

Working Paper No. 13 (Summary)

ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR EUROPE  
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION  
STATISTICAL OFFICE OF THE  
EUROPEAN COMMUNITIES (EUROSTAT)**

**Joint ECE/Eurostat work session on statistical data confidentiality**  
(Luxembourg, 7-9 April 2003)

Topic (iv): Confidentiality issues for small areas

**Neighbourhood statistics In England And Wales:  
Disclosure control problems and solutions**

**Invited paper**

Submitted by the Office for National Statistics, United Kingdom<sup>1</sup>

---

<sup>1</sup> Prepared by Phil Armitage (phil.armitage@ons.gov.uk).

## **Neighbourhood Statistics in England and Wales: disclosure control problems and solutions**

Phil Armitage, Methodology Group, Office for National Statistics, London.

The Neighbourhood Statistics project has the ambitious aim of providing a detailed profile of every neighbourhood throughout the country. It will collect together information derived from Census and Surveys conducted by the Office for National Statistics and other government departments, and also derived from administrative sources throughout central and local government. The information provided, all on a common geographic basis, will include basic demographic data on births and deaths, various indicators of the health and wealth of the local population, levels of government support (e.g. welfare benefits) provided, educational performance of the area's children, local access to key services, and many other descriptions. This project has been driven by the needs of neighbourhood renewal: a concerted effort on the part of central government to identify and remedy inequalities throughout the country at a local level in social and physical environment, resulting in great differences in opportunity for economic advancement, and in quality of life. The aim therefore is to obtain a detailed picture at local levels of deprivation of various kinds and inequalities in opportunity and performance. As deprivation is often very localised, these new statistical descriptions were required for small areas with relatively low populations: the smallest units within the project have target populations of 200-250 people. A further motivation is to monitor local changes on an ongoing basis, so that the success or otherwise of policies to remedy deprivation and inequality can be continuously assessed. This paper deals with Neighbourhood Statistics in England and Wales; separate but related similar projects are underway in Scotland and Northern Ireland.

Virtually all the outputs provided will be aggregate statistics of some kind. Nevertheless, there are substantial disclosure risks arising from the very small geographical scales involved, and the fact that multiple sets of government data covering very many facets of the local population and environment will be brought together and available for public perusal from a single web site. Many of the datasets will be from administrative sources and will therefore be effectively complete censuses, further increasing the disclosure risk. The risk from this source is not perhaps as serious as it at first sight seems, because of the biases through incomplete coverage, and the different population frames and methods used for some administrative sources, but nevertheless the risks are much greater than when reporting the results of sample surveys.

There are also potential disclosure control problems arising from the continually changing administrative units for which descriptions are required. In the United Kingdom, the Boundary Commission is continually revising administrative boundaries - which are also often electoral boundaries - to adjust for changes in population. The initial solution proposed for Neighbourhood Statistics is to provide all outputs for standard building blocks related to administrative areas fixed at one date, and to only change the geographical definition of these building blocks very occasionally. All outputs for these building blocks will have adequate disclosure control applied, so that confidentiality of individuals contained within the areas is appropriately protected. If information is required to be aggregated to different boundaries for administrative purposes, then this will be done using some synthetic apportionment method. The original data will not be aggregated to different boundaries, and so there will be little opportunity for disclosure by differencing the aggregate statistics for two closely related boundary sets.

Other issues arise because the disclosure control will often require to be exercised at source. The government departments providing the administrative data have the primary responsibility for protecting the confidentiality of the people they report on, and in many cases, are not yet legally empowered to transfer confidential data to other government departments, even for statistical purposes. So they will generally apply disclosure control to their aggregate statistics before transfer to the Neighbourhood Statistics project staff for display on the Neighbourhood Statistics web site. This means that there are substantial coordination problems to ensure adequate and reasonably uniform standards of data quality etc, and confidentiality protection, are applied.

Many of the datasets that can be reported at a local geographical level report information about households and individuals. Business information is generally restricted to small businesses, and as yet, there is little business information reported on the Neighbourhood Statistics web site. Some variates (e.g. mean or median individual income) planned for inclusion on the web site will be essentially small area estimates, and as they involve complex models of the response being reported, will carry little disclosure risk. Many of the tabulations for which some disclosure risk occurs are of frequency tables, and more attention has so far been given to improving disclosure control methods for these types of outputs. Because of the many tables describing the same population with similar classifications (e.g. gender and broad age group), there are many opportunities for linkage between tables that could substantially increase the disclosure risk.

This paper will describe methods we have devised for systematically assessing the weaknesses of one standard method - random rounding - for frequency tables arising because of table linkage. Potential weaknesses in this methodology have been known for some time, but, so far as we know, few systematic methods are available as practical tools for systematically monitoring and remedying these weaknesses. The longer term goal is to use controlled rounding with inbuilt protection levels, as advocated by Fischetti and Salazar (1998, *Journal of Official Statistics*, 14, 553-565). In the meantime, some other methods to reduce disclosure risks are required and these will be described. The paper will also briefly describe the main features of the overall approach to the management of disclosure control in this complex project.