FULL-LENGTH ARTICLE NPub.org/l2g2hs

Validation of an algorithm for identifying MS cases in administrative health claims datasets

William J. Culpepper, PhD, MA, R.A. Marrie, MD, PhD, Annette Langer-Gould, MD, PhD, et al., on behalf of the United States Multiple Sclerosis Prevalence Workgroup (MSPWG)

Correspondence
Dr. Culpepper
William.Culpepper@va.gov

Cite as: *Neurology*® 2019;92:e1016-e1028. doi:10.1212/WNL.0000000000007043

Study objective and summary result

This study tested several candidate algorithms for identifying cases of multiple sclerosis (MS) in administrative health claims (AHC) datasets. The optimal case definition required ≥3 MS-related hospitalizations or outpatient visits or MS disease modifying drug prescription, in any combination, within 1 year.

What is known and what this paper adds

Various factors make analysis of AHC datasets an attractive option for estimating disease prevalences, but case-identifying algorithms must be validated prior to such analyses. This study identifies a valid case definition for identifying cases of MS across disparate AHC datasets.

Participants and setting

This study's algorithm development datasets included (1) 3,452 individuals from the US Veterans Affairs (VA) healthcare system that had at least one encounter with an MS code; (2) 2,935 individuals insured by Kaiser-Permanente Southern California (KPSC) who had clinical encounters with MS-related diagnostic codes and (3) 1,654 Manitoba residents with MS-related diagnostic codes. The Saskatchewan validation dataset consisted of 200 individuals with confirmed MS and 200 controls randomly drawn from the Inpatient Rehabilitation Center database.

Design, size, and duration

This study first used the algorithm development datasets to test various case definitions of MS that involved inpatient and outpatient visits and, optionally, prescription records. This study used reference standard diagnoses from medical records as comparators against which to calculate the sensitivity and specificity of each candidate case definition. The best-performing case definition was then applied to the Saskatchewan validation dataset.

Primary outcome measures

The primary outcome was the best-performing case definition's diagnostic performance in the Saskatchewan dataset.

Main results and the role of chance

When the preferred case definition (≥ 3 MS-related hospitalizations, outpatient visits, or prescriptions filled, in any combination) was applied to the Saskatchewan (validation) dataset the positive predictive value was 99.0 and the negative predictive value was 96.0.

Bias, confounding, and other reasons for caution

This study's algorithms did not use the latest International Classification of Diseases codes due to the time-period of data collection.

Generalizability to other populations

Datasets from the US and Canada were chosen to enhance generalizability but may not perform as well in datasets outside of North America.

Study funding/potential competing interests

This study was funded by the National MS Society. Some authors report receiving grants, consulting fees, and committee appointments from various foundations, US and Canadian government agencies, and healthcare companies; holding endowed chairs; serving as investigators on industry-sponsored clinical studies; and being employed by KPSC or the National MS Society. Dr. Marrie serves on the editorial board for *Neurology*® Go to Neurology.org/N for full disclosures.

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

Validation of an algorithm for identifying MS cases in administrative health claims datasets

William J. Culpepper, Ruth Ann Marrie, Annette Langer-Gould, et al.

Neurology 2019;92;e1016-e1028 Published Online before print February 15, 2019
DOI 10.1212/WNL.0000000000007043