

**Dweud eich dweud**  
**Have your say**

Arolwg Cenedlaethol Cymru  
National Survey for Wales



**Comisiynydd y  
Gymraeg**  
**Welsh Language  
Commissioner**

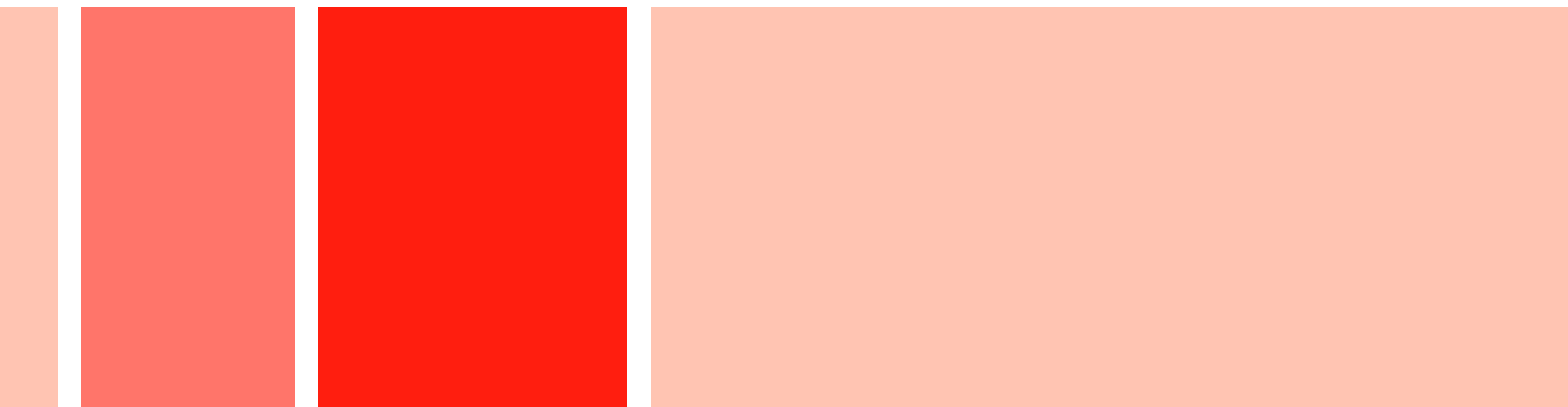


**Llywodraeth Cymru**  
**Welsh Government**

[www.cymru.gov.uk](http://www.cymru.gov.uk)

# Welsh Language Use Survey 2013-14 Technical Report

January 2015



**Tim Hanson, TNS BMRB**  
**Sam Sullivan, TNS BMRB**

Views expressed in this report are those of the researcher and not necessarily the views of the Welsh Government nor the Welsh Language Commissioner.

For further information please contact:

Lisa Walters  
Knowledge and Analytical Services  
Welsh Government  
Cathays Park  
Cardiff  
CF10 3NQ  
Tel: 029 2082 6685  
Email: [lisa.walters@wales.gsi.gov.uk](mailto:lisa.walters@wales.gsi.gov.uk)

Welsh Government Social Research, 29 January 2015

© Crown Copyright 2015



All content is available under the Open Government Licence v3.0, except where otherwise stated.

<http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

## Table of contents

1	Background and objectives .....	2
2	Sample design .....	4
3	Questionnaire content and development.....	9
4	Data collection .....	11
5	Fieldwork figures and response rates .....	17
6	Data preparation .....	20
7	Weighting .....	22

# 1 Background and objectives

## Survey history

- 1.1 The 2013-14 Welsh Language Use Survey (WLUS) is a continuation of research undertaken on behalf of the Welsh Language Board (WLB) between 2004 and 2006. Over a series of three surveys, WLB sought to obtain information on who uses Welsh and with whom, when, and how much they use it. These surveys were organised to supplement the *Living in Wales* survey (LIW), conducted between 2004 and 2006 by the Local Government Data Unit ~ Wales on behalf of the Welsh Assembly Government.
- 1.2 For the 2004-2006 surveys, the LIW survey provided the sample frame (i.e. those identified during the LIW interview as able to speak Welsh).
- 1.3 Further details on the 2004-06 survey can be found at the UK data archive.<sup>1</sup>

## Background to the 2013-14 survey

- 1.4 The WLUS was re-commissioned in 2013, on behalf of the Welsh Government and the Welsh Language Commissioner (WLC). The new wave of the survey will run across two years, from 2013 to 2015. This report provides details of the 2013-14 survey.
- 1.5 The key aim of the WLUS is to obtain information on who uses Welsh, and how, when and how much they use it.
- 1.6 It was decided that rather than run the WLUS as a separate survey, it would be more efficient to set it up as an add-on element to the National Survey for Wales.
- 1.7 The National Survey for Wales is a large-scale face-to-face random probability survey being conducted by TNS BMRB and Beaufort Research on behalf of

---

<sup>1</sup> <http://discover.ukdataservice.ac.uk/Catalogue/?sn=7477&type=Data%20catalogue>

the Welsh Government. Fieldwork began in January 2012 and runs continuously, with a new batch of assignments issued each month.

- 1.8 The National Survey was used to provide a sample for the WLUS. The process by which WLUS respondents are recruited is explained in more detail in Chapter 4.

### **The National Survey Technical Report - relevant sections**

- 1.9 Due to the way in which the survey has been administered, many of the details relating to the 2013-14 WLUS are covered by the Technical Report that was produced for the 2013-14 National Survey. Individual references are included in the following chapters. In summary, the relevant sections in the 2013-14 National Survey Technical Report are:

- Sample Design: Paragraphs 3.1 to 3.31
- Identification of Welsh Speakers: Table 4.1 and Paragraph 4.11
- Overview of the WLUS: Paragraphs 4.14 to 4.17
- Briefing interviewers about the WLUS: Paragraphs 6.5 and 5.6
- Weighting strategy: Paragraphs 7.1 to 7.22

## 2 Sample design

### Overview of National Survey sampling approach

- 2.1 A full explanation of the sampling approach carried out on the National Survey for Wales can be found in the 2013-14 Technical Report.<sup>2</sup> What follows is a summary of that approach.
- 2.2 The specification for the National Survey stated that the survey should be representative of all adults aged 16 or over living in private households. The annual sample requirement was for a stratified, single-stage random selection of addresses across Wales drawn from the small user Postcode Address File (PAF), belonging to Royal Mail. Within each sampled household, an individual interview should be conducted with one adult aged 16 or over. Where a household contains more than one adult, a single adult should be randomly selected to participate from all household members aged 16 or over.
- 2.3 In order to meet the sample requirements, it was agreed that a wholly unclustered sample of addresses would be drawn to achieve equal samples of 600 effective interviews in each local authority per annum. To do this it was calculated that an actual sample size of 662 interviews per LA would need to be achieved in each survey year. The full sample for each reporting year is drawn in one stage and allocated to each quarter first, and then each month within each quarter.
- 2.4 Within each local authority, the sample was selected systematically with a fixed interval. Apart from local authority there was no further explicit sample stratification. Because addresses are ordered by postcode within local authority, a systematic sample with a small sampling interval will be (implicitly) stratified by neighbourhood and always geographically representative. The practical consequence is that the sampling error is reduced (in comparison with a less controlled sample) for any measures that vary significantly between neighbourhoods.

---

<sup>2</sup> <http://wales.gov.uk/docs/caecd/research/2014/140917-national-survey-wales-technical-report-2013-14-en.pdf>

- 2.5 The sample was drawn from the PAF by the Office for National Statistics (ONS) based on a specification agreed between TNS BMRB and the Welsh Government.<sup>3</sup>
- 2.6 The target response rate for the 2013-14 reporting year was 70%. When calculating the number of addresses for ONS to draw in each LA, a response rate of 70% was assumed in all LAs. This figure was refined at the point of address selection to take account of response rate performance on the 2012-13 survey (see paragraph 2.8 for details).
- 2.7 Once the addresses were drawn, ONS passed the full sample file to TNS BMRB via secure file transfer. TNS BMRB was then responsible for batching and allocating the sample in preparation for fieldwork.
- 2.8 Once the full sample file was received by TNS BMRB, a smaller sample file was drawn containing addresses to issue throughout the survey year. As before, numbers were selected to achieve a flat target of 662 interviews per local authority. Response rate assumptions were based on the average responses rates achieved in each local authority across Q0-Q3 on the 2012-13 survey. One of three assumed responses rates was applied: 73% for LAs with an average rate of 73% or more; 70% for LAs with an average rate of between 68% and 72%; and 67% for LAs with an average rate of 67% or less. The assumed deadwood rate was the average within each LA across Q0-Q3.
- 2.9 The original target of 662 interviews per LA was subsequently revised in light of data obtained from the 2012-13 survey, and corresponding adjustments were made to the number of issued addresses within each LA. Full details of these adjustments can be found in the National Survey 2013-14 Technical Report.

---

<sup>3</sup> The sample was drawn by ONS using PAF and their Common User Address File, which ensured that no household would be selected for the National Survey that had previously been selected for any of a range of other large-scale ONS surveys within the last three years. This also ensured that the sampled addresses would not be selected for other major ONS surveys or waves of the National Survey for Wales for the next three years.

2.10 TNS BMRB developed a bespoke area geography in Wales in which each area unit a) covers sufficient sampled addresses to (approximately) equal one fieldwork assignment, and b) is 'geographically sensible' as an assignment.<sup>4</sup> Area units were constructed using Output Areas (OAs). Lower Layer Super Output Areas (LSOAs) were originally to be used but this would have resulted in too much variation in the number of addresses per assignment, especially in the smaller LAs.

2.11 Once the units had been created, they were sorted by area unit aggregate values.<sup>5</sup> Once stratified, they were systematically allocated to the four survey quarters of the 2013-14 reporting year (Quarters 1-4).<sup>6</sup> Any area units that had been allocated to Quarter 1 in the 2012-13 survey were allocated to Quarter 1 in the 2013-14 reporting year, with the same pattern repeated for Quarters 2-4. This approach ensures that any consecutive four quarters of data will be formally unclustered.

2.12 Once assignments were allocated to quarters they were then randomly allocated to a month within the quarter. This was done while ensuring that each LA had a roughly equal number of points in each month and that each month had approximately the same number of sample points.

2.13 As noted above, assignment sizes varied based on the bespoke sample geography developed by TNS BMRB. In 2013-14 all assignments had between 18 and 34 addresses, with the mean assignment size being 24 addresses. More than nine in ten assignments (93%) included between 20 and 28 addresses. A total of 994 assignments were issued in 2013-14.

---

<sup>4</sup> This meant that as far as possible addresses in each assignment were clustered in a way that reduced the distance interviewers were required to travel between them, while also ensuring that the number of addresses did not vary substantially between assignments.

<sup>5</sup> The area unit values used were Welsh Index of Multiple Deprivation (WIMD) scores and Urbanisation.

<sup>6</sup> This involved taking all of the assignments within each LA, sorting by the average area unit aggregated values for each assignment and allocating to quarters.



2.14 Table 2.1 shows the number of addresses within each LA across the 2013-14 survey year.

**Table 2.1 – Final issued sample for the National Survey 2013-14**

<b>Local Authority</b>	<b>Addresses issued</b>
Blaenau Gwent	1,097
Bridgend	1,102
Caerphilly	1,058
Cardiff	1,127
Carmarthenshire	1,131
Ceredigion	1,111
Conwy	1,080
Denbighshire	1,066
Flintshire	1,085
Gwynedd	1,173
Isle of Anglesey	1,150
Merthyr Tydfil	1,110
Monmouthshire	1,081
Neath Port Talbot	935
Newport	1,047
Pembrokeshire	1,202
Powys	1,069
Rhondda Cynon Taf	1,131
Swansea	1,150
The Vale of Glamorgan	1,031
Torfaen	938
Wrexham	1,024
<b>TOTAL</b>	<b>23,898</b>

2.15 At each issued address, interviewers were required to carry out three stages of selection in order to identify the correct person for interview:

- Dwelling unit selection (in addresses containing more than one dwelling unit)
- Household selection (in dwelling units containing more than one household)
- Person selection (in households containing more than one adult aged 16 or over)

2.16 Full details of how these selections were carried out can be found in the 2013-14 National Survey Technical Report.

**Original assumptions around Welsh-speaking population**

2.17 In order to gauge the likely numbers that were achievable on the survey, estimates were made on the likely proportion of Welsh speakers who would be identified through the National Survey.

2.18 Table 2.2 below shows the original projections that were made regarding the likely population of Welsh speakers, and the actual number achieved.

**Table 2.2 – Projections for Welsh-speaking population, and corresponding achieved numbers**

	<b>Original Projection</b>	<b>Achieved</b>
<b>Households interviewed on National Survey</b>	14,500	14,771
<b>Adults identified</b>	27,405	27,170
<b>Average number of adults per household</b>	1.89	1.84
<b>Welsh-speaking adults identified</b>	7,125	6,241
<b>% of adults identified who speak Welsh</b>	26%	23%
<b>3-15 year olds identified</b>	5,945	5,065
<b>Average number of 3-15 year-olds per household</b>	0.41	0.34
<b>Welsh-speaking 3-15 year-olds identified</b>	2,675	2,200
<b>% of 3-15 year-olds identified who speak Welsh</b>	45%	43%

2.19 The process for identifying Welsh speakers is explained in more detail in Chapter 4.

### **3 Questionnaire content and development**

#### **Overview**

- 3.1 The WLUS comprised two separate questionnaires – an adult questionnaire (for respondents aged 16+) and a young person questionnaire (for respondents aged 3-15). Both cover Welsh Language use in various situations. Copies of both questionnaires can be found at <http://wales.gov.uk/statistics-and-research/Welsh-language-use-survey/?lang=en>.
- 3.2 Both questionnaires were dual-language, giving respondents the option to complete the questionnaire in either Welsh or English. A single booklet was produced including the Welsh and English versions of the questions, with the questions ‘flipped’ so that each begins on a front cover and meet on the central page. The adult questionnaire comprised 12 sides (including instructions and both languages) and the children’s questionnaire 8 sides.

#### **Questionnaire content**

- 3.3 The questionnaire content was developed jointly by the Welsh Language Commissioner and the Welsh Government. The majority of the questions included were unchanged from the 2004-06 survey. These questions can be found in chapters 20 and 21 of the [report](#) published on the 2004-06 survey findings.
- 3.4 As noted above, the questionnaires were produced in both Welsh and English.

#### **Format and appearance**

- 3.5 The questionnaires were printed on different colour paper – orange for the adult questionnaire and blue for the children’s questionnaire. The intention was to make it easier for interviewers to distinguish between the different versions when handing out the questionnaires, and thus reduce the risk of a respondent being issued with the wrong version.

3.6 The design team at TNS BMRB developed the layout and style of the questionnaires in an effort to make them clearer and more engaging. The key considerations were to:

- ensure that the layout was not too cluttered, whilst making most efficient use of the available space;
- to develop a colour scheme which would help respondents to distinguish between individual questions and complete the questionnaire in full; and
- to ensure a consistent format throughout to aid completion of the questionnaire.

3.7 The final questionnaires were agreed between the Welsh Government, the Welsh Language Commissioner and TNS BMRB.

## 4 Data collection

### Selection process

- 4.1 In order to be eligible for the Welsh Language Use Survey, respondents must be aged 3 or over and be able to speak Welsh. The questionnaire completed depended on the age of the individual, with 3-to-15 year-olds completing the child questionnaire and respondents aged 16 or over completing the adult questionnaire.
- 4.2 Respondents to the survey were recruited through the National Survey for Wales. The Survey collected information about all individuals living with the respondent as part of the same household.
- 4.3 In the National Survey interview, the respondent is asked whether they can speak Welsh. Interviewers are able to code to one of three possible responses:
  1. Yes
  2. No
  3. (SPONTANEOUS ONLY) No, but have some Welsh-speaking ability
- 4.4 For the purposes of the Welsh Language Survey, any respondent selecting either Code 1 or Code 3 were regarded as Welsh speakers and therefore eligible to receive a questionnaire.
- 4.5 It should be noted that this question was changed in July 2013 from the original version of the question, which featured only 'Yes' and 'No' response options. The change was made following analysis of 2011 Census data, which indicated a lower proportion of Welsh-speaking adults (16%) when compared with results from the 2012-13 survey (24%). National Survey interviewers had been instructed to code a response of 'Yes' for respondents with any level of Welsh speaking ability in order to allow them to be routed through to subsequent, more detailed questions on their ability in Welsh and the frequency in which they speak it. It is likely that this approach contributed to a

higher proportion of National Survey respondents being identified as Welsh speakers than were found in the Census. The addition of the new code was designed to provide flexibility to allow the National Survey results to be compared more closely with the Census results (i.e. focusing on respondents answering “Yes” to the question on Welsh speaking), but also where necessary to enable analysis of National Survey topics in a way that is consistent with previous years (i.e. by combining “Yes” and “No, but have some Welsh-speaking ability”).

- 4.6 In households containing more than one person aged 3 or over, a similar question was asked to establish which other household members are able to speak Welsh:

*I now want to ask you whether other members of your household can speak Welsh*

*Can [name] speak Welsh?*

1. *Yes*
2. *No*
3. *Don't know*

- 4.7 All household members coded as ‘Yes’ were regarded as Welsh speakers and therefore eligible to receive a questionnaire.

### **Issuing questionnaires**

- 4.8 Face-to-face interviewers were responsible for handing out questionnaires to all eligible individuals at the end of the National Survey interview. A full explanation of the WLUS and the interviewer requirements were covered in the main National Survey briefing attended by all interviewers. Details of the briefing content are described in the National Survey Technical Report.
- 4.9 All interviewers working on the National Survey were provided with a number of WLUS questionnaires and freepost envelopes. The number provided to

each interviewer was determined by the expected penetration of Welsh speakers in their area.

4.10 Interviewers were instructed not to refer to the WLUS until all sections of the National Survey interview had been completed. In deciding on this approach there were two key considerations.

4.11 Firstly, there was a need to ensure that the administration of the survey did not impact on respondents' inclination to take part in a main National Survey interview. It was felt that any introduction of the WLUS prior to the main National Survey interview could create the impression of increased respondent burden and discourage people from taking part. It could also have a detrimental impact within non-Welsh-speaking households by making the National Survey seem less relevant to them.

4.12 Secondly, it was felt that by the end of the National Survey interview the interviewers will have had the opportunity to build rapport with the respondent, increasing the likelihood of them accepting the WLUS questionnaire to complete.

4.13 Immediately after the final question in the National Survey interview, interviewers were instructed to read out an introduction to the WLUS that appeared on screen:

*The Welsh Language Commissioner and the Welsh Government are conducting a survey of Welsh language use. They are asking Welsh speakers in households that have participated in the National Survey for Wales to complete a short paper questionnaire. The questionnaire is only 6 pages for adults and 4 pages for children (aged 3 or over) and just needs to be completed and returned to TNS BMRB in a freepost envelope in the next two weeks.*

4.14 On the screen, interviewers were presented with a list of all individuals within the household who were eligible to receive a WLUS questionnaire. Alongside the name of each individual, the following information was shown:

- Household serial number
- Person Number
- Which questionnaire they should receive (adult or child)

4.15 For each individual, the interviewer was asked to record whether or not a questionnaire had been accepted.

4.16 Interviewers were then required to write the individual's name, household serial number, and person number on both covers (English and Welsh) of the relevant questionnaire, before handing to the respondent. Respondents were provided with a freepost return envelope for each questionnaire issued.

4.17 A slightly different approach was taken for the first month of the survey, whereby the National Survey script included a question asking whether the respondent was willing to accept the questionnaires. In order to maximise acceptance rates, a change in approach was implemented. The question was removed and interviewers were instructed to leave questionnaires for all eligible individuals unless they were explicitly declined by the respondent.

### **Postal reminders**

4.18 As noted above, WLUS respondents are asked to complete and return their questionnaire within two weeks of receiving it. Respondents who had not returned a completed questionnaire within this period were sent a reminder pack through the post, containing a replacement questionnaire and a cover letter explaining about the survey and inviting them again to take part.

4.19 Postal reminders were sent out on a weekly basis throughout fieldwork.

### **Telephone reminders**

4.20 Respondents who did not return a questionnaire after being sent a postal reminder received an additional reminder, conducted by telephone. Naturally this relied on the National Survey respondent having provided a valid telephone number, meaning that some WLUS non-respondents did not receive a telephone reminder.



- 4.21 Telephone reminders were carried out by a member of TNS BMRB's telephone interviewer panel on a monthly basis. As a result, the interval between receiving the initial questionnaire and receiving a telephone reminder varied for each respondent depending on the date of the original National Survey interview. All telephone reminder calls took place at least two weeks after the respondent had received their postal reminder pack (i.e. four weeks after receiving the initial questionnaire). The maximum interval between receiving an initial questionnaire and receiving a telephone reminder was eight weeks.
- 4.22 Households containing more than one outstanding questionnaire only received one telephone reminder call. In all cases, the interviewer calling asked to speak to the respondent who took part in the National Survey interview, regardless of whether or not they are a Welsh speaker. The interviewer then read from a short script providing the names of all household members who had not yet returned a questionnaire.
- 4.23 There were no questions for the respondents to answer – the call simply served as a reminder about the WLUS and requested that all named respondents return their completed questionnaires as soon as possible.
- 4.24 Table 4.1 shows the return rates for all respondents, broken down by whether they received no reminder, a postal reminder only, or a telephone reminder. So, for example, 1,957 respondents were sent a postal reminder only (due to either returning a questionnaire before a telephone reminder was required, or because they provided no contact telephone number). Of this number, 1,070 subsequently returned a completed questionnaire, constituting a return rate of 55%.

**Table 4.1 – Comparative return rates for different reminder groups**

	<b>Sent</b>	<b>Received</b>	<b>Return rate</b>
<b>No reminder</b>	2,441	2,441	100%
<b>Postal reminder only</b>	1,957	1,070	55%
<b>Telephone reminder</b>	2,212	337	15%
<b>TOTAL</b>	6,610	3,848	58%

## 5 Fieldwork figures and response rates

5.1 The figures below cover the 2013-14 Welsh Language Use Survey WLUS (WLUS), which comprises all respondents recruited through National Survey interviews that took place within the survey year for 2013-14 (April 1<sup>st</sup> 2013 – March 31<sup>st</sup> 2014). In order to be included, respondents needed to return their completed questionnaire by May 31<sup>st</sup> 2014.

5.2 In total, 14,771 interviews were carried out on the 2013-14 survey, in which 33,422 individual household members were recorded. Based on responses to the relevant questions (see paragraphs 4.3 and 4.5 above) 8,441 Welsh speakers were identified.

5.3 Table 5.1 shows the acceptance, return and response rates for the adults and children WLUS questionnaires. These figures are calculated as follows:

- Acceptance rate: the proportion of eligible individuals who accepted a questionnaire;
- Return rate: the proportion of accepted questionnaires that were completed and returned; and
- Response rate: the proportion of eligible individuals who returned and completed a questionnaire. This is equal to the acceptance rate multiplied by the return rate.

**Table 5.1 – Overall 2013-14 acceptance, return and response rates (excluding ineligible respondents)**

	<b>Eligible</b>	<b>Accepted</b>	<b>Received</b>	<b>Acceptance rate</b>	<b>Return rate</b>	<b>Response rate</b>
<b>Adults</b>	6,017	4,858	3,030	81%	62%	50%
<b>Children</b>	2,173	1,752	818	81%	47%	38%
<b>TOTAL</b>	8,190	6,610	3,848	81%	58%	47%

- 5.4 Due to the nature of the recruitment process, information on Welsh speakers within the household is provided by the respondent on behalf of other household members. As a result, it is possible for questionnaires to be issued to individuals who do not consider themselves to be Welsh speakers. In order to account for this potential discrepancy, a question was included at the beginning of the questionnaire asking each WLUS respondent whether they can speak Welsh. Anyone indicating that they do not speak Welsh was instructed to return the questionnaire without answering any further questions, to indicate that they are actually ineligible for the WLUS.<sup>7</sup>
- 5.5 As can be expected, there was some variation in eligibility rates across all 22 local authorities covered by the survey. Table 5.2 shows eligibility, acceptance, return and response rates for all local authorities.

---

<sup>7</sup> Ultimately 251 respondents indicated that they were not Welsh speakers and were therefore ineligible for the survey.

**Table 5.2 – Local authority acceptance, return and response rates (excluding ineligible respondents)**

LA	Interviews	Total household members	Eligible	Accepted	Returned	Eligibility rate	Acceptance rate	Return rate	Response rate
Blaenau Gwent	764	1467	180	151	42	12%	84%	28%	23%
Bridgend	667	1444	162	120	59	11%	74%	49%	36%
Caerphilly	694	1348	185	146	64	14%	79%	44%	35%
Cardiff	646	1599	247	193	120	15%	78%	62%	49%
Carmarthenshire	659	1669	645	494	317	39%	77%	64%	49%
Ceredigion	743	1285	763	646	437	59%	85%	68%	57%
Conwy	619	1572	438	360	221	28%	82%	61%	50%
Denbighshire	691	1641	490	416	232	30%	85%	56%	47%
Flintshire	720	1558	340	283	139	22%	83%	49%	41%
Gwynedd	662	1549	1094	929	598	71%	85%	64%	55%
Isle of Anglesey	725	1591	968	785	536	61%	81%	68%	55%
Merthyr Tydfil	696	1430	177	156	67	12%	88%	43%	38%
Monmouthshire	727	1645	210	152	76	13%	72%	50%	36%
Neath Port Talbot	556	1556	229	188	101	15%	82%	54%	44%
Newport	611	1379	139	105	45	10%	76%	43%	32%
Pembrokeshire	672	1314	333	262	153	25%	79%	58%	46%
Powys	608	1640	311	234	146	19%	75%	62%	47%
Rhondda Cynon Taf	724	1423	233	191	113	16%	82%	59%	48%
Swansea	637	1644	297	199	103	18%	67%	52%	35%
Torfaen	636	1461	186	149	50	13%	80%	34%	27%
Vale of Glamorgan	634	1577	213	156	96	14%	73%	62%	45%
Wrexham	680	1630	350	295	133	21%	84%	45%	38%
<b>TOTAL</b>	<b>14771</b>	<b>33422</b>	<b>8190</b>	<b>6610</b>	<b>3848</b>	<b>25%</b>	<b>81%</b>	<b>58%</b>	<b>47%</b>

## **6 Data preparation**

### **Coding**

- 6.1 The WLUS contained no full open-ended questions, but the questionnaire did include two questions with 'Other' options, in which respondents could write in their response rather than selecting one of the pre-written responses. Both questions appeared on the adult questionnaire, and one appeared on the children's questionnaire.
- 6.2 Verbatim responses entered at these questions were coded by a member of TNS BMRB's coding team. They reviewed each response and, if appropriate, allocated the response to one of the pre-existing responses codes. Any response that could not be allocated to a pre-existing response was left in 'Other'.

### **Scanning**

- 6.3 Completed questionnaires were returned to TNS BMRB's operations centre in Warwick. Each returned questionnaire was received by the booking-in team, who recorded the serial number and person number before sending the questionnaire to be scanned.
- 6.4 The questionnaires were passed through a scanner to convert each page into a digital image file. Each image is stored on a computer drive and assigned a unique image reference that can be linked to the respondent's serial number. During scanning, all responses given in the questionnaire are recorded and converted into a separate data file.

### **Data editing**

- 6.5 Due to the unmonitored nature of self-completion questionnaires, it was necessary to subject all returned questionnaires to a data editing process to ensure that the information recorded is consistent and accurate.
- 6.6 Editing was carried out to address instances in which the questionnaire appeared to have been filled in incorrectly; for example where more than one

response was recorded at a single-coded question. There are two main reasons why this may happen. The most common explanation is that a respondent has actually selected more than one response. However, in some instances the scanner may pick up a small mark in one of the responses boxes (for example, if a cross marked in one box spills over into the neighbouring box) and mistakenly record this as a response.

- 6.7 Where there are a large number of erroneous multi-coded responses (known as ‘dirty’ responses), it is standard practice to establish a set of editing rules that are subsequently applied to all ‘dirty’ data. The Welsh Language Commissioner produced an edit specification, outlining how these kinds of responses should be edited, which was applied to the final data by the research team at TNS BMRB.
- 6.8 An additional stage of editing was applied to account for instances in which a respondent had provided inconsistent or contradictory responses. For example, a respondent may indicate in one question that they don’t speak any Welsh, but indicated in another question that they speak Welsh on a daily basis. The rules that were applied to these responses are outlined in the edit specification.

### **Data outputs**

- 6.9 The Welsh Language Commissioner provided TNS BMRB with a data specification document to determine the format and labelling to be applied to the data delivered.
- 6.10 Interim data was produced midway through the data collection period in the form of two SPSS datasets: one containing data from the adult questionnaire and another containing data from the children’s questionnaire. This allowed TNS BMRB to test output production prior to final data being produced.
- 6.11 At the end of the survey year, two final datasets were produced (adult and children). Each dataset contained weighting variables, which are explained in more detail in Chapter 7.

## 7 Weighting

- 7.1 The weighting strategy for the WLUS data incorporates weights produced for the 2013-14 data. Full details for how these weights were produced can be found in the 2013-14 Technical Report.
- 7.2 The weight given to each individual in National Survey sampled households (WalesPopWeight) served as a base-weight ( $W_{\text{base1}}$ ) for the WLUS.<sup>8</sup>
- 7.3 The base-weight ( $W_{\text{base1}}$ ) was then combined with an individual level response propensity weight ( $W_{P(\text{response})}$ ) to create the WLUS weight ( $W_{\text{WLUS}}$ ). The individual level response propensity weight ( $W_{P(\text{response})}$ ) equalled one divided by the predicted probability of response to the WLUS. Probabilities of responses were estimated by means of a logistic regression model with a set of known individual and household characteristics acting as predictor variables.
- 7.4 The definition of 'response' included both (i) completed questionnaires from eligible respondents and (ii) incomplete questionnaires returned by respondents declaring themselves to be non-eligible. Since eligibility status was unknown for non-respondents, it was necessary to combine the eligible and non-eligible respondents - so far as this model is concerned - to maintain symmetry between the responding and non-responding samples. In analysis, non-eligible respondents are, of course, excluded.
- 7.5 To decide which predictor variables to include in our model of response propensity, we tested multiple combinations of candidate variables aiming to achieve a good fit to our survey data without an excessive number of predictor variables. This approach should yield a more robust model than one based on maximum fit to this particular sample.

---

<sup>8</sup> The base-weight  $W_{\text{base1}}$  was the product of three component weights: (a) a design weight that compensates for unequal household selection probabilities in the NSW issued sample; (b) a response propensity weight that gives larger than average weights to areas with lower than average NSW response rates; and (c) a calibration weight that aligns the profile of responding NSW households to a set of relevant population data (specifically the ONS 2011 census-based population projections for 2013).



7.6 The candidate variables were:

- a) National Survey weight SamplePopWeight, used to make the sample of households reflect all-Wales characteristics
- b) Whether an individual was a respondent to the National Survey (as opposed to another household member)
- c) Number of people in the household
- d) Respondent's age
- e) Respondent's gender
- f) Respondent's tenure status
- g) Whether the household included individuals aged under 5
- h) Whether the household included individuals aged under 16
- i) Whether the household included individuals aged under 19
- j) Whether the household included individuals aged over 60
- k) Unitary authority where the household is located
- l) Urbanisation level of the area where the household is located
- m) Acorn classification of the area where the household is located
- n) Overall multiple deprivation classification of the area where the household is located
- o) Whether a household includes a National Survey respondent who stated he/she understands spoken Welsh
- p) Whether a household includes a National Survey respondent who stated he/she can read Welsh
- q) Whether a household includes a National Survey respondent who stated he/she can write Welsh
- r) Whether a household includes a National Survey respondent who stated he/she is fluent in Welsh vs. they can speak a fair amount of Welsh vs. can only speak a little Welsh vs. can say just a few words in Welsh or cannot speak Welsh at all
- s) Whether a household includes a National Survey respondent who stated he/she speaks Welsh daily vs. weekly vs. less often than weekly vs. never

- t) Interaction terms between whether an individual was a respondent to the National Survey and variables o, p, q, r, and s
- u) Whether all people in a household speak Welsh as opposed to a fraction of people

7.7 Ultimately, the following variables were excluded from the final model:

- a) Whether the household included individuals aged under 5
- b) Whether the household included individuals aged under 16
- c) Whether the household included individuals aged over 60
- d) Whether a household includes a National Survey respondent who stated he/she can read Welsh
- e) Whether a household includes a National Survey respondent who stated he/she is fluent in Welsh or they can speak a fair amount of Welsh or can only speak a little Welsh or can say just a few words in Welsh or cannot speak Welsh at all

7.8 Once the WLUS weight ( $W_{WLUS}$ ) was constructed, cases that declared their ineligibility were removed from the dataset and the Welsh-speaking population distribution was estimated.

7.9 The design effect<sup>9</sup> for the WLUS sample was estimated at 1.84<sup>10</sup>

7.10 Tables present the distribution of Welsh speakers for the over 16-year-olds and the under 16-year-olds in the weighted WLUS sample.

**Table 7.1 – WLUS weighted population distribution of Welsh-speaking individuals aged 16+**

<b>Gender</b>	<b>Population count</b>
Female	279,092
Male	221,007
<b>Total</b>	<b>500,099</b>

<sup>9</sup> The design effect cited here is an “unconditional” design effect, if “conditions” are perceived as specific question level correlations with the weights.

<sup>10</sup> The median design effect was estimated as:  $1 + (\text{Standard deviation of } W_{WLUS} / \text{mean of } W_{WLUS})^2$ .

<b>Age</b>	<b>Population count</b>
16 to 18 years old	37,840
19 to 24 years old	58,024
25 to 30 years old	47,031
31 to 35 years old	38,597
36 to 40 years old	32,241
41 to 45 years old	39,683
46 to 50 years old	37,337
51 to 55 years old	35,426
56 to 60 years old	36,063
61 to 65 years old	34,695
66 to 70 years old	33,657
71 to 75 years old	24,026
76 to 80 years old	20,167
81 to 102 years old	25,312
<b>Total</b>	<b>500,099</b>

<b>Region</b>	<b>Population count</b>
Isle of Anglesey	31,937
Gwynedd	66,485
Conwy	29,230
Denbighshire	19,687
Flintshire	20,917
Wrexham	19,525
Powys	21,744
Ceredigion	32,075
Pembrokeshire	19,072
Carmarthenshire	65,300
Swansea	39,590
Neath Port Talbot	18,265
Bridgend	9,307
Vale of Glamorgan	9,794
Cardiff	36,542
Rhondda, Cynon, Taff	20,039
Merthyr Tydfil	3,165
Caerphilly	13,790
Blaenau Gwent	3,627
Torfaen	5,081
Monmouthshire	6,347
Newport	8,578
<b>Total<sup>11</sup></b>	<b>500,097</b>

<sup>11</sup> The total population of Welsh-speaking individuals of 16+ years of age appears 500,097 due to rounding of the weighting factors.

**Table 7.2 – WLUS weighted population distribution of Welsh-speaking individuals aged 3-15 year olds**

<b>Group</b>	<b>Population count</b>
Female	83,198
Male	92,361
<b>Total</b>	<b>175,559</b>

<b>Group</b>	<b>Population count</b>
3 to 6 years old	45,589
7 to 10 years old	55,559
11 to 15 years old	74,412
<b>Total<sup>12</sup></b>	<b>175,560</b>

<b>Region</b>	<b>Population count</b>
Isle of Anglesey	7,484
Gwynedd	13,783
Conwy	6,680
Denbighshire	8,243
Flintshire	7,327
Wrexham	8,217
Powys	8,426
Ceredigion	7,415
Pembrokeshire	5,137
Carmarthenshire	17,986
Swansea	10,963
Neath Port Talbot	3,525
Bridgend	4,469
Vale of Glamorgan	6,496
Cardiff	15,136
Rhondda, Cynon, Taff	13,141
Merthyr Tydfil	3,671
Caerphilly	5,540
Blaenau Gwent	4,021
Torfaen	5,304
Monmouthshire	4,322
Newport	8,273
<b>Total</b>	<b>175,559</b>

<sup>12</sup> The total population of Welsh-speaking individuals of 3 to 15 years of age appears 175,560 due to rounding of the weighting factors.