

Response of Statiaith (Hywel Jones' website, former Welsh Language Board and Welsh Language Commissioner Statistician), 29 December, 2014.

## Gwynedd and Anglesey Housing and Language Survey

Between September and November 2013, a survey was undertaken by Gwynedd Council's Research and Analysis Service on behalf of Gwynedd Council, Isle of Anglesey County Council and Snowdonia National Park Authority. Additional financial support was provided by the Welsh Language Commissioner and Welsh Government. The [full report \(pdf\)](#) explains how the sample for the postal survey was selected:

"3.4 We sampled new houses with completed planning permissions over the period 2007-2011. It was possible to identify which of those were designated as affordable housing.

3.5 Houses sold otherwise within the 10 wards observed during 2008-2012 were also included within the sample.

3.6 Additionally, we also included a proportion (30%) of the houses not sold over that period within the 10 wards in the sample." (page 11)

The 10 wards referred to in 3.5 and 3.6 were not randomly selected, rather, they were purposefully selected (section 3.3, page 10) in an effort to select wards that could be similar to other wards. They might indeed be similar, but they cannot be considered as representative of all wards. Therefore, this is not a classic cluster sample.

All houses with completed planning permissions over the period 2007-11 represents one tier of all Gwynedd and Anglesey's housing. It is therefore possible that the responses received from those are representative of all houses with completed planning permissions over the period 2007-11.

3.5 and 3.6 explain that three tiers were defined within the 10 wards:

1. new houses with completed planning permissions over the period 2007-2011 (part of the wider sample as these were sampled over both counties);
2. houses sold otherwise during 2008-2012;
3. houses not sold over that period.

In order to calculate estimates with statistical validity for the 10 individual wards, it would be necessary to weigh up the estimates to try to ensure that the constitution of estimates would reflect the constitution of housing within the ward. This was not undertaken: all that was undertaken was adding up those who responded, regardless of whether the response came from tier 1, 2 or 3.

Subsequently, to form estimates for 'Gwynedd' and 'Anglesey', all that was undertaken was adding up the estimates for the wards within the relevant authority (6 in Gwynedd: 2 each in Arfon, Dwyfor and Meirionnydd, 4 in Anglesey. In total, there are 71 wards in Gwynedd and 40 in Anglesey). In relation to the Park, the responses which came from referrals within the National Park were added, but again without any element of weighing up.

Furthermore, it is stated: “3.10 The online questionnaire was generally available, so other households had the opportunity to take part in the survey. This means the final results include households which were not targeted as part of the above sample.” It is a mistake to assume that receiving such responses would be of assistance: it is quite the opposite, as they would be likely to exacerbate the bias of responses.

In order to form reliable estimates from the responses to a complex sample plan such as this, how to weigh up the individual responses must be carefully considered so that the results are representative of all the area’s housing - if it is intended to present the results claiming that they are representative. Unfortunately, Gwynedd Council admits that the responses were not weighed up at all. **Despite how detailed the report is, unfortunately it cannot be claimed that it has statistical validity.**

There is another methodological issue where the results on the Welsh language are concerned. The questionnaire was aimed towards the household, but the questions on language requested a response in relation to the individuals in the household. The instruction was to, “Tick one box only for each person”. It is therefore very possible that one person in the household responded on behalf of everyone. (For example, in the Cyngar ward - the ward where the highest number of responses from a household was received (157) - results on language were presented as the responses of 364 people.) That could be acceptable for a factual question such as “In what sector do you work?”. It is not an acceptable tool to collect responses to questions such as “How important is the Welsh language to you?”, “How often do you use the Welsh language in the following situations?” or “In your workplace, which language is mainly used?”