International Journal of Population Data Science

Journal Website: www.ijpds.org

Linking big data for cardiovascular health surveillance – opportunities and challenges using the CANHEART cohort

Chu, A[1], Hennessy, D[2], Johnston, S[3], Udell, J[4], Ko, D[1], Khan, A[1], Jia, J[1], and Tu, J[1]

[1] Institute for Clinical Evaluative Sciences
[2] Statistics Canada
[3] Bruyere Research Institute
[4] Women's College Hospital

Introduction

Our increasing ability to link large population-based health administrative datasets to create 'big data' cohorts offers unique opportunities to conduct health and health services surveillance at lower costs than traditional methods using surveys or primary data collection. However, comparability of findings from big data with traditional methods is unknown.

Objectives and Approach

In the CArdiovascular HEalth in Ambulatory Care Research Team (CANHEART) 'big data' initiative, we linked 19 population-based health databases to obtain baseline and 5-year follow-up health information on a cohort of 9.8 million adult residents of Ontario, Canada as of January 2008. We compared cardiovascular risk factor prevalence with results from 3500 participants in the 2007-09 Canadian Health Measures Survey (CHMS), a traditional population health surveillance survey. Additionally, we determined cardiovascular preventative care use and clinical event rates by sex and age. Planned linkages to new data sources will enable continued cohort surveillance of population health-related and care indicators.

Results

Cholesterol and glucose levels determined from the CANHEART cohort were comparable to the CHMS, whereas blood pressure values and obesity rates were substantially higher. Overall, receipt of cardiovascular preventive care in the CANHEART cohort was high, with 85.7% of males and 91.8% of females having blood pressure assessments, and 67.8% of males and 79.4% of females having weight assessments. Cholesterol and diabetes screening rates among those recommended for screening were over 75%. Incidence of myocardial infarction, stroke or cardiovascular death was 51% higher among males than females (3.8 and 2.5 events per 1000 person-years, respectively). Challenges encountered in analyzing data included treatment of repeated and time-varying measures, selection of valid diagnostic and physician billing codes, changing coding practices and handling of missing and outlying data.

Conclusion/Implications

Comparability of cardiovascular risk factor prevalence using linked administrative data with survey methods varies by indicator. Selection biases amongst survey participants and different measurement methods could explain discrepancies. The added ability to examine health care indicators longitudinally and by subgroup supports use of linked population-based data to enhance health surveillance.