Epidemiological and Statistical Secured Matching in France

Catherine Quantin[1-2], Benoît Riandey[2]

1- CHRU Dijon, Service de Biostatistique et d'Informatique Médicale (DIM), Dijon, F-21000, France. Contact author : catherine.quantin@chu-dijon.fr

2- Inserm, U866, Univ de Bourgogne, Dijon, F-21000, France

3- Institut National d'Etudes Démographiques, Paris, France

In 1969, in their innovative founding article, Ivan Felligi and Allan Sunter opened the way for the inter-sectorial processing of administrative files such as the various chapters of a survey questionnaire. With their probabilistic matching technique, they brought a solution to the knotty problems of the absence of identifiers or the legal restriction of their use.

Countries in Scandinavia and Northern Europe took advantage of the wealth of information in the social files of their social democracies as well as national confidence in their institutions to develop very effective statistics based on matching. This even allowed them to dispense with the collection of census data. The wealth of information in French administrative files gave rise to the hope that an equally effective statistics service could be created. However, in 1978, fears remaining from the period of Nazi occupation led to very restrictive legislation. In any case, it prevented the matching of administrative files, and thus led to their analysis as individual items.

The need for the efficient management of healthcare expenditure led epidemiologists to find an effective and secure solution based on the hashing of identifiers. There are many applications in epidemiology and in healthcare economics, whether it concerns the anonymous counting of patients with AIDS and other notifiable diseases, or matching care episodes for the same patient in different hospitals and healthcare establishments, or economic information related to healthcare costs and their management.

Such hashing techniques (SHA algorithm) have enabled this progress to take place, but they do have a limit: fundamentally, they can only be used if matching was planned during the initial stages of the project. Any extension of the project following a change in the analysis plan or a new participation is impossible. We have thus imagined a secure matching technique that allows controlled enlargement of the scope of investigation without impinging on the level of confidentiality. The technique associates hashing using a public key and encryption (reversible) with a secret key. This overcomes the above-mentioned limit while maintaining a high level of security. On our suggestion, this technique was implemented for healthcare statistics in Switzerland.

This technique may lead to the emergence of a new system for official French statistics, in support of the new centre for secure remote access to data (CASD) of the CREST-INSEE. Let us hope that current reflection on this strategy will quickly lead to ambitious projects.

Keywords: Record linkage, hash-coding, Security, Non-reversible encryption, Epidemiological survey