Wellbeing inequality in Britain

methodology paper

March 2017

Saamah Abdallah,
Hanna Wheatley and
Annie Quick

This paper describes the calculation of measures of wellbeing inequality for local authorities and counties, using data from the UK Annual Population Survey.

This work is an output of the What Works Centre for Wellbeing Communities Evidence Programme. The What Works Centre for Wellbeing is an independent, UK government funded, organisation set up to produce robust, relevant and accessible evidence on wellbeing. We work with individuals, communities, business and government, to enable them to use this evidence to make decisions and take action to improve wellbeing.

The centre is currently supported various partners including the ESRC to produce evidence on wellbeing in four areas: work & learning, culture & sport, community, and cross-cutting capabilities in definitions, evaluation, determinants and effects.

The Community evidence programme brings together evidence on what community-level factors determine wellbeing – focusing on place (the physical characteristics of where we live), people (the social relationships within a community), and power (participation in local decision-making). The programme aims to make wellbeing evidence usable to people from diverse sectors who are working to improve it.

The consortium involves four universities and five civil society organisations. managed by the University of Liverpool, partners include Leeds Beckett, Sheffield, and Durham Universities the New Economics Foundation, Happy City, the Centre for Local Economic Strategies, Social Life Ltd, and Locality (Goldsmiths University was also involved in the first six months of the programme).

Data used

We used six Special Licence Access datasets from the Annual Population Survey, including annual datasets for the years 2011-12, 2012-13, 2013-14 and 2014-15, and pooled datasets for the years 2011-14 and 2012-15.

We used three different geographical variables depending on availability within each dataset. For 2012-13, 2013-14, 2014-15 and 2011-14 we used Unitary Authority/Local Authority Level 2009 boundaries (UALA09), which includes unitary authorities, shire counties, metropolitan boroughs and London boroughs, but not local government districts.

For the annual dataset 2011-12 only the Unitary Authority/County Level 2009 (UACNTY09) was available, so this was used. This variable does not include the 68 metropolitan boroughs and London boroughs for that year. Instead it identifies the 6 Metropolitan Counties (e.g. Merseyside rather than Liverpool, Knowsley, etc.) and two categories for London (Inner London and Outer London).

So as to understand change in these key parts of the country, we also calculated, for all years, wellbeing inequality statistics based on these larger geographical areas, and have included them at the bottom of the datasets for 2012-13, 2013-14, and 2014-15.

For the pooled dataset 2012-15 we used Local Authority District Codes (laua), which also includes local government districts such as Allerdale and Chelmsford.

To make analyses representative of the target population and lessen potential biases, we used a wellbeing weighting provided by the Office for National Statistics for each dataset. On the Annual Population Survey, wellbeing questions are only asked of persons aged 16 and above who have a personal interview and proxy answers are not accepted. Therefore, the wellbeing weight is calculated for each individual, and is zero for respondents who were under 16 years of age or who were not present in person for the interview.

The ONS considers the data from this survey robust enough to report at these geographical scales. For example, in 2014-15, for life satisfaction, the median sample size for each local authority used in our analysis was around 840 respondents, with samples ranging from 274 for Bedford to 1,686 for Hampshire. Sample sizes were marginally different for other questions as a few respondents did not answer all four wellbeing questions. According to the ONS, for the mean of life satisfaction, the median confidence interval was 0.29 points, and intervals ranged from 0.17 for Hampshire to 0.47 for Greenwich. According to the methodology for the survey, 'weighting factors take account of … the composition of the local population by age and gender'.[1]

The data file includes the N for each of the wellbeing questions each year, by geographical area. However, note that some of our measures use a subset of this sample, e.g. average of bottom 50% will be based on half the sample. Ethnicity difference and Education difference were not calculated if the N for ethnic minority or lower levels of education respondents fell below 40 and 100 respectively.

Types of wellbeing inequality

We use 'wellbeing inequalities' to refer to two types of measure.

The first type – univariate wellbeing inequality – refers to the degree of variance in wellbeing within a population, in our case, a locality, with no reference to any demographic or socioeconomic groups. This type of inequality is analogous to standard measures of income inequality such as the Gini coefficient. In the same way the Gini coefficient is calculated purely on the basis of distribution data about income, with no reference to any other variables, a univariate wellbeing inequality measure can be calculated purely on the basis of distribution data about wellbeing, with no reference to any other variables.

The second type – bivariate wellbeing inequality - refers to differences in wellbeing between population groups defined by some other demographic factor, for example between males and females, or between ethnic groups. This type of inequality is familiar to those who work on health inequalities, which are often operationalised in terms of the difference in health outcomes between socioeconomic groups. Where differences in health inequalities are measured by, for example, difference in terms of years of life, wellbeing inequality is measured in terms of higher versus lower levels of subjective wellbeing.

How we calculated measures of wellbeing

Wellbeing measures used

We used the four wellbeing questions collected in the Annual Population Survey since 2011. People are asked to respond to the questions on a scale from 0 to 10 where 0 is 'not at all' and 10 is 'completely'. The four questions are:

- "Overall, how satisfied are you with your life nowadays?"
- "Overall, to what extent do you feel the things you do in your life are worthwhile?"
- "Overall, how happy did you feel yesterday?"
- "Overall, how anxious did you feel yesterday?"

These questions were developed as part of the ONS Measuring National Well-being Programme. The ONS sought advice from experts working in the field of subjective wellbeing and consulted with specialists on the National Statistician's Measuring National Wellbeing Advisory Forum and Technical Advisory Group. Based on this, as well as an extensive programme of question testing, four questions were designed which provide a

[1] ONS (2016) *Personal Well-being in the UK QMI*, accessed via https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/qmis/subjectivewellbeingannualpopulationsurveyapsqmi

concise and balanced approach to the measurement of subjective wellbeing.[2]

Data from these questions are technically ordinal, rather than scalar. Ordinal data values represent categories with intrinsic ranking, but the exact numbers do not have significance beyond their ability to establish a ranking. Scalar data values, on the other hand, represent ordered categories with a meaningful metric. For example, if someone is 8 years old, and another is 4 years old, the first is double the second. But if someone has a wellbeing score of 8 and another has a wellbeing score of 4, we can't necessarily say that the first has double the wellbeing of the second. However previous studies have shown that wellbeing data can be treated as scalar in multivariate regressions[3] and a study has found that differences between scores when testing individuals twice were independent of the level of reported happiness, which if generally true would support this assumption.[4] More work needs to be done to confirm the validity of this approach.

We calculated wellbeing inequalities for each measure separately (the variable names are *satis, worth, happy and anxious*). We also calculated average inequality measures, which involved taking the average of either three[5] or four inequality measures for each question, to provide summary statistics.[6] As well as calculating an average where the questions were equally weighed, we also used a study by Lord Gus O'Donnell and Professor Andrew Oswald which attempted to determine the relative importance of the four questions to the general public by asking a convenience sample of respondents to give relative weights for each question.[7]

Wellbeing inequality between groups

Education-based wellbeing inequality

In terms of bivariate analyses, we used the example of education-based wellbeing inequality in the briefing paper to illustrate what a bivariate analysis using wellbeing inequality might look like. Education-based wellbeing inequality is operationalised as the difference in wellbeing between respondents whose highest qualification is GCSE level or lower, versus those who have some form of higher education, either a degree or vocational study, using the variable *Highest qualification (detailed grouping) (hiqul11d).*[8] Those whose highest level of qualification is A-level are not included in the calculations.[9] We chose to look at level of education because it represents the best available proxy measure for socioeconomic status

---

[2] Tinkler, L. and Hicks, S. (2011). *Measuring Subjective Well-being*, Office for National Statistics.

[3] Frijters, P. and Ferrer-i-Carbonell, A. (2004). How important is methodology for the estimates of the determinants of happiness? *The Economic Journal*, 114, 641–659.

[4] Krueger, A.B. and Schkade, D.A. (2008). 'The reliability of subjective well being measures', *Journal of Public Economics*, 92(8-9), pp. 18331845. Cited in Layard, R. (2017) In-press.

[5] To provide summary statistics for 'average of bottom 50%', 'average of bottom 40%', and '80:20 difference' we calculated average inequality measures by taking the average of the inequality measures for each of the three positively worded questions, i.e. satisfied, happy and worthwhile. As the fourth question, on anxiety has a very different distribution – with higher numbers indicating low wellbeing, it would not be appropriate to include it in these summary statistics.

[6] Note we took the average of the four inequality measures calculated separately for each wellbeing measure, rather than calculating average wellbeing for each individual respondent based on the four indicators, and then calculating an inequality statistic based on that average.

[7] O'Donnell G & Oswald A (2015) *'National well-being policy and a weighted approach to human feelings'* Working paper. Warwick economics research papers series (WERPS). We used the data in Figure 3.

[8] Variable: Highest qualification (detailed grouping) (hiqul11d). We recoded 'degree or equivalent' and 'higher education' into HighEd and 'GCSE grades A*-C or equivalent', 'other qualifications', and 'no qualification' into LowEd. 'GCE, A-level or equivalent' was left out.

[9] This was because those with A-levels tended to have a level of wellbeing intermediate between those with GCSEs or less, and those with higher education.

given the survey's lacks of data on household income.

The education-based wellbeing inequality figure was calculated by subtracting the lower education group average from the higher education group average for each local authority. Where the lower education group N was less than 100, we did not calculate the difference.

It is worth recognising that there is some correlation between education level and age.[10] Table 1 shows the percentage of each age group within each education group in 2014-15.

% within Age groups									
		Highest qualification (detailed grouping)							Total
		Degree or equivalent	Higher education	GCE, A-level or equivalent	GCSE grades A*-C or equivalent	Other qualifications	No qualification	Did not know	
Age groups	16-19	0%	2%	31%	48%	5%	12%	1%	100%
	20-24	21%	6%	39%	21%	7%	5%	1%	100%
	25-29	34%	7%	23%	20%	9%	6%	1%	100%
	30-34	37%	9%	21%	18%	9%	6%	1%	100%
	35-39	37%	10%	19%	18%	9%	6%	1%	100%
	40-44	32%	10%	19%	21%	10%	7%	1%	100%
	45-49	25%	11%	19%	24%	10%	8%	1%	100%
	50-54	23%	12%	21%	24%	10%	10%	1%	100%
	55-59	21%	11%	22%	20%	11%	15%	1%	100%
	60-64	19%	12%	21%	17%	11%	19%	1%	100%
	65-69	15%	11%	20%	15%	13%	25%	1%	100%
	70-74	22%	9%	21%	13%	11%	23%	2%	100%
	75-79	21%	9%	15%	14%	11%	30%	1%	100%
	80 & over	25%	7%	17%	8%	15%	25%	2%	100%
Total		24%	10%	22%	22%	10%	12%	1%	100%

Table 1: Cross tabulation: age by education group, 2014-15

Ethnicity-based wellbeing inequality

While not included in the briefing paper, we actually calculated a second bivariate wellbeing inequality measure –ethnicity-based wellbeing inequality. However, the numbers of ethnic minority respondents in the sample were so low that in order for us to have an N of more than 40, we had to calculate ethnicity –based wellbeing inequality as the difference in wellbeing between respondents who describe themselves as White, and those who use any of the other ethnic categories in the APS,[11] in other words the difference between White respondents and all ethnic minority respondents. Given that the differences *between* ethnic minority groups, and even between different generations of migrants, could plausibly be as

[10] Given this analysis uses descriptive statistics, we have not controlled for this correlation here. However, further research into wellbeing inequalities in 2017 will.

[11] Variable: Ethnicity (11 categories) UK level (White, Gypsy/Traveller/Irish Traveller, Mixed/Multiple ethnic groups, Indian, Pakistani, Bangladeshi, Chinese, Other Asian Background, Black/African/Caribbean/Black British, Arab, Other ethnic group, Does not apply, No answer)

wide as the difference between white and ethnic minority respondents, this distinction seemed too blunt to be useful.

However, the methodology and results of this analysis are reproduced here for information.

We used the variable *Ethnicity (11 categories) UK level (ethuk11)* to enable comparison across all six datasets. This variable does not separate White Other from White British, however, summary data suggested their wellbeing was much closer to that of White British respondents than to ethnic minority respondents. For example, using the variable Ethnicity (11 categories) GB level (which does separate them out), average life satisfaction in 2014-15 for White Other respondents was 7.68, compared to 7.62 for White British and 7.46 for other ethnic groups combined.

The ethnicity-based wellbeing inequality figure was calculated by subtracting the ethnic minority average from the White average for each local authority. Where the number of ethnic minority respondents (N) was less than 40, we did not calculate the difference because the sample size would have been too small. Although we initially hoped that we could use a threshold of 100, as was done for the education groups, this excluded a very high number of local authorities, and so the threshold was lowered. This is not ideal for statistical purposes and highlights the need for larger sample sizes when undertaking sub-group calculations at local authority level.

Table 2 ranks the ten local authorities in which ethnic minority respondents saw the largest wellbeing penalties. The number represents how much higher the average wellbeing score was for those who describe themselves as White compared to those of other ethnicities.

	Ethnicity-based wellbeing inequality	
1	Tameside	0.56
2	Wandsworth	0.55
3	Greenwich	0.54
4	Croydon	0.46
5	Bradford	0.39
6	Hampshire	0.39
7	Wokingham UA	0.30
8	Stoke-on-Trent UA	0.29
9	Oxfordshire	0.28
10	Hammersmith and Fulham	0.27

Table 2: Ten local authorities in which ethnic minority respondents saw the largest wellbeing penalties in 2014-15

In terms of ethnicity, we had a significantly reduced data set given the low numbers of people from ethnic minorities many local authorities. Ethnicity-based inequality was therefore only calculated for 103 local authorities in 2014-15, excluding many of the places which displayed high wellbeing inequality based on other measures (for example Blaenau Gwent). The region where ethnic minority respondents suffered the biggest gap in wellbeing was Tameside in Greater Manchester (an average difference of 0.6 points over all four ONS questions). The biggest difference was in responses to the life satisfaction question, where white respondents scored an average of 7.4, and ethnic minority respondents scored 6.7 – a difference of 0.7.

Following Tameside, several London boroughs appear in the top ten local authorities where ethnic minority respondents saw the largest wellbeing penalties, including Wandsworth, Greenwich and Croydon, with Bradford in fifth spot.

There were also many local authorities where ethnic minorities actually had slightly higher average levels of wellbeing. The local authority where this difference was the biggest was Stockport, and the advantage was strongest in terms of anxiety (white respondents being 0.8 points more anxious than ethnic minority respondents). Similar advantages were seen in several London boroughs (including Tower Hamlets and Enfield). As with Stockport, the advantage is typically biggest for anxiety. Again, in the middle of these ends of the scale, some local authorities had little to no difference in wellbeing levels between white and ethnic minority respondents. Examples include Sandwell, Sheffield and Bolton.

Other bivariate measures

We did consider inequalities based on other demographics, but discounted them for different reasons.

Inequality between genders in terms of wellbeing is generally very low. For the UK overall in 2014-15, the difference in average score between female and male respondents for satisfied, anxious, worthwhile and happy were 0.05, 0.31, 0.23 and 0.03 respectively, with females scoring slightly higher on each measure.

The relationship between wellbeing and age is well known, with a U-shaped curve typically found for wealthy nations such as the UK. However, precisely because the relationship is U-shaped rather than linear, it is not straightforward to create a single measure of wellbeing inequality. If one is to calculate the difference between two age groups, which two age groups should they be?

Lastly, we would have liked to calculate a measure of wellbeing inequality between different income groups. However, the Annual Population Survey wellbeing data set does not include data on household income, so this is currently not possible.

Further exploration of effective measures for these and other demographic groups could be a useful area for further study.

Univariate wellbeing inequalities

We calculated five types of univariate wellbeing inequality:

- Standard deviation
- Mean pair distance
- 80:20 wellbeing difference (i.e. difference in wellbeing between the 20% highest scoring and 20% lowest scoring respondents in a locality)
- Mean of the bottom 40%
- Mean of the bottom 50%

The **standard deviation** is the default measure of variance within a single variable. It is the average difference from the mean for any individual within a dataset.[12]

Mean pair distance[13] is the average difference in score between two randomly chosen individuals within the dataset. Functionally, it is equivalent to the Gini coefficient used to measure income inequality, the Gini being the mean pair distance divided by twice the mean. The difference between the two is that the Gini coefficient is a ratio measure,

[12] In calculating the standard deviation, all differences are squared, before they are averaged, and then the square root is taken.

[13] Kalmijn, W. and Veenhoven, R. (2005) Measuring inequality of happiness in nations: In search for proper statistics, *Journal of Happiness Studies,* 6, pp.357-396.

whereas mean pair distance is an interval measure. The Gini treats the difference between £100 and £200 the same as it treats the difference between £500 and £1,000, but the mean pair distance does not treat the difference between a wellbeing score of 1 and 2, as the same as the difference between 5 and 10.

The mean pair distance for a distribution can be calculated relatively simply using a frequency table, simply working out all the differences between each pair of respondents, summing them all up, and dividing that by the number of possible combinations of respondents.

We have calculated the **80:20** wellbeing difference as analogue to the 80:20 ratio calculated for income distributions. It is the difference, for any wellbeing measure, between the average wellbeing of the 20% of respondents with the highest wellbeing, and the average wellbeing of the 20% of respondents with the lowest wellbeing. As such, it focuses on the extremes of the distribution, whereas the mean pair distance and standard deviation reflect variation across the whole distribution. The 80:20 difference can also be calculated using a frequency table, and some relatively straightforward algorithms.

Finally, we calculated the **average of the bottom 50%** and the **average of the bottom 40%** for each measure. For the positive measures (satisfied, happy and worthwhile), the bottom refers to the lower end of the scale, whereas for the negative measure (anxious), the bottom refers to the higher end of the scale, as higher levels of anxiety equate to lower wellbeing. These measures are similar to relative measures of poverty, in that they focus on the situation of individuals at the bottom of the distribution. We chose these relative measures rather than absolute measures (e.g. those living below $1 a day, or in wellbeing terms, those with a wellbeing score below a certain threshold) because of the difficultly in choosing a threshold – at what point is someone experiencing the wellbeing equivalent of poverty, and who decides where that threshold is? Instead, we chose to measure the experience of the worse-off half, and the worse-off bottom 40 per cent to make the data available for comparison to see if either of these measures tells us something different from the Mean Pair Distance, the 80:20 difference or the Standard Deviation.

Statistical tests

To be able to make meaningful comparisons using measures of wellbeing inequality, tests of statistical significance are important. These allow us to be able to say whether, for example, two places really do have different levels of wellbeing inequality, or whether in fact an apparent difference is just the result of noisy data.

When comparing means, there are plenty of recognised statistical tests. However, the same cannot be said of measures of inequality, presenting an important methodological challenge to this research area.

Bivariate measures

For the bivariate measures, two types of test are relevant.

To determine whether an inequality within a local authority is real, a t-test is required to compare the averages for the two groups in question. The raw data is required for this and, unfortunately, we only had access to that raw data for a limited period of time, so we were unable to perform such t-tests.

To determine whether there is a significant difference *between* the bivariate inequalities of two different local authorities, a two-way interaction ANOVA test is appropriate. The two independent variables would be locality (e.g. Liverpool vs. Dorset) and group (e.g. high education vs. low education). The interaction term tells one the size of the difference between the effects of the group (i.e. education) in one locality versus the other. If the term is significant that means that education plays a significantly different role in determining levels of wellbeing in the two localities.

Univariate measures

For univariate measures of inequality, the situation is a little more complex. Whilst there is no direct test to compare **standard deviations**, Levene's test of equivalence is commonly used to compare variances (which are closely related to standard deviations). The test statistic W is calculated as follows: [14]

$$W = \frac{(N-k)}{(k-1)} \frac{\sum_{i=1}^{k} N_i (Z_{i\cdot} - Z_{\cdot\cdot})^2}{\sum_{i=1}^{k} \sum_{j=1}^{N_i} (Z_{ij} - Z_{i\cdot})^2},$$

Where

- W is the result of the test,
- k is the number of different groups to which the sampled cases belong,
- N is the total number of cases in all groups,
- N_i is the number of cases in the ith group,
- Y_{ij} is the value of the measured variable for the jth case from the ith group,
- $Z_{ij} = \begin{cases} |Y_{ij} - \bar{Y}_{i\cdot}|, & \bar{Y}_{i\cdot} \text{ is a mean of the } i\text{-th group}, \\ |Y_{ij} - \tilde{Y}_{i\cdot}|, & \tilde{Y}_{i\cdot} \text{ is a median of the } i\text{-th group}. \end{cases}$

(Both definitions are in use though the second one is, strictly speaking, the Brown–Forsythe test – see below for comparison.)

- $Z_{\cdot\cdot} = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{N_i} Z_{ij}$ is the mean of all Z_{ij},
- $Z_{i\cdot} = \frac{1}{N_i} \sum_{j=1}^{N_i} Z_{ij}$ is the mean of the Z_{ij} for group i,

The significance of W is tested against $F(\alpha, k-1, N-k)$ where F is a quantile of the F-test distribution, with $k-1$ and $N-k$ its degrees of freedom, and α is the chosen level of significance (usually 0.05 or 0.01).

The Excel spreadsheet includes a tool to calculate Levene's test for equivalence as per this formula for the user's choice of two local authorities.

Note that the Levene's test can only be applied to data from a single wellbeing question – it is not possible to apply it to the overall wellbeing inequality measure made by taking the average of the standard deviations for all four questions.

What kinds of differences are significant? Typically, for two median sized areas, a difference in standard deviation of over about 0.12 is significant at the p<0.05 level, whilst a difference over 0.20 is typically significant at the p<0.01 level. Having said that, sometimes differences in standard deviation as small as 0.01 can underlie significant differences in variance, whilst much larger differences in standard deviation can fail to be significant, particularly when sample sizes are small.

---

[14] Here, W, the test statistic can be assumed to have the same distribution as F. In Excel, one can use the F.DIST.RT formula to calculate the p value associated with W.

Another way to look at this is in terms of ranking. What difference in ranking in inequality can typically be expected to be significant? For life satisfaction, a ranking difference of 30 places (out of 203) is significant 16% of the time, looking at data from 2014-15. A difference of 60 places (out of 203) is significant more often than not (77 of these differences are significant, 66 are not, at p<0.05). By the time you reach a ranking difference of 82 places, differences are significant at the p<0.01 level more often than not.

A difference in ranking of 109 places or more is significant at p<0.05 90% of the time.

With happiness and anxiety, ranking differences of such magnitudes are marginally more likely to be significant, with worthwhile, marginally less likely to be significant.

We also looked at the number of changes in inequality over time that were statistically significant. Table 3 shows that life satisfaction significantly decreased in more than half of the 143 local authorities for which we had data in 2011-12 and 2014-15, and only significantly increased in one. However, fewer differences were statistically significant for the other three wellbeing questions.

	Life Sat	Happy	Worthwhile	Anxiety
Significant increase in inequality (p<0.01)	0	1	0	4
Marginally significant increase in inequality (p<0.05)	1	1	4	1
No significant difference	64	87	80	105
Marginally significant decrease in inequality (p<0.05)	23	20	21	14
Significant decrease in inequality (p<0.01)	55	34	38	19

Table 3: Count of local authorities for which differences in inequality between 2011-12 and 2014-15 were significant (out of 143 local authorities)

Theoretically, combining the four questions in the overall wellbeing inequality measure should increase our statistical power and reduce the noise in the data, thus making smaller differences significant. However, substantial statistical work is required to be able to quantify this. For now, no statistical tests are recommended for comparing the overall wellbeing inequality measure.

For the **80:20 wellbeing difference**, a two-way interaction ANOVA test should be appropriate (as with bivariate inequality). In this case, the two independent variables would be the locality and whether a given respondent is in the bottom or top 20% of the wellbeing distribution (with those in the middle 60% excluded from the analysis). The significance of the interaction term would indicate whether the difference between the top 20% and bottom 20% in terms of wellbeing is significantly different in the two localities. However, raw data is required to calculate this, and we did not have sufficient time to do this.

To compare **averages of the bottom 50% of bottom 40%**, usual statistics for comparing means should be appropriate.

The biggest challenge is with regards to the **mean pair distance** statistic. Preliminary research suggests that complex bootstrapping methodologies are required to be able to say whether two mean pair distances are statistically significant. Further work is required to understand how to apply these methods to wellbeing data.

Discussion of measures

The 80:20 difference may be easier to communicate to a broad public, but ignores the 60% of the population in the middle of the distribution, so may miss important differences in wellbeing inequality. Both the standard deviation and mean pair distance reflect the entire distribution but differ in their methods of calculation. Whilst the standard deviation may be more familiar to those with a basic knowledge of statistics, the mean pair distance is a recognised measure as well, and we believe the decision about which measure should be used in future work depends on empirical research. Key questions that need to be answered include:

a) Which measure is the most independent from the mean – i.e. provides additional information that the mean does not offer?

b) Which measure is most methodologically robust (e.g. least vulnerable to cultural biases in response style, or 'bounded scale' effects)?

c) Which measure has more external validity – i.e. is predicted by things one would expect to affect wellbeing inequality?

d) Which measure has more predictive power – i.e. predicts other outcomes?

e) Which measures best reflect public and political priorities – i.e. highlights the parts of the distribution of most public interest or concern?

f) Which measures allow statistical tests to be conducted?

Answering these questions requires further empirical research.

In the analyses we have conducted we have found that all three measures of univariate wellbeing inequality are strongly correlated. Table 4 presents the top and bottom 10 rankings for local authorities in 2014-15 based on the three measures.

Top 10 Unequal						
	Standard deviation		**Mean Pair Distance**		**80:20 difference**	
1	Blaenau Gwent	2.5	Blaenau Gwent	2.8	Blaenau Gwent	6.3
2	Liverpool	2.4	Liverpool	2.7	West Dunbartonshire	6.2
3	Neath Port Talbot	2.4	Merthyr Tydfil	2.6	Doncaster	6.2
4	Merthyr Tydfil	2.4	Knowsley	2.6	Renfrewshire	6.2
5	Knowsley	2.4	Neath Port Talbot	2.6	Barnsley	6.2
6	Sunderland	2.4	Sunderland	2.6	South Ayrshire	6.1
7	Rotherham	2.4	Inverclyde	2.6	North Lanarkshire	6.1
8	Kingston Upon Hull	2.4	Kingston Upon Hull	2.5	Blackpool	6.1
9	Inverclyde	2.4	North Ayrshire	2.5	Wakefield	6.1

10	North Ayrshire	2.3	Rotherham	2.5	Sefton	6.1
Top 10 Equal						
	Standard deviation		**Mean Pair Distance**		**80:20 difference**	
1	Enfield	1.8	Enfield	1.9	Enfield	4.8
2	Cheshire East	1.8	Harrow	1.9	Cheshire East	4.8
3	Harrow	1.8	Falkirk	1.9	Eilean Siar, Orkney & Shetland	4.9
4	Eilean Siar, Orkney & Shetland	1.8	Cheshire East	1.9	Warwickshire	4.9
5	Warwickshire	1.9	Warwickshire	1.9	Falkirk	5.0
6	Wokingham	1.9	Eilean Siar, Orkney & Shetland	2.0	Harrow	5.0
7	Falkirk	1.9	Wokingham	2.0	Barnet	5.0
8	Lambeth	1.9	Aberdeenshire	2.0	Aberdeenshire	5.0
9	Aberdeenshire	1.9	Barnet	2.0	Wokingham	5.0
10	Barnet	1.9	Wandsworth	2.0	Wandsworth	5.1

Table 4: Top and bottom 10 rankings for local authorities in 2014-15 according to standard deviation, Mean Pair Distance and the 80:20 difference

Whilst the orders are slightly different, the top 10 for inequality on standard deviation and mean pair distance are identical. The ranking for 80:20 difference is a little different, with more Scottish local authorities such as West Dunbartonshire and Renfrewshire featuring in the top 10.

The most equal local authority on all three measures is Enfield and the top 10s for the three measures are almost identical.

Box 1: Wellbeing inequality predicts Brexit referendum decision

Perhaps the most important political decision that the UK public has made for more than a generation was the vote in June 2016 to leave the European Union. The causes of this decision are deep-rooted and complex. Economic inequality has been touted by many as the main cause,[15] whilst others have focussed on the alienation of the white working classes.

We wanted to explore whether wellbeing played a role in explaining the results. Whilst

[15] https://www.equalitytrust.org.uk/brexit-and-inequality-its-not-about-globalisation

mean wellbeing within a local authority did not predict the percentage of people who voted to leave within an area,[16] overall wellbeing inequality did. Places which had higher overall wellbeing inequality were more likely to vote to leave the European Union. The strongest relationship was for the Mean Pair Distance of life satisfaction, which is shown in Figure 1.[17]

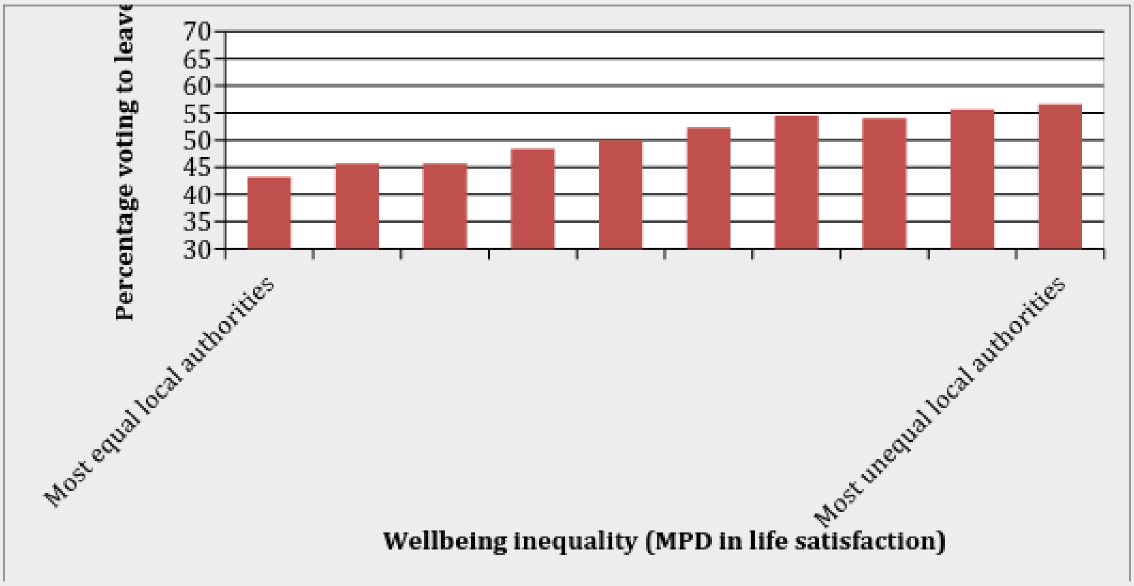

Figure 1: Voting patterns in local authorities as a function of wellbeing inequality (measured as MPD in life satisfaction)

The relationship was significant, even after we controlled for several key variables which have been discussed in relation to the referendum results including median income, income inequality, unemployment levels, education levels and ethnicity. Table 5 shows the results of this regression. Variables highlighted in yellow were marginally significant (p<0.05), those in orange were significant (p<0.01). Those that aren't highlighted weren't significant at all – in other words they did not help explain the referendum result once the preceding variables had been taken into account. The beta coefficients indicate the relative strength of each effect, with bigger numbers indicating stronger effects. Interestingly, average levels of anxiety within a locality also predicted referendum results, with places with higher levels of reported anxiety more likely to vote to remain in the European Union.

Variable	Beta coefficient
Education (% residents with higher education)	-0.377
MPD in Life Satisfaction	0.233
Mean anxiety	-0.161
Median income	-0.106
% Residents who are White British	0.099
Income inequality (80:20 income ratio)	0.092
Unemployment rate	-0.023

Table 5: Standardised beta coefficients for regression predicting percentage voting to leave, by local authority (R^2 for model – 0.45)

It is worth noting that this analysis was sensitive to the precise measure of wellbeing inequality being used. Using the standard deviation of life satisfaction, instead of the MPD, wellbeing inequality marginally fails to be significant. However both the overall MPD (i.e. the

[16] Though average wellbeing did predict turnout.
[17] Bivariate correlation had R=0.37, p=0.000

average MPD for all four wellbeing measures) and the overall standard deviation *were* significant. These subtle differences highlight the importance of carefully identifying the most appropriate measures of wellbeing inequality for different purposes.

Implications for further research

We hope the data made available with this briefing paper will allow local authorities to start looking at wellbeing inequality in their areas, while the datasheet can be used by analysts to explore wellbeing inequality patterns across the country. We also hope to open up the debate about the best ways to measure and report wellbeing inequalities. In particular, we are keen to explore which of the measures outlined in this methodology could have the best ratio of statistical robustness to public resonance.

These are very early days in the study of wellbeing inequalities, particularly at low geographical levels such as local authorities. As with any research, our study has a number of limitations. Two key questions that still need to be resolved in relation to the study of wellbeing inequality include: a) What are the best measures of wellbeing inequality, i.e. which measures best reflect the kind of inequality that matters most to people, and which measures are most robust? And b) what are the best methods for testing statistical significance, particularly where wellbeing measures have been combined?

The question of most importance however is what drives wellbeing inequality at the local level, and what can be done to reduce it.

This study is the first in a programme of work being undertaken by the What Works Centre, the ONS and the New Economics Foundation. Over 2017 we will also be producing:

- Analysis on the drivers of wellbeing inequality at local authority level, using multi-level modeling to establish causal relationships. We'll focus on a local level to help policy makers and those working in communities to understand how inequalities can be reduced.
- A review of the methodological considerations surrounding the measurement of wellbeing inequality including an assessment of the appropriateness of a range of different indicators for different uses.

We'd love to hear your thoughts, comments and ideas.

Which measure of wellbeing inequality most interests you, or could be most or least useful in your work? How might information on wellbeing inequalities be used in local decision-making? What more would you like to know? Please send comments to hanna.wheatley@neweconomics.org.