

Impact of unmeasured covariates on bias and statistical power in health administrative databases: a simulation study

Mélanie Née^{1,2,3}, *Marta Avalos*^{1,2,3}, *Ludivine Orriols*^{1,2}, *Emmanuel Lagarde*^{1,2}

¹firstname.lastname@isped.u-bordeaux2.fr, Univ. Bordeaux, ISPED, F-33000 Bordeaux, France

²INSERM, Centre INSERM U897–Epidémiologie–Biostatistique, F-33000 Bordeaux, France

³INRIA SISTM Bordeaux – Sud-Ouest, F-33405 Talence, France

French health insurance databases (SNIIRAM) cover the entire French population. These databases include demographic (age, gender, city of residence), and out-hospital reimbursement (drug dispensing and long-term diseases). The use of these administrative databases for epidemiological research has the strengths of being readily available and relatively inexpensive. Also, the large number of patients, without loss of follow-up, allows for sufficient powering of studies. Furthermore, the information is large, comprehensive and detailed, without any exclusion.

An example of use is given by the CESIR (Combination of Studies on Health and Road Safety) project. To assess the impact of medicines use on the risk of injury road traffic crashes, data from the health care insurance database were matched with data from the National police database of injurious road traffic crashes. More than seventy thousand drivers involved in an injurious crash in France, between July 2005 and May 2008, were included in the study. Their reimbursement data for drugs dispensed within six months of the road traffic crash were retrieved.

Administrative databases are not without limitations. Concerning medical drugs, no information about the use of over-the-counter drugs, prescription drug misuse or medication adherence is available. Diagnoses of chronic diseases could be inaccurate since data are collected and coded in an “administrative way”. There is also a lack of information on potential confounders.

In this work we explored different observational study designs that can address the research question of the CESIR project (case–control, matched case–control, case–crossover, case series. . .). For each study design, we aimed to evaluate the impact of unmeasured confounders on bias and statistical power through data simulation. The simulation study was set up to mimic the real CESIR data in several respects. To generate event times conditional on time-dependent covariates, we adapted the permutational algorithm implemented in the publicly available R package `PermAlgo`. We focused on two medicinal drugs: Benzodiazepines (whose effects on crash risk are well established in the literature) and Antihistamines for systemic use (whose effects are controversial). Our results allowed us to develop several recommendations to guide future analyses of the second phase of the CESIR project (comprising the period June 2008 to December 2011).

Keywords: Observational study designs, Permutational algorithm, Pharmacoepidemiology.