EUROPEAN **F**ORUM FOR **G**EOSTATISTIC**S**

SOFIA CONFERENCE 2013

23 - 25 October, Sofia, Bulgaria

"The proof of the apple is in the eating!"

EUROPEAN FORUM
FOR GEOSTATISTICS

ABSTRACT

Title: Sensitivity analysis of disclosure control measures.

Author 1: David MARTIN

University of Southampton, United Kingdom

Author 2: Klaus STEINNOCHER

Austrian Institute of Technology, Austria

Author 3: Ekkehard PETRI, Eurostat – European Commission, Luxembourg

In the course of the GEOSTAT project a number of tasks were performed in order to better understand the potential and quality of disaggregated population grids. This presentation describes a sensitivity analysis undertaken to assess the effect of various statistical disclosure control thresholds on the population grid dataset, particularly when combining specific population characteristics. The sensitivity analysis covers the full extent of the GEOSTAT 2006 population grid. This grid itself contains only population estimates, but demographic data from 2001 are available on LAU2 areal units relating to sex, age groups, current activity status, etc.

To study the effect of disclosure control, model datasets were created reflecting the range of frequency distributions of demographic characteristics that exist in different countries, particularly in very rural areas. The challenge was to redistribute the population of each LAU2 area onto the inhabited cells of the GEOSTAT 2006 grid so that the totals in each demographic sub-group are preserved within the areal units and also the population totals within the grid cells. In order to inform this modelling, auxiliary distributions for the demographic variables were generated from gridded or small area data for Norway,

NATIONAL STATISTICAL INSTITUTE
REPUBLIC OF BULGARIA

WITH THE SUPPORT OF

eurostat

EUROPEAN FORUM FOR GEOSTATISTICS
SOFIA CONFERENCE 2013
23 - 25 October, Sofia, Bulgaria
"The proof of the apple is in the eating!"

EUROPEAN FORUM
FOR GEOSTATISTICS

ABSTRACT

Austria and England. These distributions were expressed in terms of cell–level proportional offsets from LAU2–level means and were generated separately for three population size ranges for each variable. The following variables were used: pop > 65, male pop > 65, women * employed, and women * employed * in area.

The method examines each cell value, which is initially just a LAU2 mean, and applies an adjustment (up or down) drawn randomly from the relevant auxiliary distribution. After all the grid values in a LAU2 have been adjusted, they are rescaled to ensure that the correct total across the LAU2 is preserved. No effort was made to adjust LAU2s with very small numbers of cells or very small populations.

Starting from these pseudo–randomly distributed datasets, four thresholds (3, 10, 30, 50) were applied to the adjusted variables, resulting in estimates of the numbers of cells to be suppressed for confidentiality reasons using each threshold. The thresholds reflect current national practices in grid data disclosure control.

As expected, the greatest disclosure risk occurs in countries with large remote areas and within the smallest population groups. This can be seen clearly for the smallest threshold (3) and the variable "pop > 65". While the disclosure affects only 20% of the cells in Italy, it goes up to 60% of the cells in Scandinavian countries. For the "women * employed * in area" these values rise to 40% in Italy and almost 70% in Scandinavia. Looking at the largest threshold (50) values go up to 98% for "male pop > 65" in Finland, but also reach 84% in Italy. With the exception of Italy, from the threshold 10 and higher more than 50% of all grid cells showing the population breakdown 65 and older have to undergo confidentiality treatment with even higher values for combinations of breakdowns. With regard to the variables, "male pop > 65" is most affected, followed by "women * employed * in area". While the percentage of cells affected at threshold 3 ranges from 20% to 70%, for threshold 50 the range covers 70% to 98%.