

International Journal of Population Data Science

Journal Website: www.ijpds.org



Healthy Controls versus General Controls in Matched Cohort Studies – Comparison from a Post-Stroke Dementia Cohort Study in Ontario, Canada

Jiming Fang¹, Peter Austin², Raed Joundi³, and Moira Kapral⁴

¹ICES, Toronto, Canada

²ICES, Toronto, Canada; University of Toronto, Toronto, Canada

³Population Health Research Institute, Hamilton, Canada; McMaster University, Hamilton, Canada

⁴University of Toronto, Toronto, Canada; ICES, Toronto, Canada

To match disease cohorts with the general population, researchers often exclude individuals with the disease (“healthy controls”). Alternatively, they may include those disease-free at baseline who may develop it later (“general controls”). Few studies have compared these methods. We examined how control selection affects cause-specific hazard ratios (csHR) for post-stroke dementia among stroke survivors and matched controls. Using Ontario’s population-based Discharge Abstract Database (2002–2022), we identified adults hospitalized with stroke. Each stroke survivor was hard-matched 1:1 to residents without prior stroke on age, sex, rural residence, neighborhood deprivation, and vascular comorbidities. Matching was conducted separately for healthy and general control groups. Cause-specific hazard models, accounting for death as a competing risk (from the linked Registered Persons Database), estimated associations between stroke and

incident dementia. General controls who subsequently experienced stroke were censored at stroke onset. Of 175,980 stroke survivors, 174,787 (99.3%) and 174,817 (99.3%) were successfully matched to healthy and general controls, respectively. At 20 years of follow-up, the cumulative incidence of dementia among stroke survivors was similar across both match types (0.302 vs. 0.302). However, healthy controls had a higher incidence than general controls (0.249 [95% CI: 0.244–0.253] vs. 0.238 [95% CI: 0.234–0.242]). The csHR comparing strokes with healthy controls (1.694 [95% CI: 1.668–1.720]) was significantly lower than that with general controls (1.759 [95% CI: 1.731–1.787]). Using healthy controls may underestimate post-stroke dementia risk. General controls more accurately represent the population at risk and are recommended for matched cohort studies.

