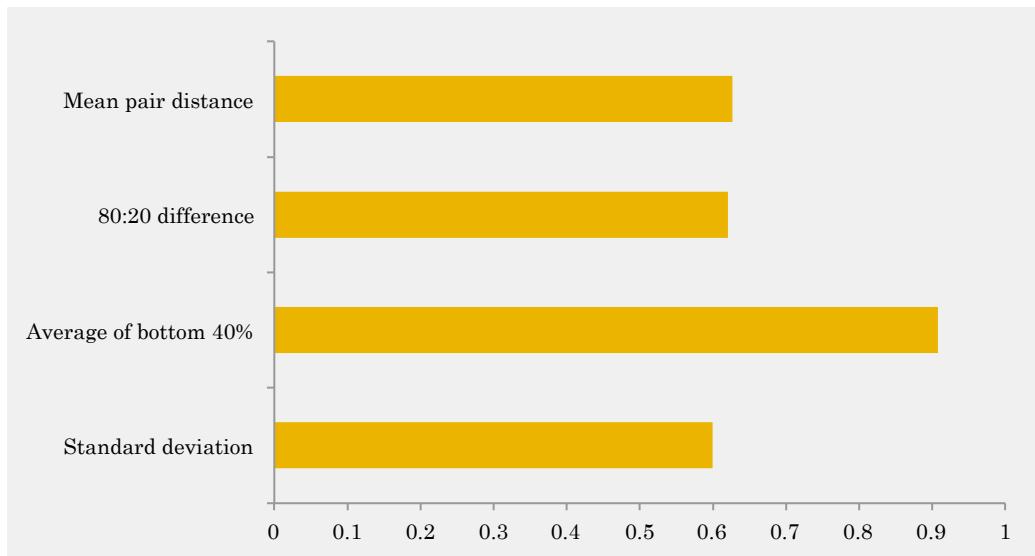


Figure 27. Correlation between mean life satisfaction and different life satisfaction inequality measures at local authority level

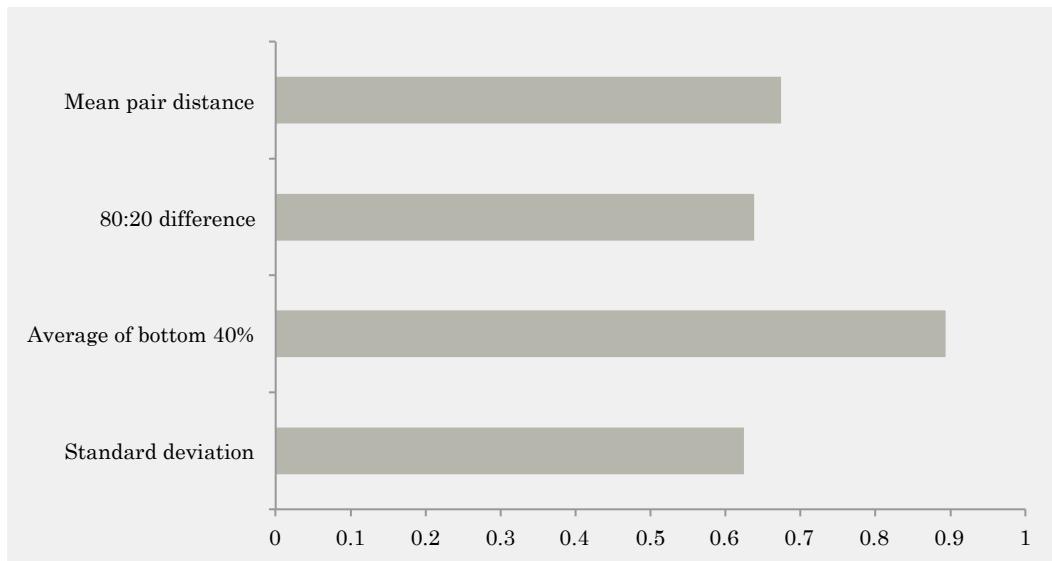


Figure 28. Correlation between mean happiness and different happiness inequality measures at local authority level

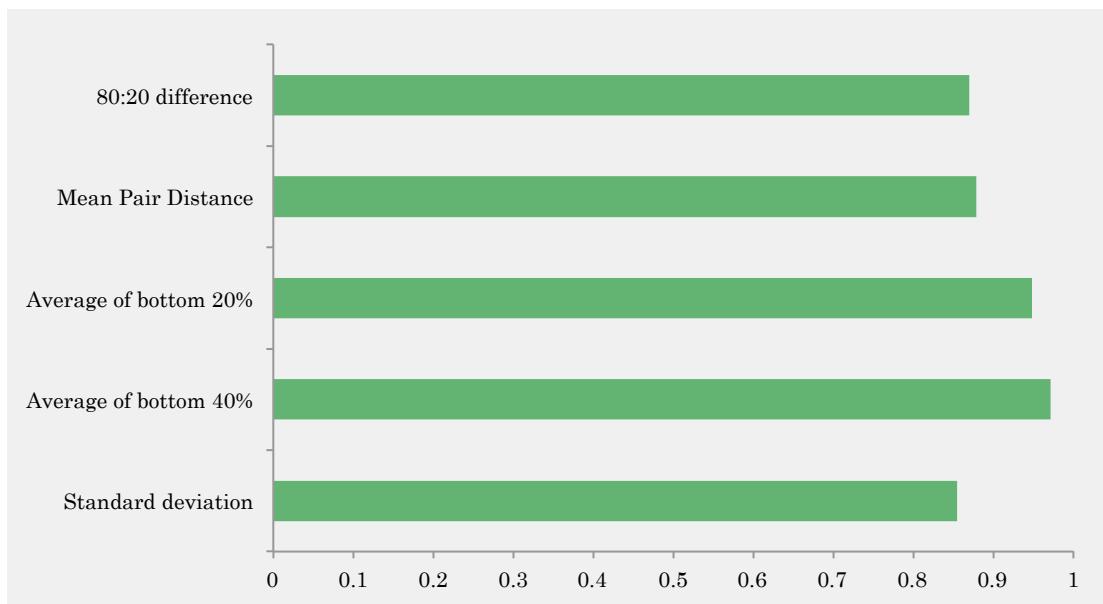


Figure 29. Correlation between mean life satisfaction and different life satisfaction inequality measures at country level

3. Which measures have greatest power in predicting other societal outcomes?

Here we are interested in whether there are differences between the inequality measures in their ability to explain the variance in some relevant dependent variable.

To do so we look at both bivariate regression models and more complex multivariate regression models and compare the coefficient of determination (R^2) when different inequality measures are substituted. Given that we are interested purely in the difference between regressions that are identical except for the inequality measure used, to some extent the exact form of that regression model is unimportant – we are not attempting to construct a rigorous, causal model. Nonetheless, measures could perform differently depending on the nature of the model so we test both a simplistic and a more complex regression model.

The R^2 calculates the proportion of total variance in the dependent variable that is accounted for by variation in the independent variables. Thus, if all of the independent variables are kept the same with the exception of the wellbeing inequality measure then we can get a sense of which measure contributes more to understanding variation in the outcome variable. Since we are comparing models with the same number of estimated parameters it is not necessary to use the adjusted R^2 .

We also look at Aikake's information criterion (AIC) and the Bayesian information criterion (BIC) as an additional measure of the goodness of fit of each model. These measures indicate relative (not absolute) model quality and allow comparison between similar models.

However, model selection is not a well-refined science so this exercise can only be considered indicative of the relative performance of different measures.

The main conclusion from this exercise is that there is relatively little difference between the measures in terms of their predictive power – there are small differences but none can be considered substantially better or worse than the others. There is no strong reason, therefore, to use, for example, standard deviation over other measures on the basis of its statistical performance.

We considered four separate regression models (with bivariate and multivariate versions of each):

- inequality of life satisfaction as a determinant of the proportion voting to leave the European Union in June 2016 at the UK local authority level
- inequality of happiness as a determinant of the proportion voting to leave the European Union in June 2016 at the UK local authority level
- inequality of life satisfaction as a determinant of levels of social trust in Europe at the country level
- inequality of life satisfaction as a determinant of levels of trust in politicians in Europe at the country level

When looking at local authority data in a regression on the proportion of voters that chose to leave the EU, life satisfaction MPD performs best in a bivariate regression, while average life satisfaction of the bottom 40% performs best in a more complex regression. For happiness, average happiness of the bottom 40% also performs best in a more complex regression, while the 80:20 difference performs best in the bivariate regression.

When looking at country level data, MPD performs best in all cases except the multivariate regression on levels of trust in politicians. In all cases the difference between measures is very small.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	leave	leave	leave	leave	leave	leave	leave	leave
mean life satisfaction					0.293	2.774	2.037	18.45***
					(2.970)	(3.059)	(3.030)	(5.581)
standard deviation of life	23.53***				10.94***			

satisfaction								
	(2.570)				(3.552)			
mean pair distance of life satisfaction		22.44***				14.56***		
		(2.215)				(3.302)		
80:20 difference of life satisfaction			8.819***				5.419***	
			(0.893)				(1.322)	
average life satisfaction of bottom 40%				-8.252***				-15.91***
								(3.415)
median income					-0.00207***	-0.00200***	-0.00202***	-0.00198***
					(0.000118)	(0.000120)	(0.000119)	(0.000120)
% Asian					-16.74***	-16.39***	-16.26***	-15.77***
					(4.767)	(4.671)	(4.706)	(4.655)
% black					-105.6***	-104.6***	-104.7***	-101.9***
					(11.38)	(11.14)	(11.27)	(11.11)
air pollution					3.625***	3.618***	3.606***	3.597***
					(0.252)	(0.248)	(0.250)	(0.249)
unemployment					0.520***	0.439**	0.469***	0.427**
					(0.177)	(0.176)	(0.176)	(0.175)
life expectancy					1.630***	1.676***	1.695***	1.633***
					(0.330)	(0.327)	(0.329)	(0.325)
constant	7.876*	7.444*	6.930	98.63***	-96.60***	-127.9***	-122.5***	-121.7***
	(4.679)	(4.279)	(4.433)	(7.247)	(33.78)	(34.47)	(34.67)	(33.24)
N	740	740	740	740	533	533	533	533
R-sq	0.098	0.114	0.111	0.050	0.602	0.611	0.608	0.612
AIC	5557.1	5543.7	5546.5	5595.7	3573.4	3561.5	3564.7	3559.4
BIC	5566.3	5552.9	5555.7	5604.9	3611.9	3600.0	3603.2	3597.9

Table 7 Comparison of measures for life satisfaction inequality in UK local authorities in regression on EU referendum result

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	leave	leave	leave	leave	leave	leave	leave	leave
mean happiness					7.105** (2.867)	8.717*** (3.067)	7.621*** (2.860)	22.17*** (5.308)
standard deviation of happiness	22.87*** (2.395)				15.80*** (3.534)			
mean pair distance of happiness		19.67*** (2.055)				15.70*** (3.353)		
80:20 difference of happiness			8.403*** (0.844)				6.133*** (1.280)	
average happiness of bottom 40%				-7.672*** (1.231)				-15.38*** (3.196)
median income					-0.00202*** (0.000121)	-0.00200*** (0.000121)	-0.00201*** (0.000121)	-0.00200*** (0.000121)
% Asian					-15.38*** (4.727)	-15.40*** (4.735)	-15.27*** (4.705)	-15.10*** (4.755)
% black					-100.3*** (11.62)	-99.50*** (11.57)	-99.83*** (11.68)	-98.64*** (11.53)
air pollution					3.640*** (0.241)	3.602*** (0.240)	3.632*** (0.239)	3.590*** (0.239)
unemployment					0.527*** (0.175)	0.497*** (0.176)	0.492*** (0.176)	0.486*** (0.176)
life expectancy					1.597*** (0.330)	1.645*** (0.329)	1.605*** (0.329)	1.641*** (0.328)
constant	1.072 (5.203)	4.667 (4.808)	1.060 (4.983)	91.07*** (6.542)	-159.6*** (35.70)	-177.6*** (37.42)	-165.9*** (35.57)	-158.5*** (34.25)
N	740	740	740	740	533	533	533	533
R-sq	0.109	0.107	0.115	0.048	0.612	0.615	0.614	0.616
AIC	5548.2	5549.9	5543.3	5597.1	3560.3	3556.4	3556.7	3554.3
BIC	5557.5	5559.1	5552.5	5606.3	3598.8	3594.9	3595.2	3592.8

Table 8 Comparison of measures for happiness inequality in UK local authorities in regression on EU referendum result

					(1)	(2)	(3)	(4)
				pplrst	pplrst	pplrst	pplrst	
mean life satisfaction				0.0593	0.0175	0.0521	0.123	
				(0.106)	(0.113)	(0.114)	(0.106)	
standard deviation of life satisfaction	-0.664***			-0.424**				
	(0.128)			(0.202)				
mean pair distance of life satisfaction		-0.570***			-0.419**			
		(0.105)			(0.179)			
80:20 difference of life satisfaction			-0.226***			-0.138*		
			(0.0437)			(0.0747)		
average life satisfaction of bottom 40%				0.185***			0.0681	
				(0.0402)			(0.0610)	
frequency of meeting friends				-0.0578	-0.0607	-0.0462	-0.0567	
				(0.119)	(0.119)	(0.120)	(0.121)	
crime victim				1.184*	1.159*	1.224*	1.640**	
				(0.664)	(0.659)	(0.668)	(0.652)	
health				-0.0499	-0.0845	-0.0571	0.0383	
				(0.342)	(0.342)	(0.345)	(0.348)	
religious				-0.0329	-0.0270	-0.0416	-0.0583	
				(0.0775)	(0.0774)	(0.0774)	(0.0774)	
age				-0.0155	-0.0149	-0.0144	-0.0148	
				(0.0289)	(0.0288)	(0.0290)	(0.0294)	
years of education				0.0193	0.0198	0.0172	0.0126	
				(0.0413)	(0.0411)	(0.0415)	(0.0418)	

unemployment					0.00159	0.00211	0.000575	-0.00215
					(0.00826)	(0.00822)	(0.00824)	(0.00812)
log GDP					0.311	0.307	0.287	0.209
					(0.377)	(0.375)	(0.379)	(0.384)
constant	6.306***	6.211***	6.204***	3.960***	1.170	1.627	1.259	-0.305
	(0.271)	(0.240)	(0.251)	(0.207)	(3.774)	(3.784)	(3.835)	(3.741)
N	155	155	155	155	147	147	147	147
R-sq	0.185	0.201	0.185	0.152	0.305	0.312	0.299	0.284
AIC	-152.9	156.0	-153.0	-146.9	-160.8	162.3	-159.4	-156.5
BIC	-146.9	149.9	-146.9	-140.8	-127.9	129.4	-126.5	-123.6

Table 9 Comparison of measures for life satisfaction inequality in European countries in regression on level of social trust

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	trustpoliti cians	trustpoliti cians	trustpoliti cians	trustpoliti cians	trustpoliti cians	trustpoliti cians	trustpoliti cians	trustpoliti cians
mean life satisfaction					-0.0502 (0.181)	-0.0793 (0.191)	-0.106 (0.192)	0.151 (0.208)
standard deviation of life satisfaction	-1.587*** (0.299)				-0.684* (0.371)			
mean pair distance of life satisfaction						-0.570* (0.323)		
80:20 difference of life satisfaction			-0.541*** (0.102)				-0.265* (0.135)	
average life satisfaction of bottom 40%				0.257** (0.0994)				-0.0494 (0.115)
frequency of meeting friends					-0.483** (0.229)	-0.487** (0.229)	-0.463** (0.228)	-0.462** (0.232)
crime victim					0.712	0.768	0.646	1.467

					(1.221)	(1.218)	(1.222)	(1.166)
health					0.944 (0.649)	0.906 (0.652)	0.915 (0.649)	1.015 (0.660)
religious					0.192 (0.146)	0.193 (0.147)	0.183 (0.145)	0.144 (0.147)
age					-0.0821 (0.0504)	-0.0816 (0.0504)	-0.0820 (0.0503)	-0.0766 (0.0511)
years of education					-0.0792 (0.0774)	-0.0795 (0.0776)	-0.0792 (0.0772)	-0.103 (0.0781)
satisfaction with the economy					0.261*** (0.0472)	0.264*** (0.0469)	0.263*** (0.0464)	0.292*** (0.0448)
constant	6.800*** (0.631)	6.454*** (0.564)	6.554*** (0.584)	2.140*** (0.511)	6.876 (4.353)	6.900 (4.413)	7.444* (4.430)	2.802 (3.833)
N	155	155	155	155	155	155	155	155
R-sq	0.193	0.193	0.193	0.053	0.535	0.534	0.537	0.522
AIC	109.3	109.1	109.2	133.9	39.69	40.08	39.09	44.16
BIC	115.4	115.2	115.3	140.0	70.12	70.52	69.52	74.59

Table 10 Comparison of measures for life satisfaction inequality in European countries in regression on level of trust in politicians