HHS Public Access

Author manuscript

Amyotroph Lateral Scler Frontotemporal Degener. Author manuscript; available in PMC 2015 September 29.

Published in final edited form as:

Amyotroph Lateral Scler Frontotemporal Degener. 2014 September ; 15(0): 433–439. doi: 10.3109/21678421.2014.887119.

Feasibility of creating a National ALS Registry using administrative data in the United States

WENDY E. KAYE[1], **MARCHELLE SANCHEZ**[2], and **JENNIFER WU**[2]

[1]McKing Consulting Corporation, Atlanta, Georgia

[2]Agency for Toxic Substances and Disease Registry/Centers for Disease Control and Prevention, Atlanta, Georgia, USA

Abstract

Uncertainty about the incidence and prevalence of amyotrophic lateral sclerosis (ALS), as well as the role of the environment in the etiology of ALS, supports the need for a surveillance system/ registry for this disease. Our aim was to evaluate the feasibility of using existing administrative data to identify cases of ALS. The Agency for Toxic Substances and Disease Registry (ATSDR) funded four pilot projects at tertiary care facilities for ALS, HMOs, and state based organizations. Data from Medicare, Medicaid, the Veterans Health Administration, and Veterans Benefits Administration were matched to data available from site-specific administrative and clinical databases for a five-year time-period (1 January 2001–31 December 2005). Review of information in the medical records by a neurologist was considered the gold standard for determining an ALS case. We developed an algorithm using variables from the administrative data that identified true cases of ALS (verified by a neurologist). Individuals could be categorized into ALS, possible ALS, and not ALS. The best algorithm had sensitivity of 87% and specificity of 85%. We concluded that administrative data can be used to develop a surveillance system/ registry for ALS. These methods can be explored for creating surveillance systems for other neurodegenerative diseases.

Keywords

Amyotrophic lateral sclerosis; prevalence studies; epidemiology; registry

Introduction

Although amyotrophic lateral sclerosis (ALS) (1) was first described by the French neurologist Jean-Martin Charcot in 1896, there is still little known about its causes.

Reports from the United States and other countries indicate an annual incidence rate of 0.2 to 2.4 per 100,000 population and a prevalence of 0.8 to 7.3 per 100,000 population (2). The

Correspondence: W. E. Kaye, McKing Consulting Corporation, 2900 Chamblee-Tucker Road, Building 10, Suite 100, Atlanta, GA 30341, USA. Fax: 770 220 0642. wek1@cdc.gov.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper. The findings and conclusions in this report are those of the author(s) and do not necessarily represent the views of the Agency for Toxic Substances and Disease Registry.

onset of ALS is age related with the highest rate of onset occurring between 55 and 75 years of age (2–4). The average survival time after onset of symptoms is approximately three years, and only a small proportion of patients survive beyond five years (2). ALS is more common in males than females by a ratio of 1.5–2:1 (4,5), but recent studies have suggested that this difference is decreasing (4,6).

Uncertainty about the incidence and prevalence of ALS, as well as the role of the environment in the etiology of ALS, supports the need for a surveillance system for this disease (7,8). In addition, such a system could provide an unbiased source from which to recruit patients to participate in research studies.

For more than 15 years, researchers have understood the importance of clinical databases in the study of ALS. A number of registries were initiated across Europe including but not limited to registries in Scotland (9,10), Ireland (11), Piemonte and Valle d'Aosta, Italy (12), and Sclerosi Laterale Amiotrofica – Puglia (SLAP) (13). In order to recruit a large population based sample of patients with ALS and to collect baseline information on newly diagnosed patients across Europe, the EURALS registry was established by merging information collected from existing national and regional registries in Italy, Ireland, England, and Scotland and collecting information on newly diagnosed patients in other countries such as Spain and Serbia where population based registries were no longer available (14). Differences in ethnic diversity, population size, and healthcare delivery model between European countries and the United States, make the methods used to develop the European registries less likely to be able to be implemented in the Unites States.

There have only been two attempts to create large databases of ALS patients in the United States. The first large such database was the ALS Patient Care Database with the primary purpose of improving the quality of care for patients with ALS. This observation database was begun in 1996 and was available to all neurologists practicing in North America (15).

The only other registry of ALS patients in the United States is the national registry of Veterans with ALS (16). Neither of these registries is enrolling any new participants.

A bill to amend the Public Health Service Act to provide for the establishment of an Amyotrophic Lateral Sclerosis Registry, S. 1382: ALS Registry Act, was signed into law on 10 October 2008 by President Bush and became Public Law No: 110-373. Given the issues of implementing methodologies developed in Europe and that the authority to make a disease reportable in the United States rests with the states which historically have reserved this designation for infectious diseases, the Agency for Toxic Substances and Disease Registry (ATSDR) decided to determine the feasibility of using administrative data to populate such a registry.

Methods

To evaluate the feasibility of using existing administrative data to identify cases of ALS, ATSDR funded four diverse pilot projects to be conducted by the Mayo Clinic (Rochester and Olmsted County, Minnesota), Emory University, the South Carolina Policy Board, and nine members of the HMO Research Network (HMORN) in California, Oregon,

Pennsylvania, Wisconsin, New Mexico, Michigan, and Massachusetts. The Mayo Clinic and Emory University were tertiary care facilities with neurologists who specialized in the diagnosis and treatment of persons with ALS; however, only cases of ALS in the defined catchment areas were included. The Mayo Clinic also had access to information about individual healthcare utilization for all residents of Rochester and Olmsted County. The South Carolina Policy Board had access to more than 18 databases covering individual health care utilization and spending for residents of the entire state. The members of the HMORN covered a large racially and ethnically diverse population throughout the country from both urban and rural areas.

Each of these four pilot projects matched data from Medicare, Medicaid, the Veterans Health Administration (VHA), and the Veterans Benefits Administration (VBA) to data from site-specific administrative and clinical databases (e.g. ALS clinic records) for a five-year time- period (1 January 2001–31 December 2005). The national databases covered approximately 90 million people. VHA data include inpatient, outpatient, and pharmacy records for veterans receiving healthcare benefits. Approximately 20% of veterans qualify for this benefit. VBA data include records for veterans receiving pensions or compensation for disabilities considered service related. During the study period, ALS was considered service related if it was diagnosed within one year of separation from active duty. Medicare data included inpatient and outpatient records. Medicare is United States government-provided insurance for people aged 65 years or older, some disabled people under age 65 years, and people of all ages with End-Stage Renal Disease. Individuals approved for the Social Security Administration Disability Insurance Benefit or Supplemental Security Income because of ALS can begin receiving Medicare without a 24-month waiting period. Medicaid data include inpatient, outpatient, and pharmacy records for individuals receiving this benefit. Medicaid is the United States health program for individuals and families with low incomes and resources.

From the national databases, ATSDR identified individual patient encounters resulting in an ICD-9 code for any MND (335.2–335.29) and VBA-specific codes for any MND for the specific project catchment area (Table I). ATSDR then provided individual encounter data including full name and social security number to the pilot projects. Many of the data sets contained up to 10 ICD-9 codes for an encounter. The MND code could be in any position, not just the primary diagnosis.

For the Mayo Clinic and the state of South Carolina, catchment area was easily defined by county or state boundaries. However, defining catchment areas for Emory and the HMORN sites was more difficult because not everyone in a defined geographic area used the specific institution/provider services. Catchment area for each site was determined based on the site investigator’s knowledge of penetration into the surrounding counties. Zip codes, which change over time, could not be used.

To identify additional potential cases of ALS, sites evaluated administrative and clinical databases available to them.

Pilot sites completed a standardized spreadsheet for each individual found in any available database. Individuals could have multiple records for a single encounter in the databases, e.g. physician and neuroimaging, and multiple encounters. All encounter records from all databases were combined into one record per individual. The spreadsheet included the years in which an individual had records, in which database(s) a record was found, ICD-9 code recorded for the encounter (ALS or other MND), prescription for riluzole, and types of providers seen (e.g. neurologist, internal medicine). Using a standardized medical records abstraction form designed by the project neurologists, accessible medical records were abstracted for signs and symptoms of ALS and EMG results when available and diagnoses were verified by a neurologist using the El Escorial criteria (17), which was considered the 'gold standard'. Those records that did not have enough information to make the determination were classified as possible ALS to be evaluated as additional data become available. A de-identified dataset was sent to ATSDR. That dataset included the spreadsheet information for all individuals who were identified in any of the databases with an ICD-9 code for any MND or VBA-specific code for any MND, and who had their medical record reviewed by a neurologist regardless of the neurologist's determination. ATSDR combined the data from the four pilot projects for this analysis.

A more detailed description of the methodology used and findings of the pilot projects conducted by the South Carolina Budget Control Board (18) and Emory University (19) can be found in the site-specific publications.

ATSDR determined that these pilot projects were not human subjects research.

Results

Approximately 24,000 individuals were identified with an MND ICD-9 code in the national and local databases. The pilot projects reviewed approximately 4700 local medical records (Table II). Except at the HMORN sites, most of the charts reviewed were for individuals identified through the national databases. The percentage of charts with MND codes identified only through site review ranged from 1% (Mayo Clinic) to 77% (HMORN). Table III shows the demographic characteristics of all the individuals identified with an ICD-9 code or VBA-specific code for any MND and those individuals for whom charts were reviewed. There was a significant difference in the racial, gender and age distribution of those whose charts were reviewed, i.e. a greater percentage of the medical records of blacks and unknown race, males, and younger cases was reviewed.

Table IV examines the demographic characteristics of those with an ALS ICD-9 code (335.20) and those with any other MND ICD-9 code (335.2, 335.21–335.24, and 335.29) for whom medical records were reviewed. Consistent with what we know about the demographics of ALS patients, those with an ALS code were more likely to be white, male, and between the ages of 40 and 79 years.

The sensitivity and specificity of individual variables and combined variables were evaluated. For example, having the ALS code in at least one encounter had a sensitivity of 96% and specificity of 52%. Using our knowledge of ALS and findings from other

algorithm building studies, variables were combined into the algorithm to maximize both sensitivity and specificity. It was possible to develop an algorithm using variables from the administrative data that identified true cases of ALS (as verified by a neurologist) (Table V). Individuals were placed into one of three categories: ALS, Possible ALS, and Not ALS. Any individual not falling into the ALS or Not ALS category was in the possible ALS category (not shown). For example, an ALS code in two years but no visit to a neurologist would be in the possible ALS category and re-evaluated as new data become available. The best algorithm had sensitivity of 87% and specificity of 85%.

Discussion

Increasingly, electronically available data collected for purposes other than research, such as claims data, are being used in epidemiological studies. A great deal of research has gone into the reliability for research purposes of coding in large datasets such as Medicare and Medicaid because the information was collected for other uses such as claims. Most of the research has focused on identifying a specific disease or procedure using codes and comparing that with the medical record, which is considered the gold standard (20–23). One study comparing the accuracy of Medicare hospital claims with the hospital records found that for Diseases of the Nervous System and Sense Organs (ICD-9-CM: 320–389), the agreement between the coding and the medical record was 91.4% (24).

Causes for an erroneous code in a claims database can range from computer entry errors to lack of sufficient clinical information to accurately code the claim (25). Changes in reimbursement based on Diagnosis Related Group (DRG) may also cause some coding inconsistencies.

Many studies use only one database. Researchers must understand the limitations of the data being used (26) and evaluate the database on a macro level to assess gaps (27). Several researchers have pointed out that obtaining accurate information relies on review of multiple encounters/reports because diagnoses can change (28) and chronic conditions may not be listed in each encounter (29). In a review of strengths and limitations of Medicaid data for epidemiologic research the authors point out that it may be hard to identify incident conditions (33). The first mention of a condition does not necessarily indicate that it was diagnosed on that date but might merely indicate the first time it was documented. Requiring documentation of additional procedures along with the diagnosis can assist in identifying incident cases (31). In a study of hip fracture, investigators developed an algorithm that defined hip fractures using both diagnosis and procedure codes and a combination of information from both hospital claims and Part B claims (outpatient) for Medicare recipients. Even for a condition that is almost certainly treated on an inpatient basis, some claims would have been missed without the outpatient information. The authors of this study also point out the importance of including information from the Veterans Administration (VA) because some Medicare recipients might receive care at a VA facility, information which would not be reflected in the Medicare files (32). Therefore, using multiple data sets, creating an algorithm to identify cases that include inpatient and outpatient information, as well as using a combination of diagnoses and procedures, can also increase the certainty of the diagnosis.

In addition, it is important in chronic diseases to look at multiple years of data. In a study by Pope et al., the prevalence of MS increased with the length of observation. The prevalence estimate for one year of claims for the privately insured population was 18 per 10,000 enrollees, 29 per 10,000 Medicare enrollees, and 53 per 10,000 Medicaid disabled enrollees. When two years of data were used to determine who had MS, the prevalence estimated increased to 24, 36 and 71 per 10,000 enrollees, respectively (33). In another study examining the accuracy of Medicare claims data for identifying Alzheimer's disease, the authors determined that a minimum of three years of data were needed to identify the patients. More years of data increased the number identified, but only slightly. In addition, hospital files alone identified only 29% of the patients, whereas physician encounters and institutional outpatient files together identified 75% of the patients. Using five years of inpatient and outpatient data, researchers identified 79% of the cases. An analysis of clinical data on the patients revealed that those with less severe disease were less likely to be identified (34).

Our experience using administrative data for the identification of true ALS cases is similar to that found by others. Using multiple years of data to ensure that the individual had a definitive diagnosis of ALS rather than as one possible diagnosis, increased the sensitivity and positive predictive value (data not shown). Linking patient's data in data sets from different sources (e.g. being able to link physician encounter data in Medicare with prescription data from the Veterans Health Administration) strengthened the ability to identify true cases. Similar results were found when data from the individual pilot projects were analyzed.

Many of the data sets contained up to 10 ICD-9 codes for an encounter. It was important to consider all these codes and not just the primary diagnosis because ALS might not be the reason for the visit to a medical provider. For example, an individual might be seen in the emergency room for difficulty breathing, which was listed as the primary diagnosis for the visit; however, difficulty breathing was secondary to ALS, which was listed among the other codes. There were concerns that individuals with ALS would have one of the other MND codes in their administrative records and that cases of ALS would be missed. However, we found that MND codes other than the specific code for ALS were often misused, particularly the codes for Progressive Muscular Atrophy and Pseudo Bulbar Palsy. For example, these codes were often used to describe symptoms attributable to other conditions such as Parkinson's disease, stroke, and post-polio syndrome. Data from Table IV support this finding, which shows a larger percentage of individuals 80+ years of age with an MND code other than ALS. The highest onset of ALS is between 55 and 75 years of age (2–4). Therefore, individuals whose administrative records included only MND codes for conditions other than ALS and never included the ALS specific code of 335.20 were considered definitely not to have ALS.

Within the HMOs, the number of charts reviewed compared with the number of records identified was small. This number is misleading because the national data could be identified only to the county level. The HMOs defined their catchment areas based on county, realizing it was not completely accurate and not everyone in a county was a member of their HMO. Therefore, we believe that the percent of possible individuals with ALS who

belonged to the HMO and for whom a chart was reviewed was much larger. In addition, the number of cases found only in the HMO databases was higher than those found in the site-specific databases of the other sites. This could be because the Medicare data are encounter data and do not include Medicare recipients who choose the HMO option. This is a limitation of Medicare data for identifying cases of specific diseases.

Conclusion

There is a public health need for accurate estimates of people affected by neurodegenerative diseases to better assess the healthcare needs of the population, detect changes in healthcare practices, and assess the burden of disease. Although the idea of a comprehensive public health surveillance system using existing data was described more than 10 years ago (35), there have been no attempts to initiate such a system on a national level.

Our results suggest that administrative data can be used to create an ALS surveillance system although it will be necessary to identify other sources of data to capture those ALS cases not covered by the administrative databases. Although these databases do not cover the entire population of the U.S., they represent a significant portion of the population most likely affected by ALS because Medicare covers individuals 65 years of age or older and those with certain diseases such as ALS. Not all ALS patients will apply for benefits; however, while only 13% of Medicare beneficiaries overall are less than 65 years of age, our data show a significant number of ALS patients (41%) identified in Medicare are less than 65 years of age. Since completion of the pilot project, Medicare Part D (prescription coverage) is now offered and the addition of these new data will be evaluated to determine how to use them in further refining the algorithm to optimize sensitivity and specificity. In addition, the six validation questions used by the VA to screen individuals for the VA ALS registry were found to be very accurate (93.4% of those who passed the screening questions were determined by a neurologist to have ALS) (36) and could be used to allow ALS patients to self-identify. These methods can be explored for creating surveillance systems for other neurodegenerative diseases.

Acknowledgments

The authors thank Mark Hornbrook and the pilot project team at the HMO Research Network, Michael Benatar and the pilot project team at Emory University, Julie Royer and the pilot project team at the South Carolina Budget Control Board, and Eric Sorenson and the pilot project team at the Mayo Clinic.

References

1. Bobowick AR, Brody JA. Epidemiology of motor neuron diseases. N Engl J Med. 1973; 288:1047–55. [PubMed: 4576955]
2. Mitsumoto, H.; Chad, DA.; Pioro, EP. Amyotrophic Lateral Sclerosis. Philadelphia: F.A. Davis Company; 1998.
3. Eisen A. Amyotrophic lateral sclerosis is a multifactorial disease. Muscle Nerve. 1995; 18:741–52. [PubMed: 7783764]
4. Chio A. Risk factors in the early diagnosis of ALS: European epidemiological studies. Amyotroph Lateral Scler Other Motor Neuron Disord. 2000; 1(Suppl 1):S13–8. [PubMed: 11464920]
5. Kondo K. Motor neuron disease: changing population patterns and clues for etiology. Adv Neurol. 1978; 19:509–43. [PubMed: 369333]

6. Maasilta P, Jokelainen M, Loytonen M, Sabel CE, Gatrell AC. Mortality from amyotrophic lateral sclerosis in Finland, 1986–1995. Acta Neurol Scand. 2001; 104:232–5. [PubMed: 11589652]
7. Litt J, Tran N, Malecki KC, Neff R, Resnick B, Burke T. Identifying priority health conditions, environmental data, and infrastructure needs: a synopsis of the Pew Environmental Health tracking project. Environ Health Perspect. 2004; 112:1414–8. [PubMed: 15471735]
8. Brown RC, Lockwood AH, Sonawane BR. Neurodegenera-tive diseases: an overview of environmental risk factors. Environ Health Perspect. 2005; 113:1250–6. [PubMed: 16140637]
9. The Scottish Motor Neuron Disease Registry: a prospective study of adult onset motor neuron disease in Scotland. Methodology, demography and clinical features of incident cases in 1989. J Neurol Neurosurg Psychiatry. 1992; 55:536–41. [PubMed: 1640227]
10. Forbes RB, Colville S, Cran GW, Swinger RF. Unexpected decline in survival from amyotrophic lateral sclerosis/motor neuron disease. J Neurol Neurosurg Psychiatry. 2004; 75:1753–5. [PubMed: 15548498]
11. Traynor BJ, Codd MB, Corr B, Forde C, Frost E, Hardiman O. Incidence and prevalence of ALS in Ireland, 1995–1997: a population based study. Neurology. 1999; 52:504–9. [PubMed: 10025778]
12. Piemonte and Valle d'Aosta Register for Amyotrophic Lateral Sclerosis (PARALS). Incidence of ALS in Italy: evidence for a uniform frequency in Western countries. Neurology. 2001; 56:239–44. [PubMed: 11160962]
13. Logroscino G, Beghi E, Zoccolella S, Palagano R, Fraddosio A, Simone IL, et al. SLAP Registry. Incidence of amyotrophic lateral sclerosis in southern Italy: a population based study. J Neurol Neurosurg Psychiatry. 2005; 76:1094–8. [PubMed: 16024886]
14. Beghi, E. On Behalf Of The European ALS Register Eurals Group. Neuromuscul Disord; 127th ENMC International Workshop: Implementation of a European Registry of ALS Naarden; The Netherlands. 8–10 October 2004; 2006. p. 46-53.
15. Miller RG, Anderson FA, Bradley WG, Brooks BR, Mitsumoto H, Munsat TL, Ringel SP. the ALS C A.R.E. Study Group. The ALS Patient Care Database: goals, design, and early results. Neurology. 2000; 54:53–7. [PubMed: 10636125]
16. Kasarkis EJ, Dominic K, Oddone EZ. The National Registry of Veterans with Amyotrophic Lateral Sclerosis: Department of Veterans Affairs Cooperative Studies Program (CSP) #500a. Amyotroph Lateral Scler Other Motor Neuron Disord. 2004; 5(Suppl 1):129–32. [PubMed: 15512895]
17. Brooks BR. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial 'Clinical limits of amyotrophic lateral sclerosis' workshop contributors. J Neurol Sci. 1994; 124(Suppl):96–107. [PubMed: 7807156]
18. Stickler DE, Royer JA, Hardin JW. Validity of hospital discharge data for identifying cases of amyotrophic lateral sclerosis. Muscle Nerve. 2011; 44:814–6. [PubMed: 22006696]
19. Benatar M, Wuu J, Usher S, Ward K. Preparing for the U.S. National ALS Registry: lessons from a pilot project in the State of Georgia. Amyotroph Lateral Scler. 2011; 12:130–5. [PubMed: 20843169]
20. Benesch C, Witter DM, Wilder AL, Duncan PW, Samsa GP, Matchar DB. Inaccuracy of the International Classification of Diseases (ICD-9-CM) in identifying the diagnosis of ischemic cerebrovascular disease. Neurology. 1997; 49:660–4. [PubMed: 9305319]
21. Tennis P, Bombardier C, Malcolm E, Downey W. Validity of rheumatoid arthritis diagnoses listed in the Saskatchewan hospital separation database. J Clin Epidemiol. 1993; 46:675–83. [PubMed: 8326356]
22. Katz JN, Barrett J, Liang MH, Bacon AM, Kaplan H, Kieval RI, et al. Sensitivity and positive predictive value of Medicare Part B physician claims for rheumatologic diagnoses and procedures. Arthritis Rheum. 1997; 40:1594–600. [PubMed: 9324013]
23. Cooper GS, Yuan Z, Stange KC, Dennis LK, Amini SB, Rimm AA. The sensitivity of Medicare claims data for case ascertainment of six common cancers. Med Care. 1999; 37:436–44. [PubMed: 10335746]

24. Fisher ES, Whaley FS, Krushat M, Malenka DJ, Fleming C, Baron JA, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. Am J Public Health. 1992; 82:243–8. [PubMed: 1739155]
25. Peabody JW, Luck J, Sharad J, Bertenthal D, Glassman P. Assessing the accuracy of administrative data in health information systems. Med Care. 2004; 11:1066–70. [PubMed: 15586833]
26. Humphries KH, Rankin JM, Carere RG, Buller CE, Kiely FM, Spinelli JJ. Comorbidity data in outcomes research: are clinical data derived from administrative databases a reliable alternative to chart review? J Clin Epidemiol. 2000; 53:343–9. [PubMed: 10785564]
27. Green J, Wintfeld N. How accurate are hospital discharge data for evaluating effectiveness of care? Med Care. 1993; 31:719–31. [PubMed: 8336511]
28. Worth RM, Mytinger RE. Medical insurance claims as a source of data for research: accuracy of diagnostic coding. Hawaii Med J. 1996; 55:9–11. [PubMed: 8786232]
29. Szeto HC, Coleman RK, Gholami P, Hoffman BB, Goldstein MK. Accuracy of computerized outpatient diagnoses in a Veterans Affairs general medicine clinic. Am J Manag Care. 2002; 8:37–43. [PubMed: 11814171]
30. Wang PS, Walker A, Tsuang M, Orav EJ, Levin R, Avorn J. Strategies for improving comorbidity measures based on Medicare and Medicaid claims data. J Clin Epidemiol. 2000; 53:571–8. [PubMed: 10880775]
31. Bright RA, Avorn J, Everitt DE. Medicaid data as a resource for epidemiologic studies: strengths and limitations. J Clin Epidemiol. 1989; 10:937–45. [PubMed: 2681546]
32. Fisher ES, Baron JA, Malenka DJ, Barrett J, Bubolz TA. Overcoming potential pitfalls in the use of Medicare data for epidemiologic research. Am J Public Health. 1990; 80:1487–90. [PubMed: 2240336]
33. Pope GC, Urato CJ, Kulas ED, Kronick R, Gilmer T. Prevalence, expenditures, utilization, and payment for persons with MS in insured populations. Neurology. 2002; 58:37–43. [PubMed: 11781403]
34. Taylor DH, Fillenbaum GG, Ezell ME. The accuracy of Medicare claims data in identifying Alzheimer's disease. J Clin Epidemiol. 2000; 55:929–37. [PubMed: 12393082]
35. Thacker SB, Stroup DF. Future directions for comprehensive public health surveillance and health information systems in the United States. Am J Epidemiol. 1994; 140:383–97. [PubMed: 8067331]
36. Allen KD, Kasarskis EJ, Bedlack RS, Rozear MP, Morgenlander JC, Sabet A, et al. The national registry of veterans with amyotrophic lateral sclerosis. Neuroepidemiology. 2008; 30:180–90. [PubMed: 18421218]

Table I

Motor neuron disease codes (ICD-9-CM, ICD-10 and VBA) used to identify potential ALS cases.

System	Code	Description
ICD-9-CM	335.2	Motor neuron disease unspecified
ICD-9-CM	335.20	Amyotrophic lateral sclerosis
ICD-9-CM	335.21	Progressive muscular atrophy
ICD-9-CM	335.22	Progressive bulbar palsy
ICD-9-CM	335.23	Pseudobulbar palsy
ICD-9-CM	335.24	Primary lateral sclerosis
ICD-9-CM	335.29	Other motor neuron disease
ICD-10	G12.2	Motor neuron disease
VBA	8017	Amyotrophic lateral sclerosis
VBA	8023	Progressive muscular atrophy
VBA	8005	Bulbar palsy

VBA, Veterans Benefits Administration.

Table II

Numbers of individuals identified with an ICD-9-CM MND code and charts reviewed by site.

	Individuals identified in databases [a]		Individuals identified only by site [b]		Total individuals identified		Total charts reviewed		Charts reviewed–database-identified individuals		Chart reviewed–site-identified individuals	
Site	***n***	**(%)**	***n***	**(%)**	***n***	**(%)**	***n***	**(%)**	***n***	**(%)**	***n***	**(%)**
Emory University	1889	78%	524	22%	2413	10%	446	9%	323	72%	123	28%
HMORN	15,360	85%	2592	15%	17,952	75%	2587	54%	592	23%	1995	77%
Mayo Clinic	1478	100%	5	< 1%	1483	6%	547	12%	544	99%	3	1%
South Carolina	1628	78%	462	22%	2090	9%	1174	25%	901	77%	273	23%
Total					23,938		4754					

MND, motor neuron disease.

[a] Individuals identified in the existing administrative datasets (Medicare, Medicaid, VBA, VHA).

[b] Individuals identified only in the site administrative and/or medical datasets.

Amyotroph Lateral Scler Frontotemporal Degener. Author manuscript; available in PMC 2015 September 29.

Table III

Demographic characteristics of individuals identified with MND code and charts reviewed.

	Individual charts reviewed		Individual charts not reviewed		*p*-value[a]	Total individuals	
Race	***n***	**(%)**	***n***	**(%)**	**< 0.0001**	***n***	**(%)**
White	3316	70%	15,579	81%		18,895	79%
Black	495	10%	1246	6%		1741	7%
Other	167	4%	998	5%		1165	5%
Unknown	776	16%	1361	7%		2137	9%
	4754		19,184			23,938	
Gender	***n***	(%)	***n***	**(%)**	**0.0001**	***n***	**(%)**
Male	2642	56%	10,182	53%		12,824	53%
Female	2085	44%	8610	45%		10,695	45%
Unknown	27	1%	392	2%		419	2%
Age	***n***	**(%)**	***n***	**(%)**	**< 0.0001**	***n***	**(%)**
< 18	5	< 1%	13	< 1%		18	< 1%
18–39	341	7%	594	3%		935	4%
40–59	1434	30%	3158	16%		4592	19%
60–79	2348	49%	9580	50%		11,928	50%
80+	590	12%	4247	22%		4837	20%
Unknown	36	1%	1592	8%		1628	7%

MND, motor neuron disease.

[a] χ^2 test.

Table IV

Demographic characteristics of chart-reviewed individuals by ALS code compared with other MND codes.

	ALS code		Other code		*p*-value[a]	Total charts	
Race	***n***	**(%)**	***n***	**(%)**	**0.0251**	***n***	**(%)**
White	2443	70%	873	68%		3316	70%
Black	337	10%	158	12%		495	10%
Other	131	4%	36	3%		167	4%
Unknown	566	16%	210	16%		776	16%
Gender	***n***	**(%)**	***n***	**(%)**	**<.00001**	***n***	**(%)**
Male	1971	57%	671	53%		2642	56%
Female	1503	43%	582	46%		2085	44%
Unknown	3	0%	24	2%		27	1%
Age	***n***	**(%)**	***n***	**(%)**	**< 0.0001**	***n***	**(%)**
< 18	2	0%	3	0%		5	0%
18–39	238	7%	103	8%		341	7%
40–59	1087	31%	347	28%		1434	30%
60–79	1769	51%	579	46%		2348	50%
80+	372	11%	218	17%		590	13%

MND, motor neuron disease.

[a] χ^2 test.

Amyotroph Lateral Scler Frontotemporal Degener. Author manuscript; available in PMC 2015 September 29.

Table V

Best algorithm for determining true cases of ALS from administrative data. [a]

		Neurologist review		
ALS	**Not ALS**	**Algorithm**	**ALS**	**Not ALS**
• A visit for ALS in ≥ 1 year + death certificate [b] or Rx for riluzole [c]; OR	• No ALS visit [e] & no Rx for riluzole [c]; OR	ALS	1385	252
• A visit for ALS in ≥ 2 years + neurologist visit [d]; OR	• A visit for ALS in 1 year & no neurologist visit [d]; OR	Not ALS	211	1474
• Age < 65, a visit for ALS in Medicare + neurologist visit; OR	• Age < 18 years; OR	Sensitivity = 0.87 Specificity = 0.85		
• A visit for ALS in ≥ 1 year + neurologist visit [d] and a visit for ALS in another database; OR	• No visit for ALS in any database; OR	PPV = 0.85 NPV = 0.87		
• A visit for ALS in ≥ 3 databases; OR	• Death certificate only			
• A visit for ALS in 1 year and ≥ 5 neurologist visits [d]				

PPV, positive predictive value; NPV, negative predictive value.

[a] National Databases include Veterans Health Administration, Veterans Benefits Administration, Medicare, and Medicaid.

[b] Death certificate includes ICD-10 code G12.2 for motor neuron disease. Death certificates are not an independent database because there is not a specific code for ALS.

[c] Riluzole is the only prescription medication specifically used to treat ALS.

[d] In the same database.

[e] One or more visits for a motor neuron disease other than ALS.