

# Data fusion in pharmaceutical marketing: new perspective from administrative data.

Paolo Mariani<sup>1</sup>

**Abstract** Through public administration digitization as prescribed in E-Gov 2012, the availability of data open to pharmaceutical marketing in Italy will increase substantially thereby permitting data fusion between administrative and survey information.

**Keywords:** administrative data, data fusion, pharmaceutical marketing.

## 1 Scenario

The pharmaceutical industry is transforming rapidly. The changes are not only due to patent expiries but above all to widespread computerization in both the public and private sectors offering a panorama of particular interest to all health actors.

By 2012, all medical services should be simplified and digitized: prescriptions, certificates, national online reservation, health records, networking of all general practitioners and NHS pediatricians. Reading, analysis and information sharing tools will have to be adapted to the changes so that companies can compete in the market.

As the main source of information will be provided mainly by administrative data rather than by ad hoc surveys, companies will have to quickly correct the power factor of their activities according to data source changes and culture reference models.

In addition, in a consumer society, we see ourselves as individuals and as free agents when we exercise consumer choice, therefore, it is not difficult for pharmaceutical companies and other privatized healthcare deliverers to convince us that it is empowering to think of ourselves not as patients but as consumers (Applbaum K., 2006). It is often said that leading drug companies now spend more on marketing than on research and development (Angell M., 2004).

## 2 Pharmaceutical Administrative Data

Administrative data is already routinely used for the study of income distribution, for data validation and evaluation of measurement errors in sample surveys. A few more contributions are found in application-oriented enterprises especially in marketing. Multiple data sources such as a sample survey and administrative records can be combined (Jenkins S.P., Lynn P., Jäckle A., Sala E., 2008). Administrative data can be defined as "information gathered and stored by public institutions for the purposes of monitoring or intervention of individuals or other entities" (Fortini M., 2000).

Information provided by administrative data can be used for various purposes. Company marketing pays particular attention to pharmaceutical pricing data collected by the Ministry of Health through Sogei<sup>2</sup> and NHS data regarding general practitioners (GPs) and pediatricians.

Appropriately treated data becomes a source of statistics in its own right as well as a solid support base for the validation of data collected in other statistical summaries. Administrative data from a simple output of administrative processes and management can assume the role of privileged information support for the decisions and acts of public administrations and private companies so that

---

<sup>1</sup> Paolo Mariani, Unimib; email: paolo.mariani@unimib.it

<sup>2</sup> Information and Communication Technology company owned by the Ministry of Economy and Finance.

they are able to provide targeted and timely responses to public needs. Thereby, they also foster the development of the entire national economic system and benefit from research as a driver (Calzaroni M., 2008). In the specific case of public administration, multiple repositories from various sources need to be integrated as none of them is complete in all research aspects but each potentially provides information and relevant specifications. (Aimetti P., Zavanella B., 2004, Martini M., 2000).

Many firms routinely compile databases reporting transactions with their customers (database marketing) which can be of great value for direct/micro marketing. When the prescription 'database marketing' (from the administrative archive) becomes available, companies will be able to extract relevant information to take competitive advantage. (Wedel M., Kamakura W., Bockenholt U., 2000).

### 3 Data Fusion in Pharmaceutical Marketing

Various terminology describes the fusion or integration of data sets, for example: multisource imputation, data attribution, data fusion, statistical record linkage, statistical matching, micro data set merging, and so on. Data fusion is generally defined as the use of techniques that gathers and combine data from multiple sources. Fused data is expected to be more informative than the original inputs. Furthermore, data fusion is integration followed by reduction or replacement whereas data integration is the combining of data.

The matching of subjects considered in all archives requires special techniques of integration of information as the data relating to the same statistical unit is contained in multiple repositories.

The general problem of data fusion can best be illustrated using a schematic representation (van der Putten, P., Kok, J.N., 2010) in Figure 1. In this representation data set A is the GPs database and contains knowledge and information (represented by J items) from physicians. Data set B contains knowledge and information (represented by J + 1 items) from a small group of physicians. The first amount of knowledge and information (represented by the first J items) for a single customer is the same in each data set. However, from the small group of physicians in data set B there is some additional knowledge and information, that is, item J + 1. The goal of this technique is to fuse the extra knowledge and information in data set B, that is, item J + 1, to data set<sup>3</sup>. As a result of this data fusion, the knowledge and information about item J + 1 becomes known for all physicians in the database, data set A. (van Hattum P. and Hoijtink H., 2008)

**Figure 1:** Data Fusion for database marketing

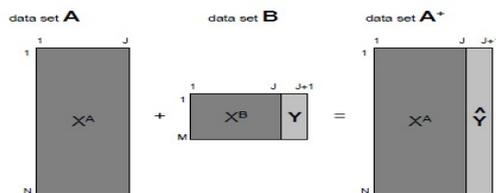


Figure 1: Schematic representation of data fusion in marketing (derived and adjusted from Van der Putten et al.<sup>5</sup>)

Source: van Hattum P. and Hoijtink H., 2008.

While awaiting the new administrative data, an example of the potential of new sources can be inferred from the use of a joint first data set A that collects administrative data of the entire population of general practitioners working in Italy and a second data set B containing the results of a survey conducted by an independent international market research on a sample of 3.000 GPs (usually 5-8% of the universe) working in Italy and drawn from the universe defined by the first data set. Both data sets do not show variables with high discriminating power, such as name and surname of the doctor, for reasons of data privacy.

The variables common to each physician within the two files are:

- the birth year
- the graduation year
- the number of patients
- the work district
- the medical specialization

<sup>3</sup> Nearest neighbor, logistic regression and based clustering approach are methods widely used.

Two very interesting variables can be found in the investigation sample: the amount prescribed in terms of packs, and the molecules of products prescribed to the last 10 patients.

## 4 Final remarks

In this paper, we started by discussing the increase in information and how data fusion can be a possible solution for data management. We presented a brief overview of the new administrative data becoming available and briefly described a marketing database enriched by data fusion.

The resulting database can provide marketing information to reach customers at various levels: as a customer/physician/patient/company/etc. Further research will be required to evaluate the validity of the fused data and to develop mining algorithms.

## References

- 1 Aimetti, P., Zavanella, B.: Qualità degli archivi amministrativi e qualità dell'informazione statistica, FrancoAngeli, Milano (2004)
- 2 Angell, M.: Over and above: Excess in the pharmaceutical industry, Canadian Medical Association Journal, vol. 171 (2004)
- 3 Applbaum, K.: Pharmaceutical Marketing and the Invention of the Medical Consumer, PLoS Medicine, Volume 3, Issue 4 (2006)
- 4 Calzaroni, M.: Le fonti Amministrative nei processi e nei prodotti della statistica ufficiale. IX Conferenza Nazionale di Statistica: Leggere il cambiamento del Paese, Roma (2008)
- 5 Fortini, M.: Linee guida metodologiche per rilevazioni statistiche. ISTAT, Roma (2000)
- 6 Jenkins S.P., Lynn P., Jäckle A., Sala E.: The Feasibility of Linking Household Survey and Administrative Record Data: New Evidence for Britain, International Journal of Social Research Methodology, 11, pp. 29-43 (2008)
- 7 Martini, M.: L'integrazione statistica di registri e indagini. Tecnologie informatiche e fonti amministrative nella produzione di dati, FrancoAngeli, Milano, pp. 291-305 (2000).
- 8 Ministero per la Pubblica Amministrazione e per l'Innovazione, Piano e-gov 2012, Available on [www.e2012.gov.it](http://www.e2012.gov.it) (8 Feb 2012)
- 9 van der Putten, P., Kok, J.N.: Using Data fusion to Enrich Customer Database with Survey Data for Database Marketing. Marketing Intelligence Systems, Springer-Verlag Berlin Heidelberg (2010)
- 10 van Hattum, P., Hoijsink, H.: The Proof of the Pudding is in the Eating Data Fusion: An Application in Marketing, Utrecht University, The Netherlands (2008)
- 11 Wedel M., Kamakura W., Bockenholt U., Marketing data, models and decisions, International Journal of Research in Marketing, 17, (2000)