International Journal of Population Data Science

Journal Website: www.ijpds.org

Validation of dementia coding in hospital admissions, death and primary care data

Wilkinson, Tim^{1*}, Schnier, Christian¹, Ly, Amanda¹, and Sudlow, Cathie¹

¹University of Edinburgh

Objectives

Dementia is a major public health concern worldwide and consequently there is an urgent need to expedite research into its causes so that preventative strategies can be sought. Given that dementia is likely to be the result of a complex interplay of many factors, large study populations are required in order to detect effects reliably. UK Biobank (UKB) is a large, population-based, prospective cohort study of over 503,000 participants aged 40-69 years when recruited between 2006 and 2010. Participant follow up is chiefly via linkage to routinely-collected health datasets such as hospital admissions, death registrations and increasingly, to primary care data. In this pilot study we sought to estimate the accuracy of using these routine data sources to identify dementia outcomes in UKB participants.

Approach

We created a list of ICD-10 and primary care (Read version 2) dementia codes, with the intention of maximising positive predictive value (PPV) over sensitivity. We identified UKB participants who were recruited in Edinburgh and had at least one dementia code in any of the three data sources. We searched the NHS Lothian electronic medical record (EMR) for each participant and extracted all relevant letters and investigation results. Participants were excluded if no EMR entry for that patient could be found. A neurologist adjudicated on whether dementia was present or not based on the extracted case record, providing the reference standard to which the coded data were compared. The PPV was then calculated for each data source individually and combined. A subgroup analysis was performed on participants who had a dementia code across more than one dataset.

Results

Among 17,000 Edinburgh-based participants (median age 57 years at recruitment in 2007/8), hospital and death data were available to 2012 with primary care data for 12,000 to 2013. 46 participants had a dementia code in at least one data source. 44 of these had available EMR data. PPVs for dementia were 41/44 (93%, 95% CI 81-99) overall, 13/15 (87%, 95% CI 60-98) for hospital admissions, 2/2 (100%, 95% CI 16-100) for death registrations, 33/34 (97%, 95% CI 85-100) for primary care, and 7/7 (100%, 95% CI 59-100) for participants with codes in ≥ 2 datasets.

Conclusion

Routinely-collected health data may be sufficiently accurate to identify dementia outcomes in UK population-based cohorts. We plan to extend this study to longer follow-up times and other regions to increase sample size, investigate dementia subtypes and assess generalisability.

*Corresponding Author:

Email Address: tim.wilkinson@ed.ac.uk (T. Wilkinson)