Meeting abstract

Using administrative data for research: the importance of appropriate statistical techniques

J Perelman* and C Mateus

Address: Escola Nacional de Saude Publica, Universidade Nova de Lisboa, Lisbon, Portugal

* Corresponding author

This abstract is available from: http://www.biomedcentral.com/1472-6963/9/S1/A17

Introduction

Administrative data routinely collected at hospitals are attractive for researchers: they are large, often exhaustive, and of relatively easy access. However, they are not intended for research, and they lack the clinical details of observational studies or clinical trials. Researchers thus face a trade-off between using large but incomplete databases versus using detailed but often poorly representative ones. One of the major limitations of missing information in administrative data is that endogeneity cannot be corrected due to the non-observability of the characteristics of some patients.

Let us suppose that we seek to evaluate the impact of a given treatment on a patient's health. The decision to treat a patient is not random in real practice, contrary to what occurs in clinical trials. In the "real world", patients are selected into treatment arms based on their expected outcomes. Hence, the explanatory variable (treatment) is endogenous, as it is explained by the dependent variable (outcome). This problem would be solved if one could control for a large array of patients' characteristics, in order to estimate the differences between the treated and the untreated. Unfortunately, this is not the case with administrative data.

In the present study, however, we postulate that appropriate statistical techniques can help reduce this problem. To do so, we examine the impact of invasive treatments for cardio-vascular disease - percutaneous coronary intervention (PCI) - and coronary artery bypass grafting (CABG) on in-patient mortality, using administrative data from Portuguese NHS hospitals. We examine how outcomes vary whether we account for endogeneity or not. Then, we examine how the selection bias spreads to other indicators, namely, the differences between men's mortality and women's mortality following invasive treatments.

Methods

We study patients admitted for cardio-vascular disease at NHS hospitals in Portugal for the 2000-2006 period (diagnoses were selected using ICD-9-CM codes). Since cardio-vascular diseases are mostly treated at NHS hospitals, this offers us an exhaustive data set representative of national patterns of treatment. Patients are selected according to their principal diagnosis and grouped according to the HCFA-DRG classification. Our final sample includes 259,519 discharges from 57 hospitals.

First, we consider a simple probit model to measure the impact of invasive treatment on in-patient mortality, with in-patient mortality as a dependent binary variable (0/1), controlling for the patient's age and comorbidities. Indeed, our data do not provide further details on the severity of disease (in particular, the ejection fraction and number and type of affected vessels). Then, we estimate the impact of treatment, controlling for endogeneity through the use of a recursive bivariate model, which consists in assuming that allocation to treatment is non-random and endogenous to mortality.

The basic idea of the model is that mortality and treatment can be thought of as two latent variables from a bivariate normal distribution. Hence, we assume from the

start that there is a correlation between the error terms of both variables, i.e., that there are unobservable variables that affect both mortality and treatment. Then, we compare the findings between the simplest model and the recursive bivariate model.

Results

Without accounting for endogeneity, we observe that patients treated by PCI have a 51% likelihood of dying during hospitalization. When controlling for endogeneity, the reduction in in-patient mortality increases to 87%. As regards CABG, treated patients have a 12% lower mortality ratio on average with the simple binomial model, and a 76% lower mortality ratio using the recursive bivariate model. In both cases, the discrepancy in results indicates that the endogeneity bias is large, and that treated patients have some characteristics which make them more likely to die. Hence, the impact of treatment is under-estimated using the simple model.

As regards the differences between men and women, we observe a similar pattern. Women have a 3% higher likelihood of dying during hospitalization after PCI according to the simplest model, for a 6% lower mortality ratio when controlling for endogeneity. In this case, the discrepancy in results is even more dramatic, since the sign of the inequality is reversed. Similar variations are observed for CABG.

Conclusion

Our study indicates the relevance of using appropriate statistical techniques when relying on administrative data for clinical research. However, our outcomes also show that, when using more sophisticated techniques, we obtain results with administrative data that are comparable in sign and magnitude to those obtained from observational studies. This should encourage us to pursue investigation using administrative data, but with a proper adjustment for the lack of detailed patients' characteristics.