

Department of Work and Pensions: Benefit Data

This is a best practice example of data dissemination by the Department of Work and Pensions (DWP). This particular case study is most relevant to anyone who wants or needs to disseminate flexible tabular data to a public audience.

When reading this case study it is important to remember that the anonymisation assessment, recommendations and actions are specific to the data examined here although they will have relevance to other similar data and data context.

- **Summary Information**

Organisation disseminating data: Department of Work and Pensions.

Data: Housing Benefit claimant data including the number of people in receipt of benefit and the amount of benefit received, a breakdown of local authority, tenancy type, passported benefits and age group.

Data type: Flexible tabulations, open data.

Summary of sharing, dissemination, publication practices: Disseminated online, at <https://stat-xplore.dwp.gov.uk>, there are no access restrictions.

Risk type: Identification disclosure, disclosure by “differencing” between tables.

Features of risk: Prevalence of small cell counts.

Disclosure Control Methods (SDC): Introduction of random error using a technique called *Random Adjustment*.

Disclosure risk checks: Random adjustment protects against differencing and repeated table requests.

- **Background**

Housing benefit claimant data is administrative data. The data relates to living identifiable individuals and is therefore personal data. Prior to this data being disseminated by the DWP it is anonymised. We outline here some of the core features of this anonymisation process.

The DWP is responsible for welfare and pension policy. Most government data have traditionally been made available in static pre-defined table format, which reduced ease of use and ability for ad-hoc query and answer. To meet customers’ needs DWP developed **Stat-Xplore** which allows housing benefit claimant data to be both more freely available and flexible whilst ensuring that data confidentiality is protected.

- **The Anonymisation Problem**

Small cell counts are a potential feature of this type of data. The presence of small cell counts increases the risk that an individual, within an anonymised dataset, could be re-identified. This is because it is *potentially* easier for a data user to, for example, recognise someone they know such as a neighbour or friend from small cell counts if the following conditions are met: (i) they know that the person is in the dataset and (ii) they know key information about that person that would support recognition and confirm identification with a high degree of certainty.

The problem of small cell counts occurs because many classifications used with the tables have an uneven distribution of data throughout their categories, in particular across geographical areas. When geographical area is cross-tabulated with other breakdowns, such as age/gender/family type, the number in the table cell could be small. Even when variables are more evenly distributed in the classifications, the problem still occurs. The more detailed the classifications, and the more of them that are applied in constructing a table, the greater the incidence of very small cells.

- **Anonymisation Practices**

To address the problem of small cell counts the DWP uses a technique to introduce just enough error into the data to prevent any identifiable data being exposed whilst maximising the utility the data. Selection of this method considered the confidentiality principles associated with the Code of Practice for Official Statistics specifying which disclosure risks are relevant.

This technique *random adjustment* is applied to the aggregate data. What it does is assign a permanent numeric key to each unit record. These keys are used to generate consistent values for the perturbation that is applied to the cells in the table. Whenever the same contributors appear in a cell, the same perturbation will be applied. Perturbation is applied to all cells, to protect against differencing and overcoming the risk of a user undoing protection via repeated requests for the same table.

Random adjustment allows the automatic creation of confidentialised large tables even though they may contain very small numbers (it can do this because there is no certainty as to the exact count). It is not possible to determine which individual figures have been affected by random error adjustments, but the small variance which may be associated with derived totals can, for the most part, be ignored.

Utility/Impact of random adjustment - Introduced random error is an attractive method for protecting data in this case as it allows consistency across tables and allows *any* table to be produced from the microdata whilst still protecting data

confidentiality. The disadvantages of this method are that it generally perturbs all cells (albeit the perturbation could be zero) and is not targeted towards risk so in practice may protect more cells than is strictly required. Moreover introduced random error requires an additivity module to be applied, if additivity is desirable.

The statistical effects are dependent on which distribution has been chosen for adding error. In the case of these data, many different classifications are used in Stat-Xplore tables and the tables are produced for a variety of geographical areas. The effect of the introduced random error is minimised if the statistic required is found direct from a tabulation rather than from aggregating more finely classified data. When calculating proportions, percentages or ratios from cross-classified or small area tables, the random error introduced can be ignored except when very small cells are involved, in which case the impact on percentages and ratios can be significant.

- **Data sharing, Dissemination, Publication**

The data is accessible online using Stat-Xplore which is a web based product developed working with Space-Time Research. It allows users to obtain more detailed data than previously available.

0

Department for Work & Pensions | Stat-Xplore

Home Data Catalogue Explore

Open selection

- Housing Benefit
 - Housing Benefit claimants (14)
 - Table 01 - Region by caseload and award
 - Table 02a - Local Authority by caseload
 - Table 02b - Local Authority by average award
 - Table 03 - Family Type by Age Band
 - Table 04 - Family type by Gender
 - Table 05 - Bedroom entitlement for LHA
 - Table 06 - Employment status
 - Table 07a - Westminster ParIC caseload
 - Table 07b - Westminster ParIC average award
 - Table 08 - Scottish ParIC
 - Table 09a - Travel to Work Areas by caseload
 - Table 09b - Travel to Work Areas by average amount
 - Table 10 - Payment Destination by Region
 - Table 11 - Child Dependents by Housing Sector

Welcome to Stat-Xplore

Stat-Xplore provides a guided way to explore DWP benefit statistics, currently holding data relating to Housing Benefit claimants and in future will include data on a wider set of DWP benefits and programmes.

Getting Started

Use Stat-Xplore to create your own bespoke tabulations and view the results in interactive charts. Simply expand the benefit folder you would like to view on the left hand side. User guidance is available in the 'help' section.

Visualisations

Some pre-prepared interactive visualisations are also available. These can be customised to users needs. To view these, click on the visualisations below:

Contact Us

We are keen to receive feedback and comments on Stat-Xplore. If you have any queries or would like to leave feedback please contact: StatXplore@dwp.gsi.gov.uk

For further information please contact: Elaine Mackey at admin@ukanon.net

