Small area income estimates: quality assurance of administrative data, Oct 2016

The level of quality assurance required for the administrative data used in the production of the small area income estimates (SAIE), which provide estimates of mean weekly household income for middle layer super output areas (MSOAs) in England and Wales.

Contact:

Release date:
24 October 2016

Next release:
To be announced

Table of contents

1 . Introduction

This article briefly describes the level of quality assurance required for the administrative data used in the production of the small area income estimates (SAIE). The SAIE provide estimates of mean weekly household income for middle layer super output areas (MSOAs) in England and Wales. The latest SAIE are for the financial year ending 2012 (April 2011 to March 2012).

This quality assurance article draws on the UK Statistics Authority's toolkit for the quality assurance of administrative data (QAAD) as a base for outlining the level of quality assurance for the administrative data required as part of the production of the SAIE.

An outline of the main features of the different administrative data sources is provided in section 3 of this article together with information about how these features impact the required level of quality assurance. Descriptions of the overall required level of quality assurance relate to both the latest SAIE and the provisional level of quality assurance anticipated for the production of future SAIE products. However, there are some instances of where quality assurance activities for future SAIE products will be different to previous outputs. These differences are described where applicable.

Whilst this article provides an overview of the required level of quality assurance of administrative data, a more detailed description of quality management actions specific to each SAIE product will be available as part of each output's supporting documentation. These will also include descriptions of census data, which is not an administrative data source but does comprise a significant part of the input data for the modelled SAIE.

2 . Level of assurance

The quality assurance of administrative data (QAAD) toolkit includes a risk and profile matrix which allows for a subjective judgement about the administrative data sources used in the statistics with regards to their level of risk of quality concerns and their public interest profile (Appendix A). Table 1 outlines the risk and profile matrix for the administrative data sources used for the production of the small area income estimates (SAIE) for the financial year ending 2012.

Table 1: Risk and profile matrix assessment of administrative data sources

Data source	Level of risk of quality concerns	Public interest[1]	Risk and profile matrix position	Level of assurance
Department for Work and Pensions - benefit claimant data	Low	Medium	A1/A2	A1: Basic assurance
Valuation Office Agency - Council Tax band data	Medium	Medium	A2	A2: Enhanced assurance
Her Majesty's Revenue and Customs - Child and Working Tax Credit data	Low	Medium	A1/A2	A1: Basic assurance
Office for National Statistics - house price statistics[2]	Low	Medium	A1/A2	A1: Basic assurance
Department of Energy and Climate Change[3] - domestic electricity consumption	Medium	Medium	A2	A2: Enhanced assurance

Notes:

1. Public interest refers to the level of public interest in the overall SAIE output.

2. The house price statistics used are now produced by the Office for National Statistics, but were formerly produced by the Department for Communities and Local Government in the production of the financial year ending 2012 SAIE.

3. Now part of the Department for Business, Energy and Industrial Strategy (BEIS).

In addition to the quality assurance carried out by both suppliers of the administrative data and by us, the method of modelling the SAIE enables some of the potential biases of the data to be overcome. In particular, the method of multiple linear regression which is used to produce the SAIE requires some assumptions about the input data (known as covariates) to be met. For example, the method assumes that the covariates are normally distributed, but some of the covariate data sources are not normally distributed. For these covariates, the quality is maintained by transforming the data to a logit and centred scale before including them in the models. This transformation ensures the covariates meet the assumption of normal distribution, which helps to produce a more robust model of income.

3 . Level of risk of quality concerns

Data on the Neighbourhood Statistics Service website

Previous releases of the small area income estimates (SAIE) have used administrative data published on the Neighbourhood Statistics Service (NeSS) website. This website is run by the Office for National Statistics (ONS) which carries out quality assurance procedures on the data supplied by other organisations. Additional quality assurance for data sources which are not designated as National Statistics are carried out before publication on the NeSS website. The details of these are described on the data processing standards and requirements page of the NeSS website.

Quality assurance carried out by the NeSS team helps to significantly reduce the level of risk of quality concerns because any potential quality concerns are identified and investigated with the data supplier ahead of the data's publication on the NeSS website. Primarily, these checks are focused on aggregation of geographical areas to larger areas, derivation checks for data expressed as percentages and trend checks to identify potentially anomalous differences in the data over time.

In addition to the checks carried out by the NeSS team, each producer of administrative data also has quality assurance procedures in place.

The level of risk of quality concerns is also reduced by us ensuring that the administrative data meets the requirements of the modelling process for the SAIE and provide appropriate measures for the covariates used in the models. These are described in the remainder of section 3.

Department for Work and Pensions – benefit claimant data

The Department for Work and Pensions (DWP) publishes National Statistics on the number of claimants of various types of benefits, at the middle layer super output area (MSOA) level. The data come principally from the DWP's statistical summaries product. Benefits for which statistics are available include:

- Disability Living Allowance

- Pension Credit

- Incapacity and Severe Disablement Benefit

- Income Support

- Jobseeker's Allowance

These statistics are produced from the DWP customer information system which is an administrative database of benefit claimant records. The statistics comprise part of the DWP's reporting of unemployment and benefit claimants, and have numerous quality assurance procedures in place which are described in more detail within the publication.

The dataset provides counts of benefit claimants categorised by their statistical group (their main reason for interacting with the benefit). Claimants may be claiming more than one benefit and are therefore categorised according to a benefit hierarchy. The data refer to a snapshot in time and these snapshots are taken at quarterly intervals at the end of February, May, August and November.

One of the main strengths of the benefits data is that, in addition to counts of claimants, data are available as a percentage of the total number of working age claimants. This makes the data more useful for the SAIE as it gives clearer comparisons within and between areas.

Although the information is collected primarily for administrative purposes, the data are used to perform a range of statistical and research analyses, and some operational purposes, to enable further opportunities to evaluate the effectiveness of the benefits system. Another strength of this dataset from an SAIE perspective is that the double counting of claimants of multiple benefits has been removed so that the modelled estimates can draw upon a more accurate picture of benefit-claiming at a small area level. Therefore these data are appropriate for inclusion in the SAIE.

The quality standards applied in their production meet the requirements of the SAIE and they are produced and published on a regular basis. DWP employ various quality assurance procedures between obtaining the record level data at the start of statistical production process, right through to when the summary statistics are published. The initial focus of the quality assurance is on the plausibility of the number of benefits cases and whether changes in the number of claimants represent genuine changes.

In addition to employing general quality assurance procedures, DWP has improved the quality of both short-term benefit claimant statistics and the quality of housing benefit claimant statistics. For short-term benefit claimants, the record level data are now extracted every 2 or 6 weeks (depending on benefit type), which is a significant improvement on the 3-monthly extract used previously. This reduces the likelihood of short-term benefit claims not being included in the published statistics.

One of the limitations of the DWP benefits data is that, for housing benefit claims, all claims in a local authority can be excluded from the data if the local authority does not submit data on time. However, to avoid reporting erroneous figures for missing local authorities, DWP now replaces missing local authorities with plausible, previous data until the problem is resolved. This reduces the impact of missing data and helps ensure the statistics meet the requirements of the SAIE modelling process.

For more information about the quality of the DWP benefits statistics, see the DWP Quality Statement. Overall there is a low level of risk of data quality concerns.

Valuation Office Agency – Council Tax band data

The Valuation Office Agency (VOA) publishes official statistics on the number of dwellings within each Council Tax band at the middle layer super output area (MSOA) level. These statistics are produced using administrative data held on the VOA operational database, which is an administrative database of residential dwellings, constructed for the purpose of assessing and maintaining each dwelling's correct Council Tax band. The statistics are rounded to the nearest 10 to provide an appropriate indication of accuracy.

It is the statutory requirement of the VOA to maintain valuation lists for Council Tax. In order to do this, billing authorities (which are usually local or county councils) are required to notify the VOA of any changes to the dwelling stock. Consequently, one of the main strengths of these data is that they provide an indication of dwelling stock by Council Tax band.

The VOA statistics are not designated as National Statistics which means they have not been subject to assessment by the UK Statistics Authority against the Code of Practice for Official Statistics.

There is a small amount of variability in the length of time taken for a billing authority to provide an update to the valuation list. Despite the limitation of there being variable lengths of time between billing authorities providing these updates, given the relatively large number of dwellings in each MSOA, the potential impact on the Council Tax band distribution and therefore the SAIE, is not considered sufficiently large to affect the estimates.

However, given the potential variability of valuation list updates it is useful to provide an enhanced level of quality assurance of the VOA data specifically for the purpose of their use in the SAIE. Therefore, some additional quality management actions and data supplier communications will be conducted as part of future SAIE production processes. For example, the distribution of Council Tax bands at the MSOA level could be checked against record level Council Tax data received by ONS from VOA. These checks and subsequent discussion with VOA will help ensure an enhanced level of assurance that the MSOA level Council Tax band distribution figures are suitable for use in the SAIE production process.

Overall, the Council Tax band data provide a sufficiently reliable and timely data source for the SAIE.

Her Majesty's Revenue and Customs – child and working tax credit data

Her Majesty's Revenue and Customs (HMRC) publishes National Statistics on the number of families and children in families in receipt of tax credits as at 31 August each year, in a series going back to the financial year ending 2004. These statistics provide a breakdown of the proportion of households receiving tax credits by tax credit element as well as whether the family was benefitting from help with child care costs. The statistics are available for the 2011 Census MSOA geography, which makes them geographically compatible with the SAIE.

Tax credits are a flexible system of financial support designed to deliver support as and when a family needs it, tailored to their specific circumstances. They are part of wider government policy to provide support to parents returning to work, reduce child poverty and increase financial support for all families. The flexibility of the design of the system means that as families' circumstances change, so (daily) entitlement to tax credits changes. This means tax credits can respond quickly to families' changing circumstances providing support to those that need it most.

The published HMRC data provide statistics on the number and percentage of families and children of tax credit claimants by profile position. One of the main strengths of these data from an SAIE perspective is that the statistics are derived from 100% of administrative records and therefore it is not subject to sampling error. The data does, however, exclude any cases where the claimants live outside the UK or where the region or area of the claim cannot be allocated. This makes the statistics particularly suitable for consideration for inclusion in the modelled income estimates.

The small area data is published during the summer, around one year following completion of the entitlement year in question. The delay in publication is the result of the finalisation process built into the tax credits system as well as the time taken to produce and quality assure the statistics.

Most families have until 31 July following the end of the entitlement year to renew their award, reporting their finalised income for the year in question. However, families that report income from Self Assessment have until 31 January of the following year to finalise their income. As a result, the full picture is not known until at least February the year after the entitlement year ends. For example, the 2013 small area data is based on the 2013 to 2014 finalised awards data, but only awards live as at 31 August 2013 are selected for inclusion. This is not a limitation for the SAIE production because the delay in publication is not longer than the time needed to access the published Family Resources Survey data, which provides the income data for the SAIE models.

The overall level of risk of quality concerns for the HMRC Child and Working Tax Credits data is low.

Office for National Statistics – house price statistics

House Price Statistics for Small Areas (HPSSAs) have been produced by us (ONS) since February 2015. Elements of these statistics were formerly produced by the Department for Communities and Local Government (DCLG). These statistics report the count, median price, mean price, lower quartile price and 10th percentile price of all residential dwellings sold and registered since 1995, at the middle layer super output area (MSOA) level. They are calculated using open data from the Land Registry (LR), a source of comprehensive record level administrative data on residential property transactions.

The main strength of the LR Price Paid Data used in the production of these statistics is that they capture the transactions of individual residential properties which have sold for full market value and cover both cash sales and those involving a mortgage. Therefore the data provide a highly reliable statistic for average property price which is suitable for inclusion in the SAIE modelling process.

A further strength of the data is that the registration of a property transaction with the LR is compulsory for all changes of ownership except leases with less than 7 years to run. Solicitors therefore register the transaction as quickly as possible after completion which helps to ensure the timeliness of the data and therefore its quality.

The LR Price Paid Data used in the HPSSAs are taken directly from the sale contract and are audited. Deliberate misreporting of price in this documentation would in most cases be fraud. To minimise errors occurring in transcription, the LR employs quality control procedures to check for exceptions in the data capture process such as price band and application type.

The LR Price Paid Data contain records of each completed and registered residential dwelling transaction. One limitation of the data is that recent transactions can be registered with the LR after the publication of the Price Paid Data and so are not included in the HPSSAs. This is known as registration lag and is predominantly the case for transactions which occurred in the most recent quarter. However, when the next data are published all previous periods are updated to reflect any additional registrations for transactions which took place during those previous periods. This means the effect of registration lag on the HPSSAs is minimised as much as possible and does not affect the SAIE process because of the reference period for these being further back in time than any effects of registration lag.

In addition to quality assurance procedures carried out by LR, we also conduct a number of checks as part of the HPSSA production process. For example, information on duplicate records, geographical coverage and unmatched records is also produced alongside the main HPSSA output to ensure the data do not have errors or significant gaps. This further strengthens the view that the LR data is of sufficient quality to include in the SAIE modelling process.

For more information on what is included in the LR Price Paid Data, see the user guide and for more information on the HPSSAs, see the Quality and Methodology Information. Overall, the HPSSA data are a reliable indicator of actual prices paid and provide a suitable data source for inclusion in the SAIE modelling process.

Department for Business, Energy and Industrial Strategy – domestic energy consumption

The Department for Business, Energy and Industrial Strategy (BEIS) produces annual National Statistics on the consumption of domestic gas and electricity in Great Britain, at the MSOA level. These were formerly produced by the Department of Energy and Climate Change (DECC) until July 2016.

These statistics use administrative data from the gas and electricity supply networks and systems to enable councils and others to monitor and target small areas for further interventions as part of their local energy strategies, and enhance implementation of energy efficiency programmes and thus reduce carbon dioxide emissions. This level of detail in terms of both the statistics and their geographical coverage make the domestic energy consumption data suitable for inclusion in the SAIE modelling process.

One limitation of the energy consumption data at the MSOA level is that it does not include consumption from meters which cannot be allocated to an MSOA. This is mainly as a result of BEIS receiving either a partial postcode or no postcode from the data suppliers. However, we check that the energy consumption data which can't be allocated to an MSOA are not biased, either geographically or in terms of the effect they have on the mix of energy types. This ensures that the remaining energy consumption data are suitable for inclusion in the SAIE.

For some MSOAs, the energy consumption data has been merged with another MSOA to avoid the publication of potentially disclosive data. When an MSOA needs to be merged, the priority is to merge it with other disclosive MSOAs in the same local authority so that as much of the data as possible are retained within a local authority. When this is not possible, it is merged with the non-disclosive MSOA in that local authority with the fewest number of meters. It should be noted that the proximity of the MSOAs is not taken into account in this process.

Another limitation of the data is that the number of gas or electricity meters in an MSOA is not always the same as the number of households in that MSOA. The assumption that the number of meters is representative of the number of households at the MSOA level can be tested to ensure the energy consumption data are suitable for inclusion in the SAIE models.

This test checks that there is a strong positive correlation between the number of ordinary electricity meters and the estimate of the number of households in the most recent census. The number of electricity meters is used in this check because the number of gas meters depends partly on whether an area receives mains gas, so the number of gas meters is not likely to be representative of the number of households in such areas. The check shows that the MSOA level is sufficiently large for average energy consumption per meter to be broadly representative of average household energy consumption and so the data are suitable for inclusion in the SAIE models. Should that assumption not be met, the likelihood is that the data would not meet other inclusion criteria in the models and so wouldn't pass further quality assurance checks during the SAIE production. These checks are described in the SAIE Technical Report.

Improvements to the recording of domestic energy consumption through BEIS's smart meters data is likely to further improve the suitability of the data for inclusion in the SAIE production.

4 . Public interest profile

An assessment of medium public interest has been made in relation to the overall small area income estimates (SAIE) output. This assessment is based on a recognition that whilst the topic of household income and inequality more generally, has an appreciably high public profile, the geographically specific nature of the model-based SAIE product has a relatively narrow user profile. The SAIE have therefore been assessed as having medium public interest to reflect this balance.

The public interest of the administrative data sources used in the production of the SAIE is not considered to exceed that of the SAIE product itself and so the overall level of assurance and associated activities and quality management actions reflect this medium level of public interest.

5 . Communication with data supply partners

For previous releases of the small area income estimates (SAIE), we have used the Neighbourhood Statistics Service (NeSS) website to collate all administrative data and so communication with data suppliers has largely been done via the well-established communication channels used by the NeSS team. However, from the 2013 to 2014 SAIE onwards, the administrative data collection process will source the data from each supplier via the relevant organisation's part of the GOV.UK website.

To reflect the new approach to the collection of administrative data, our communication with data suppliers will now solely be done directly. Initially, communication will be focused on ascertaining the availability of suitable data and its publication schedule (if not already known). Once this has been established, there will be ongoing communication regarding the details of the quality assurance aspects described in section 2 as well as further communication to understand any unforeseen quality findings that emerge during the data collation stage and the subsequent SAIE production.

6 . Appendices

Appendix A – Risk and profile matrix

The critical judgement about the suitability of the administrative data for use in producing official statistics should be pragmatic and proportionate, made in the light of an evaluation of the likelihood of quality issues arising in the data that may affect the quality of the statistics and of the nature of the public interest served by the statistics, as shown in Table 2.

Table 2: Risk and profile matrix

Level of risk of quality concerns	Public interest profile		
	Lower	Medium	Higher
Low	Statistics of lower quality concern and lower public interest profile (A1)	Statistics of low quality concern and medium public interest profile (A1/A2)	Statistics of low quality concern and higher public interest profile (A1/A2)
Medium	Statistics of medium quality concern and lower public interest profile (A1/A2)	Statistics of medium quality concern and medium public interest profile (A2)	Statistics of medium quality concern and higher public interest profile (A2/A3)
High	Statistics of higher quality concern and lower public interest profile (A1/A2/A3)	Statistics of higher quality concern and medium public interest profile (A3)	Statistics of higher quality concern and higher public interest profile (A3)