

SOLUTIONS FOR USING SENSITIVE MICRODATA IN INTERNATIONAL REGISTER-BASED STUDIES

Reijo Sund¹

¹ Centre for Research Methods, Department of Social Research,
University of Helsinki, Finland

Population-based registers may contain nationwide data on microlevel events with direct identifiers. Registers are an invaluable source of data for researchers although the secondary nature of register data and requirements for confidentiality pose some problems for utilization and study designs.

These problems magnify when data from several countries are to be used simultaneously. First step is always to prepare common standardized data sets for each country. It may be very challenging to find a satisfying compromise for the data structure, but usually it is the only way to proceed.

If confidentiality was not an issue, the simplest way for combined analysis would be to collect all data in the pooled database and use microdata as such. Another option is to keep all data separately, but to conduct a controlled meta-analysis by performing exactly same analyses in each country by shared analysis scripts and by pooling only the required summary statistics for the actual meta-analysis.

An in-between possibility is to share only sufficient statistics or other suitable reduced-form data for certain type of models. One approach, enabled by the additivity of the log-likelihood contributions and requiring specialized computational solutions to connect the different databases, is to iteratively share the values of log-likelihood functions at requested parameter values, without sharing any confidential microdata. Another and usually simpler way is to transform data to a suitable tabulated form such as the usual data format for Poisson regression analysis or to some other more general summary of time-specific event histories (person-moments) suitable for fitting other specific types of models.

In conclusion, if a common standardized data format across countries for microdata of interest is found, there is usually a possibility to do the required confirmatory analyses as well. Free form exploratory analyses are more difficult and often an unfeasible approach unless data can be pooled. Remote access systems may help to pool sensitive data, but may also prevent the pooling totally if country specific data are available only in country specific remote access systems.

Keywords: Registers, Confidentiality, Data Sharing.