

Exploring Alternative Designs using ‘Big’ Administrative Data

Roos, L¹, Wall-Wieler, E¹, and Torabi, M¹¹University of Manitoba

Introduction

Large population-based data sets present similar analytic issues across such fields as: population health, clinical epidemiology, education, justice, and children’s services. Step-wise approaches and generalized tools can bring together several pillars: big (typically administrative) data, programming, and study design/analysis. How can we improve efficiency and explore alternative designs?

Conclusion/Implications

Presenting design alternatives is straightforward and applicable across disciplines. Ongoing work is facilitating comparisons of ‘experimental’ and control groups. Literature-derived guidelines and simulation-based techniques should lead to better design decisions. Automated model assessment can help analyze robustness, statistical power, residuals, and bias, suggesting artificial intelligence approaches.

Objectives and Approach

Linked data sets typically contain: 1) files presenting longitudinal histories 2) substantive files noting various events (concussions, burns, loss of a loved one, public housing entry) and several possible covariates and outcomes.

Step-wise approaches enable automating tasks by developing general tools (decreasing programmer input) and facilitating alternative designs. Macros improve upon the classic ‘one design, one data set’ perspective. Two case studies highlight tradeoffs in retrospective cohort studies (quasi-experiments) among sample size, length of follow-up, and the number of time periods.

Results

Study 1: Step 1 calculated the number of mothers with a child placed in care during various index years. Taking 1 year before and after placement generated 5,991 eligible mothers; selecting 5 years before/after decreased the N to 2,281. Step 2 selected appropriate in-province residents. Step 3 handled missing covariates and outcomes, while Step 4 ran alternative designs.

One example (of several) compared maternal mental health outcomes using 8 time periods (in 2 years) before/after the event with outcomes using 16 time periods (in 4 years) before/after. Besides showing increasing maternal problems, the 4-year follow-up sometimes produced different statistically significant periods than the 2-year follow-up.

Study 2. Swedish/Canadian comparisons of mothers with children placed in foster care highlighted growing differences in maternal pharmaceutical use.

