



Version 1.0
Prepared by Tony Cramp
Data Orchard CIC
October 2018

HEREFORDSHIRE QUALITY OF LIFE SURVEY 2018 - TECHNICAL REPORT



Contents

.....	1
Contents	2
Introduction	3
Version history	3
Sampling methodology	4
Weighting Scheme.....	5
Confidence intervals	7
Appendix 1 – Attributes of postcode associated with each response	9

Introduction

This report supplements other reports on the Herefordshire Quality of Life Survey conducted by Data Orchard CIC in 2018.

It provides details about certain technical aspects of the survey and its analysis, specifically regarding the sampling methodology, the weighting scheme applied to the results and the calculation of confidence intervals applied to the results.

The survey design and the technical aspects referred to above are based on the guidance provided for the Place Survey 2008-9. This was a national survey conducted by local authorities in England in order to measure and compare opinions on a range of quality of life aspects of residents across the country.

Version history

Version 1.0 – Initial release.

Sampling methodology

A list of all residential addresses in the county of Herefordshire was procured from Arc en Ciel Ltd in April 2018 who extracted the data from Royal Mail's national data set. This provided a total of 84,532 addresses.

The Intelligence Unit at Herefordshire Council provided a "postcode lookup" file of all postcodes in the county¹, together with a range of further attributes of each postcode, such as parish, ward, census output area, Defra Urban/Rural category, Deprivation ranking. (See Appendix 1 for full list).

To provide the opportunity for analysis of the survey results with these postcode attributes, the data for each postcode was merged with the list of addresses.

The postcode for 106 of the 84,532 addresses (0.13%) were not found in the postcode lookup file, these were deleted prior to the sampling process, leaving 84,426 addresses. These addresses were ordered alphanumerically first by Postcode and then by the first line of the address.

It was decided that a sample size of 4,125 households would be used for the survey. This was considered a reasonable balance to provide enough responses to provide a useful, sufficiently accurate estimation of the results for the county as a whole and to enable some more detailed analysis of sub groups of the population.

Consideration was given to stratify the survey, for example by locality, to ensure sufficient representation from each area. It was decided that, on balance, the benefits of doing this - an improved understanding of views in each locality - would be outweighed by additional inaccuracies introduced by the additional weighting then necessary to estimate the results for the county as a whole.

To achieve a random sample of 4,125 addresses from the 84,426 address one in every 20.5 ($= 4,125 / 84,426$) addresses was selected starting from a randomised start point.

These selected addresses were assigned a reference number from 1 to 4,125 which was printed on the questionnaire and enabled the returned questionnaire to be matched with the address. This enabled the attributes of the postcode to be associated with each response and used during the analysis. It also meant that reminder letters were only sent to those addresses where no response had yet been forthcoming.

The questionnaires were posted out with a covering letter, which explained the rationale for the survey, instructions with a deadline and a Freepost envelope to use to post a response back. They were mailed out in April 2018 and the final responses were accepted up until early June 2018.

¹ Postcode Directory, Herefordshire, derived from All Fields Postcode Directory, ONS. December 2017

Weighting Scheme

Summary

The weighting is based on the scheme used on the National Place Survey in 2008 – 9 which had three purposes:

- To account for different sample rates (e.g. due to stratification)
- To adjust for lower chance of inclusion of a people in larger households, and
- To compensate for different likelihood of response and coverage.

Since the sampling in the 2018 Quality of Life Survey did not utilise stratification or different sampling rates, this element of the weighting scheme is not applicable here.

This survey asks for one adult in each household to return a completed questionnaire regardless of the number of people in that household. This means that the higher the number of people in a household, the lower the chance each person in that household has to participate in the survey. A weight factor of the number of people in the household is used to compensate for this.

It is understood that different groups of people (by age, ethnicity etc.) tend to be more or less likely to respond to surveys. The weighting scheme compares the achieved response rate of certain groups to estimates of that group in the target population and adjusts the weight factor for the response to compensate. So the weight factor for a group that is under-represented in the survey response compared to the total population is greater than one, and for those over represented, it is less than one.

The above weight factors for each response are multiplied together to give a combined weight for the response and then normalised so that the total responses after weighting is the same as that before.

As recommended in the Statistical Review of the Place Survey², to reduce the impact of particularly high weights, any weight factor of five or over is capped at five. Finally, the weight factor for each response is again normalised by multiplying each by the total number of responses divided by the sum of all weight factors to give, in effect, an average weight factor equal to one.

Calculation of weight factors

The number of people (27) who selected the options "Other white" or "Any other ethnic group" in response to question 43 which asked "How would you describe your ethnic group?" were considered too few to be used for weighting purposes, so ethnic group was not used as a factor in the weighting process.

Gender, or more accurately, the response to question 41, "Do you identify: As a man, As a woman, In some other way or Prefer not to say" was also considered as a factor for weighting purposes. However, due to the layout on this question in the questionnaire, there was some concern about the reliability of the responses to this question. For this reason, gender was not used in the calculation of the weighting factors for the responses.

The weight factors were therefore based on the number or people in the household (question 29) and the respondents' age band (question 38).

² Published by Department for Communities and Local Government, Feb 2010, ISBN: 978-1-4098-2232-5.

Imputing missing data

A number of respondents had not replied to question 29, “How many adults aged 16 or over are living here?” and / or question 38, “What was your age on your last birthday?”. In order to calculate the weight factors, these variables were imputed in the cases where they are missing. A method called “Hotdecking” was used for this process – see the Place Survey 2008-9 weighting scheme for full details. It is important to note that these imputed values are used for the purposes of calculating the weight factors only and are not assigned to these responses when analysing the results.

Number of adults in household

As the following table shows, 35 respondents had not indicated how many adults lived in their household and a number was imputed for these responses.

Number in Household	Before imputing		After imputing	
	Number	%	Number	%
1	210	32.6%	220	34.2%
2	246	38.2%	257	39.9%
3	116	18.0%	127	19.7%
4	26	4.0%	28	4.3%
More than 4	11	1.7%	12	1.9%
Not answered	35	5.4%	0	0.0%
Total	644		644	

Age band

Respondents were asked for their age on their last birthday (question 38). For the purpose of imputing, these were initially categorised into the age bands shown below. A total of 35 respondents did not answer the question - some were the same as those above who chose not to answer the number in household question, but not all. The results of the imputation process are shown in the table below.

Age band	Before imputing		After imputing	
	Number	%	Number	%
16 to 17	0	0.0%	0	0.0%
18 to 24	5	0.8%	5	0.8%
25 to 34	29	4.5%	31	4.8%
35 to 54	124	19.3%	135	21.0%
55 to 64	143	22.2%	147	22.8%
65 or over	308	47.8%	326	50.6%
Not answered	35	5.4%	0	0.0%
Total	644		644	

Calculating weight factors – Number in Household

The weight factor for each response is simply the number of adults aged 16 or over living in the household, except that where the respondent has selected more than four, this has been treated as five. For the purpose of calculating the weight factors, imputed values (see the previous section) were used in the cases where a respondent had not answered the question.

To normalise the result, these weight factors were then multiplied by a scaling factor equal to the total number of responses divided by sum of the household size weighting factor for all responses. This factor equalled 0.5003885.

Calculating weights factors – Age band

When calculating weight factors for age, four age bands were used: 16 to 34 years, 35 to 54, 55 to 64 and 65 and over. The number of respondents in each age band (using imputed data where missing) was compared to the number of people in Herefordshire in each age band using the Herefordshire 2016 mid-year estimates³. The ratio of the population estimate for this age band divided by the number of responses in the achieved sample provides the age-based weighting factor for each age band to compensate for over or under-representation of each age group in the survey. This is shown in the table below.

Age band (years)	Proportion in achieved sample	Proportion in population	Weight factor for age
16 to 34	5.6%	24.3%	4.3501
35 to 54	21.0%	30.8%	1.4690
55 to 64	22.8%	16.4%	0.7204
65 or over	50.6%	28.4%	0.5619

Calculating final weight factors

To calculate the final weight factor applied to each response, the weight factor for age and the household size weight factors were multiplied together and scaled to normalise the weights. This led to a few high value weights – two responses had a combined weight of 10.7 and two of 6.4 while no other responses had a combined weight of more than 5.0. The weighting factor for these four responses were capped at 5.0 and the weights again scaled to normalise to provide a sum of 644 responses, the size of the achieved sample.

Confidence intervals

This survey uses the views of the responses achieved from those sampled to estimate the views of the population of a wider population, such as all adults aged over 16 in the county, or in some cases sub groups of the wider adult population, such as those adults who live in rural areas of the county. As such there is a margin of uncertainty about the result, known as a confidence interval and which can be calculated.

³ Herefordshire Population figures 2016 mye, ONS, Population Estimates Unit Annual Mid-Year Population Estimates for the UK, Office for National Statistics © Crown Copyright 2016.

As recommended by the Statistical Review of the Place Survey referenced above, this confidence interval has been inflated to account for the effects of weighting and the effective sample size.

The worst case inflated confidence interval for this survey when estimating the views of the adult population of Herefordshire = + / - 5.32%.

This means that there is a 95% probability that the true figure for the views of the adult population lies within the range of the estimated figure plus or minus 5.32%.

The inflated confidence intervals have been calculated as follows:

$$\text{C.I.} = \text{C.I. inflation factor} \times 1.96 \times \text{Sqrt} (p \times (1-p) / (n-1)) \times \text{Sqrt} (1-(n-1)/\text{Pop})$$

where

p = proportion of the achieved sample giving the response

n = the achieved sample for the specific question

Pop = size of the population whose views are being estimated by this response

CI inflation factor = 1.3786 - see below for the calculation of this factor.

Confidence interval inflation factor

The effective sample size has been calculated as equal to the square of sum of weighting factors divided by the sum of the squared weighting factors = $414736 / 1152 = 360$.

Based on the achieved sample size of 644, the worst case confidence interval for the survey when estimating the views of the adult population of Herefordshire is + / - 3.86%.

Based on the effective sample size of 360, the worst case confidence interval for the survey when estimating the views of the adult population of Herefordshire is + / - 5.32%.

Hence the confidence interval inflation factor = $5.32 / 3.86 = 1.3786$.

Appendix 1 – Attributes of postcode associated with each response

Ward
Ward Name
2011 Census output area
2011 Census lower layer super output area (LSOA)
LSOA Name
2011 Census middle layer super output area (MSOA)
MSOA Name
2003 Ward according to allocated OA
Parish according to allocated OA
2011 Urban/Rural classification
Children centres (after merge)
Locality
Sub locality
Index of Multiple Deprivation (IMD) Decile (where 1 is most deprived 10% of LSOAs in England)
IMD15: national percentile
HEREFORDSHIRE RANK (1=most deprived; 116=least deprived)
IMD15: Herefordshire percentile

***** End of Report *****