

Technical report: logistic regression and latent class analysis of loneliness using the Community Life Survey August 2016 to March 2017

Description of the statistical methods and techniques which underpin the article 'Loneliness - What characteristics and circumstances are associated with feeling lonely?'

Contact:
Edward Pyle and Dani Evans
qualityoflife@ons.gsi.gov.uk
+44 (0) 1329 447141

Release date:
10 April 2018

Next release:
To be announced

Table of contents

1. [Introduction](#)
2. [The Community Life Survey 2016 to 2017 data](#)
3. [Logistic regression](#)
4. [Latent class analysis](#)
5. [Appendix 1: Logistic regression – statistical explanations and tables](#)
6. [Appendix 2: Latent class analysis R output](#)

1 . Introduction

This technical report accompanies [Loneliness – What characteristics and circumstances are associated with feeling lonely?](#), an exploration of factors associated with loneliness. Using data from the [Community Life Survey August 2016 to March 2017](#), bivariate analysis was initially carried out to explore possible associations between a range of individual characteristics and circumstances and self-reported loneliness. This was followed by further, more in-depth analyses to explore the nature and relative strength of these relationships with loneliness. The aim has been to produce in-depth insights to help decision makers target initiatives to alleviate loneliness more effectively.

The research reported here used an iterative research programme involving descriptive analysis followed by logistic regression and finally, latent class analysis (LCA). The logistic regression and LCA analysis approach the exploration of loneliness from two different, but complementary, standpoints. Whilst the logistic regression seeks to isolate single factors that impact on the likelihood of loneliness, LCA seeks to identify combinations of factors that frequently appear together among those who report loneliness. This helps to provide a more holistic picture and highlights that, in practice, it may be a combination of multiple characteristics and circumstances that together shape our experiences and perceptions of loneliness. This article provides technical information about how these techniques were applied.

2 . The Community Life Survey 2016 to 2017 data

The research relied on data from the annual Community Life Survey (CLS), a nationally representative household survey of adults (aged 16 and over) in England. The CLS 2016 to 2017 dataset contains data for 10,256 adults for the period August 2016 to March 2017. For further information see the [Community Life Online and Paper Survey Technical Report 2016 to 2017](#).

The CLS 2016 to 2017 dataset was selected for analysis because the survey asked respondents about their frequency of loneliness. The survey also solicited information about the respondents' socio-demographic characteristics, behaviours, attitudes, community engagement and circumstances, which were used as explanatory variables.

Loneliness: the outcome variable

Central to the analysis was the question included in the CLS 2016 to 2017, which asked respondents: How often do you feel lonely?

1. Often/always
2. Some of the time
3. Occasionally
4. Hardly ever
5. Never

For the purposes of this report this is referred to as “the loneliness question”.

(Re)coding variables for analysis

Dichotomising loneliness

A binary version of the loneliness variable was used for the logistic regression and LCA. Responses of “often /always”, “some of the time”, and “occasionally” were collapsed into a single category of “more often lonely”, and those of “hardly ever” or “never” into another of “hardly ever or never lonely”. Whilst dichotomising the outcome variable in this way obscures some differentiation between frequency categories of reported loneliness, it was necessary for the logistic regression and LCA techniques. Reasons for recoding loneliness in this way are detailed in this section.

There is a relatively small sample size. The CLS 2016 to 2017 dataset contains responses from 10,256 individuals and, of these, 10,057 cases have valid data for the loneliness question. For a case to be included in the LCA model there must be valid data for every variable included in the model. With inclusions of each additional variable there is greater likelihood that any given case will become ineligible due to missing data and so be excluded from the model. In the final logistic model and LCA specification (see section 3 and 4 respectively), the sample size was reduced to 6,414 and 6,149 respectively because of missing data.

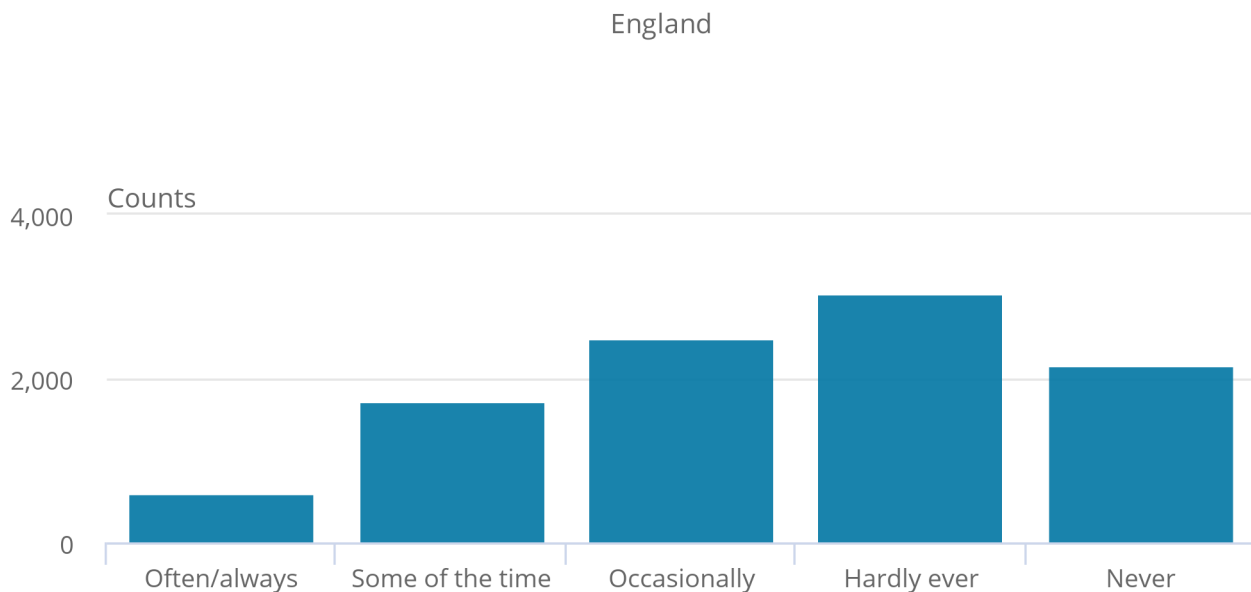
For reasons of statistical quality, it was decided that explanatory variables should, ideally, be tabulated with the binary loneliness variable so that wherever possible all (unweighted) cell counts are at least 100. This “100 minimum cell count” rule was relatively arbitrary but it was decided that some sort of minimum count was needed. This rule was achieved in all variables except for economic activity where, due to relatively small numbers of unemployed in the sample, 60 (unweighted) cases reported unemployment and that they experienced loneliness “hardly ever” or “never”.

Whilst it was necessary to recode variables to have fewer categories, ideally recoding should preserve the underlying distribution whilst having fewer categories¹. The distribution of responses to the loneliness question is shown in Figure 1.

Figure 1: Distribution of responses to the loneliness question (unweighted counts), August 2016 to March 2017

England

Figure 1: Distribution of responses to the loneliness question (unweighted counts), August 2016 to March 2017



Source: Community Life Survey 2016 to 2017', Department for Digital, Cultural, Media & Sport

This shows that the frequency of loneliness is skewed towards the “hardly ever” and “never” end of the response scale. By dichotomising the loneliness variable as described previously, categories had broadly similar frequencies of respondents thereby broadly preserving the distribution of the original variable: 4,841 were “more often lonely” and 5,216 as “hardly ever or never lonely”. With a larger sample size, it may have been possible to include more categories of loneliness thereby aiding greater differentiation in terms of loneliness frequency.

Another reason is consistency between coding for the logistic regression and LCA. As the LCA (for the reasons described previously) required a binary version of the loneliness variable, for consistency of results it made sense to apply a form of logistic regression that uses binary coding. Additionally, while it is possible to conduct multinomial logistic regression with multiple categorical outcomes, logistic regression with binary outcomes (for example, “lonely” compared with “not lonely”) is also easier to interpret and explain.

Recoding (and deriving) explanatory or independent variables

In many instances, independent or explanatory variables needed further preparation before inclusion in the models.

As noted earlier, it is better to preserve the original distribution of variables as much as possible when recoding for LCA and this was taken into consideration when recoding explanatory variables. Also, (as noted earlier) missing data is problematic. Therefore, variables that had more than 3,000 missing cases were excluded.

Small cell counts can produce poor quality analysis. As noted earlier, to ensure that when each explanatory variable was tabulated with the loneliness variable there was a minimum cell count of 100, categories were collapsed and, where appropriate, some categories were recoded as missing, thereby removing those cases from analysis. After recoding, and as already noted, only economic status broke this rule due to a relatively small number of unemployed people in the sample.

Greater importance, though, was given to producing recodes that were useful for meaningful interpretation – categories were only collapsed where the new category made sense. For example, it would not have been meaningful to collapse unemployed people into any other economic category.

Missing data and bias

As noted, cases with missing data for variables included in the LCA model are excluded from analysis. Missing data can produce biased estimates and invalid conclusions, particularly if data are not “missing at random” or, in other words, if there is some (unknown) patterning to that “missingness” (Graham, 2009)².

We have not examined missing data in our analysis and we do not know if, or to what extent, some people with particular characteristics may fail to provide responses more than people with different characteristics. We did not use any techniques for dealing with missing data (for example, imputation). Consequently, we cannot know if or how the patterning of missing data impacted on our findings.

Notes for: The Community Life Survey 2016 to 2017 data

1. Strait, DS, Moniz, MA and Strait, PT (1996), ‘Finite mixture coding: a new approach to coding continuous characters’, *Systematic Biology*, Volume 45, Issue 1, pages 67 to 78.
2. Graham, JW (2009), ‘Missing data analysis: Making it work in the real world’, *Annual review of psychology*, Volume 60, pages 549 to 576

3 . Logistic regression

Logistic regression analysis allows for the relationship between an explanatory variable and the outcome variable to be examined, whilst at the same time taking into consideration other explanatory variables that influence the outcome. Logistic regression is used as it is suitable when looking at categorical outcomes (which is the form taken by most of the Community Life Survey (CLS) variables). While it is possible to conduct multinomial logistic regression with multiple categorical outcomes, logistic regression with binary outcomes (for example, “lonely” compared with “not lonely”) was chosen. This was chosen to increase ease of understanding (with the predicted outcomes being either “lonely” or “not lonely”); and for consistency with the LCA.

Procedure

This analysis has been carried out in SAS 9.3. All variables have been treated as categorical variables. The sample size for the logistic regression analysis is 6,414. Backwards logistic regression was used to create the final model. The contribution of each variable is assessed by looking at the significance value of the t-test for each predictor. If there is at least one non-significant variable, the variable with the highest p-value is removed from the model. This procedure is repeated, until the all the remaining variables are significant at the 0.05 level.

There are multiple ways in which variables could be entered in to the model. Forward, backwards and stepwise models were tried and it was found that most of the variables were the same in each case. The backward logistic regression method was used for the final model as it produced a model with the lowest Akaike Information Criterion (AIC); additionally, forward approaches often allow for important variables to be missed due to other variables being entered in to the model first (“suppressor effects”).

Multicollinearity

Many of the variables collected in the Community Life Survey are correlated with one another. Multicollinearity (also known as collinearity) is where one or more explanatory variables in a regression model are highly correlated such that they linearly predict each other with a high degree of accuracy. However, an important assumption of multivariate regression is that explanatory variables are not too highly correlated with one another. Too high a degree of correlation between predictor variables in a regression model can affect the stability and interpretation of the regression estimates.

In the final model, there were a few variables that were correlated, however, their absolute Pearson’s Correlation value was less than 0.5 and the model performs better including these variables and so they have remained in the model. These are disability and health (Pearson’s correlation figure of negative 0.46463), and chatting to neighbours, belonging to the neighbourhood and satisfaction with the local area (Pearson’s correlation figure of 0.31267 for chatting to neighbours and belonging to the neighbourhood, 0.16419 for satisfaction with the local area and chatting to neighbours, and 0.39001 for belonging to the neighbourhood and satisfaction with the local area).

Goodness of fit

Goodness of fit describes how well a model fits the data from which it is generated. It can be used to assess how well the data that the model predicts and corresponds to the data that have been collected. There are various measurements used to assess the model fit. The first two, AIC and Schwarz Criterion (SC) are deviants of negative two times the log-likelihood (-2 Log L). AIC and SC penalize the log-likelihood by the number of predictors in the model. AIC and SC are used for the comparison of non-nested models on the same sample. Ultimately, the model with the smallest AIC and SC are considered the best, although the AIC and SC value itself is not meaningful.

The Likelihood Ratio (LR) Chi-Square test, the Score Chi-Square Test and the Wald Chi-Square Test all test that at least one of the predictors’ regression coefficient is not equal to zero in the model. The Residual Chi-Square Test shows the Chi-Square test statistic, the degrees of freedom (DF) and the associated p-value ($P > \text{ChiSq}$) corresponding to the specific test that all of the predictors are simultaneously equal to zero. A small p-value from all three tests leads to the conclusion that at least one of the regression coefficients in the model is not equal to zero.

Interaction effects

Interactions can be used to test for the joint effect of two or more predictor variables on an outcome variable. It allows us to explore how the relationships between dependent and independent variables differ by context. Some interactions were identified as being significant, however, there is no prior evidence to support the link with loneliness. Some of the interactions appeared to be counter intuitive and did not have a large improvement to the model in terms of improving the AIC. Additionally, adding an interaction term to a model drastically changed the interpretation of all of the coefficients in the model. It was decided, for the purpose of this analysis, to remove interactions for the benefit of identifying individual impacts of each variable.

Causality

Regression analysis can identify relationships between factors; however, it cannot tell us about causality. While, for some factors, causality is fairly clear based on prior knowledge (for example, loneliness does not cause someone to become widowed, however, becoming widowed can cause loneliness), for others the relationship between cause and effect is more blurred (for example, ill health can cause loneliness, but also loneliness can cause ill health). Therefore, where prior knowledge does not make the direction of causality clear it's important to note that causality can operate in either direction (or both).

Weighting

The results of the Community Life Survey are weighted to compensate for unequal selection probabilities and differential non-response (that is, to ensure that the age and sex distribution of the final dataset matches that of the population of England). Our regression models take the weights into account.

Interpretation of the results

The odds ratio is the usual output from logistic regression. The odds ratio for each variable in the model is obtained by exponentiating the estimate. The odds ratio can be interpreted as follows: for a one-unit change in the predictor variable, the odds ratio for a positive outcome is expected to change by the respective coefficient, given the other variables in the model are held constant.

The 95% Wald Confidence Limits are provided for each odds ratio. For a given predictor variable with a level of 95% confidence, that upon repeated trials, 95% of the confidence interval (CIs) would include the "true" population odds ratio. The CI is equivalent to the Chi-Square test statistic: if the CI includes one, the null hypothesis that a particular regression coefficient is equal to zero and the odds ratio is equal to one, given the other predictors are in the model would fail to be rejected. An advantage of a CI is that it is illustrative; it provides information on where the "true" parameter may lie and the precision of the point estimate for the odds ratio.

4 . Latent class analysis

Latent class analysis (LCA) is a statistical technique used to identify sub-groups within a population. Applied to survey data, LCA classifies individuals into groups or "types" based on patterns of characteristics represented as categorical variables. LCA was used in the [loneliness article](#) to group individuals with similar patterns of characteristics including reported experience of loneliness. By employing LCA as reported here, combinations of characteristics that "go with" experience of loneliness are revealed.

Some combinations were found to characterise groups that were more frequently lonely (these factors may be risky in terms of loneliness) whilst other characteristics were found to characterise groups that were less frequently (or never) lonely (these factors may be more protective against loneliness). It is reasonable to think of these characteristics in terms of profiles. Using LCA in this way can aid the identification of groups in the general population who exhibit combinations of characteristics that put them at greater risk of loneliness and others with characteristics more protective in terms of loneliness.

LCA approach taken

The loneliness variable was included within the model along with other variables and then, by adding and taking away variables one-by-one, the aim was to produce a model with good separation (particularly on the loneliness variable). Another method would have been to split our dataset in terms of responses to the loneliness question prior to developing a LCA model. For example, a subset of the data could have been taken to include only those who reported feeling lonely “often/always” and then tested some variables for good separation – this may have produced various groups with different similar characteristics all of which were most frequently lonely. Similarly, a subset of data could have included only those cases in the LCA model who report being less lonely (for example, never).

However, these approaches were not taken for two main reasons. Firstly, use of the full dataset (rather than a subset) allows for better comparisons between people with different characteristics across all variables including the loneliness variable. Secondly, the relatively small sample size would have been reduced further leading to poorer quality results.

Selection of explanatory variables for the final LCA specification

The logistic regression highlighted characteristics that significantly increase or decrease likelihood of loneliness if all other factors are held constant. As a starting point in building the LCA specification, these were used to build LCA models¹. Through trial and error, adding and taking away one variable at a time and re-running the algorithm, a model specification was produced using the variables pertaining to the following:

Loneliness frequency:

- 1 = Often/always, Some of the time; Occasionally
- 2 = Hardly ever; Never

Marital status:

- 1 = single, that is, never married and never registered in a same-sex civil partnership; Separated/divorced
- 2 = Living with partner in a marriage or civil partnership (and not separated)
- 3 = widowed

General health:

- 1= Very good or good
- 2 = Fair
- 3 = Very bad or bad

Housing Tenure:

- 1 = Own outright/buying with mortgage/loan/part buy part rent
- 2 = Renting

Presence or absence of a physical or mental health condition/illness lasting or expected to last 12 months or more:

- 1 = Yes
- 2 = No

Lives alone or does not live alone:

- 1 = Lives alone
- 2 = Does not live alone

Age grouped into three categories:

- 16 to 34
- 35 to 64
- 65 and over

Identifying lonely groups or profiles

LCA is undertaken to produce groups of individuals with different characteristics so that individuals within groups are more similar to each other while, at the same time, distinct from other groups. Table 1 presents figures for the final LCA model.

A model with better separation has less equal distribution between each group in terms of variable categories – in general, values approaching 100% indicate clearer delineation between groups². As our focus was loneliness, it was important that our LCA output showed good separation in terms of the loneliness variable. For example, in Table 1 Group C shows the best separation of all with 85% of individuals reporting “hardly ever” or “never” feeling lonely and 15% who reported feeling lonely “often/always”, “some of the time” or “occasionally”. Of course, a more useful model also provides good separation in terms of other variables included – unequal distributions and deviations from the mean are particularly worth noting because this suggests characteristics that differ from the average and/or other groups.

Based on our data, a deviation from the mean of 15% was chosen for identifying lonely and non-lonely groups. As shown in Table 1, there are four groups that fulfil this criterion: groups A, C, D and E. In the [main loneliness article](#), we only report on these groups because these had distributions of loneliness most different from the mean. For transparency, Table 1 presents all seven groups produced by the LCA model. For the raw LCA data, see Appendix 2.

In the accompanying [loneliness article](#), we refer to:

- Group A as the Widowed older homeowners living alone with long-term health conditions group
- Group C as the Married homeowners in good health living with others group
- Group D as the Unmarried, middle-agers, with long-term health conditions group
- Group E as the Younger renters with little trust and sense of belonging to their group

Optimal number of groups

The LCA process involves running the algorithm with different numbers of groups specified. The analyst first specifies one group, then two groups, then three and so on. With each run a goodness of fit statistic, the Bayes Information Criterion (BIC), is produced. In exploratory LCA, the BIC coefficient is used to identify the optimal number of classes (Lin and Dayton 1997³) and in line with this, the number of groups with the lowest BIC coefficient was chosen as the best model. A model with seven classes was identified to be best – see Appendix 2 for the BIC coefficients of models with one through to eight classes.

In Table 1, groups A, C, D and E show good separation in terms of loneliness. These groups have loneliness responses that differ from the mean proportion of the sample by at least 15% in terms of loneliness. Looking at the whole sample, 46% of people fall into the “more often lonely” category whilst in group A, for example, 69% of people fall into the “more often lonely” category – a much higher proportion than the sample’s average.

Table 1: Groups and characteristics included in the latent class analysis model

England		Group %							Counts
	Category	A	B	C	D	E	F	G	Mean
Loneliness	More often lonely	69	38	15	81	61	57	57	46
	Hardly ever or never lonely	31	62	85	19	39	43	43	54
Marital status	Single, separated or divorced	26	16	2	92	76	95	2	35
	Married or civil partnership	4	83	98	3	24	2	96	61
	Widowed	70	1	1	4	0	3	2	4
General health	Very good or good	48	91	86	16	89	93	7	74
	Fair	41	9	14	51	10	7	66	20
	Very bad or bad	11	0	0	33	1	0	28	6
Tenure	Homeowner	88	86	97	40	22	74	80	72
	Renting	12	14	3	60	78	26	20	28
Long-term health condition	Yes	65	13	37	90	11	15	92	33
	No	35	87	63	10	89	85	8	67
Lives alone	Yes	88	0	2	54	3	90	0	16
	No	12	100	98	46	97	10	100	84
Age group	16-34	0	15	0	18	80	15	3	22
	35-64	4	81	36	71	20	70	56	55
	65+	96	4	64	10	0	16	41	23

Source: 'Community Life Survey 2016 to 2017', Department for Digital, Cultural, Media & Sport

1. Single, separated or divorced includes those who have never married or registered in a same-sex civil partnership, and those who are divorced or separated (but may still be legally married).
2. Homeowner includes those own their home outright, are buying with a mortgage/loan, and part buy, part rent.
3. Renting includes those renting
4. Those who reported living rent free or occupy in any other way are coded as missing due to small cell counts.
5. A long-term health condition means that respondents reported having a "physical or mental health conditions or illnesses lasting or expected to last for 12 months or more"

The groups identified are dependent on the variables included in the LCA model. Had other variables been included then the groups produced would have been different. Unlike some other statistical techniques (for example, logistic regression), variable selection is less automated by the algorithm and more dependent on the choices of the analyst. The absence or presence of a single variable can change whether good separation is achieved or not, and/or how any groups are found optimal. There are practically countless combinations of variables and codes and it is not possible to test them all.

Additional descriptive statistics

In the final LCA algorithm, only the variables and categories as shown in Table 1 were included. In general, with additional variables included in the model there was poorer separation in terms of loneliness across clusters. Good separation in terms of loneliness was the main focus. However, when fewer variables were included, the LCA model became less informative because there was less differentiation in terms of other characteristics, simply because these variables were not included in the model. It is therefore a balance between producing good separation on loneliness and with including more variables that can contribute to, and can be used to describe, the groups. Table 2 presents the characteristics of all seven groups in terms of additional descriptive statistics.

Table 2: Groups and additional characteristics**England**

		Group							Average
		A	B	C	D	E	F	G	
Median age (years)		73	46	70	49	28	53	62	49
Sex (%)	Male	33	49	61	44	40	47	53	48
	Female	67	51	39	56	60	53	47	52
Paid job (%)	Yes	9	82	12	48	76	73	35	61
	No	91	18	88	52	24	27	65	38
Living as a couple? (%)	Yes	5	94	99	15	53	2	98	71
	No	95	6	1	85	47	98	2	29
Life satisfaction (mean)		6.87	7.37	8.26	5.29	6.97	6.98	6.36	7.1
Happiness (mean)		7.07	7.37	8.24	5.13	7	7.07	6.29	7.1
Anxiety (mean)		3.28	3.3	2.25	5.06	3.7	3.19	3.98	3.44
Worthwhile (mean)		6.94	7.61	8.21	5.55	7.11	7.2	6.59	7.28
Economic status (%)	Employed	10	82	15	49	76	75	37	62
	Unemployed	0	2	0	5	4	3	2	2
	Inactive	90	16	85	46	20	22	61	35
Limiting long-term health condition (%)	Yes	53	7	16	79	5	7	79	22
	No	47	93	84	21	95	93	21	78
Neighbourhood strength of belonging (%)	More strongly	67	67	79	47	45	53	66	62
	Less strongly	33	33	21	53	55	47	34	38
Trust in people living in neighbourhood (%)	"Many people can be trusted"	61	50	68	27	25	41	41	45
	"Some can be trusted"	23	32	22	33	37	36	34	32
	"A few can be trusted"	16	16	9	29	31	19	22	19
	"None can be trusted"	0	2	0	11	7	4	2	3
English Index of Multiple Deprivation 2015 (LSOA) (%)	Bottom 50%	38	43	27	69	71	53	44	49
	Top 50%	62	57	73	31	29	47	56	51

Source: 'Community Life Survey 2016 to 2017', Department for Digital, Cultural, Media & Sport

Notes for: Latent class analysis

1. However, the variables that were tested in the LCA model were not restricted only to these variables. It is important to keep in mind that variables which are not significant may still contribute to good separation and so produce meaningful groups.
2. Celeux and Soromenho (1996), 'An entropy criterion for assessing the number of clusters in a mixture model'.
3. Lin TH and Dayton CM (1997), 'Model selection information criteria for non-nested latent class models', Journal of Educational and Behavioral Statistics, Volume 22, Issue 3, pages 249 to 264.

5 . Appendix 1: Logistic regression – statistical explanations and tables

Initial list of variables considered:

- Mode of Interview
- Age group
- Sex
- Ethnicity
- Relationship status
- Income
- Urban or rural classification
- Region
- Housing tenure
- Disability
- General health
- Education
- Digital skills
- Employment status
- Number of adults
- Number of children
- Volunteering
- Caring responsibilities
- Agree people in neighbourhood pull together
- Whether chat to neighbours more than just to say Hello
- Trust people in neighbourhood

- Belong to neighbourhood
- Religion (even if not practicing)
- Satisfaction with local area as a place to live
- Has area got better or worse in last two years
- Years lived in neighbourhood
- Number of services and amenities in local area
- Index of Multiple Deprivation
- National Statistics Socio-economic Classification (NS-SEC)
- This local area is a place where people from different backgrounds get on well together?
- How often meet up in person with family members or friends
- How often speak on the phone or video or audio call via the internet with family members or friends
- How often email or write to family members or friends
- How often exchange text messages or instant messages with family members or friends

Variables removed as not being significant predictors on their own:

- Religion was removed as it is not correlated with loneliness using the Pearson product-moment correlations. The correlations range from negative 1 to positive 1, and the Pearson product-moment correlation between religion and loneliness is 0.00148 (p equals 0.8827).

Variables removed as not being significant predictors when part of a regression model:

- Mode of interview
- Ethnicity
- Urban or rural classification
- Region
- Housing tenure
- Education
- Digital skills
- Employment status
- Number of children
- Volunteering
- Agree people in neighbourhood pull together
- Trust people in neighbourhood
- Has area got better or worse in last two years
- Number of services and amenities in local area
- Index of Multiple Deprivation
- NS-SEC
- This local area is a place where people from different backgrounds get on well together?
- How often speak on the phone or video or audio call via the internet with family members or friends
- How often email or write to family members or friends
- How often exchange text messages or instant messages with family members or friends

Final model

The SURVEYLOGISTIC Procedure

Table 3: Logistic Regression output - Response Profile

Ordered Value	Lonely	Total Frequency	Total Weight
1	0	3339	16194897
2	1	3075	13721262

Source: 'Community Life Survey 2016 to 2017', Department for Digital, Cultural, Media & Sport

Table 4: Logistic Regression output: Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	5333515.28	37	<.0001
Score	4916532.05	37	<.0001
Wald	658.0485	37	<.0001

Source: 'Community Life Survey 2016 to 2017', Department for Digital, Cultural, Media & Sport

Table 5: Logistic Regression output: Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Rage9_recode	6	52.1036	<.0001
Sex	1	48.9797	<.0001
MarStatg2	3	46.7504	<.0001
ZIncomhh1	2	16.7081	0.0002
dill2	1	27.6347	<.0001
ghealth2	1	48.9124	<.0001
nadults	4	46.9456	<.0001
RCare	1	10.643	0.0011
chat2neigh	1	9.0925	0.0026
SBeNeigh	3	17.8962	0.0005
Slocsat	4	31.3641	<.0001
yearsLived	5	17.1413	0.0042
meetupnew	5	33.3145	<.0001

Source: 'Community Life Survey 2016 to 2017', Department for Digital, Cultural, Media & Sport

Table 6: Logistic Regression output: Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	0.4182	0.1081	14.9774	0.0001
Rage9_recode		2	1 0.418	0.0921	20.5841	<.0001
Rage9_recode		3	1 0.0972	0.0817	1.4147	0.2343
Rage9_recode		4	1 0.0374	0.0837	0.1994	0.6552
Rage9_recode		5	1 0.0514	0.0874	0.3457	0.5566
Rage9_recode		6	1 -0.5309	0.0912	33.8912	<.0001
Rage9_recode		7	1 -0.5271	0.1412	13.9349	0.0002
Sex	Female	1	0.2425	0.0347	48.9797	<.0001
MarStatg2		1	1 -0.4894	0.0772	40.1797	<.0001
MarStatg2		2	1 -0.1122	0.0985	1.2971	0.2547
MarStatg2		3	1 0.7944	0.14	32.1764	<.0001
ZIncomhh1		1	1 -0.0865	0.0465	3.4509	0.0632
ZIncomhh1		2	1 -0.161	0.0612	6.9128	0.0086
dill2		0	1 0.2224	0.0423	27.6347	<.0001
ghealth2		1	0.6291	0.0899	48.9124	<.0001
nadults		2	1 -0.1246	0.0711	3.0671	0.0799
nadults		3	1 -0.1122	0.0868	1.6695	0.1963
nadults		4	1 -0.074	0.0911	0.6603	0.4165
nadults		5	1 -0.3339	0.1736	3.6988	0.0545
RCare	Yes	1	0.1573	0.0482	10.643	0.0011
chat2neigh		2	1 0.1798	0.0596	9.0925	0.0026
SBeNeigh	Fairly strongly	1	0.0165	0.0569	0.0842	0.7717
SBeNeigh	Not at all strongly	1	0.1541	0.0963	2.5626	0.1094
SBeNeigh	Not very strongly	1	0.1447	0.0598	5.8572	0.0155
Slocsats	Fairly dissatisfied	1	0.1061	0.1213	0.7644	0.382
Slocsats	Fairly satisfied	1	-0.0654	0.0724	0.815	0.3667
Slocsats	Neither satisfied nor dissatisfied	1	0.2445	0.0957	6.5244	0.0106
Slocsats	Very dissatisfied	1	0.0968	0.1885	0.264	0.6074
yearsLived		2	1 0.0255	0.0843	0.0917	0.762
yearsLived		3	1 -0.0169	0.0909	0.0345	0.8527
yearsLived		4	1 -0.1321	0.0898	2.1642	0.1413
yearsLived		5	1 0.1558	0.0926	2.8314	0.0924
yearsLived		6	1 -0.2356	0.0924	6.4985	0.0108
meetupnew		1	1 -0.2097	0.0636	10.8763	0.001
meetupnew		2	1 -0.1702	0.0677	6.3285	0.0119
meetupnew		3	1 -0.0139	0.0923	0.0227	0.8802

meetupnew	4	1	0.1778	0.1005	3.1281	0.077
meetupnew	5	1	0.413	0.0952	18.8238	<.0001

Source: 'Community Life Survey 2016 to 2017', Department for Digital, Cultural, Media & Sport

Table 7: Logistic Regression Output: Odds Ratio Estimates

Effect	Point Estimate	95% Wald	Confidence Limits
Rage9_recode 2.00 vs 1.00	0.964	0.72	1.292
Rage9_recode 3.00 vs 1.00	0.7	0.516	0.95
Rage9_recode 4.00 vs 1.00	0.659	0.48	0.905
Rage9_recode 5.00 vs 1.00	0.668	0.476	0.939
Rage9_recode 6.00 vs 1.00	0.373	0.262	0.533
Rage9_recode 7.00 vs 1.00	0.375	0.242	0.58
Sex Female vs Male	1.624	1.418	1.861
MarStatg2 1.00 vs 0.00	0.743	0.61	0.906
MarStatg2 2.00 vs 0.00	1.084	0.821	1.431
MarStatg2 3.00 vs 0.00	2.683	1.814	3.969
ZIncomhh1 1.00 vs 0.00	0.716	0.601	0.853
ZIncomhh1 2.00 vs 0.00	0.665	0.533	0.829
dill2 0.00 vs 1.00	1.56	1.322	1.841
ghealth2	1.876	1.573	2.237
nadults 2.00 vs 1.00	0.463	0.368	0.584
nadults 3.00 vs 1.00	0.469	0.358	0.615
nadults 4.00 vs 1.00	0.487	0.367	0.648
nadults 5.00 vs 1.00	0.376	0.235	0.602
RCare Yes vs No	1.37	1.134	1.655
chat2neigh 2.00 vs 1.00	1.433	1.134	1.81
SBeNeigh Fairly strongly vs Very strongly	1.393	1.14	1.703
SBeNeigh Not at all strongly vs Very strongly	1.599	1.18	2.166
SBeNeigh Not very strongly vs Very strongly	1.584	1.273	1.971
Slocsat Fairly dissatisfied vs Very satisfied	1.629	1.196	2.218
Slocsat Fairly satisfied vs Very satisfied	1.372	1.169	1.611
Slocsat Neither satisfied nor dissatisfied vs Very satisfied	1.871	1.476	2.372
Slocsat Very dissatisfied vs Very satisfied	1.614	0.999	2.608
yearsLived 2.00 vs 1.00	0.837	0.676	1.037
yearsLived 3.00 vs 1.00	0.802	0.632	1.019
yearsLived 4.00 vs 1.00	0.715	0.564	0.907
yearsLived 5.00 vs 1.00	0.954	0.744	1.223
yearsLived 6.00 vs 1.00	0.645	0.501	0.83
meetupnew 1 vs 0	0.987	0.789	1.236
meetupnew 2 vs 0	1.027	0.813	1.298
meetupnew 3 vs 0	1.201	0.907	1.59
meetupnew 4 vs 0	1.455	1.082	1.955

Source: 'Community Life Survey 2016 to 2017', Department for Digital, Cultural, Media & Sport

6 . Appendix 2: Latent class analysis R output

Table 8: LCA output: Loneliness frequency

Group	Often to occasionally	Hardly ever or never
A	0.6937	0.3063
B	0.3849	0.6151
C	0.1528	0.8472
D	0.811	0.189
E	0.6133	0.3867
F	0.5669	0.4331
G	0.5706	0.4294

Source: 'Community Life Survey 2016 to 2017', Department for Digital, Cultural, Media & Sport

Notes:

1. Often to occasionally includes the following responses: "Often/always", "Some of the time" and "Occasionally".

Table 9: LCA output: Marital status

Group	Single, separated or divorced	Married or civil partnership	Widowed
A	0.261	0.0425	0.6965
B	0.1619	0.8322	0.006
C	0.0166	0.9751	0.0083
D	0.9233	0.0336	0.0431
E	0.7612	0.2369	0.0018
F	0.9455	0.0236	0.0309
G	0.0197	0.9554	0.0249

Source: 'Community Life Survey 2016 to 2017', Department for Digital, Cultural, Media & Sport

Notes:

1. Single includes those never married or registered in a civil partnership
2. Separated includes those still legally married

Table 10: LCA output: General health

Group	Very good or good	Fair	Very bad or bad
A	0.4817	0.413	0.1054
B	0.9084	0.0916	0
C	0.859	0.141	0
D	0.1607	0.5079	0.3314
E	0.8853	0.1039	0.0108
F	0.9277	0.0723	0
G	0.0653	0.6551	0.2796

Source: 'Community Life Survey 2016 to 2017', Department for Digital, Cultural, Media & Sport

Table 11: LCA output: Tenure

Group	Homeowner	Renting
A	0.8783	0.1217
B	0.8591	0.1409
C	0.9659	0.0341
D	0.4024	0.5976
E	0.2176	0.7824
F	0.7445	0.2555
G	0.8024	0.1976

Source: 'Community Life Survey 2016 to 2017', Department for Digital, Cultural, Media & Sport

1. Homeowner includes own outright, buying with mortgage or loan, and part buy/part rent

Table 12: LCA output: Long-term physical/mental health condition

Group	Long-term physical /mental health condition	No long-term physical /mental health condition
A	0.6461	0.3539
B	0.1297	0.8703
C	0.374	0.626
D	0.8961	0.1039
E	0.113	0.887
F	0.1478	0.8522
G	0.924	0.076

Source: 'Community Life Survey 2016 to 2017', Department for Digital, Cultural, Media & Sport

Table 13: LCA output: Lives alone or with others

Group	Lives alone	Lives with others
A	0.8794	0.1206
B	0	1
C	0.0219	0.9781
D	0.5353	0.4647
E	0.0302	0.9698
F	0.8959	0.1041
G	0	1

Source: 'Community Life Survey 2016 to 2017', Department for Digital, Cultural, Media & Sport

Table 14: LCA output: Age Group

Group	16 to 34	35 to 64	65+
A	0	0.0354	0.9646
B	0.1503	0.8091	0.0406
C	0	0.3575	0.6425
D	0.1834	0.7127	0.1038
E	0.8017	0.1983	0
F	0.1467	0.6974	0.1559
G	0.0306	0.5608	0.4087

Source: 'Community Life Survey 2016 to 2017', Department for Digital, Cultural, Media & Sport

Table 15: Bayes Information Criterion coefficients for models with 1 through to 8 classes

Number of classes	Bayes Information Criterion (BIC)
1	60013.12
2	57058.24
3	55265.58
4	54367.77
5	53991.97
6	53708.09
7	53609.42
8	53617.31

Source: 'Community Life Survey 2016 to 2017', Department for Digital, Cultural, Media & Sport

