International Journal of Population Data Science

Journal Website: www.ijpds.org

Developing and Validating Electronic Medical Record Based Case Definitions for Liver Diseases and Comorbidities

Xu, Yuan^{1*}, Li, Ning², Lu, Mingshan¹, Myers, Robert¹, Dixon, Elijah¹, Walker, Robin¹, and Quan, Hude¹

¹University of Calgary

²Capital Medical University

Background

China has collected high volume of electronic health record (EMR) data. The rich information in EMR data could be used for health services research. The first challenge is developing methods for extracting study variables. Our study aimed to develop and validate data extraction methods for defining clinical conditions.

Approach

The EMRs were from Beijing You-An Hospital, a leading liver diseases specialized teaching hospital affiliated with Capital Medical University in China. We developed EMR based case definitions for extracting common liver diseases and comorbidities in Charlson and Elixhauser comorbidity algorithms. We developed the EMR case definitions based on the fundamental EMR structure and clinical expertise. To determine validity of the EMR case definitions, we calculated the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for each case definition based on the "gold standard": randomly selected 450 charts review conducted by two hepatologists. The agreement between the two reviewers was assessed by kappa value.

Results

In total, 69,864 EMRs for adult patients with liver disease admitted between 2010 and 2015 were included in this study. Among these patients, we identified 13,763 (19.7%) PLC, 18,654 (26.7%) Hepatitis B Virus (HBV), 4681 (6.7%) Hepatitis C Virus (HCV), 18,933 (27.1%) cirrhosis, 7,964 (11.4%) diabetes, 8,523 (12.2%) hypertension, 3,283 (4.7%) Fluid and Electrolyte Disorder (FED), 3,563 (5.1%) renal disease and 1,258

(1.8%) Cerebrovascular Disease (CEVD). Between the two reviewers, the Kappa values fell between 0.65 and 1.00. Of the 450 EMRs reviewed, the two reviewers identified 95 (21.1%) PLC, 197 (43.8%) HBV, 40 (8.9%) HCV, 146 (32.4%) cirrhosis, 39 (8.7%) diabetes, 45 (10.0%) hypertension, 25 (5.6%) FED, 26 (5.8%) renal disease and 8 (1.8%) CEVD. Compared to chart review, the sensitivity, specificity, PPV and NPV of the algorithms for above conditions are respectively: (PLC) 100%, 99.44%, 97.94%, 100%; (HBV) 61.93%, 99.21%, 98.39%, 76.99%; (HCV) 72.50%, 99.51%, 93.55%, 97.37%; (cirrhosis) 78.77%, 98.36%, 95.83%, 90.61%; (diabetes) 97.44%, 98.05%, 82.61%, 99.75%; (hypertension) 95.00%, 99.72%, 97.44%, 99.45%; (FED) 71.43%, 100.00%, 100.00%, 99.09%; (renal disease) 96.15%, 100.00%, 100.00%, 99.76%; (CEVD) 100.00%, 99.77%, 88.89%, 100.00%.

Conclusions

Our EMR case definitions of above conditions had high validity and could be applied to extract clinical variables including major liver diseases and comorbidities from Chinese EMR. This extracting method could be modified slightly for extracting other medical conditions from Chinese EMRs.

*Corresponding Author:

Email Address: shawnxu724@gmail.com (Y. Xu)