Ministry of Housing, Communities & Local Government

English Housing Survey
Methodology Paper

Findings from the 2014-15 Surveyor Variability Study

50th Anniversary
ENGLISH HOUSING SURVEY
1967 - 2017

Contents

Chapter 1: Introduction and main findings 3

Chapter 2: Methodology 6

Chapter 3: Results 11
- Chance agreement 11
- Correlated surveyor variance 14
- Combined analysis taking Kappa Scores and CSVs into account 16
- Bias adjustments, standard errors and confidence intervals 26
- Comparison of surveyor variability in 2009-10 and 2014-15 29

Annex 1: Detailed results and cross-tabulations 32

Annex 2: Adjustment to original complex standard errors 34

Chapter 1
Introduction and main findings

1.1 The English Housing Survey (EHS) is a national survey of people's housing circumstances and the condition and energy efficiency of housing in England. It is commissioned by the Ministry of Housing, Communities and Local Government (MHCLG) and run by a consortium of contractors led by NatCen Social Research. From time to time, MHCLG commissions work to improve EHS data collection and methodology. This report gives the findings from one such piece of work: a surveyor variability study (SVS) conducted using the 2014-15 EHS to assess the effects of surveyor variability on the precision of estimates from the EHS physical survey.

1.2 Like all estimates based on surveys, the findings of the EHS are subject to various sources of error. This is because survey findings are based on a sample of a population rather than the whole population. The total error in a survey estimate is the difference between the estimate derived from the data collected and the (unknown) true value for the population. The main sources of error are random error, measurement error, and systematic error. Information on these types of error can be found in the EHS Quality Report[1].

1.3 The EHS has two components: an interview undertaken with approximately 13,300 households each year and a physical survey of a random sample of the dwellings of about 6,000 of those households as well as about 200 vacant dwellings. This report into surveyor variability focuses on a particular form of measurement error, the error relating to the assessments made by surveyors who carried out the physical surveys for the 2014-15 EHS. Experience has shown that surveyor variability cannot be completely eliminated or even reduced to an insignificant level, but precautions are taken on the EHS to control its impact. These precautions include:

- using a large number of surveyors and setting limits of 65 surveys per year in total by any one surveyor. In addition, 90% of the surveyors carry out no more than 45 surveys in any one region (outside of North East England, where surveyors carry out no more than 35 surveys).
- Providing surveyors with a rigorous and uniform training in the form of: a five day face-to-face briefing designed to minimise subjectivity; survey manuals; supervision in the field; refresher briefings; and desk-based exercises designed to calibrate the assessments made by the surveyors.

1.4 Despite taking these precautions, it is natural that a degree of personal judgement and subjectivity will still affect surveyors' assessments. For

[1] https://www.gov.uk/government/publications/english-housing-survey-quality-report

example, although the EHS approach to assessing hazards in the home using the Housing Health and Safety Rating System (HHSRS) provides surveyors with a systematic approach with which to make their judgements some surveyor variability is to be expected in the assessments of the potential severity of hazards in the home. This between-surveyor variability introduces an additional source of variance to the estimates of the EHS physical survey data. The additional variance can be measured by carrying out statistical analyses of the data collected by the surveyors.

1.5 A surveyor variability study (SVS) was conducted in 2014-15 to assess the effects of surveyor variability on the precision of estimates from the EHS physical survey. Fieldwork took place between April and September, 2014. The initial 'parent' surveys were carried out in waves 1 and 3 of the 2014-15 survey. The repeat 'child' surveys were carried out one wave later, in waves 2 and 4 respectively. This study was a repeat of a similar exercise carried out on the EHS in 2009-10[2] and on the English House Condition Survey (EHCS) in 2003-04[3].

1.6 The 2014-15 study involved a call-back exercise in which 303 dwellings were re-surveyed by a second surveyor and the results were compared. The objectives of this study were to:

- compare variability between surveyors;
- highlight key survey measures on the EHS which were subject to high variability (low levels of agreement), or low variability (high levels of agreement);
- produce evidence to improve training of surveyors and to improve physical survey form design;
- assess impact of surveyor variability on complex standard errors (and confidence intervals); and
- compare the levels of variability in the EHS over time.

Main findings

1.7 The 2014-15 study found that **overall, there was a high level of agreement between surveyors' assessments of the dwellings:** 85% (61 of the 73 survey measures included in the study had a high or acceptable level of agreement after allowing for chance agreement.

[2] See here for details on the 2009-10 study: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/211310/Surveyor_variability.pdf.
[3] See here for details on the 2003-04 study: http://webarchive.nationalarchives.gov.uk/20120919132719/www.communities.gov.uk/publications/housing/ehsttechnicalreport2007.

1.8 Four survey measures were identified as having potentially low levels of agreement and seven measures with an acceptable level of agreement were also flagged. The 11 survey measures found to have a low level of agreement tended to be items related to the external environments (six measures) or the amenities and services aspects of the dwelling (four measures); one related to the condition of the stock. They typically required a surveyor to exercise a subjective judgement, for example give an opinion of the local area or the condition of the dwelling.

1.9 The overall level of surveyor variability found in the 2014-15 study was in line with the results from the 2009-10 study. We therefore conclude that **variability has remained consistent across time**.

Acknowledgements and further queries

1.10 Each year the English Housing Survey relies on the contributions of a large number of people and organisations. The Ministry of Housing, Communities and Local Government (MHCLG) would particularly like to thank the following people and organisations, without whom the 2014-15 Surveyor Variability Study and this report, would not have been possible: all the households who gave up their time to take part in the survey, the surveyors responsible for carrying out the surveys, NatCen Social Research, the Building Research Establishment (BRE) and CADS Housing Surveys.

1.11 This report was produced by Klaudia Lubian, NatCen Social Research and MHCLG.

1.12 If you have any queries about this report, would like any further information or have suggestions for analyses you would like to see included in future EHS reports, please contact ehs@communities.gov.uk.

1.13 The responsible analyst for this report is: Stephanie Freeth, Housing and Planning Analysis Division, MHCLG. Contact via ehs@communities.gov.uk

Chapter 2
Methodology

2.1 The 2014-15 Surveyor Variability Study (SVS) was the third study of this kind on the EHS. It adopted a 'call back design' similar to those utilised in the previous two studies. This design essentially involved arranging for a second surveyor to revisit a subset of the originally sampled dwellings selected at random from the first two quarters of EHS 2014-15. Analysis of variability was conducted by comparing selected survey measures from the dwellings in the original sample (parent surveys) with the corresponding measures from the surveys carried out during the revisits (child surveys). In total, the analysis compared survey measures from 303 dwellings.

2.2 In line with the previous variability studies, the 2014-15 study focussed on 73 key survey measures included in the physical survey component of the EHS.

2.3 Variability between surveyors was assessed using four approaches to give a holistic view of surveyor variability. The approaches were:

- percentage level of observed agreement;
- chance agreement measured using Kappa Scores;
- correlated surveyor variance; and
- bias adjustments calculated to take account of the impact of the correlated surveyor variance on the existing EHS complex standard errors.

Percentage level of observed agreement

2.4 The first approach assessed the percentage of cases where both the parent survey and child survey were in agreement with one another. This was useful to get a general view of the level of observed agreement for each survey measure. It is calculated by:

$$P_0 = \frac{N_a}{N}$$

where P_0 = observed agreement, N_a = agreement between surveyors, and N = total number of properties.

2.5 This approach does not take account of the level of agreement that could be expected by chance (i.e. if both surveyors guessed their assessment, it is possible they would both agree by chance).

Chance agreement measured using Kappa Scores

2.6 The second approach, the Kappa Score, allows us to produce a statistic identifying the level of agreement which takes into account chance agreement. It is the ratio of the difference between the observed and the expected agreement, to the maximum possible agreement. It is calculated by:

$$K = \frac{P_o - P_e}{1 - P_e}$$

Where K = kappa coefficient, P_o = observed count of agreement; and P_e = expected count of agreement.

2.7 The Kappa coefficient can have any value between -1 and +1:

- +1 indicates the theoretical limit of maximum agreement between surveyors
- 0 indicates agreement at a chance level
- -1 indicates agreement smaller than chance (although a value less than 0 is unlikely)

2.8 Kappa Scores can be grouped to provide insight into the strength of agreement, Table 2.1.

Table 2.1: Groupings of Kappa Scores for indicating the strength of agreement

strength of agreement	value of Kappa Score
poor	<0.00
slight	0.00 - 0.20
fair	0.21 - 0.40
moderate	0.41 - 0.60
substantial	0.61 - 0.80
almost perfect	0.81 - 1.00

Source: Landis, J. and Koch, G. (1977) "The measurement of observer agreement for categorical data" in Biometrics. Vol. 33, pp. 159–174.

2.9 It should be noted that Kappa is an index that considers observed agreement with respect to a baseline agreement. The baseline is frequently described as

the agreement due to chance, and it can be questioned whether it is relevant for the particular research question. Moreover, Kappa = 0 when the observed allocation is apparently random, regardless of the quantity of disagreement as constrained by the marginal totals. Therefore, it can be claimed that the Kappa Score does not take into account the degree of disagreement between surveyors.

2.10 In addition, it is known that, for some of the measures, a very low Kappa Score may not be a reliable indicator of high variability. This is related to the chi-square statistic used to compute the Kappa Score. As a measure based on the chi-square statistic, the Kappa Score fails to convey meaningful information when most of the responses fall into one response category thus creating some categories with very few responses. In a cross-tabulation of child and parent cases, when a significant proportion of cells have expected count of less than 5, measures based on chi-square statistics (including the Kappa Score) are not reliable measures of agreement. To judge the absolute effect of the impact of surveyor variability, a Kappa Score needs to be looked at alongside the 'raw' percentage of observed agreement produced by cross-tabulating the parent and child cases. A low raw percentage of agreement combined with a low Kappa score indicates that surveyors are disagreeing over a substantial number of marginal cases.

Correlated surveyor variance

2.11 The third approach, the correlated surveyor variance (CSV), refers to the tendency of an individual surveyor to make assessments that are consistent with the other assessments he/she has made but different from the average assessment of all surveyors. There is the possibility that a surveyor is more likely to assess a dwelling towards one response consistently, when compared with other surveyors. In order to calculate CSVs, multi-level (ML) modelling was conducted multiple times with each individual response category of the key survey measures investigated in this study as a dependent variable. [4]

2.12 ML models are an extension of standard regression models, allowing for analysis of hierarchically structured data. Modelling of each response category resulted in estimation of one of the assessments' total variance: variance due to surveyors. This was done by regressing a binary variable associated with the response categories indicating whether the response category was selected or not with an indicator of each dwelling (which paired up the parent and child cases) and by specifying random effects at the

[4] The response categories "does not apply", "no answer" and "section not applicable" were not included in the calculation of CSVs. Also, response categories with no observed counts were also excluded. Finally, CSVs were not produced for response categories where less than 12 observations were found. The reason for this is that the estimate for correlated surveyor variance is already subject to a degree of variability, and any cases with less than 12 observations were felt not to be reliable for inclusion in this analysis.

surveyor level (level 3) and household level (level 2). The correlated surveyor variance is calculated by dividing the estimate of the variance due to surveyor by the total variance. The latter is computed using an analysis of variance (ANOVA) for each individual response category. ML modelling for this study was carried out in Stata using the runmlwin[5] command. Further information on the model used in this analysis can be obtained by contacting the EHS team at MHCLG (ehs@communities.gov.uk).

2.13 The process for calculating CSV described here matches the process adopted in the 2009-10 SVS but a different statistical package was used in 2009-10 (SAS). We replicate the results from 2009-10 on a few selected outcomes to assure us that we would be able to conduct a comparable analysis using Stata.

2.14 In terms of interpreting the CSV:

- a score of zero indicates no variability (perfect agreement); and
- a score of 1 indicates the theoretical limit of high variability (low agreement).

2.15 For the purpose of this report, and in keeping with previous studies, any CSV greater than or equal to 0.1 will be considered as having substantial levels of variability (low agreement). It should be noted that these CSVs do not provide a measure of the extent of any disagreement between surveyors.

Bias adjustments and revised standard errors

2.16 Complex standard errors (CSE) were calculated on the combined 2013-14 and 2014-15 EHS data to take account of the complex design of the survey. However, these only partly reflect the effect of between-surveyor variability. Based on the assumption that we anticipate some correlation between individual observations from the same surveyor, there is a chance that the complex standard errors produced are underestimated. Therefore, bias adjustments were calculated to take into account the impact of CSVs on the existing EHS complex standard errors. Bias adjustments were produced for variables that were found to be particularly prone to surveyor error[6] and a few not affected by it to provide a benchmark. The bias adjustment was calculated for all response categories of the variables presented in Appendix 5.3[7], using the following formula:

[5] http://www.bristol.ac.uk/cmm/software/runmlwin/
[6] Variables usually used for classification in EHS reports, such as nature of the local area, number of dwellings in the local area, visual quality of the local area and poor quality environment were excluded from the bias adjustment analysis.
[7] Appendix 5.3 is provided as a separate excel document.

$$bias\ adjustment = a * \overline{b^2} * (1 + c^2) * CSV * v$$

where a is number of surveyors working in 2013-14 and 2014-15 EHS, $\bar{b}$ is the average proportion of the sample allocated to each surveyor, c is the coefficient of variation of these proportions, CSV is the estimated correlated surveyor variance and v is the total variance measured from the study. This formula replicates the process adopted in 2009-10. A bias adjustment was calculated based on the number of surveyors (157) who worked on the EHS in 2013-14 and 2014-15, and the total number of all core cases with a physical survey (12,297). The constant factor $a * \overline{b^2} * (1 + c^2)$ used in the formula was 0.007.

2.17 In order to add our bias adjustment to produce a revised standard error taking surveyor variability into account, several stages were required.

- Converting complex standard errors (CSE) into variance.
- Converting bias adjustment from a proportion estimate to a percentage estimate. This is because CSE is based on percentage estimates, whilst the SVS bias adjustment is based on proportions. This is achieved by multiplying the bias adjustment by 1002.
- Adding our variance from the EHS standard error to our revised bias adjustment.
- Turning this new variance into our revised standard error (RSE).
- Applying RSE in the calculation of revised confidence interval to enable comparison of the impact of CSV on precision of the estimates.

2.18 We need to remember that this revised estimate is subject to a degree of variance itself, as well as other confounding factors which may not have been captured in this analysis. Therefore, users should treat these revised confidence intervals with the appropriate level of caution.

Chapter 3
Results

3.1 The results from this study have been split into the following sections of analysis:

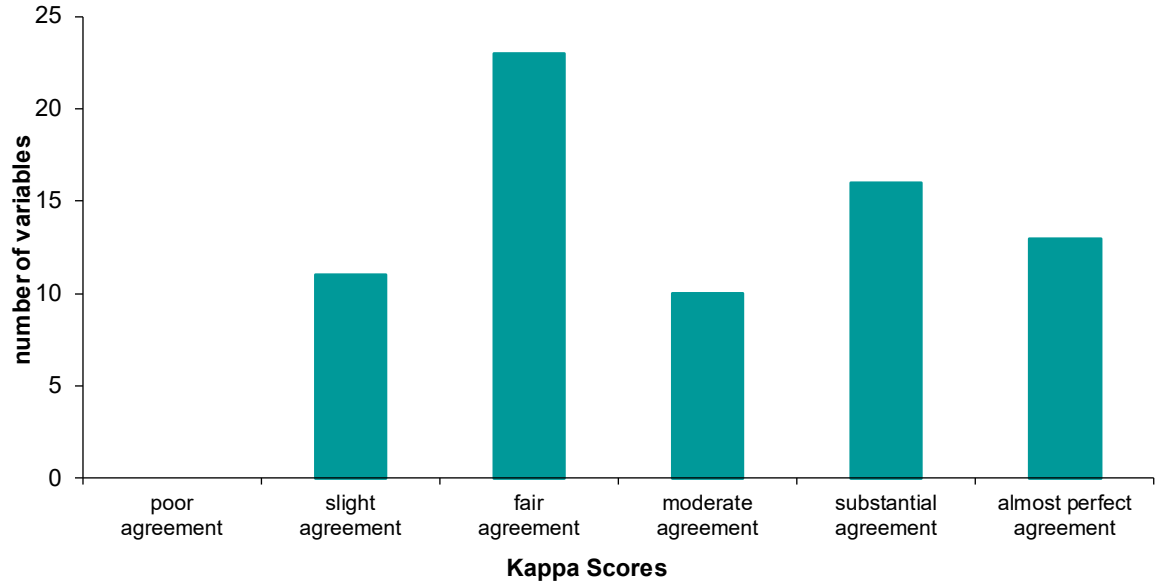
- chance agreement (explored using Kappa Scores and raw percentage of agreement);
- correlated surveyor variance;
- combined analysis (taking Kappa and correlated surveyor variance into account);
- bias adjustments, standard errors and confidence intervals; and
- comparison of surveyor variability in EHS across time.

3.2 Further data and results can be found in the annex tables published alongside this report, including: spreadsheets listing all the variables included in this study; the key statistics referred to in this section of the report; and cross-tabulations of observations from parent and child surveys for all 73 measures included in the study, Annex 2.

Chance agreement

3.3 Kappa Scores were produced for the 73 key survey measures under investigation in the report to show the level of agreement observed between surveyors, Figure 3.1. Scores with poor or slight agreement are collectively defined as having ‘low agreement’, Table 2.1.

Figure 3.1: Distribution of Kappa scores of the 73 survey measures



Note: one measure has a Kappa Score of 1 indicating a perfect agreement

3.4 The top ten survey measures with high variability/low agreement (based on Kappa Scores only) are those describing: wall structure faults; wall structure urgent repairs; Decent Homes repair criterion; damp problems; poor quality environment; number of dwellings in the area; foundation settlement; differential movement; action required on the boiler; and structural defects. It is known that for some of the measures a very low Kappa Score may not be a reliable indicator of high variability. This is related to the chi-square statistic used to compute the Kappa Score. As a measure based on the chi-square statistic, the Kappa Score fails to convey meaningful information when most of the responses fall into one response category thus creating some categories with very few responses. In a cross-tabulation of child and parent cases, when a significant proportion of cells have expected count of less than 5, measures based on chi-square statistics (including the Kappa Score) are not reliable measures of agreement. To judge the absolute effect of the impact of surveyor variability, the Kappa needs to be looked at alongside the 'raw' percentage of agreement produced by cross-tabulating the parent and child cases. A low raw percentage of agreement combined with a low Kappa score indicates that surveyors are disagreeing over a substantial number of marginal cases. Taking account of both the Kappa scores and raw percentage of agreement, five survey measures can be considered to have potentially high variability, they are: number of dwellings in the area; defects in foundation settlement; defects in differential movement; action required on the boiler; and structural defects, Table 3.1.

Table 3.1: Top 10 survey measures with high variability/low agreement based on the Kappa Score

survey measure	EHS variable associated with measure	Kappa Score	extent of the variability taking account of the Kappa Score and raw percentage of observed agreement
faults – wall structure	Fexwsfl	0.037	High level of observed agreement (92.1%) but Kappa Score indicates potentially high variability.
urgent repair – wall structure	Fexwsur	0.044	High level of observed agreement (92.1%). 24 out of 303 dwellings had some disagreement but Kappa Score indicates potentially high variability.
decent homes repair criterion	dhdisrx	0.058	High level of observed agreement (93.4%) but Kappa Score indicates potentially high variability.
damp – problem present	Dampalf	0.120	High level of observed agreement (92.7%) but Kappa Score indicates potentially high variability.
poor quality environment – utilisation problems	lv3utilx	0.123	Very high level of observed agreement (96%) but Kappa Score indicates potentially high variability.
number of dwellings in area	Fardwell	0.124	Observed agreement of 31%. 209 out of 303 dwellings had some level of disagreement. Therefore variability is potentially high.
foundation settlement – defect	Fstfoude	0.185	Observed level of agreement of 88.4%. 35 out of 303 dwellings had some kind of disagreement. Therefore variability is potentially high.
differential movement – defect	Fstmovde	0.186	Observed level of agreement of 88.4%. 35 out of 300 dwellings had some kind of disagreement. Therefore variability is potentially high.
boiler – action required	Finchbac	0.192	Very high level of observed agreement (95%). Potentially high variability given low prevalence of action required.
structural defects	Fstpres	0.200	Observed level of agreement of 88.8%. 34 out of 300 dwellings had some kind of disagreement. Therefore variability is potentially high.

3.5 The top ten survey measures with low variability / high agreement (based on Kappa scores) are variables describing: the dwelling type (in details as well as distinguishing whether the dwelling is a house or a flat); number of floors above ground; main fuel type; main heating type; predominant type of wall finish; presence of mains gas supply; main heating system; tenure; and presence of an attic. The raw percentage of agreement of those measures are also high, Table 3.2.

Table 3.2: Top 10 survey measures with low variability/high agreement based on the Kappa Score

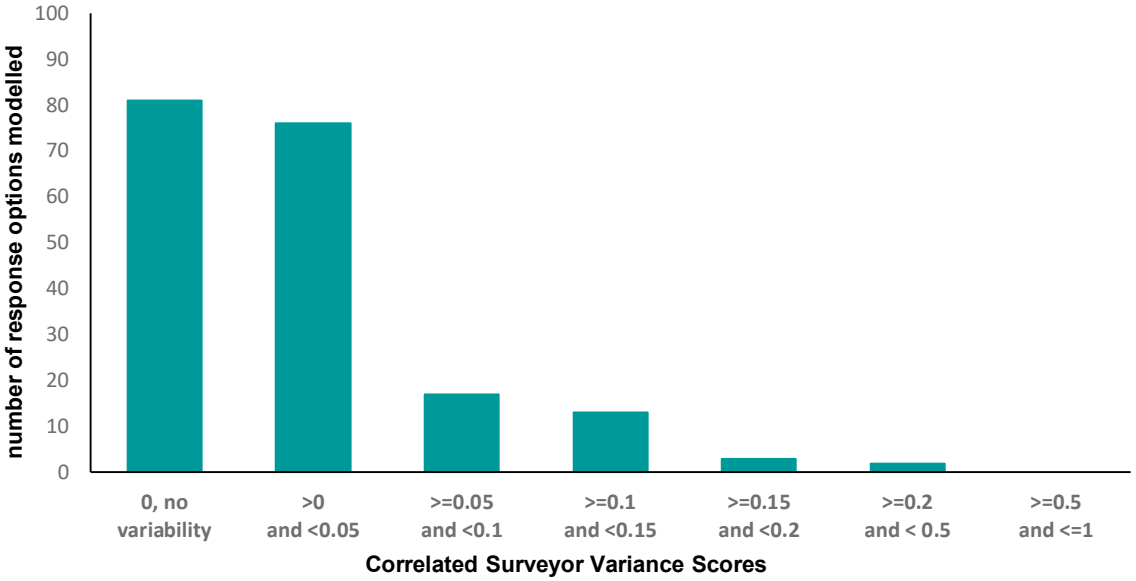
survey measure	EHS variable associated with measure	Kappa Score	extent of the variability taking into account Kappa score and raw percentage of observed agreement
dwelling type distinguishing a house from a flat	Foddtype	1.000	Perfect level of agreement (100% observed agreement)
detailed dwelling type	dwtypenx	0.988	Near perfect level of agreement (99% observed agreement).
number of floors above ground in the house or module	Storeyx	0.975	Near perfect level of agreement (98.7% observed agreement).
main fuel type	Fuelx	0.928	Near perfect level of agreement (98.3% observed agreement).
main heating fuel	Finmhfue	0.916	Near perfect level of agreement (98% observed agreement).
predominant type of wall finish	typewfin	0.915	Very high level of agreement (96.4% observed agreement).
mains gas supply present	Fingasms	0.904	Near perfect level of agreement (98% observed agreement).
main heating system	heat4x	0.903	Near perfect level of agreement (98.7% observed agreement).
tenure	Fodtenur	0.892	Very high level of agreement (92.1% observed agreement)
attic present in dwelling	Attic	0.889	Near perfect level of agreement (98.3% observed agreement).

Correlated surveyor variance

3.6 Correlated surveyor variance (CSV) was calculated for all the response categories of the 73 survey measures that had at least 12 observations. Thirty-six categories had fewer than 12 observations and these were excluded from the analysis. In total, 192 estimates of CSV were obtained from the analysis: 81 response categories had a CSV of 0, indicating no variability; 93 had a CSV of less than 0.1, with the remaining 18 having a CSV of greater than 0.1, Figure 3.2. Estimates of the rest of the categories (15) could not be

estimated due to computational issues related to the random parameters in the multilevel modelling[8].

Figure 3.2: Distribution of Correlated Surveyor Variance Scores



3.7 The 18 response options with high CSVs (greater than 0.1) included: number of dwellings in area; kitchen last refurbished in the 2000s; bathroom last refurbished in 2000s; bathroom last refurbished in the 1980s; windows and doors not fully secure; windows and doors secure; area was rural residential; area was suburban residential; area was 'other urban centre'; no problems with poor quality environment; problems with poor quality environment; and no urgent repair to wall finish needed.

3.8 Regional analysis has been conducted on response categories which had a CSV equal or higher than 0.1, to see if the variability observed happened only in specific regions. Three regional groupings have been used in this report, which have been derived using the following Government Office Regions:

- North (formed by combining North East, North West, Yorkshire and the Humber Government Office Regions),
- London and the South East (formed by combining London, South East England Government Office Regions),
- Rest of England (formed by combining East Midlands, West Midlands, Eastern England, South West England Government Office Regions).

[8] Sometimes the modelling of complex models cannot continue because all the random parameters obtain zero values during the iterations.

3.9 The analysis revealed that most of the variability was found in London and the South East, Table 3.3.

Table 3.3: Regional analysis of CSVs

response category	EHS variable associated with response category	CSV all regions	CSV North	CSV London and South East	CSV rest of England
date kitchen last refurbished – 2000s	FinKitlrR4	1.688	0	0.140	0
number of dwellings in area – 500+	FardwellR6	0.327	0	0.336	0.305
date bathroom last refurbished – 2000s	FinbatlrR4	0.183	0	0.378	6.982
date bathroom last refurbished – 1980s	FinbatlrR2	0.106	0.488	0.097	0.015
windows and doors not fully secure	SecureR0	0.178	0.195	0.199	0.134
secure windows and doors – secure	SecureR1	0.178	0.195	0.200	0.134
nature of area – rural residential	ArnatxR4	0.136	0	0	0.158
nature of area – suburban residential	ArnatxR3	0.128	0	0.383	0.086
nature of area – other urban centre	ArnatxR2	0.109	0.019	0.278	0.061
any problems with poor quality environment – no	LvanyxR0	0.132	0.135	0.211	0.058
any problems with poor quality environment – yes	LvanyxR1	0.132	0.135	0.211	0.058
predominant type of roof covering – asphalt	typercovR5	0.131	0	0.217	0
nature of area – Suburban	FarnaturR2	0.128	0	0.383	0.086
nature of area – Urban	FarnaturR1	0.109	0.007	0.322	0.090
poor quality environment – upkeep problems – no	lv1upkpxR0	0.124	0.104	0.218	0.023
poor quality environment – upkeep problems – yes	lv1upkpxR1	0.124	0.104	0.218	0.023
doorsets and Circulation meet Part M of Building Regulations	FincircuR1	0.117	0.008	0.146	0.190
urgent repair – wall finish – No	FexwfurR2	0.100	0.135	0.124	0.049

Combined analysis taking Kappa Scores and CSVs into account

3.10 When identifying areas of high variability, it is important to take into account both the Kappa and CSV scores, in order to get a holistic picture of the impact

of surveyor variability on the EHS. For the reasons explained above (Paragraph 3.4), for some survey measures the Kappa Score is not a reliable indicator of variability. CSV scores offer an alternative assessment of the level of agreement between surveyors and help in the final judgment. In order to aid interpretation of the statistics produced in this study, a RAG (red, amber and green) scoring system has been developed. A similar process was adopted in the 2009-10 and 2003-04 surveyor variability studies, Table 3.4.

Table 3.4: RAG scoring system for combined analysis

score		
	**Red**	This indicates a survey measure with a low level of agreement between surveyors. Survey measures scored red will have a kappa score <= 0.40, and a CSV >= 0.1 for all individual response options.
	**Amber**	This indicates a survey measure with a moderate/ acceptable level of agreement between variables. Survey measures scored amber will have a kappa score between 0.41 and 0.60. ***OR*** A kappa score which indicates high level of agreement (>= 0.61), however one or more response options with a CSV>= 0.1. ***OR*** A kappa score indicating a low level of agreement (<=0.40), however a one or more response options with a CSV<0.1.
	**Green**	This indicates an overall high level of agreement. Survey measures scored green will have a kappa >= 0.61, and CSVs <0.1 for all individual response options.

3.11 We found that overall, there was a high level of agreement between surveyors' assessments of the dwellings: 28 of the 73 survey measures (38%) included in the study got a Green RAG score indicating a definite high level of agreement. A further 41 measures (56%) have been awarded an Amber RAG score indicating an acceptable level of agreement. This category includes also measures that required further investigation as one of the composite indicators suggested a rather low level of agreement. The results of more in-depth analysis are presented in the following sections. **Only 4 survey measures (5%) were identified as having definitely low levels of agreement (Red RAG score)**. All of them describe amenities and services or external environment, Table 3.5.

Table 3.5: Overall RAG rating of the measures

survey measure	Red		Amber		Green		all	
	count	%	count	%	count	%	count	%
stock profile	0	0	1	6	15	94	**16**	**100**
amenities and services	2	25	2	25	4	50	**8**	**100**
external environments	2	20	6	60	2	20	**10**	**100**
stock condition	0	0	27	96	1	4	**28**	**100**
energy performance	0	0	5	45	6	55	**11**	**100**
all	**4**	**5**	**41**	**56**	**28**	**38**	**73**	**100**

Green scoring variables

3.12 Of the 73 measures, 28 were classed as having a good level of agreement and therefore not of concern. Their description and EHS variable names are presented in Table 3.6.

Table 3.6: Measures scored overall 'Green' on RAG scale

survey measure	EHS variable associated with the measure
attic present in dwelling	Attic
basement present in dwelling	basement
dwelling age	dwage6x
dwelling type	dwtypenx
material and construction of house/module – PARENT	Fmtconst
construction date	fodconst
dwelling type (house/flat)	FODDTYPE
type of occupancy	Fodishmo
tenure	Fodtenur
predominant type of roof structure	typerstr
predominant type of wall finish	typewfin
predominant type of window	typewin
predominant type of wall structure	typewstr
type of wall	wallcavx
type of wall and insulation	wallinsx
whether shared facilities exist – PARENT	Ffcshare
accessibility – shower or bath at entrance level?	Finbaten
interior space heating present	Fincheat
mains gas supply present	Fingasms
total useable floor area m2	Floorx
no of floors above ground in the house/module	Storeyx
HHSRS – cold homes	hsrcld
type of primary heating system– PARENT	Finchtyp
main heating fuel	Finmhfue
roof insulation above living space present– PARENT	Fliinsul
main fuel type	Fuelx
main heating system	heat4x
energy efficiency (SAP05) rating	Sap05

Amber scoring variables

3.13 Forty-one survey measures had been awarded an Amber score. In order to decide whether a variable with an Amber score was of concern or not it was necessary to take into account the findings of all three indicators. This is because Kappa Scores underestimate the level of agreement between the parent and child interviews in variables where most of the responses fall into one category. An example is variable Fexwsfl (Wall structure) with Kappa Score estimated at

0.04 (very low) while the percentage level of observed agreement between the two interviews is very high: 92.1%, Table 3.7.

Table 3.7: Wall structure (EHS variable name: Fexwsfl), responses of parent and child survey

faults – wall structure, parent	faults – wall structure, child yes	no	all
yes	1	10	**11**
no	14	278	**292**
all	**15**	**288**	**303**

3.14 The variability of seven measures with an Amber score was considered problematic: date bathroom last refurbished (Finbatlr); date kitchen last refurbished (FinKitlr); nature of area (Arnatx and Farnatur); number of dwellings in area (Fardwell); visual quality of local area (Farquali); and wall finish needing urgent repair (Fexwfur), Table 3.8.

Table 3.8: Measures scored overall 'Amber' on the RAG scale

survey measure	EHS variable associated to measure	extent of the variability	whether between-surveyor variability was a concern
date bathroom last refurbished	Finbatlr	Low Kappa Scores and two response categories with high CSV (however, the CSV score for this variable is potentially unreliable, see Table 2.3)	yes
date kitchen last refurbished	FinKitlr	Low Kappa Score and one response category with high CSV (however, the CSV score for this variable is potentially unreliable, see Table 2.3)	yes
urgent repair – Wall finish	Fexwfur	Relatively low Kappa, one CSV score suggest low level of agreement	yes
nature of area	Arnatx	Moderate Kappa Score, three response categories with CSV indicating high level of variability	yes
number of dwellings in area	Fardwell	Low Kappa Score, low agreement. One response category with very high CSV and several others with rather high CSV.	yes
nature of area	Farnatur	Moderate Kappa, but two response categories with high CSV	yes
visual quality of local area	Farquali	Reasonably low Kappa, but CSV suggest good level of agreement	yes
poor quality environment – traffic problems	lv2trafx	Low Kappa, but CSV suggest high level of agreement	no
poor quality environment – utilisation problems	lv3utilx	Low Kappa, but very high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
damp – problem present?	Dampalf	Low Kappa, but high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
decent homes repair criterion	dhdisrx	Very low Kappa, high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
decent homes HHSRS 15 criterion	dhhhsrsx	Relatively low Kappa, high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
Ddecent homes HHSRS 26 criterion	dhhhsrsy	Relatively low Kappa, high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no

decent homes modern facilities criterion	dhmodx	Relatively low Kappa, very high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
decent homes – HHSRS 15 model	dhomesy	Relatively low Kappa. CSV suggest high level of agreement	no
decent homes – HHSRS 26 model	dhomesz	Relatively low Kappa. CSV suggest high level of agreement	no
decent homes thermal comfort criterion	dhthermy	Moderate kappa, high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
faults – external doors	Fexdffl	Relatively low Kappa, quite high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
urgent repair – external doors	Fexdfur	Low Kappa, but high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
faults – roof covering	fexrcfl	Low Kappa, but high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
urgent repairs – roof covering	Fexrcur	Low Kappa, but high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
faults – roof structure	fexrsfl	Low Kappa, but high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
urgent repair – roof structure	Fexrsur	Low Kappa, but high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
faults – wall finish	Fexwffl	Moderate Kappa. CSV suggest high level of agreement	no
faults – windows	Fexwnfl	Moderate Kappa, high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
urgent repair – windows	Fexwnur	Relatively low Kappa, high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
faults – wall structure	Fexwsfl	Very low Kappa, but high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
urgent repair – wall structure	Fexwsur	Very low Kappa, but high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no

boiler – action required	Finchbac	Low Kappa, very high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
boiler – age ranges	finchbag	Moderate Kappa. CSV suggest high level of agreement	no
foundation settlement – defect	Fstfoude	Low Kappa. CSV suggest high level of agreement	no
differential movement – defect	Fstmovde	Low Kappa, but relatively high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
structural defects	Fstpres	Low Kappa. CSV suggest high level of agreement	no
overall all 26 hazards	HSRALL_26	Relatively low Kappa, high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
HHSRS falls category 1	hsrfalls	Low Kappa, however, high agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
extent of double glazing	Dblglaz4	Moderate Kappa. CSV suggest high level of agreement	no
EPC energy efficiency band	epceeb05e3R	Moderate Kappa. CSV suggest high level of agreement	no
cavity wall insulation present	Fhqcavit	Moderate Kappa, moderate level of agreement based on the raw percentage of agreement. CSV suggest high level of agreement	no
loft insulation thickness	Flithick	Moderate Kappa. CSV suggest high level of agreement	no
loft insulation thickness	loftins6	Moderate Kappa. CSV suggest high level of agreement	no
predominant type of roof covering	typercov	Only one response category causing issues (asphalt), and then with only a very small number of cases	no

Red scoring variables

3.14 Four of the 73 key survey measures were found to have a red RAG rating. They relate to: secure windows and doors; accessibility**;** poor quality environment; and poor quality environment related to upkeep problems, Table 3.9.

Table 3.9: Measures scored overall 'Red' on the RAG scale

survey measure	EHS variable associated with measure	extent of the variability	whether between-surveyor variability was a concern
secure windows and doors	Secure	Low Kappa Score (0.40 or lower) and a CSV of or higher than 0.1 for all individual response options	yes
accessibility - doorsets and circulation meet part M of Building Regulations	Fincircu	Low Kappa Score (0.40 or lower) and a CSV of or higher than 0.1 for all individual response options	yes
poor quality environment – upkeep problems	lv1upkpx	Low Kappa Score (0.40 or lower) and a CSV of or higher than 0.1 for all individual response options	yes
poor quality environment – any problems	lvanyx	Low Kappa Score (0.40 or lower) and a CSV of or higher than 0.1 for all individual response options	yes

Applying the RAG ratings to help reduce surveyor variability

3.15 Surveyor variability cannot be eliminated completely but actions, including rigorous surveyor training and supervision, can reduce it. The findings of this study can inform the development of surveyor training and supervision by highlighting the survey measures that have relatively high surveyor variability.

3.16 Taking account of all the approaches this study has used to assess variability, we conclude that 11 of the 73 measures assessed have relatively high surveyor variability. They are the measures with a Red RAG score and the seven measures with an Amber RAG score and highlighted as 'of concern' in Table 3.8.

3.17 To help prioritize actions going forward, we have indicated their relative level of variability showing the RAG rating as well as the exact values of their Kappa and CSV Scores and raw percentage of observed agreement in Table 3.10. These should be interpreted as follows:

- The Kappa coefficient ranges from 0 (agreement by chance) to +1 (a theoretical maximum agreement between surveyors;;
- CSV: a score of 1 indicates the theoretical limit of high variability (low agreement), a score of zero indicates no variability (perfect agreement); and

- Observed agreement: close to 0 indicates no agreement, close to 100 indicates perfect agreement.

In some cases, just some categories of a measure proved to have higher surveyor variability only so action targeted at those measures needs to focus on the categories showing variability rather than on the whole measure.

Table 3.10: Survey measures found to have a low level of agreement or to require further action

survey measure and response category	EHS variable associated with measure	observed agreement	Kappa Score	CSV	overall RAG
accessibility: doorsets and circulation meet part M of Building Regulations	Fincircu	74.9	0.380		
yes				0.117	
no				*	
secure windows and doors	Secure	78.9	0.279		
secure				0.178	
not fully secure				0.178	
poor quality environment – upkeep problems	lv1upkpx	87.8	0.284		
yes				0.124	
no				0.124	
poor quality environment – any problems	lvanyx	82.5	0.289		
yes				0.132	
no				0.132	
date bathroom last refurbished	Finbatlr	63.0	0.238		
pre 1980				0.010	
1990s				0.067	
1980s				0.106	
2000s				0.183	
date kitchen last refurbished	FinKitlr	75.0	0.398		
Pre 1980				0.000	
1990s				0.051	
1980s				0.039	
2000s				0.217	
nature of area	Arnatx	70.6	0.420		
city centre				0.077	
suburban residential				0.128	
other urban centre				0.109	
rural				0.071	
rural residential				0.136	
number of dwellings in area	Fardwell	31.0	0.124		
under 25				0.070	
50–99				0.075	
25–49				0.092	
300–499				0.091	

100–299				0.062
isolated				–
500+				0.327
nature of area	Farnatur	75.2	0.487	
urban				0.109
rural				0.058
suburban				0.128
visual quality of local area	Farquali	80.9	0.379	
good quality				0.050
worst Quality				0.028
average quality				0.043
urgent repair – wall finish	Fexwfur	76.2	0.315	
yes				0.066
no				0.100

Note: * = value not reported as multi-level modelling produced on fewer than 12 responses to this variable; or all the random parameters in the multilevel modelling model obtained value zero, the estimation could not continue. Results too variable to produce confident estimates, and therefore excluded.

Bias adjustments, standard errors and confidence intervals

3.18 One of the aims of the SVS was to assess the impact of surveyor variability on the complex standard errors and confidence intervals of the EHS variables. Complex standard errors (CSE) were calculated on the combined 2014-15 and 2013-14 EHS data to take account of the complex design of the survey. However, these only partly reflect the effect of between-surveyor variability. Based on the assumption that we anticipate some correlation between individual observations from the same surveyor, there is a chance that the complex standard errors produced are underestimated. As a part of the analysis carried out on the SVS, bias adjustments were calculated to take into account the impact of Correlated Surveyor Variances on the existing EHS complex standard errors. Revised Standard Errors (RSE) were produced for six measures in the combined 2013-14 and 2014-15 EHS dwelling sample identified by this study as particularly prone to surveyor error and three benchmark measure drawn from those that were not particularly affected by surveyor error, Table 3.11. Those RSEs have been compared with the existing 'Complex Standard Errors' (CSE) produced by accounting for sampling and grossing only, Annex 2.

Table 3.11: Variables selected for the standard error review

survey measure	EHS variable associated with measure	RAG score	whether between–surveyor variability was a concern	extent of the variability
secure windows and doors	Secure		yes	High level of surveyor variability for both categories of window and door security, Kappa of 0.279 indicates 'fair' agreement
accessibility: doorsets and circulation meet part M of Building Regulations	Fincircu		yes	High variability in the assessment of doorsets regulations. Kappa of 0.380 shows 'fair' agreement.
date bathroom last refurbished	Finbatlr		yes	Low Kappa and two response categories with high CSV (however, CSV score for this measure is potentially unreliable)
date kitchen last refurbished	FinKitlr		yes	Low Kappa and one response category with high CSV (however, CSV score for this measure is potentially unreliable)
urgent repair – wall finish	Fexwfur		yes	Relatively low kappa, CSV for one category suggest low level of agreement
decent homes HHSRS 15 criterion	dhhhsrsx		no, included as a benchmark	Relatively low Kappa, high agreement based on the raw percentage of agreement. CSV suggest high level of agreement
total useable floor area m2	Floorx		no, included as a benchmark	High level of agreement
main heating system	heat4x		no, included as a benchmark	High level of agreement
energy efficiency (SAP05) rating	Sap05		no, included as a benchmark	High level of agreement

3.19 A total of 25 individual response categories from the nine survey measures had a bias adjustment created. Five of them had a bias adjustment of 0 calculated, due to the CSV being close or equal to 0. It was not possible to calculate bias adjustment for a further three categories because CSV was not available due to small sample size or to computational issues in the multilevel modelling.

3.20 This section looks at the Confidence Intervals (CIs) produced from the CSE and RSE to create a picture of the impact of the variability of the physical survey data on the EHS estimates. The reason for analysing CIs rather than standard errors, is that they provide a practical example of the impact variability is having on the survey estimates. For this report, we have identified any individual response category where the CI has increased by more than 1 percentage point as a significant impact in variability.

3.21 As expected, the revisions to CIs for the measures scoring Green on RAG score were minimal. Among those scoring Amber or Red, seven response categories were found to have increased the confidence interval by at least one percentage point:

- Date bathroom last refurbished (Finbatlr): 1980s, change by 0.97 percentage point[9]
- Date bathroom last refurbished (Finbatlr): 1990s, change by 0.96 percentage point
- Date bathroom last refurbished (Finbatlr): 2000s, change by 2.30 percentage points[10]
- Date kitchen last refurbished (FinKitlr): 2000s, change by 8.17 percentage points[11]
- Doorsets and Circulation meet Part M regulations (Fincircu): Yes, change by 1.69 percentage points
- Secure windows and doors (Secure): Not fully secure, change by 1.89 percentage points
- Secure windows and doors (Secure): Secure, change by 1.89 percentage points
- Urgent repair – wall finish (Fexwfur): No, change by 1.17 percentage point.

[9] Result should be treated with caution due to lack of stability in estimates coming from models with different number of iterations.
[10] Result should be treated with caution due to lack of stability in estimates coming from models with different number of iterations.
[11] Result should be treated with caution due to lack of stability in estimates coming from models with different number of iterations.

Annex 2 shows the RSE for all the variables and response categories explored in the study against the CSE produced by taking into account sampling and grossing only.

3.22 From among the reliable estimates, the response option with the largest increase to the confidence interval was having secure windows and doors. Among the parent cases, 82.5% of households were assessed as having secure windows and doors. The original CI surrounding this estimate (based on CSE) is ±0.87%, suggesting we can be confident 95% of the time that our estimate would fall between 81.65% and 83.38%. However, when taking our bias adjustment into account, the revised CI (based on RSE) surrounding our estimate would increase to ±2.76%. This suggests that we can be confident 95% of the time that our estimate would fall between 79.76% and 85.28%.

Comparison of surveyor variability in 2009-10 and 2014-15

3.23 Comparisons of the CSVs were made with the most recent SVS, 2009-10, as this best illustrates the current and recent impact of surveyor variability. Results from the 2003-04 study will be included on selected tables for information, but not included in the commentary.

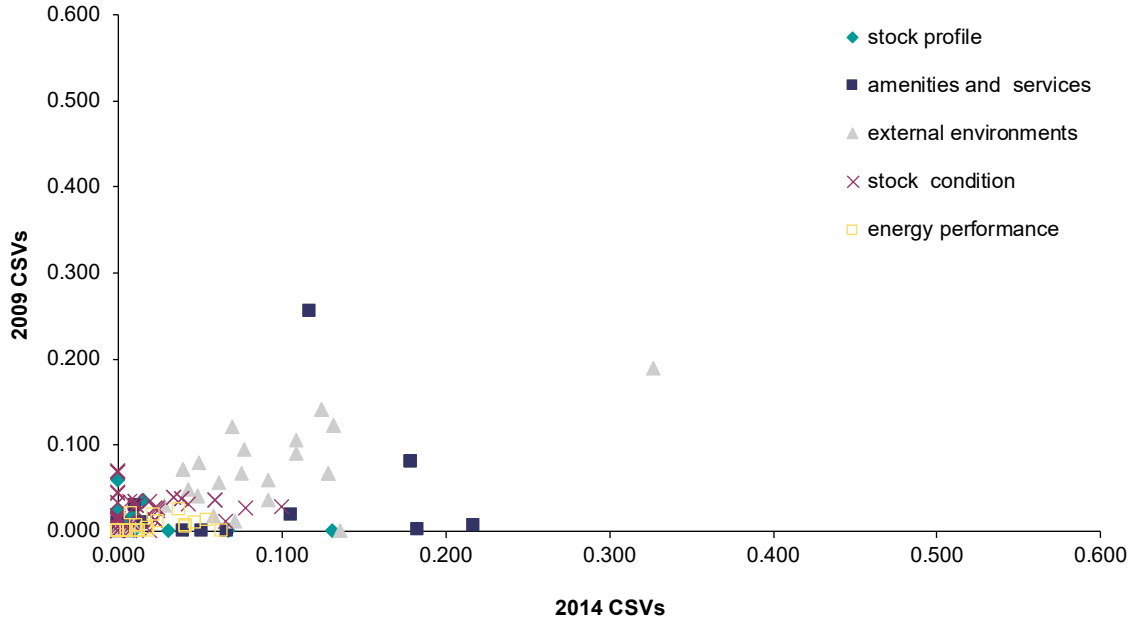
3.24 The CSVs produced in the 2014-15 SVS ranged from a minimum value of 0 to a maximum value of 0.33, lower than for the 2009-10 EHS where the maximum value was 0.54. The mean value was 0.03 (compared with 0.02 in 2009-10). The 2014-15 SVS had a slightly higher CSV score than in 2009-10 but the standard deviation was similar across the two years, Table 3.12. This means that surveyor variability (as measured by CSV) has remained consistent across time.

Table 3.12: CSV statistics, 2003-04, 2009-10 and 2014-15 surveyor variability studies

	2014-15	2009-10	2003-04
number of variables total	243	251	428
number of variables calculated	192	214	374
mean	0.03	0.02	0.06
standard deviation	0.05	0.05	0.09
minimum	0	0	0
maximum	0.33	0.54	0.48

3.25 The level of variability in the 2014-15 and 2009-10 EHS is broadly similar, Figure 3.3.

Figure 3.3: Correlated Surveyor Variance (CSV) scores, 2009-10 and 2014-15



Note: excludes response options for which CSV scores could not be calculated

3.26 A comparison of the value of the CSV scores for 2009-10 and 2014-15 shows that, of the 184 response options for which CSV scores are available:

- 165 had CSVs under 0.1 in both years, indicating that variability is stable for those response options between the two years.
- For the 19 response options with a CSV greater than 0.1 in either 2009-10 or 2014-15, eight scored over 0.1 in 2009-10 and all but one of those have remained so in 2014-15, Table 3.13.
- Eleven response options (6%) had scores above 0.1 in 2014-15 but not in 2009-10.

Table 3.13: Response categories with CSV scores above 0.1, 2009-10 and 2014-15

survey measure and response category	EHS variable associated with measure	2014-15 Correlated Surveyor Variance (CSV)	2009–10 Correlated Surveyor Variance (CSV)
predominant type of roof covering – asphalt	typercovR5	0.131	0.000
bathroom last refurbished in the 1980s	FinbatlrR2	0.106	0.019
bathroom last refurbished in the 2000s	FinbatlrR4	0.183	0.001
doorsets and Circulation meet Part M of Building Regulations	FincircuR1	0.117	0.255
kitchen last refurbished in the 2000s	FinKitlrR4	0.217	0.006
windows and doors not fully secure	SecureR0	0.178	0.080
windows and doors secure	SecureR1	0.178	0.080
nature of area – other urban centre	ArnatxR2	0.109	0.105
nature of area – suburban residential	ArnatxR3	0.128	0.066
nature of area – rural residential	ArnatxR4	0.136	0.000
number of dwellings in area – under 25	FardwellR1	0.070	0.121
number of dwellings in area – 500+	FardwellR6	0.327	0.188
nature of area – urban	FarnaturR1	0.109	0.089
nature of area – Suburban	FarnaturR2	0.128	0.066
poor quality environment – upkeep problems – no	lv1upkpxR0	0.124	0.141
poor quality environment – upkeep problems – yes	lv1upkpxR1	0.124	0.141
any problems with poor quality environment – no	LvanyxR0	0.132	0.121
any problems with poor quality environment – yes	LvanyxR1	0.132	0.121
urgent repair to wall finish – No	FexwfurR2	0.100	0.027

Annex 1
Detailed results and cross-tabulations

1. Tables showing the detailed results for the 73 variables (and individual response options) explored in the 2014-15 SVS are available on the web pages providing technical advice on the EHS: https://www.gov.uk/government/collections/english-housing-survey-technical-advice.

2. The tables are grouped by topic area and contain information on the following statistics:

    - Descriptive Analysis:
    - Observed Agreement
    - Chance Agreement
    - Kappa Score
    - Standard Error of Kappa
    - Multi-Level Modelling:
    - # obs : Number of observations for a given response option
    - Correlated Surveyor Variance
    - RAG Scores:
    - Descriptive (kappa)
    - Correlated Surveyor Variance
    - Overall.

3. The tables in Appendix 1 also contain cross-tabulations of the child and parent cases. Section 4.1 of the report points out that the interpretation of variability for each key survey measure should refer to the cross-tabulations of the child and parent cases. This is because measures based in chi-square statistics (including the Kappa Score) are not reliable measures of agreement when a significant proportion of cells have expected count of less than 5. To judge the absolute effect of the impact of surveyor variability, the Kappa scores need to be looked at alongside the 'raw' percentage of agreement produced by cross-tabulating the parent and child cases. A low 'raw' percentage of agreement combined with a low Kappa score indicates that surveyors are disagreeing over a substantial number of marginal cases. Cross-tabulations also show where the mismatch between

child and parent bases occurred to inform the development of material for training surveyors in the future.

Annex 2
Adjustment to original complex standard errors

1. The original complex standard errors (and subsequent confidence intervals) which take sampling and weighting into account, and revised standard errors (and confidence intervals) which adjust to take into account surveyor variability are available in a separate excel document, on the web pages providing technical advice on the EHS: https://www.gov.uk/government/collections/english-housing-survey-technical-advice.

© *Crown copyright, 2020*

Copyright in the typographical arrangement rests with the Crown.

You may re-use this information (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence. To view this licence,http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/

This document/publication is also available on our website at www.gov.uk/mhclg

If you have any enquiries regarding this document/publication, complete the form at http://forms.communities.gov.uk/ or write to us at:

Ministry of Housing, Communities and Local Government
Fry Building
2 Marsham Street
London
SW1P 4DF
Telephone: 030 3444 0000
Email: ehs@communities.gov.uk

For all our latest news and updates follow us on Twitter: https://twitter.com/mhclg

July 2020