Proposal for the National ALS Registry
Version 23e: December 21, 2017
(includes amendments 42-46)

Table of Contents

INTRODUCTION

Residents of communities living near hazardous waste sites have expressed concerns about elevated rates of selected neurological diseases such as amyotrophic lateral sclerosis (ALS). The absence of population-based surveillance systems or registries for neurological diseases from which estimates of prevalence and incidence could be obtained, as well as enumerating cases within a specific community, makes it difficult to address these concerns. Attempts to address these questions have resulted in efforts to identify all the cases within a specific geographic area by working with local neurologists and other medical care providers. This is both time consuming and costly. Because of previous experience by the Agency for Toxic Substances and Disease Registry (ATSDR), as well as other researchers trying to obtain information directly from medical care providers, it was decided to first identify existing surveillance systems, registries, and databases for selected neurological diseases and then use this information to examine additional ways to identify affected individuals to be included in a population-based surveillance system/registry. Four pilot projects were conducted which evaluated the feasibility of accurate identification of ALS cases using administrative data from the Centers for Medicaid and Medicare Services (CMS), the Veterans Health Administration (VHA), and the Veterans Benefits Administration (VBA) when compared with the medical records.

This protocol describes the methodology for developing the National ALS Registry/population-based surveillance using existing administrative data and self-registration of affected individuals. The primary objective of this surveillance system/registry is to obtain reliable information on the incidence and prevalence of ALS and to better describe the demographic characteristics (age, race, sex, and geographic location) of those with ALS.

BACKGROUND

Disease Description and Epidemiology

In 1869, the French neurologist Jean-Martin Charcot described a unique condition characterized by deterioration of both lower and upper motor neurons, and this condition was termed amyotrophic lateral sclerosis (ALS).[1] Many people know ALS as Lou Gehrig's disease, named after the famous baseball player who, in 1939, retired because of his illness.

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 onset of ALS is age-related with the highest rate of onset occurring between 55 and 75 years of age.[2-4] Prognosis also appears to be age-related with slightly better survival occurring among those with a younger age at onset. 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 to 1,[4, 5] but recent studies have suggested that this sex difference is decreasing over time.[4, 6]

In addition to ALS, several other less common conditions are classified under the general term of motor neuron disease, but ALS accounts for 85 percent or more of all motor neuron cases. Most individuals who are initially diagnosed with these other conditions will ultimately progress to include both upper and lower motor neurons and thus will be diagnosed as having ALS.[2, 7]

Differential diagnosis of ALS requires a neurological exam as well as neurophysiological tests and other tests to rule out non-motor neuron diseases and other motor neuron diseases with

restricted presentations. False-negative rates can be high in the early stages of the disease[8, 9] and false-positive misdiagnoses have been shown to occur in 7 to 8 percent of cases.[10, 11] The diagnosis of ALS will become more uniform worldwide as the World Federation of Neurology El Escorial criteria and its subsequent revision are utilized.[12, 13]

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 these diseases.[14, 15] In addition, such a system could provide an unbiased source from which to recruit patients to participate in future research studies.

Surveillance

Public health surveillance is defined as "the ongoing, systematic collection, analysis, and interpretation of health data essential to the planning, implementation, and evaluation of public health practice, closely integrated with the timely dissemination of these data to those who need to know. The final link of the surveillance chain is the application of these data to prevention and control. A surveillance system includes a functional capacity for data collection, analysis, and dissemination linked to public health programs."[16] Surveillance is important to monitor changes in incidence and prevalence of a condition as well as to provide a source for patients with specific conditions that can be asked to participate in research studies. Surveillance data can also be used in planning for health care needs, detecting changes in health practices, and assessing the burden of disease. For chronic diseases, monitoring the burden of disease (morbidity, disability, and mortality) may be very important.[17] To date, national disease surveillance systems have been related primarily to infectious diseases with cancer and birth defects being the two exceptions.

In 1992, directors of the World Health Organization (WHO) non-communicable disease collaborating centers and key officials in centers for non-communicable diseases advocated for the increased surveillance of non-communicable diseases. This recommendation was based on the lack of incidence data for non-communicable diseases.[18]

Traditionally, surveillance systems have relied on physicians and other health care providers "reporting" to a specified entity, usually the state or local health department; that information can then be relayed to the next level as appropriate. The designation of "reportable" is conferred by the Council of State and Territorial Epidemiologists (CSTE), which was established in the 1950s. Once a disease has been designated reportable, each state must decide if current health department authorities would include the new disease or whether new legislation must be sought. Historically no non-communicable diseases have been made reportable by CSTE, including cancer. Cancer is reportable in most states; however, this was accomplished by Congress passing Public Law 102-515, Cancer Registries Amendment Act. This legislation required the authorization of a statewide registry under state law before receiving Federal funds.

Unfortunately, physicians have historically been poor reporters of disease; for that reason laboratory and hospital reporting have been built into surveillance systems.[19] Because physicians do not make good "reporters," and the history of making diseases nationally reportable has mostly excluded non-communicable diseases, this does not appear to be the best strategy for surveillance of selected neurological diseases.

As a more feasible strategy, Thacker and Stroup describe a comprehensive public health

surveillance system which would be a network of health information systems linked electronically. Data for this system would be collected from many sources including population-based systems (e.g., vital statistics), provider-based systems (e.g., physician, laboratory, and hospital records), and payer systems (e.g., Medicare or Medicaid).[20]

Sources of data and reliability of coding

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 of coding in large datasets such as Medicare and Medicaid. The issue is how reliable are these data for studies given that 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.[21-24] 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 was 91.4%.[25]

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.[26] Changes in reimbursement using Diagnosis Related Group (DRG) may also cause some coding inconsistencies. When multiple codes are allowed, it is important to understand the uses of each and ascertain if there are differences in reliability. It is also important to understand the value of using multiple codes. For example, many chronic conditions such as diabetes might not be the "reason" for the encounter but may be listed as a contributing or comorbid condition.[27] Likewise, individuals with

ALS might be seen for symptoms such as trouble walking or swallowing.

Usually only one database is used for a particular study. It is important to understand the limitations of the dataset being used.[28] It is also important to evaluate the dataset on a macro level to assess gaps.[29] Several researchers have pointed out that to obtain accurate information it is important not to rely on just one encounter/report because there can be changes in diagnosis[30] and existing chronic conditions may not be listed in each encounter.[31] It could be necessary for some types of research to obtain additional information from another dataset. 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.[32] The first mention of a condition does not necessarily indicate that it was diagnosed on that date but might just merely indicate the first time it was documented. If additional procedures are required to be documented along with a diagnosis, this can assist in determining whether the case is incident.[33] 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 as an inpatient, some claims would have been missed without using the outpatient information. The authors also point out the importance of including information from the Veterans Administration (VA) because some Medicare recipients might receive care at a VA facility which would not be reflected in the Medicare files.[34] Therefore, using multiple data sets, creating an algorithm to identify cases which includes 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. The prevalence estimated increased with two years of data to 24, 36 and 71 per 10,000 enrollees respectively.[35] 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 only identified 29% of the patients, whereas only physician encounters and institutional outpatient files identified 75% of the patients. Using 5 years of inpatient and outpatient data, 79% of the cases were identified. An analysis of clinical data on the patients revealed that those with less severe disease were less likely to be identified.[36]

The particular use of data from Health Maintenance Organizations (HMO) comes from studies conducted by the Kaiser Permanente Center for Health Research (CHR). CHR has participated in several surveillance projects related to asthma and infectious diseases. In one such study, an algorithm is being developed and validated for identifying individuals with prevalent (pre-existing) asthma and an algorithm for identifying cases of incident asthma from the subset of members who do not already have prevalent asthma. The study will estimate the incidence of asthma in various age-sex strata, and the cost of an ongoing surveillance system for incident cases based on these tools.[37]

Legislative Mandate

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 October 10, 2008 by President Bush and became Public Law No: 110-373. The purpose of the registry as described in the bill, is to: (1) better describe the incidence and prevalence of ALS in the United States; (2) examine appropriate factors, such as environmental and occupational, that might be associated with the disease; (3) better outline key demographic factors (such as age, race or ethnicity, gender, and family history of individuals who are diagnosed with the disease) associated with the disease; and (4) better examine the connection between ALS and other motor neuron disorders that can be confused with ALS, misdiagnosed as ALS, and in some cases progress to ALS. The registry will collect personal health information that may provide a basis for further scientific studies of potential risks for developing ALS.

RATIONALE

A number of private databases have been created to study ALS which have included a large amount of clinical data and has been used to study the natural history of the disease as well as to monitor health care decisions. Each database was created to answer a specific research question; therefore, it is likely to be difficult to get agreement on what clinical information should be universally collected.

In general, existing databases are valuable for research and for surveillance as long as the researcher recognizes the limitations of the data, limitations in quality of the original data collection, and any biases that might arise in variations in ascertainment or treatment of the disease being studied.[38] The use of multiple existing databases such as Medicare, Medicaid, and Veterans Administration should be a feasible way to identify patients with ALS.

The dilemma of what clinical information to collect and how to standardize the collection and classification of this information could be eliminated by creating a surveillance system that collected a minimal amount of information on each patient. The more detailed information about the patient with ALS would reside with the source of the information. The minimal data would be used to describe the prevalence and incidence of ALS. It would also be used to identify cases to contact for consent to participate in research studies. It would be necessary to request the additional information from the source of the case if a patient agreed to participate in a study. Some data not available in administrative/claims data will be collected through voluntary surveys. (See Survey section)

PILOT PROJECTS

To evaluate the feasibility of using existing administrative data to identify cases of ALS, ATSDR funded four geographically diverse pilot projects including tertiary care facilities for ALS, HMOs, and state based organizations. These four pilot projects matched data from Medicare, Medicaid, the Veterans Health Administration, and Veterans Benefits Administration to data available within the four pilot project sites administrative and clinical databases for a 5-year time period (January 1, 2001 – December 31, 2005). ATSDR provided the pilot projects with individual encounters with an ICD-9 code for any MND (335.2-335.29) for the specific project catchment area. Pilot projects completed a standardized spreadsheet for each individual found in any database indicating in which database(s) a record was located, ICD-9 code recorded for the encounter, as well as the years and types of providers seen. Medical records were abstracted and diagnoses verified. A deidentified dataset was sent to ATSDR for analysis. All individuals who were identified with a possible ALS diagnosis, as indicated by ICD-9 code for any MND, and

had their medical record reviewed by a neurologist from the four pilot projects were combined. Approximately 4400 medical records were reviewed. It was possible to develop algorithms using variables from the administrative data that identified true cases of ALS (verified by a neurologist) (Table 1). Similar results were found in the individual pilot project analyses. These pilot projects were determined to be not human subjects research at ATSDR because only deidentified data were provided by the partners.

Table 1 – Best algorithm for determining true cases of ALS from administrative data[1]

ALS	Not ALS	Algorithm		Neurologist Review	
• ALS in ≥ 1 year + Death Certificate[2] or Rilutek[3] • ALS in > 2 years + Neurologist visit* • Age < 65, ALS in Medicare + Neurologist visit • ALS in > 1 years + Neurologist visit* & ALS in another database • ALS in > 3 databases * In same database	• No ALS visit[4] & no Rx for Rilutek • ALS in 1 year & no Neurologist visit* • Age < 18 years • No ALS in any database • Death Certificate only			ALS	Not ALS
			ALS	1282	265
			Not ALS	233	1531
			Sensitivity = 0.85		
			Specificity = 0.85		
			PPV = 0.83		
			NPV = 0.87		

1. National Databases include Veterans Health Administration, Veterans Benefits Administration, Medicare, and Medicaid.
2. Death Certificate includes ICD-10 code G12.2 for MND. Death Certificates are not an independent database because there is not a specific code for ALS.
3. Rilutek is the only prescription medication specifically used to treat ALS.
4. One or more visits for an MND other than ALS.

Any individual not falling into the ALS or Not ALS category is in the possible ALS category. For example, an ALS code in 2 years but no visit to a neurologist would be in the possible ALS category and reevaluated as new data become available.

OBJECTIVES

The objective of this project is to develop a population-based surveillance system/registry for ALS. The primary goal of the surveillance system/registry is to obtain reliable information on

the incidence and prevalence of ALS and to better describe the demographic characteristics (age, race, sex, and geographic location) of those with ALS. The secondary goal of the surveillance system/registry is to collect additional information on potential risk factors for ALS including, but not limited to, family history of ALS, smoking history, and military service.

SURVEILLANCE DESIGN AND METHODS

A population-based surveillance system/registry for ALS will be created by identifying persons with ALS from existing administrative databases and self-registration by interested ALS patients. A flowchart illustrating surveillance system/registry inclusion can be found in Appendix A. A minimal amount of data will be contained in the surveillance system/registry including:

Name
City
State
Last 5 digits of the Social Security Number
Month and Year of birth
Sex
Race
Date first in database
Database reporting/self-registration
Vital status
Date of death

Data Base Descriptions

A number of databases will be used including Medicare, Medicaid, Veterans Health Administration, and Veterans Benefits Administration. These national databases cover approximately 90 million people.

Veterans Health Administration data includes inpatient, outpatient, and pharmacy records for veterans receiving health care benefits. Approximately 20% of veterans qualify for this benefit.

Veterans Benefits Administration data includes records for veterans receiving pensions or compensation for disabilities considered service related. ALS is considered service related if it is diagnosed within 1 year of separation from active duty.

Medicare data includes inpatient, outpatient, pharmacy, and long-term care records for individuals receiving this benefit. Medicare is United States government provided insurance for people age 65 or older, some disabled people under age 65, and people of all ages with End-Stage Renal Disease (permanent kidney failure treated with dialysis or a transplant). Individuals approved for 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 includes 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. It is an entitlement program that is jointly funded by the states and federal government, and is managed by the states.

Initial Identification of Possible Cases of ALS

Anyone in an administrative data base with ICD-9-CM codes of 335.2 -335.29 will be identified. All individuals with any Motor Neuron Disease (MND) will be evaluated because other MNDs can be confused with ALS. Because of the short clinical course of ALS, it is necessary to differentiate between individuals seen once to rule out ALS and those with only one occurrence

of an ALS code because they died soon after diagnosis.

Algorithm

Using data from the pilot projects, ATSDR developed algorithms which will be used on each database to identify "cases" of ALS from all those identified by the methods mentioned previously. Applying the algorithm to possible cases will sort individuals into three categories, ALS, possible ALS, and definitely not ALS. Only those identified as having ALS will be entered into the registry. As new data become available, the algorithm will be rerun looking for new cases and reevaluating the possible ALS cases to see if the new information clarifies their case status. It is anticipated that new data will be available on a yearly basis and that the algorithms will be rerun with those individuals identified as possible ALS cases to see if case status is clarified.

Self-registration

All cases of ALS are not identified from existing administrative data for a variety of reasons, including eligibility requirements for the various entitlement programs, miscoding, and misdiagnosis. The preliminary data analysis of the pilot data suggests that individuals with ALS can be identified from existing administrative datasets. However, younger, self-payees, and those who died sooner after diagnosis are more likely to be missed in the existing data, therefore, individuals will be allowed to self-identify for inclusion in the registry. We would like all ALS patients to register so that they can take part in the collection of data not available in administrative data (See Survey Section). ALS patients will be consented prior to registering. The VA ALS registry used six validation questions to screen possible individuals with ALS. Of

the 98.7% of the veterans who passed the screening, 93.4% were confirmed to have ALS or another MND based on neurologist medical record review.[39] We will use these six questions as part of the registration process (Appendix B). Once a person self-identifies, he/she will be asked to complete the validation questions. If the answers indicate the person has ALS, he/she will be allowed to create an account to become part of the registry (Appendix C) after consenting (Appendix D). Technical assistance will be available for ALS patients because they may have difficulty with the computer because of physical disabilities related to their disease. The information collected during the registration process is primarily used to make sure there are not duplicates in the registry given that case ascertainment will come from multiple sources. The matching algorithm relies primarily on Social Security Number (SSN). In an evaluation of more than 300,000 records received from CMS, no duplicates were identified using the last 5 digits of the SSN. In addition to SSN, we will have first and last name which should make duplicates easily identifiable.

It is unknown who many individuals will self-register, however this is a motivated group. The VA ALS registry allowed self-referral and 18% of the individuals evaluated for inclusion in the registry were from self-identification, although a number of these individuals had also been identified via records review.[39] Self-registration will be encourage so that the registrant can participate in additional data collection activities. Two advocacy groups, the ALS Association and the Muscular Dystrophy Association, ALS Division, have expressed their intention to advertise the existence of the registry to their constituents.

Death Data

The National Death Index (NDI) is a central computerized index of death record information on file in the State vital statistics offices. The NDI is a national file of identifying death record information (beginning with 1979 deaths) compiled from computer files submitted by State vital statistics offices. Death records are added to the NDI file annually, approximately 12 months after the end of a particular calendar year. Cases of ALS identified by the registry will be sent to the NDI to determine vital status. Information on vital status will be maintained in the registry.

On a yearly basis ATSDR will ask the Mortality Statistics Branch at the National Center for Health Statistic to search their data for death certificates mentioning ALS. Since 2003, searches can be made on the text of the death certificate so that the search will be specific for ALS and synonyms for ALS such as Lou Gehrig's disease, and is not dependent on the ICD-10 code of G12.2. Currently ICD-10 codes do not include a specific code for ALS. Rather ICD-10 G12.2 codes for all MNDs. The death certificate number will be submitted to the state and a death certificate purchased. This information will be used to verify diagnosis as well as identify cases that may have been missed by the registry.

Surveys

Congress anticipated that two additional purposes for the registry would be to examine appropriate factors, such as environmental and occupational, that might be associated with the disease and to better outline key demographic factors (such as age, race or ethnicity, gender, and family history of individuals who are diagnosed with the disease) associated with the disease. The information necessary to examine these demographic and potential risk factors for disease are not usually part of a registry or public health surveillance system. To enable the collection of

additional information from registrants who volunteer, a series of short survey modules will be available for completion via a secure web portal. We are using survey questions validated by the ALS Consortium of Epidemiologic Studies (ACES).[41] The survey has been divided in to two parts: 1) Essential Questionnaire and 2) Follow-up Questions to capture characteristics of ALS registrants more efficiently. Essential Questionnaire is designed to be completed and submitted at the time of registration and consist of questions pulled from the original ACES survey design to capture various attributes (Demography, Lifestyle Factors, Environmental Factors, and ALS-associated and Clinical Factors) of respondents. Based on the responses provided for the Essential Questionnaire, registrants will be presented with Follow-up Questions as short modules because of the physical limitations of the study population (Appendix E). All surveys are designed to be answered only once except for the symptoms survey which can be answered at 3 and 6 months the first year and every 6 months thereafter. The ALS Functional Rating Scale-Revised (ALSFRS-R) is a standard set of questions used by physician ns to measure functioning overtime. Researchers have developed and tested a self-administered version of the ALSFRS-R which showed excellent reliability to change over time.[40] The published version of the self-administered ALSFRS-R was slightly modified to make the question responses more user friendly. Individuals will be consented prior to registering with the National ALS Registry and completing any survey modules. (See human subjects section) A webinar outlining how to complete the survey and demonstrating the different takes of response options will be available on the National ALS Registry website (Appendix I)

Although the generalizability of the survey data will be dependent on the number of individuals who choose to participate, these data can be used to inform risk factor specific study protocols.

Policy for handling surveys with a status of In-Progress.

The Agency for Toxic Substances and Disease Registry (ATSDR) has identified a number of National Amyotrophic Lateral Sclerosis (ALS) Registry participant survey(s) that were initiated, but not submitted for analysis. It is presumed that registry participants with survey(s) designated as “In-Progress” for an extended period of time are either deceased or have been unable to do so because of disease progression.

ATSDR has created a policy to address surveys that are complete and have been “In-Progress” for an extended period of time. The policy would allow ATSDR to change those surveys with a status of “In-Progress” for a period longer than one-year and with no activity during that period to a status of “Complete”. This would allow ATSDR to include this survey data with completed data submitted by registry participants.

ATSDR will not modify or change the progress of a registrant’s survey if it is partially complete regardless of whether or not the person has completed other surveys. ATSDR will assume that registrants in this situation decided not to complete the risk factor survey.

If it is determined that a participant is alive or would like to complete a survey affected by this policy; ATSDR will reactivate the registrant’s account and change the status of the survey(s) back to “In Progress”, allowing the registrant to complete the survey(s).

ALS Research Notification

The National ALS Registry is the largest group of persons identified with ALS in the United States. One of the purposes of the registry as stated in the Congressional language “is to

facilitate research." Researchers outside of ATSDR/CDC would like to be able to contact persons in the registry about studies for which they might be eligible. The registry does not allow ATSDR to release identifiable information therefore; ATSDR has developed a system where ATSDR does the notification. A request form for researchers who want to contact registrants can be found in Appendix M. The investigator must provide documentation of IRB approval for the study and the recruitment materials.

ATSDR modified the website to allow registrants to indicate whether they wish to be notified of research opportunities (see Human Subjects Protection Section). ATSDR will only send the information to those who have agreed. There is a general webpage for this activity that acts as the gateway for researchers and participants. From this webpage, potential participants are directed to a second webpage containing specific information about patient participation. (Appendix N). ATSDR will create a webinar available on the National ALS Registry website that will explain the process to interested researchers (Appendix M)

Global Unique Identifier

Researchers in ALS would like to be able to use data from different studies without including PII. The Global Rare Diseases Registry (GRDR) Global Unique Identifier (GUID) is a subject identifier used to protect the confidentiality of a research subject. When submitting data, an investigator who has appropriate access to a subject's personally identifiable information (PII) uses the GUID tool to create a unique identifier for each subject in their study. Although a GUID is based on PII, a subject's PII never leaves the local research site. Information from particular fields on the subject's birth certificate is entered into the software, and one-way hash codes are generated based on this input. The GUID for a subject is the same regardless of when or where it is generated as long as it is on the same server. If the same subject enrolls in another

investigator's project or provides a biological specimen for a repository, the same information from his or her birth certificate is entered into the software by the second investigator and the same GUID is generated. In this way, data from a de-identified individual subject can be aggregated, tracked and linked across projects, time, databases, and biobanks, so when studying one aspect of patient's presentation, an investigator can relate it to other data from the same patient.

The purpose of the GRDR GUID is to create an identifying code that allows a single research participant's data to be compiled, even if the data is collected at different locations or by different studies. The GUID is created using personally identifying information such as a subject name, date of birth and city of birth, although this personally identifying information is never actually transferred to GRDR nor stored within the GRDR database. GRDR provides software to help users generate GUIDs for research participants. Given the appropriate personally identifying information, the software can generate a GRDR GUID for each participant. In cases where not enough personally identifying information is available to generate a GUID, a placeholder pseudo-GUID may be assigned. In addition, de-identified data are given a random unique code, known as a Global Unique Identifier (GUID). Use of the GUID allows GRDR to link together all submitted information on a single participant, giving researchers access to information that may have been collected elsewhere, without using names, addresses, or other identifying information.

Other groups such as NeuroBANK™ also create a GUID. Although the same data are used to create the GUID, the ID generated is different because it is server dependent. To make the Registry data more useful to researchers, we would ask participants if they would like to have

GUID(s) generated on their behalf. If they agree, they will be asked to provide the nine standard items required for generation. (Appendix N) As long as these nine items are entered exactly the same on the same GUID server, they will return the same unique number.

Human Subjects Protection

ATSDR is requesting a waiver of consent for including the data obtained from the existing data sources, including Medicare, Medicaid, Veterans Health Administration, and Veterans Benefits Administration. The research involves no more than minimal risk to the subjects because the information has already be collected, the waiver will not adversely affect the rights and welfare of the subjects as there is no interaction with the participants; the research could not practicably be carried out without the waiver because there are more than 15,000 individuals who would need to be contacted and contact information is not up-to-date; and the ATSDR website being created for the ALS registry will provide additional pertinent information after participation.

ATSDR is also requesting a waiver of consent for the six validation questions. The research involves no more than minimal risk to the subjects because the information other than date of diagnosis is not retained and only used to determine eligibility, the waiver will not adversely affect the rights and welfare of the subjects as there is penalty for not providing the information and we will not know who chooses not to participate; the research could not practicably be carried out without the waiver because we would be obtaining consent for people who were not eligible and the would be the only information collected about them; and the ATSDR website being created for the ALS registry will provide additional pertinent information after participation.

Registration and participation in providing additional information by completing surveys is entirely voluntary. Prior to determining eligibility, a privacy statement is displayed explaining the reasons for the data collection (Appendix F). A consent statement is also shown (Appendix D). If the person decides to proceed, he/she will follow the procedures outlined in self-registration. Once registered, an individual has the opportunity to provide additional information through survey modules. Because all information is collected electronically, the individual will have to agree with the consent statement prior to proceeding with registration. We are requesting a waiver of documentation of consent for this project. The research presents no more than minimal risk of harm to subjects and involves only the collection of survey data. The greatest risk to participants is from a breach of confidentiality. To minimize this risk, we have developed extensive data security procedures outlined in the data security section. Collection of this type of survey information (smoking history, family history of disease, occupation history, etc.) is often collected outside of the research context without consent.

Some patients with ALS have severe motor impairment making computer usage difficult. Whenever ATSDR is asked how he/she can register or take surveys, we recommend he/she ask a family member, friend, or volunteer at the advocacy group to assist. In some limited number of individuals, there is no one who can assist them. ATSDR would like to have the ALS System Administrator assist them. ATSDR would only assist those who contact them directly and have no one else to assist. This project has already been granted a waiver of documentation of consent so that would not change. The individual assisting with registration already has access to the data once it is entered so there is no impact on confidentiality.

A number of potential participants have contacted the registry staff or advocacy groups with questions about how to register when they do not have computer access or a family member with computer access who could assist them. Many of these individuals are active in the ALS community and attend conferences and events. We would like to be able to offer computer access at these events. Participants will not be recruited but rather provided the capability to register if they express an interest. In addition, individuals will be available to answer technical questions about the registration process in real-time. The computers will be in a private area with dividers separating the machines.

ATSDR would like to be able to communicate with individuals about their accounts but does not want to compromise their privacy. Therefore, we will add a box to the registration page (Appendix C) which would give us permission to contact them via email. Drafts of email communication are presented in Appendix H. We will only email participants who have checked the box. Those taking surveys might start a survey but not submit it which would cause the survey to be pending. One communication would go to participants with surveys pending. This communication would not be sent until a survey was pending for at least one month and only sent every 6 months for up to 2 years from registration.

ATSDR added a consent to be contacted about research studies by outside researchers to the ALS website (Appendix N). Only those agreeing will be contacted. When an investigator submits a request to ATSDR, it will be reviewed by the PI and other team members to make sure it is complete and appropriate. ATSDR will send the complete application to members of the

scientific review group. The application requires submitting documentation of IRB approval of the study and recruitment materials. If the scientific review committee approves the study, ATSDR sends one of two standard emails that have been approved by the IRB (page 83-84) and attaches a copy of the IRB approved recruitment material. When the application is approved by the scientific review committee, an email is sent to Registry participants who have agreed to receive clinical notifications. Individuals could receive up to three emails about the same study based on the recruitment requirements for the specific study. It will be up to each individual to contact the study if interested to obtain additional information.

Data security

Creating an account

External Users (ALS Patients / External Researchers) must self-register before accessing the ALS Web Portal. Personal information is collected during this registration process and users are allowed to create their own unique username and password. Users are also required to answer security questions which are used as alternative authentication credentials if their password is forgotten. Upon successful registration, users are required to login into their account using their username and password. External Users are authenticated against a backend SQL encrypted database.

Internal Users (CDC Employees / System Administrators) are required to be pre-approved by ATSDR management before accessing the ALS Intranet Web Portal. Once a user is approved, ATSDR management sends a request to the System Administrator to create a user account. The request must include the user's CDC User ID, First Name, Last Name, Gender, City, State, Country, and Email in order for the System Administrator to add the user to the ALS System.

Users must first log into the CDC network to access the ALS Intranet Web Portal and are authenticated using Active Directory. No login is required.

The ALS system creates a sequential unique identifier in the database every time a user account is created. This unique identifier identifies each user and is used to link user information inside the system. Another unique identifier (Last 5 of SSN) will be used to verify patient data outside of the ALS system.

Login procedures

For authentication purposes, users will be verified using their unique username along with their password. External Users are allowed to self-register online and create their own username. Duplicate checks are implemented during registration to ensure uniqueness of usernames and emails.

Password management

External users are allowed to change or reset their passwords, but are not allowed to retrieve their password. Passwords can be changed via the user's account after the user has been authenticated by providing the old password and can only be changed once every 6 days. If a user forgets his/her password, the password can be reset by providing alternate authentication credentials. These credentials include the user's username, registered email address, and a security question. Passwords are required to be reset every 60 days. Users will be given a 2 week email notice before their password expires. Users will be directed to reset their expired password if they attempt to login after their password has expired.

Usernames are unique and cannot be changed. Users must contact the System Administrator by

phone to retrieve their username. The System Administrator is required to ask verification questions before releasing any information to the user; which can include the user's first and last name, month & day of birth, City, State, Country, and two security questions.

The status of an account will change to inactive if the user has not logged into his/her account in 6 or more months. Users will be given a 2 week email notice before their account is inactivated. Users will be required to contact the System Administrator by phone to re-activate their account. The System Administrator will be required to verify the user by asking verification questions which include the user's First and Last Name, Date of Birth: Month & Year (ALS Patients only), Address: City, Province/State, Country, and 2 security questions.

No personal information or credentials can be sent to a user's email, only notices or confirmations.

User accounts cannot be removed and remain in the database permanently. Only the account status can change.

Encryption

Information in Identifiable Form (IIF) fields will be masked on the Graphical User Interface because of the sensitivity of the data. For example, month and year of birth will be masked.

All Private Identifying Information (PII) data which includes the last 5 digits of the SSN will be encrypted using AES_256 (Advance Encryption Standard 256 bit) encryption, the strongest encryption standard supported by SQL Server 2005.

To encrypt/decrypt data in database columns designed to hold PII data, a user must be given access to open and close a symmetric key. _

Minimize collection of identifiable information

The information required for registration has been limited to only that needed to make sure that an individual truly has ALS and is not already part of the registry. Address information has been limited to city and state, birth information has been limited to month and year of birth, and only the last five digits of the SSN will be collected.

Physical Controls

Production and test servers are stored in a server room secured by the CDC. Access tools are in place to secure entry into CDC buildings (Guards, ID Badges, Key Card, Cipher Locks, and Closed Circuit TV).

Data management

On a quarterly basis, data will be downloaded from the web-based portal and provided to ATSDR. ATSDR will merge the self-identified individuals into the registry after first checking for duplicates. The registry will be maintained on a secure server or stand-alone hard-drive. Access to the data will be limited to approved study personnel. Deidentified data sets will be used for data analysis.

Once it is determined that registry data are representative of ALS patients, ATSDR will provide back to the web-based portal on a quarterly basis a dataset which is deidentified, including only state, age, race, and sex. This dataset will be used to populate the surveillance/registry map available on the website (Appendix G)

Promotion/Recruitment

To make persons with ALS and their caregivers aware of the National ALS Registry, ATSDR will undertake a variety of activities. These will include an introduction on the ATSDR/CDC ALS webpage, a video introduction to the Registry, ads in magazines read by ALS patients and caregivers, factsheet, patient guide, and promotional giveaways. Giveaways will include small items valued at less than $3 and only have the name and web address of the National ALS Registry. Text of promotional materials and pictures of giveaways can found in Appendix I. ATSDR will also promote the registry through the use of social media such as Facebook, Twitter, and YouTube. Posts will be sent to individuals that follow CDC. Proposed posts can be found in Appendix J. The CDC YouTube page will have a link to the IRB approved video. In addition, a poster/flyer has been created for use by physicians who treat ALS patients telling them about the registry and where to obtain additional information. (Appendix K) Additional resources which will be posted to the National ALS Registry website to assist with registration and answer frequently asked questions can be found in Appendix L.

ATSDR has entered into a contract with the National Amyotrophic Lateral Sclerosis (ALS) Association (ALSA) to engage in regular outreach activities with ALS Association Chapters, Centers and Clinics and with people with ALS and their families. Outreach activities include providing education, assistance and guidance to Chapters on best practices to promote the National ALS Registry and how to assist patients with enrollment. ALSA has developed a Toolkit for Chapters to assist with these activities for the National ALS Registry (Appendix 0). In addition, ALSA will partner with Minor League Baseball because of the association of ALS with the famous baseball player, Lou Gehrig. ALSA has developed a Toolkit for Minor League

Baseball (Appendix P) that will be distributed to Minor League Baseball teams throughout the US. The National ALS Association created a website to provide National ALS Registry materials to PALS (Appendix Q).

ATSDR has also entered into a contract with the Muscular Dystrophy Association (MDA) ALS Division to engage in outreach to ALS patients who attend MDA clinics. The MDA specific materials call be found in Appendix R.

Biorepository

ATSDR has determined that it is feasible to add a biorepository component to the Registry. The specifics of how the biorepository will function can be found in Appendix S.

Evaluation

It will be important to evaluate the completeness of the surveillance/registry. Information will be captured each time a case is identified and from which database so that capture-recapture statistical techniques can be used to estimate the number of people missed.[42, 43] It is also important to identify the level of service, e.g., primary practice or neurologist, for each identified record as it assists in the evaluation of the reliability of the diagnosis. If the evaluation identifies groups of individuals underrepresented in the registry, additional case finding strategies will be developed.

CONCLUSION

There is a public health need for accurate estimates of people affected by neurodegenerative diseases to better assess the health care needs of the population, detect changes in health care

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,[20] there have been no attempts to initiate such a system on a national level. This endeavor will provide needed information on ALS which can be used by others.

REFERENCES

1. Bobowick AR, Brody JA. Epidemiology of motor-neuron diseases. *N Engl J Med* 1973;288:1047-55.

2. Mitsumoto H, Chad DA, Pioro EP. (1998). *Amyotrophic Lateral Sclerosis*. Philadelphia, PA: F.A. Davis Company.

3. Eisen A. Amyotrophic lateral sclerosis is a multifactorial disease. *Muscle Nerve* 1995;18:741-52.

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.

5. Kondo K. Motor neuron disease: changing population patterns and clues for etiology. *Adv Neurol* 1978;19:509-43.

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.

7. Norris F, Shepherd R, Denys E, U K, Mukai E, Elias L, Holden D, Norris H. Onset, natural history and outcome in idiopathic adult motor neuron disease. *J Neurol Sci* 1993;118:48-55.

8. Belsh JM. Diagnostic challenges in ALS. *Neurology* 1999;53:S26-30; discussion S35-6.

9. Belsh JM, Schiffman PL. Misdiagnosis in patients with amyotrophic lateral sclerosis. *Arch of Intern Med* 1990;150:2301-5.

10. Davenport RJ, Swingler RJ, Chancellor AM, Warlow CP. Avoiding false positive diagnoses of motor neuron disease: lessons from the Scottish Motor Neuron Disease Register. *J Neurol Neurosurg Psychiatry* 1996;60:147-51.

11. Traynor BJ, Codd MB, Corr B, Forde C, Frost E, Hardiman O. Amyotrophic lateral sclerosis mimic syndromes: a population-based study. *Arch Neurol* 2000;57:109-13.

12. 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.

13. Miller RG, Munsat TL, Swash M, Brooks BR. Consensus guidelines for the design and implementation of clinical trials in ALS. World Federation of Neurology committee on Research. *J Neurol Sci* 1999;169:2-12.

14. 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.

15. Brown RC, Lockwood AH, Sonawane BR. Neurodegenerative diseases: an overview of environmental risk factors. *Environ Health Perspect* 2005;113:1250-6.

16. Centers for Disease Control and Prevention (CDC). Comprehensive plan for epidemiologic surveillance: Centers for Disease Control, August 1986. Atlanta, GA: CDC, 1986.

17. Thacker SB, Stroup DF, Rothenberg RB. Public health surveillance for chronic conditions: a scientific basis for decisions. *Stat Med* 1995;14:629-41.

18. Shanghai declaration on non-communicable diseases. WHO Directors of Non-Communicable Disease Collaborating Centres and Key Officials. *BMJ* 1993;306:588.

19. Birkhead GS and Maylahn CM. (2000). State and Local Public Health Surveillance. In: Teutsch SM and Churchill RE (Eds.), *Principles and Practice of Public Health Surveillance – Second Edition* (pp. 253-286). New York: Oxford University Press.

20. Thacker SB, Stroup DF. Future directions for comprehensive public health surveillance and health information systems in the United States. *Am J Epidemio*. 1994;140:383-97.

21. 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.

22. 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.

23. Katz JN, Barrett J, Liang MH, Bacon AM, Kaplan H, Kieval RI, Lindsey SM, Roberts WN, Sheff DM, Spencer RT, Weaver AL, Baron JA. Sensitivity and positive predictive value of Medicare Part B physician claims for rheumatologic diagnoses and procedures. *Arthritis Rheum* 1997;40:1594-1600.

24. 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.

25. Fisher ES, Whaley FS, Krushat M, Malenka DJ, Fleming C, Baron JA, Hsia DC. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. *Am J Public Health* 1992;82:243-8.

26. 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.

27. Humphries KH, Rankin JM, Carere RG, Buller CE, Kiely FM. Spinelli JJ. Co-morbidity data in outcomes research: Are clinical data derived from administrative databases a reliable alternative to chart review? *J Clin Epidemiol.* 2000;53:343-9.

28. Green J, Wintfeld N. How accurate are hospital discharge data for evaluating effectiveness of care? *Med Care* 1993;31:719-31.

29. Hennessy S, Bilker WB, Weber A, Strom BL. Descriptive analyses of the integrity of a US Medicaid claims database. *Pharmacoepidemiol Drug Saf* 2003;12:103-111.

30. 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.

31. 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.

32. 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* 200;53:571-8.

33. Bright RA, Avorn J, Everitt DE. Medicaid data as a resource for epidemiologic studies: strengths and limitations *J Clin Epidemiol*. 1989;10:937-45.

34. 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.

35. 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.

36. Taylor DH, Fillenbaum GG, Ezell ME. The accuracy of medicare claims data in identifying Alzheimer's disease. *J Clin Epidemiol* 2000;55:929-37.

37. Kaiser Permanente Center for Health Research Website. Available at http://www.kpchr.org/public/default.asp. Accessed March 29, 2006.

38. Kuller H. The use of existing databases in morbidity and mortality studies (editorial). *Am J Public Health* 1995;85:1198-1200.

39. Allen KD, Kasarskis EJ, Bedlack RS, Rozear MP, Morgenlander JC, Sabet A, Sams L, Lindquist JH, Harrelson ML, Coffman CJ, Oddone EZ. The national registry of veterans with amyotrophic lateral sclerosis. *Neuroepidemiology* 2008;30:180-190.

40. Montes J, Levy G, Albert S, Kaufmann P, Buchsbaum R, Gordon PH, Mitsumoto H. Development and evaluation of a self-administered version of the ALSFRS-R. *Neurology* 2006;67:1294-6.

41. See http://aces.stanford.edu/

42. Coffman CJ, Horner RD, Grambow SC, Lindquist J. Estimating the Occurrence of Amyotrophic Lateral Sclerosis among Gulf War (1990-1991) Veterans Using Capture-Recapture Methods: An Assessment of Case Ascertainment Bias. *Neuroepidemiology* 2005;24:141-150.

43. Preux P, Druet-Cabanac M, Couratier P, Debrock C, Truong T, Marcharia W, Vallat J, Dumas M, Boutros-Toni F. Estimation of the amyotrophic lateral sclerosis incidence by capture-recapture method in Limousin region of France. *J Clin Epidemiol.* 2000;53:1025-29.