Journal of Critical Care 37 (2016) 202–205

Contents lists available at ScienceDirect

Journal of Critical Care

journal homepage: www.jccjournal.org

Derivation and validation of the automated search algorithms to identify cognitive impairment and dementia in electronic health records ☆

Sakusic Amra, MD [a,b,c,*], John C O'Horo, MD, MPH [a,d,e], Tarun D Singh, MBBS [d,f], Gregory A Wilson, RRT [d], Rahul Kashyap, MBBS [d,g], Ronald Petersen, MD, PhD [f,h,i], Rosebud O Roberts [f,h,i], John D Fryer [j], Alejandro A Rabinstein, MD [d,f], Ognjen Gajic, MD, MSc [a,d]

[a] Department of Medicine, Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Rochester, MN
[b] Department of Internal Medicine, Tuzla University Medical Center, Bosnia and Herzegovina
[c] Department of Pulmonary Medicine, Tuzla University Medical Center, Bosnia and Herzegovina
[d] Department of Medicine, Multidisciplinary Epidemiology and Translational Research in Intensive Care, Emergency and Perioperative Medicine (METRIC), Mayo Clinic, Rochester, MN
[e] Department of Medicine, Division of Infectious Diseases, Mayo Clinic, Rochester, MN
[f] Department of Neurology, Mayo, Rochester, MN
[g] Department of Anesthesia, Mayo Clinic, Rochester, MN
[h] Department of Health Sciences Research, MN
[i] Department of Neurology, MN
[j] Department of Neuroscience, Mayo Clinic, Jacksonville, FL

ARTICLE INFO

Available online xxxx

Keywords:
Dementia
Cognitive impairment
Cognitive decline
Critical illness
Electronic search strategy
Electronic medical records

ABSTRACT

Purpose: Long-term cognitive impairment is a common and important problem in survivors of critical illness. We developed electronic search algorithms to identify cognitive impairment and dementia from the electronic medical records (EMRs) that provide opportunity for big data analysis.

Materials and methods: Eligible patients met 2 criteria. First, they had a formal cognitive evaluation by The Mayo Clinic Study of Aging. Second, they were hospitalized in intensive care unit at our institution between 2006 and 2014. The "criterion standard" for diagnosis was formal cognitive evaluation supplemented by input from an expert neurologist. Using all available EMR data, we developed and improved our algorithms in the derivation cohort and validated them in the independent validation cohort.

Results: Of 993 participants who underwent formal cognitive testing and were hospitalized in intensive care unit, we selected 151 participants at random to form the derivation and validation cohorts. The automated electronic search algorithm for cognitive impairment was 94.3% sensitive and 93.0% specific. The search algorithms for dementia achieved respective sensitivity and specificity of 97% and 99%. EMR search algorithms significantly outperformed *International Classification of Diseases* codes.

Conclusions: Automated EMR data extractions for cognitive impairment and dementia are reliable and accurate and can serve as acceptable and efficient alternatives to time-consuming manual data review.

1. Introduction

The aging population has led to an alarming increase in the incidence of cognitive decline, and modifiable risk factors are poorly understood. It has been noticed that a significant number of patients after critical illness experience cognitively decline [1,2,3]. More than 5 million patients are admitted to intensive care units (ICUs) in the United States each year, with approximately 55 000 of them are treated everyday [4]. Cognitive impairment following critical illness affects not only the patients but their families and society at large. The financial impact on the family is considerable; up to 20% of patient's family members have to quit their jobs to care for their cognitively impaired relatives, 29% of families lose a major source of income, and 31% of patients declared that they lost most of their family savings [5]. It has been calculated that the annual cost burden is $15 022 for mild cognitive impairment and $34 515 for severe cognitive impairments (dementia, Alzheimer disease) per patient per year [6]. This comes to an annual total estimate of $18 billion of lost income and direct costs of care [7]. High-quality research on a large cohort of critically ill patients is necessary to better characterize the frequency and severity of cognitive decline in this population.

☆ Financial disclosure and conflicts of interest: This publication was made possible in part by funding from The Mayo Critical Care Research Committee and The Alzheimer's Disease Research Center. The authors declare that they have no conflict of interest.

* Corresponding author at: Mayo Clinic, 200 First St SW, Rochester, MN, 55905. Tel.: +1 507 210 7318.
E-mail address: Sakusic.Amra@mayo.edu (S. Amra).

http://dx.doi.org/10.1016/j.jcrc.2016.09.026

Traditional paper chart forms have been rapidly replaced by electronic medical records (EMRs). These have provided many advantages to health care management, clinical practice, and research. Within research, EMRs have brought medicine into the era of "Big Data," where an unprecedented amount of information can allow evaluation and identification of risk factors on a population level with relative ease.

Big data is often defined as the "4 V′s"—volume, velocity, variability, and veracity. EMR data clearly meet the first 3 criteria, but the fourth is context sensitive. Charting and billing data's accuracy has to be confirmed before big-data approaches can be applied to a given problem. Because of the volume of data involved, veracity often involves an "electronic search algorithm" (ESA), where multiple types of data are used to identify an exposure or condition of interest. Properly validated, an ESA allows for scanning of an entire population to find previously unknown risk factors in near real time.

Our primary aim was to develop and validate a reliable electronic search algorithm to define cognitive impairment and dementia that will enable pragmatic research on large cohorts of patients using existing EMRs in future studies. We hypothesized that automated ESA for dementia and cognitive impairment will be reliable and accurate.

The secondary aim was to compare the performance of developed search algorithms to *International Classification of Diseases, Ninth Revision* (ICD-9) code and the criterion standard of the Mayo Clinic Study of Aging (MCSA) and manual medical record review.

2. Materials and methods

2.1. Study population

Our study was approved by the Mayo Clinic Institutional Review Board for the use of existing medical records of patients or their relatives who gave prior research authorization.

We included consecutive adult patients who were admitted to one of the ICUs at Mayo Clinic in Rochester, MN, between January 1, 2006, and December 31, 2014, and were cognitively evaluated by the MCSA.

The subjects of our study also had been admitted to one of the following ICUs at Mayo Clinic: medical ICU, coronary care unit, 2 mixed medical-surgical ICUs, and the cardiosurgical ICU. We excluded patients admitted to the neurological ICU because of overrepresentation of other neurologic disorders. Because of age criteria, we also excluded all pediatric and neonatal ICUs.

The derivation and validation cohorts each consisted of 151 randomly selected participants. Both cohorts included participants diagnosed with normal cognition, mild cognitive impairment, and dementia.

2.2. Data extraction strategies

We manually reviewed EMRs of the derivation cohort to identify different ways that cognitive impairment was documented by health care providers at Mayo Clinic.

Following this manual review, the research fellow (AS) developed electronic search strategies for cognitive impairment and dementia through several iterations of evaluation and refinement of the ESA. Once the derivation cohort was finalized, the search strategies were ran using the validation cohort.

2.3. Data sources

We developed and executed our ESAs on Mayo Clinic's Advanced Cohort Explorer (ACE). ACE accesses Mayo Clinic's Unified Data Platform (UDP). ACE is a powerful search tool that provides access to millions of patient's records that contain data regarding demographics, diagnosis, laboratory results, clinical notes, pathology notes, flow sheets, imaging results, patient provided information, admission/discharge transfer data, etc. ACE also enables the search of EMR by specific text phrases or terms in specific parts of clinical notes. All data extracted by ACE can be exported to Excel to enable further statistical analysis. The UDP is a comprehensive repository of Mayo Clinic's patient's EMR.

Valuable data source for our study was the MCSA, an ongoing prospective population-based cohort study designed to evaluate the prevalence, incidence, and risk factors for dementia and mild cognitive impairment. The subjects in this study underwent a comprehensive in-person evaluation including neurological and neuropsychological testing and Dementia Rating Scale [8]. A consensus diagnosis was made using previously published criteria [9]. Participants were diagnosed with normal cognition, mild cognitive impairment, or dementia.

2.4. Statistical analysis

We calculated sensitivity and specificity based on the comparison of automated search algorithm to the criterion standard using JMP statistical software version 9.0 (Statistical Analysis System Institute, Inc).

The initial criterion standard for comparison to the digital search strategies was the comprehensive in-person cognitive evaluation by the MCSA. Forty-one mismatches in the derivation cohort between the ESA and the MCSA evaluation were manually rechecked and compared by the research fellow. We noticed that the ESA found 28 patients that had been clinically diagnosed with cognitive impairment but were not diagnosed by the MCSA. The reason for this apparent discrepancy was determined to be that these patients had not come back for their 15 month follow-up cognitive evaluation of the study. From the remaining 13 mismatches, 9 were diagnosed with mild cognitive impairment by the MCSA but the ESA did not register them as cognitively impaired, whereas 4 were diagnosed as cognitively normal by the MCSA and the ESA detected them as cognitively impaired. To overcome this barrier, we had discordances adjudicated by an expert neurologist (AAR) blinded to the ESA and the MCSA diagnoses. Among 41 mismatches in the derivation cohort, 10 were diagnosed as normal and 31 were diagnosed as cognitively impaired on manual review by an expert neurologist. This adjudicated data set formed our final criterion standard.

3. Results

Between January 1, 2006, and December 31, 2014, 80 624 adult patients with research authorization on file were admitted to participating ICUs. Among these, 993 patients were participants of the MCSA. Per protocol, 151 of these patients were randomized to the derivation cohort, and 151 were randomized to validation (Fig. 1). Both derivation and validation populations consisted of participants diagnosed with normal cognition, mild cognitive impairment, and dementia.

The algorithms were created through multiple steps using the derivation cohort. (Fig. 1).

Our initial search strategy for cognitive impairment ESA used simple keywords for *dementia* and *mild cognitive impairment* in all sections of clinical notes, resulting in specificity of 72.3% and sensitivity of 74% (Table 1). Through several iterations of refinement, our final search strategy was restricted to sections of the note describing diagnoses, instructions for continuing care, ongoing care orders, and review of symptoms. Keywords in the ESA were as follows: *dementia*, *cognitive impairment*, *cognitive deficit*, *cognitive decline*, *mild cognitive impairment*, *impaired memory*, *impaired judgment*, *impaired orientation*, *difficulty concentrating*, *patient is not independent in handling finances* (Fig. 2).

Our automated text search combination for cognitive impairment in the derivation cohort achieved a final sensitivity of 96% and specificity of 96%. On the validation cohort, the electronic searched algorithm yielded sensitivity of 94.3% and specificity of 93% (Table 1). We also compared our search strategy to a traditional billing code–based search for cognitive impairment and dementia ICD-9 codes. This approach was less sensitive at 52.8% (Table 3, individual code diagnostic performance in Supplemental Table 1).

When creating electronic search algorithm for dementia, our initial search strategy was to use the word *dementia* in any part of the clinical