Accuracy of identifying incident stroke cases from linked health care data in UK Biobank

Kristiina Rannikmae, PhD, Kenneth Ngoh, Kathryn Bush, et al.

Cite as: Neurology® 2020;95:e697-e707. doi:10.1212/WNL.0000000000009924

Correspondence
Dr. Rannikmae
kristiina.rannikmae@ed.ac.uk

Study objective
This study examined the accuracy of using linked national healthcare data to identify cases of incident stroke in the UK Biobank cohort.

What is known and what this paper adds
In UK Biobank (UKB), a large population-based prospective study, cases of many diseases are ascertained through linkage to routinely collected, coded national health datasets. This investigation’s results show that such methods are accurate for research purposes.

Participants and setting
These analyses included data from 225 enrollees in the UK Biobank (49% female; median age at recruitment, 63 years; range, 41–70 years) who lived in the Lothian region of southeast Scotland, which encompasses Edinburgh, and had at least 1 code in their linked hospital admission, primary care, or death records that indicated a stroke diagnosis between the date of recruitment to the cohort and September 2015.

Design, size, and duration
Stroke physicians blinded to the relevant codes reviewed full electronic patient records and generated reference standard diagnoses and evaluated the number and proportion of cases that were true positives (i.e., positive predictive value, PPV) for all codes combined and by code source and type.

Primary outcome measures
The primary outcome was the PPVs of the linkage-derived codes for stroke in general and ischemic stroke specifically.

Main results and the role of chance
Of 232 incident stroke-coded cases, 97% had EPR information available. Data sources were: 30% hospital admission only; 39% primary care only; 28% hospital and primary care; 3% death records only. While 42% of cases were coded as unspecified stroke type, review of EPRs enabled a pathologic type to be assigned in >99%. PPVs (95% confidence intervals) were: 79% (95% CI, 73%–84%) for any stroke (89% for hospital admission codes, 80% for primary care codes) and 83% (95% CI, 74%–90%) for ischemic stroke. PPVs for small numbers of death record and hemorrhagic stroke codes were low but imprecise.

Bias, confounding, and other reasons for caution
The stroke physicians who provided reference diagnoses lacked access to the primary care system’s full free-text electronic patient records. The investigators could not easily differentiate definite false positives from uncertain cases.

Study funding/potential competing interests
This study was funded by the UK Biobank, Health Data Research UK, Dementias Platform UK, the UK Medical Research Council, and the Scottish Funding Council. Some authors report additional competing interests. Go to Neurology.org/N for full disclosures.

Table
PPVs for stroke codes from specific sources

<table>
<thead>
<tr>
<th>Code source</th>
<th>PPV (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital admissions data</td>
<td>89% (82%–94%)</td>
</tr>
<tr>
<td>Primary care data</td>
<td>80% (72%–86%)</td>
</tr>
<tr>
<td>Death records</td>
<td>57% (18%–90%)</td>
</tr>
</tbody>
</table>

Abbreviations: CI = confidence interval; PPV = positive predictive value.

A draft of the short-form article was written by M. Dalefield, a writer with Editage, a division of Cactus Communications. The corresponding author(s) of the full-length article and the journal editors edited and approved the final version.

Copyright © 2020 American Academy of Neurology