Working Paper No. 14 (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

Disclosure control problems associated with geography

Invited paper

Submitted by the Office for National Statistics, United Kingdom[1]

[1] Prepared by David Brown (david.brown@ons.gov.uk).

Disclosure Control Problems Associated with Geography

David Brown, Methodology Group, Office for National Statistics, London

Geography is frequently assigned a special status in disclosure risk evaluations and during development of disclosure control treatments. There are two main reasons for this in the United Kingdom, which apply to other countries to varying extents, depending on how administrative areas are defined, and whether statistics for small areas are routinely produced.

First, there are frequent changes in the administrative boundaries used for production of government statistics in the UK, especially at 'ward' level (including populations of between 1,000 and 10,000 people). Administrative authorities always require tabular statistics for the current boundaries for each set of data. But they also need to assess changes in time, so they usually require tabular statistics derived from the same microdata aggregated to both the old and the new boundaries. As the boundaries change gradually from year to year, this frequently results in many cases of disclosure by differencing the statistics aggregated to the two closely related boundary sets. In particular when the area defined by one boundary, say A, is a subset of the other, say A + dA, either for the whole population or a subclass, tabular statistics for the difference dA can be derived. If the disclosure control treatment is one that focusses particularly on low frequencies, then small frequencies (often resulting in potential disclosure) for the small area dA can often be obtained by differencing the two tables.

Secondly, statistical tables are routinely produced in the UK for areas with target populations as low as 200-250 people, what in the UK are called 'census output areas'. This raises problems because low frequencies are frequently obtained, resulting in potential for disclosure if precautions are not taken. It is more difficult to adequately protect such tables using standard disclosure control methods, especially without doing substantial damage to the tables. So there will be more instances of potential disclosure occurring for these two reasons. Once a potential disclosure has occurred, this disclosure can be more easily converted into an actual disclosure since the area to be searched in in order to identify the potentially disclosive case is frequently a small compact area. It will frequently happen that detailed information on most of the population contained with such a small area will be within a single person's knowledge. Furthermore, in cases where the intruder does not already posses the knowledge, it is a relatively easy matter to collect information about the residents of such an area, merely by visiting the area, and conducting an investigation. So the risks of converting the potential disclosure into an actual disclosure are much greater. Therefore in order to manage the total risk adequately, we need to reduce the incidence of potential disclosure to a much lower level.

There are various possible approaches to solving these disclosure problems relating to geography, two of which are discussed in this paper. One involves providing adequately disclosure protected tabular outputs for a standard set of building blocks, and obtaining *estimates* of the outputs for other areas by synthetic means (using methodology related to small area estimation techniques). The second involves a pretabulation method involving perturbations of the geographical position of sets of households, similar to record swapping applied within the US and UK Censuses. Initial investigations of the method's properties are briefly discussed.