International Journal of Population Data Science

Journal Website: www.ijpds.org

Analysing complex linked administrative data in health services research: Issues and solutions

Moorin, R¹, Youens, D¹, and Harris, M¹

¹Curtin University

Introduction

Linked administrative data are increasingly being used to evaluate the impact of health policy on health-service use/cost because they can comprehensively capture whole of population interactions with the health system. These analyses are complex comprising unbalanced panels and are at risk of endogeneity and associated problems.

Conclusion/Implications

Availability of linked data, together with increases in computing power, has vastly increased its potential for use. This has also increased the complexity of analyses being undertaken necessitating recognizing and addressing problems, such as endogeneity, that arise due to the observational nature of the studies undertaken.

Objectives and Approach

We evaluated the impact of changes in regularity of general practitioner contact on diabetes related hospitalisation before and after care coordination policies using whole of population, person-level linked primary care, hospital, Electoral Roll and death records. Complex panel random-effects modelling techniques were required due to the unbalanced structure of the data (individuals could exit and re-enter the study repeatedly), over-dispersion and high proportion of zeros, changes in availability of tests (ascertainment bias), the likelihood of prior health service use influencing the dependent variable (initial conditions and simultaneity/reverse causality bias) and likely correlation of observed and unobserved variables.

Results

Multivariable zero-inflated negative binomial and Cragg-hurdle clustered robust regression, which include separate components to model zero and non-zero outcomes, were required for these data. Mundlak variables (group-means of time-varying variables) were used to relax the assumption in the random-effects estimator that the observed variables were uncorrelated with the unobserved ones. Prior health service use was adjusted for using 4-year lags of GP contact and one-year lag of hospitalisation. The initial value of the dependent variable resolved the “initial condition” problem. Ascertainment bias was addressed using the number of years available for identification for each person as a covariate. AIC/BIC values were used to identify the best model. We found that more regular GP contact was associated with fewer hospitalisations, however this attenuated over time.