

## Household Analysis Review Group (HARG)

### Quality Assurance of Small Area Household Estimates Data

#### Purpose

This paper briefly outlines the quality assurance (QA) procedure National Records of Scotland's (NRS's) Household Estimates branch uses for its small area households estimates (SAHE) data collection. Household Analysis Review Group (HARG) members are asked for their comments on this process.

#### 1. Background

Since 2007 NRS has published statistics on occupied and vacant dwellings at data zone level on the Scottish Neighbourhood Statistics (SNS) website ([www.sns.gov.uk](http://www.sns.gov.uk)). These are referred to as 'Small Area Household Estimates' (SAHE). The information available on the SNS website consists of percentages of dwellings which are occupied, vacant, second homes, have an occupied exemption or have a 25 per cent Council Tax discount due to single occupancy (or because only one adult in the household is liable for council tax). We also publish information on occupied and vacant dwellings for the 6-fold Scottish Government urban-rural classification, Scottish Index of Multiple Deprivation (SIMD) deciles, National Parks and Strategic Development Plan (SDP) areas in '[Estimates of Households and Dwellings in Scotland](#)', our annual National Statistics publication which can be found on the NRS website. The figures for these geographies are built up (in part for SDP areas) from the data zone level household estimates data.

The SAHE information comes from Council Tax billing systems. As many HARG members will be aware, we produced a specification for Local Authorities (LAs) information technology (IT) providers in 2005/2006 to allow extracts of the data we require for SAHE to be produced. A full data collection was first carried out in 2007 and then in each successive year. We ask that the extracts are produced on the same day as the Council Tax Base (Ctaxbase) return is generated for the Scottish Government. LAs then e-mail their SAHE data directly to National Records of Scotland. The Ctaxbase information is the main source of household estimates at Scotland and local authority level although for certain LAs we require the SAHE figures for a particular aspect of the household estimate calculation.

#### 2. Quality Assurance Involving LAs

The most crucial part of our QA process involves producing summaries of the SAHE data that each local authority has provided and seeking their comments on the summaries. This is especially important when we believe there are problems with the data.

QA information is sent to the person in each local authority who provided the data (usually members of finance teams) and to the Population and Migration Statistics committee local authority side rep (Population and Migration Statistics (PAMS) rep). It consists of an e-mail containing an Excel file presenting their data in various ways, a Word file containing data zone maps of selected categories of dwelling and a written summary of the key

points about their data. The presentation that accompanies this paper will give some examples of the QA files.

The Excel file contains three worksheets and two chart sheets. The first worksheet compares the totals from the data zone level information provided for the most recent year with the Ctaxbase return for that year. In theory the two should be identical as they should be taken from Council Tax billing systems at the same time. Discrepancies can indicate a problem with the data. This worksheet also includes data zone and Ctaxbase totals going back to 2007 and year on year change (both level change and percentage change). Unusual and unexplained changes over time can also be an indication of data quality issues.

The second worksheet shows the data zone level figures for each category and previous years' data zone totals. It's difficult to come to a conclusion about the quality of the data when it's presented in this way so the main purpose of this worksheet is to produce the two charts. The first chart plots the total number of dwellings in each data zone in the latest year against that in the previous year. The second chart plots the data zone totals for the latest year from the SAHE data against those from an extract of the Assessors' Portal taken close to the beginning of that year. Both of these charts are scatter plots and they should show a 'strong' positive relationship i.e. most of the points in the plot should form an approximate straight line which increases from left to right and which lies close to the line which would represent equal numbers of dwellings in each year. Some variation (scatter) is to be expected, particularly in areas where there has been new building or demolition work. Large amounts of scatter or an increase in the amount of scatter compared to earlier years' charts can indicate a problem with the data.

The final worksheet contains information on postcodes which have not been automatically allocated to a data zone. In some cases NRS are able to manually allocate postcodes to data zones by either correcting typographical errors or using information from external sources. LAs may be able to allocate some of the remaining postcodes to data zones or offer suggestions on the suitability of NRS's manual allocations.

The Word file contains maps showing the number of dwellings in each data zone which are vacant, second homes, have 'occupied exemptions' (e.g. all-student households are exempt from council tax) or which have a 25 per cent discount because they are occupied by a single adult (or only one adult in the household is liable for council tax). NRS statisticians visually compare these maps with those of the previous years' data to check for any large unexplained differences. However this is an area of the QA process that can really benefit from the more detailed local knowledge our local authority contacts have, both in explaining any anomalies highlighted by NRS and spotting potential issues.

In the e-mail containing these files, NRS provides a written summary highlighting some key points about the data and flagging potential issues. Where there are issues, we ask LAs if they can provide an explanation or some advice on how best to deal with them. However, even for LAs where there are no apparent issues, it would still be extremely useful if our contacts could look over the data we've sent. They may be able to spot issues that we, with our more limited local knowledge, have missed. Where we have flagged up issues, we'd appreciate it if LAs could investigate. In such cases it is still helpful if the local authority looks at all of the data, not just the area of concern.

### **3. Issues with the QA Process**

This QA process has been in place for a number of years. We really appreciate the input we get from our local authority contacts. However, it can sometimes take a considerable amount of effort to get responses, even when there appear to be serious problems with data. We know that our contacts are busy people and this data collection is not the main part of their job. So we are keen to find out if there is any way we can improve the information we're sending out or help LAs to understand what we're looking for.

### **4. Conclusion**

HARG members are asked to comment on the suitability of the information we provide to LAs as part of the QA process. Are we explaining what we hope to get out of this process clearly enough? Should we provide guidance on what LAs should be watching out for when looking at the data? Any suggestions for how this process could be improved and how the burden on LAs could be reduced without compromising the quality of the data would be much appreciated.