

'Capture-Recapture' — estimating the prevalence of disease using linked data

Capture-recapture techniques are common in ecology and can be used to estimate the population size of different animals. These techniques might involve 'capturing', marking, and releasing animals. This process is repeated, and by observing the number of 'marked' animals in each sample it is possible to estimate the total population.

A similar concept can be applied to linked datasets in epidemiological studies. By observing the overlap in records of the disease between different data sources it is possible to use capture-recapture techniques to estimate how many cases of disease may have been missed.

We used this technique to estimate the prevalence of dementia in the 1921-26 cohort of the Australian Longitudinal Study on Women's Health (also known as Women's Health Australia). We had five sources of data containing information on dementia which we were able to link together; self-reported surveys, aged-care data, cause of death data, pharmaceutical benefits scheme data, and hospital data. We identified all dementia records from these data sources and then assessed how these data sources overlapped.

Overall 2534 (20.4%) women had a record of dementia from at least one of the data sources. The aged-care data was the largest source (2010 dementia records). Using capture-recapture techniques it was estimated that there were an additional 695 women with dementia in the cohort. This increased our estimate of dementia prevalence in the cohort from 20.4% to 26.0%.

These estimates are consistent with those from previous international studies of dementia in women, indicating that the estimates gained through linkage of multiple datasets are credible. The methods demonstrated provide a way to estimate disease prevalence in a study cohort using existing

datasets and with minimal additional subject participation burden. The use of linked datasets to estimate disease prevalence are particularly useful in older study cohorts where declining health might prevent continued survey participation.

A full description of our methods and results is available in the following open access article.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5327574/>

Citation:

Waller M, Mishra G, and Dobson AJ. *Estimating the prevalence of dementia using multiple linked administrative health records and capture-recapture methodology*, *Emerging Themes in Epidemiology*, 2017 14:3
<https://doi.org/10.1186/s12982-017-0057-3>

Contact:

Dr Michael Waller
Email: m.waller@uq.edu.au

The research on which this case study is based was conducted as part of the Australian Longitudinal Study on Women's Health by the University of Queensland and the University of Newcastle. We are grateful to the Australian Government Department of Health for funding and to the women who provided the survey data.



THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA



THE UNIVERSITY OF
NEWCASTLE
AUSTRALIA

Australian Longitudinal
Study on Women's Health