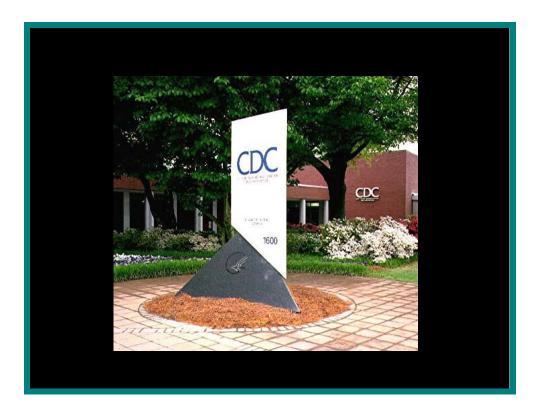# Centers for Disease Control and Prevention (CDC)
Agency for Toxic Substances and Disease Registry (ATSDR)

Summary Report: February 26, 2008

Amyotrophic Lateral Sclerosis and
Multiple Sclerosis Surveillance Annual Meeting:

Issues Related to Developing a National Surveillance System and Registries for Amyotrophic Lateral Sclerosis and Multiple Sclerosis

Amyotrophic Lateral Sclerosis and Multiple Sclerosis Surveillance Annual Meeting:

Issues Related to Developing a National Surveillance System and Registries for Amyotrophic Lateral Sclerosis and Multiple Sclerosis

Introductions, Welcome, and Overview of the Project and Goals

Introductions and Logistics

Wendy E. Kaye, PhD
Senior Epidemiologist
McKing Consulting Corporation

Dr. Kaye called the meeting to order, thanking those present for their attendance. Following a review of housekeeping issues, she led participants in a round of introductions.

Welcome and Opening Remarks

Anne Sowell, PhD
Associate Director for Science, Division of Health Studies
Agency for Toxic Substances and Disease Registry
Centers for Disease Control and Prevention

Dr. Sowell extended her welcome to the attendees, thanking them for taking time out of their busy schedules to share their work in amyotrophic lateral sclerosis and multiple sclerosis surveillance and to offer guidance to CDC pertaining to the important next stages of the surveillance projects. She explained that the Agency for Toxic Substances and Disease Registry (ATSDR) is administratively part of the Centers for Disease Control and Prevention (CDC), although it is a somewhat separate entity with respect to budget and regulations. The Division of Health Studies (DHS) is made up of three components: 1) Surveillance and Registries Branch, in which the ALS and MS surveillance projects are housed; 2) Geospacial Research, Analysis, and Services Program, that deals with GIS issues, a very big group that does a lot of work for the rest of CDC, as well as some work for the Department of Health and Human Services (DHHS); and 3) Health Investigations Branch, which examines health effects caused by Superfund sites and other locations where there are environmental toxins that may be impacting the health of communities. While DHS is fairly small, their work is very important and is of exceptional quality. Moreover, DHS is very excited about having the ALS and MS projects as part of this division and believes that this will prove to be a very interesting, long-term activity.

Overview of the Project and Goals

Oleg Muravov, MD, PhD
Medical Epidemiologist
Principle Investigator ALS and MS Pilot Projects
Agency for Toxic Substances and Disease Registry
Centers for Disease Control and Prevention

Dr. Muravov pointed out that throughout the years, ATSDR has funded various investigations to address community concerns regarding possible increases in autoimmune and immunological diseases. The major limitation of the types of investigations that have been conducted is a lack of baseline prevalence estimates, which makes it impossible to determine whether there is an increase of these diseases in the community. Given this, ATSDR funded several studies to determine the baseline prevalence for ALS and MS in several geographic areas. The completed studies resulted in useful prevalence data; however, the studies also showed that gaining access to neurologists' records is a time-consuming and costly endeavor. In addition, not all neurologists would grant access to their records, creating the potential for bias. In March 2006, ATSDR convened a workshop to discuss the creation of a national surveillance system for selected neurological and autoimmune diseases. Discussion also included the identification of existing registries and databases, selecting disease(s) for surveillance, and developing and testing methodology.

The decision was made to start with ALS and MS, secure funding for pilot projects, obtain access to existing national data sets (e.g., Center for Medicare and Medicaid Services, Veterans Health Administration, Veterans Benefits Administration, and National Multiple Sclerosis Society), and develop and fund pilot projects. Since the meeting in March 2006, ATSDR obtained funds from two different sources to fund both ALS and MS pilot projects, developed statements of work for the pilot projects, funded six pilot projects (e.g., four projects on ALS and 2 projects on MS), and held a workshop in February 2007 on Institutional Review Board (IRB) and Health Insurance Portability and Accountability Act (HIPAA) issues related to data sharing. The pilot projects for MS include: 1) New York State Multiple Sclerosis Consortium Registry; and 2) North American Research Consortium on Multiple Sclerosis Registry. The ALS projects include: 1) Mayo Clinic; 2) Emory University; 3) South Carolina Budget and Control Board; and 4) the HMO Research Network.

The advantages of conducting projects in ALS and MS simultaneously include: 1) Monetary savings (e.g., data from CMS would have been twice as expensive if requested separately; preparing the data requests are expensive and time consuming: one versus two requests); 2) More people are working on similar issues who can share ideas; 3) The ability to discuss progress and limitations related to the development of surveillance systems and registries for ALS and MS; and 4) The ability to discuss strategies for developing the surveillance systems and registries within the various medical settings and available administrative records.

In conclusion, Dr. Muravov stated that the purpose of this meeting was to discuss the progress and limitations related to the development of surveillance systems and registries for ALS and MS; discuss strategies for developing the surveillance systems and registries within the various medical settings and available administrative records; and discuss next steps. He strongly encouraged the pilot projects to candidly discuss the limitations, given that they are critically important to the success of these projects.

Discussion

- Dr. Hornbrook inquired as to whether ATSDR could brief the attendees on the Congressional mandate that began this process, and whether the motivation pertained to the burden on society.

- Dr. Kaye responded that the process was not begun by Congressional mandate. The original meeting, convened to examine the potential for creating surveillance for selected autoimmune and neurological diseases, was planned and scheduled prior to the ALS legislation. It was fortunate timing, which helped to keep the process moving forward. The motivation was that ATSDR and others at CDC continued to receive questions regarding clusters of diseases. The problem, particularly when dealing with hazardous wastes and environmental health, is that there are numerous questions pertaining to health effects. Such questions could not be answered because they did not have good prevalence data. Thus, the goal was to obtain solid prevalence and incidence data, which could subsequently be utilized to examine risk factors, especially with respect to environmental exposure issues.

Case Ascertainment from Extant Databases: Issues and Examples

William Joel Culpepper II, MA, PhDc
Associate Director for Epidemiology & Outcomes
MS Center of Excellence, East Baltimore Veterans Affairs Medical Center (VAMC)
Department of Epidemiology & Preventive Medicine
School of Medicine, University of Maryland, Baltimore

Mr. Culpepper defined case ascertainment as the methods and procedures designed to identify persons with target conditions from a large data repository (e.g., hospital, system, population) for purposes of conducting research and / or surveillance. There is no accepted standard, so case ascertainment will vary depending upon the condition of interest and the available data sources. The purpose of the case ascertainment dictates the minimum required accuracy of the case ascertainment method necessary to meet one's objectives. When cases are needed for epidemiologic / surveillance purposes, methods with high specificity (e.g., confidence in eliminating a case as being the condition of interest, but inclusion of some questionables and false positives) are required that will err on the side of being over-inclusive. This is acceptable with respect to population studies. When cases are needed for a specific study, such as a randomized controlled trial (RCT), methods with high sensitivity are required that accurately classify patients and that will err on the side of being over-exclusive to ensure that the cases entered are real cases and that extensive effort does not have to be spent on case adjudication.

Once the method is determined, a decision must be made regarding which data sources will be utilized to identify cases: If clinic charts are used, are they paper or electronic? If a database is used, how detailed is that information? Is it summary information or is it at the patient or encounter level? Once these issues are mapped out, the definition of "caseness" can be operationalized. Depending upon whether the interest is in surveillance or research, consideration must be given to whether cases are probable versus definite, whether there is an interest in one subtype of MS versus another (e.g., relapsing versus progressives or all of them), and whether disease duration and severity are part of the equation. These are study- and project-specific types of questions. Moreover, MS presents some unique dilemmas. MS is a clinical diagnosis requiring an accumulation of evidence. There is no single definitive test /
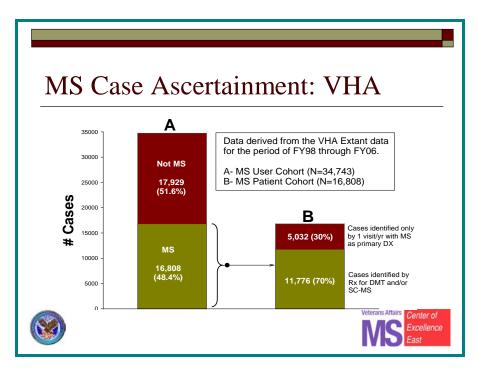
procedure or biomarker for making the diagnosis. The ICD-9 coding system provides only one 3-digit code to capture all MS diagnostic possibilities (340): Possible (rule-out), Probable (diagnostic evaluation), Definite. The ICD-9 coding system does not provide coding for subtype or disease severity.

Mr. Culpepper shared two examples of East Baltimore Veterans Affairs Medical Center's (VAMC) experience with the VHA MS Surveillance Registry (MSSR), which includes extant data collected prospectively, and the National MS Society Quality Indicators project run by RTI International. Data from both examples are encounter-based, organized by level and setting of care (e.g., individual visits and in-patient admissions), provide diagnosis and procedure codes, and provide pharmaceutical codes. The MS Center of Excellence was established in 2001 by a Congressional mandate, and was operational by 2003. The MS Center of Excellence was tasked with improving the quality of and access to MS specialty care throughout the VHA; providing an annual description of the Veteran MS population, healthcare use, and costs; establishing quality indicators and benchmarks for quantifying program enhancements; and conducting basic and clinical research to further the basic understanding of MS and its treatment.

The VA has a nationwide electronic record. There is no paper record any longer. The nationwide electronic record has a number of advantages. For example, a veteran's complete medical record can be pulled up in a VA facility in any state regardless of where he or she lives. Every week, every VA facility conducts a summary process during which every healthcare encounter is downloaded and transferred to the Austin, Texas automation center. That is then distilled down further into a series of SAS data files which are organized by fiscal year and in-patient / out-patient encounters. There are two different sources of health care costing data in the VA. The VA relies predominantly on the Health Economic Resource Center, which standardizes costs across the country and provides national numbers that allows for the comparison of a VA facility on the East Coast to one on the West Coast. This is an excellent resource, although it does not include the ideal clinical detail.

The first step was to review all in-patient and out-patient encounters (FY1998 – FY2006) and to select all patients with MS (ICD-9 340) listed as the primary diagnosis. This group was labeled the MS User Cohort. This cohort consisted of 34,743 patients and yielded an MS prevalence 4-fold greater than any previously reported prevalence, which quickly indicated that there was an issue and that more work was needed. To delve in deeper, the second step in the process was to review the Veterans Benefit Administration records for all veterans awarded a service-connected disability for MS. If symptom onset of MS occurs while on active duty or up to 7 years after discharge from active duty, a soldier can submit a claim for having a service-connected disability for MS. This entails a detailed and thorough medical work-up, review by a panel of experts, and a determination regarding whether the individual has the disease in question, whether it occurred within the active duty window, and what percentage of disability the condition should warrant. Those receiving 100% disability would receive 100% medical care, medications, prosthetics, et cetera. There is also a monetary incentive in that the individual would receive 80% of their salary at the time of discharge for the rest of their life. The fact that this process required a detailed evaluation by a panel of experts gave the MS Center of Excellence another indicator of a highly reliable case of MS.

The next step was to review the out-patient pharmacy records for all prescriptions of disease modifying therapies (DMT) and determine whether patients in the MS User Cohort (step one) had at least one encounter per year with MS listed as the primary diagnosis. At the prescription level, the VA maintains all of the out-patient pharmacy utilization from 1998 through the present. This allowed for the identification of all patients who had ever been prescribed DMT (e.g., Avonex, Copaxone, Rebif, Betaseron, Mitoxantrone, and Tysabri). While all of these efforts assisted in case ascertainment, concerns remained that there may still be primary progressive MS cases which were actually true cases that were not being captured by service connected disability status or DMT use. The VA system requires veteran's to be seen a minimum of once per year to maintain their eligibility. There is a specific evaluation they go through to determine whether a co-pay is indicated. With that in mind, the investigators argued that if this was the case and a person had MS, at a minimum they should have had one visit per year with MS listed as their primary diagnosis during the years they were in the cohort. The operational definition of "caseness" required one or more of the following criteria: Service-connected disability for MS; Prescribed DMT; At least one encounter per year with MS as primary diagnosis.

MS Case Ascertainment: VHA

Data derived from the VHA Extant data for the period of FY98 through FY06.

A- MS User Cohort (N=34,743)
B- MS Patient Cohort (N=16,808)

A — Not MS 17,929 (51.6%); MS 16,808 (48.4%)

B — Cases identified only by 1 visit/yr with MS as primary DX 5,032 (30%); Cases identified by Rx for DMT and/or SC-MS 11,776 (70%)

The results of this algorithm were compared with chart review determinations of MS diagnosis in VISN 5 (N=682), which includes Washington, DC; Baltimore, Maryland; and Martinsburg, West Virginia. Comparing the 682 VISN 5 cases to chart reviews, 296 were found not to be MS and 386 were found to be MS. The primary interest was to eliminate those cases that clearly were not MS; however, because a proportion of patients with MS present with a clinically isolated syndrome, they did not want to eliminate them inadvertently. Sensitivity was 0.93, specificity was 0.90, and positive predictive value was 0.88. The overall accuracy of this method was 92% (e.g., roughly 92 out of 100 were accurately classified as MS versus non MS with this scenario). While somewhat over-inclusive, the investigators felt comfortable with this methodology because through time, further validations, and additional studies, they would touch a number of these individuals and would conduct further evaluations and screenings and rule them in or out. Roughly 35,000 individuals were identified only by having at least one MS encounter. However, once the algorithm was applied, 17,929 (51.6%) were dropped as not being MS. This saves an

enormous amount of time, money, and resources once specific studies are begun. Of the 16,808 (48.4%) cases remaining (termed as the MS patient cohort), 5,032 (30%) were identified only by one visit per year with MS as the primary diagnosis and 11,776 (70,000) were identified by a prescription for DMT and / or Service-connected disability for MS (SC / MS).

With respect to the RTI MS case ascertainment example, Mr. Culpepper indicated that RTI was contracted by the National Multiple Sclerosis Society (NMSS) to convene an expert panel to establish quality indicators (QI) for MS care, during which 17 QIs were developed; and to evaluate how feasible these QIs could be measured in extant data. The Medstat Claims database was used, which covers 4.5 million enrollees across 10 states. Data sources included in-patient claims, out-patient and physician office claims, pharmaceutical claims, and demographics. Data consisted of ICD-9 diagnosis and procedure codes and detailed pharmaceutical codes. The major difference from the VA data is that Medstat does not have service-connected disability. Thus, five operational definitions were evaluated: 1) D1 required one claim with any MS diagnosis IP or OP *or* DMT; 2) D2 required D1 *and* 18+ years old (because for those under 18, 340 is the strep throat code); 3) D3 required 18+ years old *and* 2 DX codes for MS separated by at least 1 month *or* DMT; 4) D4 required 18+ years old *and* 2 DX codes for MS separated by at least 1 month, *or* 1 DX code for MS and DMT; and 5) D5 required 18+ years old *and* 2 DX codes for MS separated by at least 1 month. RTI chose the D3 population from 2002 as their "baseline" denominator ("MS cases"), given that it "represents a middle ground in terms of restrictiveness for MS case finding and has the benefits of the larger 2002 sample size." Unfortunately, no validation (e.g., chart review) was performed. Thus, reliability and accuracy of this case ascertainment method is unknown:

MS Case Ascertainment: RTI

Denominator	2001	2002
D1	6,860	9,888
D2	6,745	9,786
D3	4,810	7,221
D4	4,317	6,265
D5	4,055	5,736

To recap, the intent of the VHA case ascertainment was to identify MS cases for studies. A focused algorithm was utilized to identify non-MS cases to avoid being too over-exclusive, and this required validation efforts to assess the accuracy and reliability of the algorithm. The intent of the RTI case ascertainment was to identify cases for screening / surveillance for QIs. The algorithm was likely over-inclusive, but no validation was performed to confirm this. With respect to lessons learned, it is important to define the purpose for case ascertainment, know

the nuances of the data, and incorporate strategies to validate case ascertainment algorithm in at least a sub-sample.

In terms of future work, laboratory (e.g., IgG) and radiographic findings (e.g., MRI) that the VA is trying to bring on line will be included. That should help to add some precision to case identification. The bottom line with case ascertainment from extant databases is that "you can pay now" as in the VHA strategy by designing ascertainment efforts that are front-loaded and performing the algorithm and validation early in the process. With this method, the emphasis is on accuracy in order to avoid the time and expense of direct patient screening for "caseness." Or, "you can pay later as in the RTI strategy by designing ascertainment efforts that are end-loaded, with validation not performed or performed later via more laborious and time consuming methods (e.g., in clinic evaluation). With this method, the emphasis is on identifying cases at the population level for screening / surveillance with validation to be done later.

Discussion

- Dr. Hornbrook inquired as to what would occur if someone in the VA system got cancer. For example, if a veteran's primary diagnosis was MS, but then he developed prostate cancer, would MS be removed as the primary diagnosis?

- Mr. Culpepper responded that if the person died and came out of the cohort, for the time period they were included, they would have to have had at least one encounter with MS as the primary diagnosis. A veteran with prostate cancer would see their oncologist, at which time prostate cancer would appear as the primary diagnosis. However, when they visited their neurologist, MS would be the primary diagnosis and would be captured that way.

- With regard to MS case ascertainment, it was noted that perhaps there were a lot of misdiagnoses.

- Mr. Culpepper replied that there appears to be a lot of misdiagnosis occurring. There is a longitudinal study underway, which is sponsored by the MS Center East, which covers 11 sites throughout the Eastern Seaboard. Part of that process included sending patient lists to each center for validation of their respective cases. Their numbers were almost identical, plus or minus a few percentage points. Typically, upon chart review, those individuals who have been seen only occasionally are found to have been subsequently diagnosed with Parkinson's or some other disease. Given this, it is possible that the algorithm may need to be revised somewhat.

- Dr. Kasarskis requested further information about how the VHA cases dropped out versus the three criteria specified, pointing out that the business of electronic coding of one encounter per year by somebody as a diagnosis of MS is just one mouse click away from an error.

- Mr. Culpepper responded that those who were identified as being cases were identified by the service-connected disability and DMT. As mentioned, 5,032 (30%) were identified only by one visit per year with MS as the primary diagnosis and 11,776 (70%) were identified by a prescription for DMT and / or Service connected-disability for MS (SC / MS). The 17,929 (51.6%) who were excluded from the overall cohort met none of the criteria (e.g., they were not seen at least one time per year with MS as the primary diagnosis, they had no service-connected disability for MS, and they were not prescribed a DMT). The algorithm was

applied originally to the entire cohort of 35,000. The 682 cases from one area were used to validate the algorithm, which was then applied to the entire 35,000 once validated.

- An inquiry was posed regarding whether investigators learned anything from the algorithm and then subsequently re-taught it.

- Mr. Culpepper replied that this step had not been done yet. They are currently reviewing all of the validation data to determine what other pieces of information can be included in the algorithm. The VA data systems are in the process of including increasingly more laboratory findings into their laboratory data files. Currently, this does not include cerebral spinal fluid analysis. Plans are in the work to do radiographic data in order to have summary reports components. It is unknown whether they will ever have imaging available with that type of system, but those are a couple of components that are in the mill that certainly in time investigators will be able to take advantage of. Mr. Culpepper recently completed a national survey to a sub-sample of the MS patient cohort of 16,808. Approximately 1,700 responses have been returned. Part of that survey was to acquire some of the clinical detail that could not be obtained from the administrative data. In that survey, questions were posed about sub-type, date of diagnosis, symptom onset, chronological DMT history (e.g., If you stopped why you stopped?; If you switched, what did you switch to?; and quality of life and functional pieces). The investigators are in the process of analyzing those data and comparing them to what has already been done in order to triangulate this. This may also offer clues regarding other data elements to consider in order to make this process more precise.

- It was suggested that the investigators use this to teach the informatics system how to keep up with disease understanding. Otherwise, they will be stuck in medieval times.

- Dr. Bejaoui indicated that the Marshfield Clinic has an MS Center with an MS specialist who sees over 1,000 patients per year, many of whom are presenting for a second opinion. Of those, approximately 50% do not have MS.

- Mr. Culpepper responded that he was not surprised given the VA experience. The VA may be somewhat different in that they do not have MS specialization in every medical center across the VA. Some of the determinations are being made by general practitioners, urologists, neurologists, and physiatrists who do not have MS specialty training, so there are errors in fully appreciating and recognizing symptoms for what they might or might not be. Another issue the MS Center is attempting to address is how to disseminate information and consultation to the general clinics so that when a general practitioner sees someone who has MS or is suspected of having MS, there is a way to obtain more detailed information and improve the diagnostic process.

- Dr. Hoffman wondered whether there was a disincentive, as there is outside the VA, by including MS as a rule out diagnosis. If that is done in order to attempt to get a private payer to pay for an MRI and other testing and nothing other than a symptom code is used, frequently they will pay for it.

- Mr. Culpepper replied that this was a non-issue, given that the VA is using that as a rule out. In many instances, because of the way the VA is structured, they can go right to testing without having to acquire prior authorization, whether it will be covered by the insurance agency, only covered at 20%, et cetera.
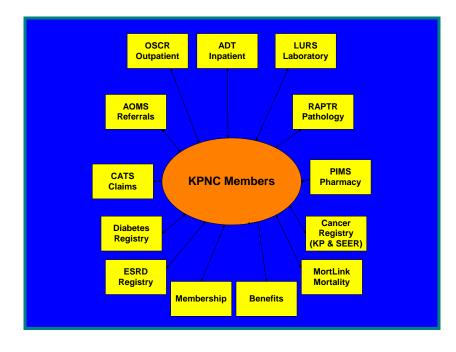
- Dr. Hornbrook thought that coding incentives would be a theme throughout the day. There are systems such as the VA and HMOs that do not require pre-authorization for testing, while there are other systems that do require pre-authorization resulting in very different diagnostic coding by physicians.

- Mr. Culpepper pointed out that when ICD-11 comes out, there will be more precision in the diagnostic codes, with a decimal code addition to identify whether something is a rule out, suspected, possible, or definite.

- Dr. Nelson noted that the VA cohort would be missing those patients who do not have at least one visit per year and who do not have any DMTs. While this may not be a large subgroup, it is increasingly being recognized as under-ascertained in studies in general.

- Mr. Culpepper responded that this is what they have done to date, realizing that there was a possibility that those individuals would be under-ascertained. There are mild, atypical cases that represent a unique group to consider beyond traditional MS patients. The hope is that information received from the survey will help to identify those types of individuals, will help to determine whether the methods are too restrictive, and / or will guide them toward another avenue that will help to identify these under-ascertained patients.

- Dr. Nelson pointed out that they are not, however, sending the survey to the entire cohort of 35,743 possible MS patients.

- Mr. Culpepper responded that the VA is in the process of rolling out a web-based portal known as "My Health Vet," which allows individuals to review their own medical records, change and update information, et cetera. We plan to have a link to this site that will take patients to the MS Center of Excellence site and the survey, which can be voluntarily completed and they can consent to be included in the surveillance registry (e.g., the mail-based respondent surveys to date). That may also capture some of the under-ascertained individuals. He has also considered that some individuals may be seen in 1998, for example, do not return to the VA for the next two years, but then appear again after that. The algorithm would ignore those patients.

- Dr. Tyry indicated that in the North American Research Committee on MS (NARCOMS) database, many individuals have both VA and private insurance. It seemed to her that some of the individuals in the full cohort (34,743) may be diagnosed in the VA system, may have cancer, and may stay in the system because they see their oncologist. However, they may go through the private insurance to see their MS practitioner.

- Mr. Culpepper agreed, noting that they recently requested CMS data for the MS cohort to determine whether switching is taking place. It is known from other chronic conditions that distance to a facility has a major impact; that is, the further away someone is from a VA facility, the more likely they are to use Medicare services, and vice versa. The plan is to mine the CMS data for additional information.

- Dr. Hornbrook inquired as to whether the VA examined the data with regard to whether the veterans were dual eligible for Medicaid. He also wondered whether Mr. Culpepper had begun working with National Language Processing (NLP) to process the notes.

- Mr. Culpepper responded that a data element that is completed at each encounter that specifies who the payer is (e.g., VA or Medicaid). While the investigators began reviewing

this information, similar to the private sector, diagnosis codes are sometimes entered because it results in better payment. This scenario was occurring frequently in the VA. Some veterans who had high co-pay and were eligible for Medicare, would get entered so that Medicare would pick up the bill. If the VA paid the bill, the veteran would have a large co-pay for that visit. Therefore, the investigators were not convinced that this would be a reliable indicator. Once they have the CMS data, they will know with greater certainty who is dual eligible. They do have a funded grant to conduct such an analysis. With respect to processing the notes in the charts, although they have a group that is interested in doing that, it has not been done yet. They have engaged in discussion about this with the Colon Cancer Registry representatives and the VA; however, the Colon Cancer Registry has a VA mandate and an entire programming cadre to assist them with this.

Lessons in Case Ascertainment and Confirmation from the Kaiser Permanente Case-Control Study of ALS

Stephen K. Van Den Eeden, PhD
Senior Investigator, Division of Research
Kaiser Permanente—Northern California

Dr. Van Den Eeden indicated that the goal of the Kaiser Permanente project was an etiological case-control study of ALS. The investigators wanted to study genes among other things, so it was important to be as accurate about diagnosis as possible and to use newly diagnosed / incident cases rather than trying to find all prevalent cases and then sort them out. Kaiser Permanente has a host of data systems which they can search about a single member within Kaiser:

The original goal was to cast a wider net, and they had reasonable good resources from recent National Institutes of Health (NIH) funding to sort through the wheat to find the kernels for which

they were looking. The data sources were all computerized and included complete information for in-patient, out-patient, provider, department, laboratory tests and results, and pharmacy (prescription drugs dispensed). While the investigators often knew that a procedure was done, they did not always have electronic access to the results. Clinic and in-patient records were available as needed in hardcopy.

The Kaiser Permanente—Northern California Region (KPNC) is a group practice HMO with approximately 3.1 million members in the San Francisco Bay and Sacramento areas. About 25-30% of the geographic population is covered in KPNC. The practice setting includes 19 hospital medical centers and 20 medical office buildings. There are approximately 65 neurologists and physicians who work exclusively within Kaiser Permanente (KP). Virtually all care is provided within KP and 100% of ALS cases are referred to neurologists. This differs from the VA in that once someone is a member of Kaiser, they do receive all of their care there.

The investigators used a standardized case definition, which included modified criteria proposed by the World Federation of Neurology Research Committee Group on Neuromuscular Disease [World Federation of Neurology Research Committee Research Group on Neuromuscular Diseases. *J Neurol Sci* 1988;86:333-60]. This includes evidence of *Lower Motor Neuron (LMN) degeneration* by clinical, electrophysiological, or neuropathologic examination; *Upper Motor Neuron (UMN) degeneration* by clinical examination, and *progressive spread of symptoms or signs* within a region or to other regions, as determined by history or examination together with *absence of: Electrophysiological and pathological evidence of other disease processes* that might explain the signs of LMN and/or UMN degeneration, and *neuroimaging evidence of other disease processes* that might explain the observed clinical and electrophysiological signs.

Searching the in-patient, out-patient, and pharmacy records, the investigators looked for anyone who was coded with an ALS diagnosis or motor neuron disease at any point in the electronic databases. The parameter was to search for cases from January 1, 1996 through March 31, 2000. Anyone who had an ALS diagnosis or motor neuron disease prior to 1996 was immediately counted as a prevalent case. From the January 1, 1996 through March 31, 2000 ascertainment period, 457 ALS diagnoses of motor neuron cases were found and sent to chart review. The chart review was an ascertainment process during which any notes were copied related to neurology, function, and / or physical therapy that might be related. The chart reviewers were trained to recognize the full range of symptoms of ALS. The results were given to a movement disorder specialist at the Parkinson's Institute, who applied the case definition using all of the available information. Those for which the information was insufficient were put on a wait list until the next time they presented, at which time the investigators acquired the new information. Of the 457 cases, 217 were deemed eligible for enrollment in the study. The data systems at the time, especially in the out-patient setting, were not particularly robust to rule out or work up diagnoses. The electronic medical record has solved this issue. The remaining suspect cases were re-reviewed for the full two years following their diagnosis in order to be confident that the ascertainment process at the beginning of their cases was accurate. Of the 217, 142 were probable cases, 51 were possible, 13 were suspect, 5 had other neurologic diseases, 2 had insufficient information, and 1 had onset prior to enrolling in Kaiser. No cases were called "definite" because it is largely a clinical diagnosis, there is no biomarker, and those who died did not go to autopsy. Some of the possible and suspect cases lacked information as well, particularly given that some of the subjects died after they were found to have been diagnosed, so there was no ability to accumulate a clinical record to meet all of the criteria. This is one reason some individuals wound up in the "possible" or "suspect" categories. They also linked to all of the California death certificate records to determine causes of death. Of the 142 probable cases, 9 were alive or had no evidence of a death certificate, 110 had a death

certificate that said ALS, 18 had no mention of ALS on their death certificate, and they knew 5 were dead but could not find a death certificate linkage for them. Of the 51 possible cases, 7 were alive, 33 had a death certificate that said ALS, 9 had no mention of ALS on their death certificate, and they knew 5 were dead but could not find a death certificate linkage for them. All of the remaining individuals were alive.

The investigators are attempting post hoc to develop an algorithm that will allow them to acquire further information about how many diagnoses subjects had, whether they received DMTs, et cetera. Also important to consider in ALS and MS are alternative diagnoses. The pattern in a clinical record may be to see ALS several times, but then to see an alternative diagnosis. Sometimes this has nothing to do with the disease process, but instead pertains to how something is recorded in the system. With the electronic medical record at Kaiser, this will be less of an issue, but historically it was a problem.

Dr. Van Den Eeden concluded that even in a comprehensive and integrated data and service environment, this process is not easy or inexpensive. The best determination includes a provision of follow-up, ideally for two or more years. Critical decisions are needed for early deaths among potential cases, equivocal cases, and death certificate cases.

Discussion

- Dr. Hornbrook asked what Dr. Van Den Eeden thought about the notion that ALS was the last disease anyone wanted to attach to a patient, so that at least in the Kaiser systems, there was a sense of not putting this down until it was the only thing left to do.

- Dr. Van Den Eeden responded that it is definitely a clinical process. When this study was begun, investigators were identifying individuals who met the case definition. However, their IRB required them to obtain physician permission prior to contact. Physicians were essentially writing back, "not yet" even though people were clearly meeting the case criteria. They also found that some of the Kaiser neurologists would send their patients for case confirmation to a clinical expert in San Francisco, for which Kaiser was paying, so it was in the system.

- Dr. Kasarskis asked how many of these individuals actually had a second opinion and what the use of Rilutek® was, the only drug approved for ALS in the Kaiser population. For example, if they applied a Rilutek® filter to the potential cases, he wondered how the data looked.

- Dr. Van Den Eeden responded that Rilutek® is routinely used. The investigators applied a Rilutek® filter, aggregated across the sites, as part of the case ascertainment process. Anyone who had a single prescription for Rilutek® was evaluated for whether they had a diagnosis. In terms of the second opinion, there are no barriers. However, the investigators would not know about a second opinion for those who self-paid for one.

- Mr. Culpepper noted that once the drill downs were done for the cohorts, they were still hitting about a 45%-48% that were actually cases by definition. In the end, this saves a lot of work and resources in screening all of the potential cohort members.

- Dr. Van Den Eeden agreed, pointing out that the electronic medical record should improve this as well.

Overview of the MS Surveillance Projects

New York State

Barbara Teter, PhD, MPH, CHES
Director of Research & Development
Coordinating Office at The Jacobs Neurological Institute
New York State Multiple Sclerosis Consortium

Dr. Teter explained that the New York State Multiple Sclerosis Consortium (NYSMSC) was established in 1994. It is a durable database of demographic and clinical data to promote MS research and enhance patient care. Membership includes 16 sites in New York State, 14 of which are currently active. The project includes 27 investigators who are MDs or PhDs and 30 research or data coordinators who are NPs, RNs, RAs, MSWs, and PhDs. Governance is by, an Executive–Financial Committee, a Scientific Review Committee, an Executive Director, and a Director of Research and Development and is governed by procedures and policy. Revisions require membership quorum. Currently in the database are approximately 8,500 registered MS patients with about 15,000 follow-up records. Both demographic and clinical data are collected, so half of a registration or half of a follow-up questionnaire is for the patient to complete and the other half is for physicians to complete. At the current time NYSMSC and NARCOMS are comparing their two databases on demographic and clinical descriptive data.

For the NYSMSC MS project, the deliverable dataset last name, first name, middle initial, maiden name, Social Security Number, day of birth, address, and zip code will be used for matching and linking. The remaining variables will be turned in to ATSDR at the end of the project. NYSMSC will construct a database specific to this pilot project in which every person from 2001-2005 will have a unique ID number (e.g., every person from VBA, CMS, NMSS, and NARCOMS), so this will be an extremely large dataset. Also included will be the first time an individual is identified in the database, race, sex, age, year and month of birth, and a field that indicates which of these databases a person appeared in. Upon completion, the investigators will compare those in the NYSMSC dataset who were in one of the other databases to those who are not. They will also compare CMS to NARCOMS, CMS to VBA, et cetera. That is, everyone will be compared to everyone else. NYSMSC currently has datasets for CMS (Medicare), VBA, VHA, and NMSS. They do not yet have the NARCOMS dataset, given that they continue to work on some IRB issues. At the end of each record a field will be provided to indicate the certainty of each match. NYSMSC will also be reporting to ATSDR frequencies on race, sex, age groups and mean and standard deviation, number of times registered in each dataset, by year in question. This will probably be reported per dataset and on the whole dataset.

NYSMSC greatest issue is the IRB / HIPAA waiver situation. This was the topic of the meeting last year. Although access to the data is allowed without written authorization for limited data sets stripped of HIPAA identifiers for public health practice versus research (research being defined as the systematic collection and analysis of identified health data for public authority to generate knowledge to benefit those beyond the participating community), academic IRBs less often than not do not distinguish practice from research, and because NYSMSC data were collected under IRB oversight, all of the sites will be required to ask their individual IRB to review and grant waivers if required. The use of the data represents minimal risk to database registrants and will not adversely affect the rights and welfare of registrants. Relevant data generated from this project will be disseminated through publications and presentations.

Minimal amounts of data will be used to provide information about prevalence and basic demographics, which is a characteristic of surveillance. Thus, they should not have any problems acquiring waivers.

With respect to linking the databases together, some of the problems encountered, resolved, or anticipated include unflattened databases; standard state codes had to be looked up; the name and address fields had to be built for the VBA dataset because they are all in one field; there were missing fields in some of the datasets; and Soundex indexing versus visual is an issue. NYSMSC will probably ask a programmer to look at data records who are similar by last name and birth date, and then they will be pulled out and examined one-by-one in order to make a judgment. For protection of private health information, IRB HIPAA waivers will be obtained according to individual consortium site IRB reveiw. One problem encountered is going through the Kaleida Health System of Western New York, which is where the coordinating office for the NYSMSC is located, if Dr. Teter writes an email to their sites that includes the words "data," "data management," or "spreadsheet" even in the subject line (while the email contains absolutely zero PHI), it is encrypted, intercepted, and returned to her—sometimes two days later. Hence, they have had to find interesting ways in which to talk to one another. Keeping the passwords that were downloaded to CDs, she hand carried the actual datasets to the analysts at the data management site.

Regarding the linking test, the matching algorithm is scored as follows:

- last name = 1
- year of birth = 1
- year of birth +/- 1 year =1/2
- sex = 1
- city = 1
- state = 1
- month of birth = 1
- day of birth = ½
- first name = ½
- 7 is a perfect match

A score of 4 might be a good match, but there are several possibilities. Last name, year of birth, month of birth, and city is a score of 4, but it is not clear whether that is really as good as last name, year of birth, sex, and city—also a score of 4. Or, is last name, sex, city, and state better? It is not unusual for brothers or sisters to both have MS, so last name, sex, city, and state is probably not of the same quality. With year of birth, it is not clear whether +/- 1 year is truly worth ½. Using year of birth +/- 1 year can actually reach a qualitative value as high as 5. However, without first name this can reach a value of 4.5. This could be a good match. The match of last name = 1 alone is not working for this project because when they output the data of the matches on 1, it is merely a list of people who have the same last name. Thus, our analytic team will examine this further.

A preliminary test was conducted on a small subset of NYSMSC data with three datasets (VA, CMS, and NMSS) (n=1850) and subset with SS # (pre-HIPAA):

MS Linking Test Progress

Test with three datasets with subset NYSMSC names (n=1850) and subset with SS # (pre-HIPAA) :

Match Value	% VA	% CMS	% NMSS
1	15.4		23.3
1.5			7.0
2	< 1.0		2.8
3	< 1.0		14.6
4			
5			20.6
6	<1.0		1.0
6.5			< 1.0
7	1.1		2.1
Total test site (n=1850)	17.4		72.2
Total db	4.0		16.3

NYSMSC -BE Teter 02/26/08

CMS was removed from this table because it showed 99.7% of having a match, regardless of the number. A large majority of these were scores of 1; however, they did not have time prior to this meeting to take out the 1s in order to give revised percentages. Another issue is that NYSMSC is comparing the data to 2001-2005 data in the NYSMSC dataset, but because the data from the other datasets given to them include those dates, NYSMSC is going to include all of the dates in their dataset in the case where someone leaves the neurology clinic and uses their VA benefits to go to a VA center in New York State.

Dr. Teter explained that leverage learning meant that ultimately, in addition to public health practice or the national surveillance system, the knowledge gained from this surveillance pilot also supports a new scientific tool (e.g., a great contribution to information technology). Dr. Barry Smith, SUNY-UB Informatics, wants to scientifically minimize overlap and maximize efficiency in the registries, for example: 1) Through an MS ontology-based integration (MSO) to logically define terms to enable retrieval and analysis of data, which is measurably more useful than queries without the MSO; and 2) To serve as a controlled vocabulary for expression in MS research and public health practice. That is, any database anywhere would be able to look at this MSO and use it for analysis. With that in mind, the investigators examined the NARCOMS and NYSMSC consortiums databases and found them to be surprisingly similar with respect to sex, marital status, race, educational attainment, most recent insurance type, family history, smoking, clinical characteristics, PDDS, EDSS, and most recent disease type. Race is almost entirely Caucasian in the NYSMSC registry population. , African Americans in the NYSMSC are 5.9% and some analyses are underway for this population. NYSMSC plans to change who collects insurance information, which is currently reported by subjects, because it appears that people may be confused about the combination of private and Medicaid insurance. Some people have Medicaid insurance that is being administered by an HMO, so they believe that they have combination insurance. Registrants in NARCOMS seem to be somewhat older at age of on-set compared with NYSMSC registrants.

With regard to the NYSMSC surveillance pilot time-line, Dr. Teter reported that the project is on target and the investigators hope to complete the analyses, submit reports to ATSDR by September 2008, and participate in decisions about dissemination of results and whether these will be presented as a platform for going forward with the next steps.

CMSC / NARCOMS Patient Registry

Tuula Tyry, PhD, MAED
Program Manager
North American Research Committee on MS Registry

Dr. Tyry indicated that NARCOMS is a long-term research project of a consortium of multiple sclerosis centers, while CMCS is a non-profit professional organization for clinicians and researchers who are conducting either clinical or research work in multiple sclerosis. The NARCOMS project has been operating since 1993, and the administrative office is currently located in Phoenix. NARCOMS collects self-report data from patients throughout the US and from 50 countries through mail or on-line surveys, which people are asked to update every six months. For those who follow through, there is a fairly good longitudinal record of disease progression, treatment patterns, et cetera. However, since it is a voluntary registry, not everyone updates every year, so there are gaps in the data.

NACROMS overlaps with other data sets that they have been given (e.g., NMSS, VA, and Medicare data) to be matched with NARCOMS participants in Georgia, Minnesota, and South Carolina). They were also asked to compare NARCOMS data with the NYSMSC. NARCOMS is using the following variables for the preliminary matching process: First name, last name, middle name / initial, maiden name, date of birth, state of residence, city, sex, and race / ethnicity. Dr. Tyry stressed that the numbers she was sharing were preliminary, given that NARCOMS is a "living" database. Numbers change as people report and there are new enrollments every day. Important to remember about NARCOMS' datasets is that they are both retrospective and prospective, so when people enroll they provide information for the past 5 to 50 years and also provide updates.

Overall, the NARCOMS database includes about 32,000 individuals. Approximately 1,700 of those are in Georgia, Minnesota, and South Carolina. In these three states, 687 participants report currently being on Medicare and 271 who report being veterans. When they first examined the NMSS DATA, they received a list of 23,536 (Georgia, Minnesota, and South Carolina) of whom 1,206 were also in NARCOMS. That means that the NMSS list captured about 70% of NARCOMS participants in these three states and missed 505. The Medicare dataset included over 16, 625 from Georgia, Minnesota, and South Carolina. Of those, 641 were also in NARCOMS. This means that Medicare data captured 93% of NARCOMS participants and missed only 49 participants in these three states. The VA dataset included 1,599 from the three states. Of those, 179 were also in NARCOMS meaning that that list captured 66% of NARCOMS participants and missed only 92 veterans in these three states. When the three lists are combined and compared to NARCOMS, there were 1,316 who were also in NARCOMS meaning that together those three databases captured 77% of NARCOMS participants in these three states and missed 457. For each of the states, the investigators then examined whether there were any major differences between those people who matched with the list and those who were only in NARCOMS. Of interest for Georgia is that the missed group, the ones who were only in NARCOMS, are slightly younger, have had the disease for slightly less time, and their disability level is less. This was similar for Minnesota and South Carolina as well.

Also in progress is the comparison of NARCOMS data to the NYSMSC data. Approximately 1,970 NARCOMS participants are from the state of New York. The cohort includes the following: Female (72%), Caucasian (92%), African American (4%), College Educated (60%), Currently Married (48%), Private Insurance (75%), and Medicare / Medicaid only (19%). This

supports the notion that NARCOMS data may biased toward those who have higher educations or have some college education, so they may be missing people at lower levels of education. The age is similar to the overall age structure in the database. They do not have a lot of newly diagnosed individuals, which is an issue that may have to be addressed in order to understand true prevalence.
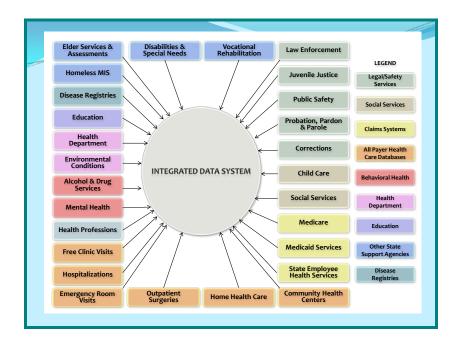
Discussion

- Dr. LaRocca asked for clarification regarding whether the consortium was using the Medicaid dataset as well.

- Dr. Tyry responded that currently they only have Medicare data.

- Dr. Kaye added that Medicaid data is a work in progress. The data are enormous because although CMS has a filter on Medicare data so that data can be ordered just by diagnostic disease code, they do not have the same option for Medicaid. Therefore, Medicaid data must be ordered by state. The statistician on the project, Jenny Wu, has been reviewing all of the data files (e.g., in-patient, out-patient, long-term care, and pharmacy) and is not finding that many people, at least with the ALS. In addition, an executive decision was made not to review anybody under 20 in order to be able to manage the datasets, which more than halved the amount of data.

- An inquiry was posed regarding whether the plan was to operationalize only those individuals who are matched across all of the datasets.

- Dr. Kaye replied that that would be a decision for another day, given that other characteristics must be determined first (e.g., DMT, how many visits they have had, whether they have been to a neurologist, et cetera). All of that is waiting until the MS data can be combined from both sites and the ALS data can be combined from the four locations.

- Referring to the overlap in the VA dataset versus the NARCOMS dataset, Dr. Kasarskis wondered whether the investigators were asking whether participants received their health care within the VA system or if they were veterans.

- Dr. Tyry responded that the 66% referred to the match with people who indicated that they were veterans. About half of those stated that they also received their care from somewhere else. These are likely the people who may be in the VA system because they have some other condition for which they make a required annual visit in order to keep their benefits, but they actually see a private neurologist. All of these data can be analyzed in more detail. Part of the reason she chose to present the summaries during this meeting was because it depended upon how they did the match. She is comparing notes with Dr. Teter to ensure that they are not duplicating efforts. It is possible that different algorithms are needed for the different datasets because some of them are stronger in some areas and some in others.

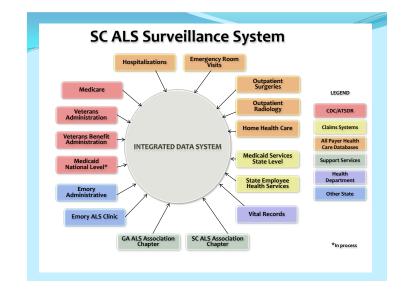## Overview of the ALS Surveillance and Registry Projects

South Carolina

Julie Royer, MSPH
Principle Investigator
South Carolina Budget and Control Board
Office of Research Statistics

Ms. Royer explained that the Office of Research Statistics (ORS) is under the South Carolina Budget and Control Board and is a service agency in South Carolina. She is in the Health and Demographics Section of the ORS, which collects, processes, analyzes, and interprets health, demographic, and Census data in South Carolina. They build off of existing systems to do this. ORS has existing partnerships with over 40 state agencies and other private entities. The data arrives at ORS with identifiers, but goes through a unique identification process to link data systems. A unique per individual identifier is assigned to each observation in a data file, and then identifiers are removed from the data leaving statistical data files. They have an integrated data system that allows them to track people across all of the various data sources , while protecting confidentiality at the same time. In addition to confidentiality, one of the hallmarks of this system is the data are always "owned" by the originating agency. For anyone who wants to use the data, ORS must seek permission from the originating agency to use and link any of their data.

ORS collects uniform billing discharge data from all South Carolina in-patient hospitalizations, all emergency department visits, out-patient surgical centers, birth and death certificate data from the health department, claims data from Medicare / Medicaid, and State Employee Health Services plan (Medicaid and State Employee Health plan includes prescription data). The following data systems are used for the ALS surveillance system:

With most of these systems, ICD-9 code 335.2 to 335.29 as either primary or secondary diagnosis was used to identify cases for years 2001-2005. Cases were already identified for ORS in the data files from ATSDR, Emory, and the South Carolina and Georgia ALS Association chapters. Ms. Royer asked for all of the encounters from the datasets that they received from ATSDR and Emory. In 2005, there were approximately 557,000 in-patient discharges from South Carolina hospitals, 1.8 million emergency department visits, approximately 413,000 Medicaid service recipients over the age of 18, and about 350,000 individuals on the state health plan insurance system. The number of ALS cases identified from ORS data sources included 479 from hospitalization, 170 emergency department, 207 other outpatient, 296 home health care, 239 Medicaid, 271 state health plan, and 354 from vital records. From the Emory and ALS Association datasets, the number of cases identified included 28 from Emory Administrative, 12 Emory Clinic, 14 Georgia ALS Chapter, and 26 from the South Carolina ALS Chapter. From CDC / ATSDR data, the number of cases identified included 1,426 from Medicare, 132 VHA, and 44 from VBA.

The South Carolina Cohort from 2001-2005 had total cases of 2,060 unique individuals. Of these, 53 cases were removed who were younger than 18 years old the first time Ms. Royer encountered their motor neuron disease ICD-9 code. Most of these were found in Medicaid and the diagnosis code was progressive muscular atrophy or other motor neuron disease. Another 27 cases were removed because they were out-of-state and for purposes of this presentation, she removed 36 cases that were missing date of birth (VBA) and 4 that were missing a diagnosis (VHA). The remaining South Carolina Cohort includes 1,940 unique individuals. Almost immediately when the investigators began examining the diagnosis codes for ALS they realized they had a problem. They thought that some of the non-specific ALS diagnosis codes (e.g., the progressive muscular atrophies, unspecified MND) would be coded on a patient's record while they were being diagnosed and then it would eventually end up as code 335.20 once that case was confirmed. However, non-ALS MND disease codes were being used to

21

describe symptoms of other diseases. Therefore, they developed a coding scheme to best-identify true ALS cases. The first seven are straightforward, although ALS is the last code found when sorted by date of code:

Description	ICD9/ICD10 Codes	Total
Unspecified MND	3352	16
ALS	33520	563
PMA	33521	147
PBP	33522	90
PP	33523	153
PLS	33524	77
Other MND	33529	479
ALS Late Diagnosis	multiple codes, last code=33520	207
ALS Unsure	multiple codes, last code≠33520	90
MND Unsure, Not ALS	multiple codes, no code≠33520	112
ALS change to PLS	33520, 33524	6

The South Carolina cohort, with all MND diagnosis codes are 51% male (980), 49% female (960), 76% Caucasian (1,478), 19% African American (357), and 5% of other / unknown race. With respect to age, 2% (44) are 19-34, 5% (95) are 31-40, 13% (247) are 41-50, 17% (331) are 51-60, 24% (479) are 61-70, 26% (499) are 71-80, and 13% (245) are 80+. Studies have shown that in younger age groups, more men than women are affected, but as age increases numbers even out somewhat. The next slide shows the South Carolina ALS Cohort age-by-sex distribution for ALS or late ALS coded cases.

Age	Male	Female	Total
19-30	7	6	13
31-40	24	12	36
41-50	56	39	95
51-60	72	47	119
61-70	122	89	211
71-80	119	94	213
80+	40	43	83
Total	440	330	770

Using all MND diagnosis codes 1,129 cases were identified in one dataset, 370 in two datasets, 210 in three datasets, 123 in four datasets, 82 in five datasets, 23 in six datasets, and 3 in seven datasets. When selecting only cases coded as ALS or late ALS only 265 were identified in one dataset (30 UB, 17 Medicaid, 13 SHP, 6 HH, 161 Medicare, 24 VA, 8 GA Chapter, 6 SC Chapter), 159 in two datasets, 144 in three datasets, 108 in four datasets, 70 in five datasets, 22 in six datasets, and 2 in seven datasets. With respect to prevalence by year, including only those cases coded as ALS or late ALS in the analysis seems to more closely reflect the prevalence estimate that would be expected. Unconfirmed prevalence per 100,000, from the South Carolina ALS Cohort data sources, was 8.4 in 2001; 7.8 in 2002; 8.1 in 2003; 83 in 2004; and 8.3 in 2005.

All of the cases, regardless of source, were linked to the South Carolina in-patient and emergency department discharge billing data files. If a South Carolina in-patient or emergency department record existed for a case and it was either following the initial motor neuron disease diagnosis or within a six month time period prior to that initial diagnosis date, it was requested for abstraction. Cases could have more than one record, so an algorithm was used to determine the best record for review. Sometimes there was a tie, so Ms. Royer requested a few more records than she had individual cases. Almost half of the records requested were the primary source that identified the case for inclusion in the cohort. In total 1,186 records were requested on 1,172 individuals (888 Inpatient records, 298 ED records, 547 primary source of MND code). The following slide reflects the cases identified and the number of records requested:

Description	Total Cases	# Records Requested
Unspecified MND	16	5
ALS	563	400
PMA	147	112
PBP	90	77
PP	153	120
PLS	77	42
Other MND	479	136
ALS Late Diagnosis	207	157
ALS Unsure	90	57
MND Unsure, Not ALS	112	77
ALS change to PLS	6	3

Early on, they realized that codes 335.29 and 335.2 were not being used for suspected ALS cases before a definitive diagnosis was made, as one would expect. Thus, not as many records as available were requested in those two categories. Instead a random sample was selected. If a record existed for the other MND codes (335.20, 335.22, 335.24, ALS late diagnosis or ALS unsure), it was requested for abstraction.

Regarding preliminary results, not all of the records abstracted are in electronic form and not all have been reviewed by a neurologist at this point. South Carolina did not use El Escorial

criteria to categorize cases for this presentation because there is not enough clinical data available in the in-patient and ED charts to consistently do so. Instead ORS categorized cases based on Dr. David Stickler "gestalt" review of the case. Dr. Stickler is the director of the ALS clinic at the Medical University of South Carolina.   An example of a "not enough information" case would be a patient who is admitted for pneumonia, ALS is documented in the chart but details regarding who made the diagnosis are not available, a neurology consultation was not obtained during the admission and Rilutek® is not listed under medications. An example of a "possible/probable" ALS case for ORS would be a patient who presented to the emergency department for a fall, ALS is documented in the chart under past medical history, and Rilutek® is listed in the medication list. Instances when clinical information is available or a patient is followed in an ALS clinic, or a patient is seen by a neurologist and carries a diagnosis of ALS, they are classified as "true" ALS cases.   Again in-patient and emergency department visit records lacked clinical information needed to categorize the majority of patients using El Escorial criteria. Patients presented to South Carolina hospitals because of end-of-life issues, falls or other injuries or because they were having a PEG or trach placed. One hundred and thirty-five abstracted records coded in ORS data sources as ALS or late ALS have been reviewed by a neurologist. Of those, 51 were deemed "no" or "not enough data to confirm," 36 were "possible" or "probable," and 48 were considered "true" cases. The 51 "no" cases came from the following datasets: 1 VHA; 26 Medicare; 10 from a combination of Medicare, Medicaid and SHP; 14 from a combination of uniformed billing (UB), Medicare, Medicaid, SHP, VHA, VBA, and SC Chapter; a diagnosis of ALS was found in the chart when the record was abstracted in 9/14 (64%) instances when the case was identified from a UB data source. In the other 37 instances the case was identified from a non-UB data source and a diagnosis of ALS was found only one time (3%) in the abstracted record.

For progressive muscular atrophy (335.21), 39 cases have been reviewed to date. None of these appear to be potential ALS cases. Other diagnoses listed in the medical record include: post-polio syndrome, muscular dystrophy, CVA, brain and spinal tumors, and progressive supranuclear palsy. For progressive bulbar palsy (335.22), 27 cases have been reviewed to date with 2 possible ALS cases. Other diagnoses listed in the medical record include: bulbar neuropathy, CVA, myasthenia gravis, and progressive supranuclear palsy. For pseudobulbar palsy (335.23), 48 cases have been reviewed. Other diagnoses listed in the medical record include CVA, dementia, and progressive supranuclear palsy. For primary lateral sclerosis (335.24), 17 cases have been reviewed, with 6 deemed probable PLS. Other diagnoses include: Parkinson's, spinal hematoma, limb pain, and falls. For motor neuron disease other (335.29), 83 cases have been reviewed. Diagnoses listed in the medical record include multiple sclerosis, Parkinson's, post-polio syndrome, dementia, fibromyalgia, migraines, depression, and seizure disorders. For "ALS unsure" (multiple codes identified a case including 335.20, last code = 335.20), 13 cases have been reviewed with 5 potential true ALS cases and 1 PLS. Other diagnoses include CVA, progressive supranuclear palsy, dementia, and neuropathy. 60 "unsure MND not ALS" cases have been reviewed. Findings include the following chart diagnoses: CVA, multiple sclerosis, progressive supranuclear palsy, neuropathy, lupus, migraines, fibromyalgia, radiculopathy, depression, Parkinson's, and benign tremor.

Discussion

- Dr. Kasarskis noted that the field of frontal temporal dementia and motor neuron disease was evolving. Some of the pseudobulbar palsy with dementia actually may have been motor neuron disease. The presentation highlighted the fact that with respect to in-patient and emergency department records, the percentage of people who are actually hospitalized

is probably the minority. Most of the work in many areas is done on an out-patient basis and will never be seen in a hospital code unless someone is suffering some concomitant of the disease (e.g., death, respiratory failure, fractures, et cetera).

- An inquiry was posed regarding whether there were any roadblocks to obtaining information from the state.

- Ms. Royer responded that there were no problems accessing the data. While the investigators had to complete applications for each of the data systems used, they work extensively with these agencies providing data on other projects, so they already understand the confidentiality measures that are in place and ORS' ability to link and track cases within and across systems without identifiersl.

- An inquiry was posed regarding how the individuals were identified for the chart abstraction.

- Ms. Royer replied that if they had a South Carolina in-patient or emergency department visit, she requested them for review, except that they sampled the other motor neuron disease cases. There is a statutory law in South Carolina that allows the health department to review medical records for endemic diseases, which is the law they utilized to review records further. For this project they used in-patient and emergency department visits but the law also allows access to physicians' records with their permission.

-
- Dr. Kaye clarified that the issues of data access had been less problematic for this project because all data requests have been made without asking that people be allowed to be contacted. Therefore, all datasets can be reviewed, matched at the local sites, and sent back to ATSDR without identifiers. How well people will cooperate at the next level remains to be seen. The idea was to determine how well this type of algorithm would work, and then if it did, to move to the next step. Sources have been more cooperative given the confidentiality of not sharing information and because there would be no contact with subjects.

Georgia

Michael Benatar, MD
Principal Investigator
Emory Department of Neurology

Dr. Benatar reported that the objective of the Emory pilot project was to evaluate the feasibility of using existing administrative and clinical databases to identify patients with ALS in the State of Georgia; and determine the predictive value of ALS ICD codes for chart-reviewed validated diagnoses of ALS. Data sources include: 1) Emory (e.g., Healthcare Administration, which tracks in-patient and out-patient records by ICD-9 codes; their large multi-disciplinary ALS Clinic, which has approximately 250 patient under active care at any given time and which is a clinical database that does not keep track of multiple visits, but simply captures that they are in the ALS clinic; EMG Laboratory, which tracks people based on their referring diagnosis and what the diagnosis seems to be based on the electrodiagnostic study; and Grady Hospital, which is linked to Emory, although they have not had a lot of success acquiring these data); 2) other academic medical centers (e.g., Medical College of Georgia and Mayo Jacksonville); 3) a few community neurologists were targeted to determine the feasibility of getting into their offices; 4) Centers for Medicare & Medicaid Services (CMS); 5) Veterans Administration (VA); 6)

ALS Association; 7) Muscular Dystrophy Association; and 8) death certificate data. The advantage of the Emory ALS clinic and the EMG laboratory diagnoses is that it is known that these individuals have been seen by a neurologist, so there is a conclusion reached that they have ALS. This is perhaps more robust than simply pulling them out of an administrative database.

The data from Emory have been the easiest to obtain. For the administrative database, the search has been by ICD-9 code, from which 1,512 patients have been identified. This is based simply on the ICD-9 code, state of residence, and the years of interest for this study (2001-2005). From the ALS clinical database search, 1,013 patients have been identified. The EMG laboratory database is relatively recent. It has not been running for the full duration of the period of interest; however, this search yielded 157 addition patients from 2004-2005. As noted, no data have been obtained from Grady thus far. While the investigators have approval from Grady's research oversight committee, given the Grady crisis, they have not been able to search their records to provide the investigators with that information.

They have experienced difficulty in acquiring data out of some of the academic medical centers. Medical College of Georgia has theoretically been on board for almost 18 months, there have been numerous encounters with the IRB, and various administrative hurdles. However, there has been a recent agreement to proceed. It is not clear whether the pilot phase will come to an end before data are actually received from Grady, but they are still working on it. Mayo Jacksonville is interested, willing, and has expressed an intent to proceed; however, there has been no progress yet on the IRB submission and they do not seem to want Emory's assistance in preparing that.

Three community neurologists were approached, all of whom agreed to provide data, but only two of which they heard back from in terms of getting data. Each of the two searched their practice administrative databases via ICD-9 codes, with 51 patients identified from practice #1 (49 charts abstracted) and 169 identified from practice #2 (chart abstraction pending).

The Medicare data have been received for Georgia via the CDC / ATSDR. The data are based on an ICD-9 code search for ALS. This yielded 2,599 cases resident in Georgia between 2001 and 2005. The VA data for Georgia were also obtained via the CDC / ATSDR and include multiple sources of data (e.g., administrative, benefits, pharmacy). The data are based on an ICD-9 code search for ALS. This yielded 191 cases resident in Georgia between 2001 and 2005.

From the ALS Association, approval was granted to search the ALSA database under the auspices of the Emory IRB. This yielded 190 subjects with ALS in Georgia between 2001-2005. Unfortunately, there are no charts for abstraction because the ALS Association is not providing care. However, it seems reasonable to conclude that these subjects likely have ALS. The MDA as an organization did not have the ability to share public health information; however, they did refer investigators to all of the individual MDA clinics within the state (e.g., Emory, Medical College of Georgia, and Roosevelt Warm Springs). IRB approval was obtained from Roosevelt Warm Springs. The Roosevelt Warm Springs database search only yielded six patients, and all of those charts have been abstracted.

Death certificate data have also been obtained. This was a lesson in government office, the various hierarchies, and the organization. Eventually, approval was obtained for data sharing from the Georgia Department of Human Resources IRB. The study was classified as exempt and did not really require IRB approval. Permission was granted to search death certificate data

to identify cases with ALS listed as a cause of death (ICD-10). This was a relatively straightforward process, with approval obtained within 6 months. This search identified 1,435 cases.

Not a great deal of data abstraction has been done, probably about 250 charts or so, and not all of those have yet been verified by a neurologist. A uniform data abstraction form was used that was developed collectively by the sites involved in the pilot project. There have been a number of methodological difficulties, such as limited information in the chart (e.g., stated diagnosis of ALS, but no data to permit independent verification); single notes with no ability to document progression of symptoms / signs that is essential to verification of the diagnosis; and uncertainty regarding whether verification should be based on application of formal diagnostic criteria or a neurologist's gestalt of the case.

Dr. Benatar acknowledged the data management staff at the Georgia Cancer Registry for the tremendous assistance they provided in synthesizing and integrating the data from all of these sources and linking them based on their standard algorithm. One of the difficulties with the diverse data sources is that very different data are received from each source. For example, the clinical databases do not have Social Security Numbers or multiple dates of visit and the EMG database does not have a date of birth. The breakdown of cases appearing in only a single database include 585 administrative, 323 clinical, 102 EMG, 98 ALSA, 152 VA, 111 Peachtree North, 1,672 Medicare, and 678 mortality. Most striking to Dr. Benatar is how relatively unique each database is. There is some degree of overlap, but the majority of subjects in each group seem to be unique, which is somewhat troubling in terms of practicality of developing a registry and identifying everybody with the disease. This suggests a need to search every data source, which is of significant concern in terms of getting into the neurologists' offices as that may be challenging.

Not a lot of analyses have been done to date, given that the investigators are still in the process of abstracting data, but to offer some sense of the 180 charts that have been abstracted and verified, Dr. Benatar reported that about three quarters of the diagnoses are verified as ALS, which seemed consistent with much of the other data presented during this meeting. Interesting is that for the administrative database, there is a high proportion of people with an alternate diagnosis. However, there are very small numbers of people in the clinical and EMG data who do not have confirmed diagnoses of ALS. For those appearing in a single data source, approximately 30% of people have a verified diagnosis. For those appearing in four to five datasets, approximately 93-94% have a verified diagnosis. Thus, there is utility in examining that. Other components that have been examined for those datasets that would permit it include care multiple times, specific ICD codes, EMG, et cetera. One thing that was somewhat disturbing is that 323 people appeared in the ALS clinic database who did not appear in the Emory health care administrative database, which raises issues about the reliability of coding.

Dr. Benatar's preliminary conclusions were that it is important to identify cases from multiple sources, given that data sources are largely independent with respect to identification of unique cases; there is tremendous variability in the ease of data acquisition from different sources; there are methodological difficulties of data abstraction for case verification; there is potential importance of case verification given the high frequency of "other" diagnoses; clinical database diagnoses of ALS is more robust than administrative data; and accuracy of diagnosis of ALS varies by the number of data sources used to identify a particular individual case.

Discussion

- Dr. Strickland suggested that this looked like an ideal set-up for a Capture-Recapture estimation, and he wondered if the Emory investigators had done this.

- Dr. Benatar responded that they have a Masters of Public Health student who wants to work with this dataset to do just that.

Mayo Clinic

Eric Sorenson, MD
Principal Investigator
Mayo Clinic

Dr. Sorenson noted that the Mayo Clinic pilot project is probably further along than many of the other pilots, given that they have fewer patients' data to deal with. With respect to the VA data, 82 subjects were residing in Minnesota at the time of their VA encounter. Of those, 29 were Mayo patients and 53 were other patients. For those seen at Mayo, the charts and data were very easy to capture. Of the 29 Mayo subjects, 16 were confirmed to have ALS by some formation of the *El Escorial* criteria, which is about half of that dataset. Of the 16, 7 were definite, 4 were probable, and 5 were possible. Fifteen cases were male and one case was female, which was not surprising coming from the VA dataset. This marked gender imbalance raises a potential source of bias in the VA dataset. There were also 13 subjects in the VA dataset who did not have ALS. Many were misdiagnosed / miscoded. The most common misdiagnosis was Parkinson's disease, which is ICD-9 code 332. ALS is ICD-9 code 335.2 so it is possible that there is miscoding or typographical errors could have occurred. There were three other cases with ICD-9 codes similar to ALS, but which were off by a digit or so, which may have caused some data entry errors.

From non-Mayo data sources, 52 cases of the total 82 were identified. For the non-Mayo patients, the Mayo investigators partnered with the ALS Association in Minneapolis and were able to review their records. As noted, there are no medical records to review within the ALS Association data. The only thing that can be reviewed is whether individuals appear in this database as a case of ALS. Presumably, someone somewhere has given these individuals a diagnosis of ALS. Dr. Sorenson said he tended to agree with Dr. Benatar that if they were in this dataset, it was highly likely that they do have ALS. From the ALS Association dataset, 19 cases were included as confirmed. For the other 34 cases identified from non-Mayo data sources, there was no clinical information whatsoever. Some of them may have ALS, but many of them probably do not. In Olmstead County, from the years 2001-2005, there were 12 incident cases of ALS. Only one of those was detected in the VA dataset. Thus, it appears that a high number of cases are being missed by that dataset. Using the certainty of match scale that scores from 1-7, in the VA dataset on which this has been done, the median score for certainty of match was 7 out of 7. Thus, matching the patients up and making sure that the patients were correct based on the dataset received from CDC / ATSDR was not a problem. The range was 3.5-7. Where they lost points was not because the data were different, but was almost always because a data field was missing.

With respect to the Mayo experience with Medicare / Medicaid data, a dataset was received from CDC / ATSDR which included 1,325 subjects identified in Minnesota. Of these, 536 were seen at Mayo and 789 were not. Comparing the distribution of Mayo / Non-Mayo patients in Minnesota for the Medicare / Medicaid (40%) and the VA (35%), the proportion is very similar.

For the 536 cases seen at Mayo, 296 were confirmed as having ALS, 142 did not have ALS, and 98 were unknown given that there were not enough data to make a classification (these were Mayo cases, but they were not seen in the years 2001-2005). Of the 296 cases that had codes for ALS, 192 were definite, 42 were probable, and 62 were possible. There were 10 prevalent cases of ALS (definite or probable) from Olmsted County, and 4 prevalent cases of possible ALS. Of those Olmstead cases, 6 were incident cases in that five-year period in the databases, compared to 12 incident cases over the five years in Olmstead County total. So, the yield was only about 50% of the cases that appeared in the Medicare / Medicaid database. For the 142 cases seen at Mayo who did not have ALS, the most common diagnoses were Parkinsonism, PSP, and MSA. Of these 142 cases, 10 were from Olmsted County, which means that for many of the cases appearing in the dataset, the diagnostic yield is very low.

An analysis was conducted regarding who was in and who was out—who was the database missing? Not surprisingly, those who were not in the Medicare / Medicaid database were younger than those who were in it. That is one potential source of bias in using Medicare / Medicaid data, particularly given that one of the major prognosticators for ALS is age of onset. The only other demographic difference between the two groups (those in and those out of Medicare / Medicaid data) was their date of onset of the disease. If someone is younger at age of onset, the time their disease progresses to the point that they are disabled and eligible for Medicare / Medicaid benefits could be delayed by a year or two. That is, the missing people did not have enough time to appear in the Medicare / Medicaid dataset.

The ALS Association dataset has not yet been abstracted for the Medicare / Medicaid set, but by extrapolation based on the VA rate, it is likely that 96 additional cases will be obtained from this database. That would give them a total number of prevalent cases in Minnesota for that five-year period of 392. The population in the 2000 Census for Minnesota is approximately 4.9 million people, so roughly 500 incident cases of ALS would be expected over a five-year period. Thus, they might be able to yield up to 80% of the cases using the Medicare / Medicaid dataset by itself. The Olmsted County yield was lower than that at 50%, but again, it may take a while for some of them to appear in the database. The reality is probably somewhere in the middle of those two numbers, between 50-79%. The Medicare / Medicaid dataset was far better than the VA dataset, which by Olmsted County yielded 8% and by the State of Minnesota yielded 7%. Comparing the VA dataset to the Medicare / Medicaid dataset, 4 cases were unique to the VA database, 12 cases were common to both, and 284 cases were unique to Medicare / Medicaid.

Ongoing challenges are that up to half of the subjects in the datasets do not have ALS. These are large numbers of patients, a huge amount of work to go through, and somehow to make a registry, the dataset must be better than what it is. Moreover, there are many common diagnostic errors of other neurodegenerative disorders, most often Parkinson's disease. With respect to Olmsted County data, 10 of the 24 cases in the dataset have no evidence of ALS, or even ALS-like syndromes. This means that 42% of cases in the Olmstead County dataset do not have ALS (95% CI: 24-61%). One of the criticisms of Olmstead County that is true is that there is a low number of cases. Thus, somewhere between 25-60% of the cases are simply not going to have ALS. That is a major challenge when dealing with such large numbers and such a large number of resources that are required to filter through all of this.

Discussion

- Dr. Hornbrook pointed out that just because someone does not have ALS does not mean that they do not have health care needs. A lot of symptoms were seen for these patients, so this is still a health care challenge that needs to be met.

- Dr. Sorenson replied that while this was true, it was not a health care challenge that was to be met within an ALS registry.

Northwest / Hawaii / Southeast

Mark Hornbrook, PhD
Chief Scientist
The Center for Health Research
Northwest / Hawaii / Southeast

Dr. Hornbrook explained that the HMO Research Network (HMORN) is a consortium of 15 Health Maintenance Organization (HMO) delivery systems that have formal, recognized research capabilities. These research centers are professionally autonomous from their health plans. That is, the corporations do not drive the research agendas or control what the investigators do. The investigators' work is conducted in the public domain, and public domain work is the dominant share of their professional activities. HMORN aims to be recognized as the Nation's premier resource for population-based health and health care research; contribute to national and global dialogues on health research and policy; promote and establish the network as a preferred research partner; foster network-led collaborative studies; and share methodologies, best practices, and consultative expertise. This strongly resonates with the NIH's Clinical and Translational Science Awards (CTSAs) which replace the General Clinical Research Centers (GCRCs). Every one of the HMORNs has some kind of tie to a CTSA.

Collaborating HMORN sites in the registry project include Fallon Health System / Meyers Primary Care Institute; Marshfield Clinic; Lovelace Health System; Geisinger Health System; Kaiser Permanente Northwest; Henry Ford Health System; Henry Ford Health System; Kaiser Permanente Southern California; Kaiser Permanente Northern California; and Harvard Pilgrim Health Care. All of these HMOs have very important data systems, but this poses data collection challenges, given that there are nearly 10 million members across 9 participating health plans. Therefore, consideration must be given to how to identify which charts to pull. Smaller HMORN sites can review medical records for 100% of cases identified via computerized search algorithms. However, large HMORN sites have multiple chart rooms located many miles apart. For them, the staff time and travel costs are too high to pull 100% of the suspected charts to review. Thus, medical record resources must be allocated to cases that require additional data to determine eligibility. Electronic data sources are multiple, and lifetime health record numbers enable linkage of clinical records over time for specific individuals, including: Hospital discharge abstracts, electronic medical records for ambulatory care, ambulatory encounter abstracts, ambulatory dispensing records, home health and hospice records, professional and facility claims, imaging information systems, laboratory information systems, and death certificates.

The methods used for the pilot project were to find all mentions of ALS / MND diagnoses and Riluzole dispenses in the electronic data systems for the period 01/01/2001 – 12/31/2005; eliminate "rule-out" diagnoses; keep in death certificate diagnoses; classify cases by data patterns into three levels of inclusion (highly likely, possible, unlikely); and validate the case classification algorithm against medical record data through expert review of charts. The Center for Health Research has been working for years to develop a virtual data warehouse (VDW), which makes it efficient to compare diagnosis and procedure data across HMORN sites. This is a distributed data model (DDM), which is a decentralized virtual data warehouse. All sites map

local data files into standardized SAS variables (labels, formats, codes). All the data sits locally under localized control. A DDM was created to support all research projects at the participating sites. Programs can be written centrally, distributed, and run by site programmers on local standardized data files. The Vaccine Safety Link Data Project (VSD), supported by CDC, takes this one step further. Under the public health availability of data disclosure under HIPAA, the VSD actually runs these standardized programs automatically, so when the data program is placed on the website, the various participating sites go to that website, download the program automatically, run it, re-post the results automatically, the originating site downloads this off of the website, and creates the data files with one programmer's effort. Cancer and ALS have not gone that far because there is still a lot of nervousness on the part of medical groups with respect to allowing data to flow that easily without somebody actually putting their eyes on the data before they leave a site.

A validation sub-study will be conducted to test the hypotheses that there are some types of data patterns that should not be included as potential ALS cases, while other data patterns should be included. That is, only 1 or 2 ALS / MND diagnoses should be excluded. Many ALS / MND diagnoses should be included. Procedures could identify suspect cases with no diagnosis. While there is a standardized abstract, for those who have electronic medical records, that represents a high level of labor for someone to read the chart, transfer it onto an abstract form, key in the abstract form, and then run the analysis. The Center for Health Research would like to open the charts electronically, print off key neurology and other records, and send that to the adjudicator. This would be the actual original chart, with no other data processing that could have introduced error. The sub-study will help to define the boundaries of the "possible" group to be abstracted and validated. With regard to adjudication, the abstractors will be trained to find the appropriate sections, just as they would be trained to complete the abstract form. The adjudicator will view the raw data from the chart, including EMG test results and graphs. This method reduces abstractor recording and coding error. However, Dr. Hornbrook noted that not all of the sites had electronic medical records during the time of this study. Thus, some sites will photocopy the charts and / or will do abstraction to bring the data forward. They are the ones who have to go through the process and meet the criteria for sensitivity and specificity.

Regarding matching, data will be received from the CDC / ATSDR for the HMO catchment areas, which will be matched to The Center for Health Research's case records using fuzzy logic methods. Each case listing from ATSDR and the HMO can be treated as a random draw from the total disease population. Tests will also determine whether both listings were likely to have been drawn from different populations.

For preliminary tabulations, counts were combined of patients with any mention of ALS / MND diagnoses and / or Riluzole dispensings over the study period from eight of the nine participating health plans. This includes three of the four health plans waiting to be brought in by new subcontracts. This is a two-phase site, which started with five health plans and then added another four. With respect to patients classified by any ALS / MND diagnosis and prescription, 35 patients had access to the drug but did not have a diagnosis. In fact, one person had 10 dispensing or more of Riluzole in the absence of a motor neuron disease. In 149 cases, there were more than 10 mentions of a diagnosis and more than 10 dispensings, which is a predominance of evidence for those individuals. There 2,199 cases identified on the basis of diagnosis alone and no dispensings. Riluzole is in the formulary in all of the health plans, so it is not a question of access to the drug. This does not even look at the issue of complication codes and symptom codes that would be in a cluster of disease progression. In terms of patients identified by ALS diagnosis alone and dispensings, 474 cases had no ALS or Riluzole

dispensing, so these are other MND cases. Six hundred ninety-four cases have only one mention of ALS diagnosis. Encounters before and after these mentions are still examined to determine whether other corroborating symptoms or procedures will allow for classification of these cases in the ALS / MND group. Data from one health system (Health System i) presented only whether a patient had one or more ALS / MND diagnoses or Riluzole dispenses, but not the actual counts of "hits." Therefore, it could not be combined with the other two tables.

Patients by age and gender, pooling data from eight of nine of the participating health plans, were as follows:

Patients by Age & Gender

| | Age Group at Reference Date | | | | | |
	15-29	30-44	45-59	60-74	75+	Total
Female	30	134	356	487	319	1,326
Male	38	205	546	619	363	1,771
Total	68	339	902	1,106	682	3,097

8 of 9 health plans

16

There were 68 cases among patients under age 30. It is clear to see that there is an age progression.

Patients by year of first diagnosis for eight of nine health plans were as follows:

Patients by Year of First Dx

	Reference Year	# Sites	# ALS/MND cases
	1996	1	27
	1997	5	61
	1998	5	57
	1999	6	82
	2000	7	190
Case Accrual Period	2001	8	547
	2002	8	495
	2003	8	534
	2004	8	536
	2005	8	566
	Total	8	3,095

8 of 9 health plans

17

These are unique cases classified by the year of first mention of ALS / MND diagnosis or Riluzole dispensing. All cases had to have an ALS / MND diagnosis or Riluzole hit during the study observation period (2001-2005). Then the investigators looked backwards to determine when individuals had their first mention of the disease. There is also a set of hits in 2006, 2007, and 2008 so these patients do live for more than a year.

Dr. Hornbrook indicated that the next steps included the execution of the subcontracts for the four new sites; IRB approvals for the new sites; development and validation of the computerized case identification algorithm at all nine sites; creation of final case listings for each site; receipt of data from ATSDR; matching of ATSDR-HMO case files; analyzing sources of mis-matches; and reporting to ATSDR. Dr. Hornbrook emphasized the fact that there is a time lag for Medicare. Medicare eligibility is automatically conferred on an ALS diagnosis. The problem is that people have to wait to receive the diagnosis, which becomes a public policy issue with respect to depriving patients who have ALS of Medicare benefits by postponing the diagnosis. For some individuals, that represents a major increase in benefits. Medicare could be the safety net that finds every single ALS case if people pursue their Medicare eligibility. However, some people will not apply because they believe it to be welfare. With that in mind, he viewed this as a health service research project because it deals with access to health insurance and how people get into Medicare and the VA, and why they are there and why they are not.

Discussion

- Dr. Benatar said it seemed that a lot of the discussion during the day had been focused on positive predictive value of the ICD-9 code for a diagnosis of ALS. It seemed that The Center for Health Research was in a unique position to get a negative predictive value as well, particularly with respect to the Riluzole data. Most of those engaged in the pilot projects do not have a mechanism to predict the negatives. With respect to why people would be on Riluzole and not be coded with ALS, he said that anecdotally Emory has an on-going predictive study of people at risk for ALS based on harboring mutations in the gene for SOD1 protein. They have run across a few people who are carrying on-going prescriptions for Riluzole based on their lifetime risk of close to 100%. While that might not explain all of Dr. Hornbrook's subjects on Riluzole who have no diagnosis, it could be one explanation.

- Dr. Hornbrook agreed that this was very important. They would like to pursue this, and if they monitor their resources carefully, they should be able to say something about that. It is important to be able to predict the negatives just as strongly as the positives.

- Dr. Van Den Eeden added that anecdotally, there is a period of a year or two in which people with Parkinson 's disease were interested in Riluzole, so people were taking it off-label.

- Dr. Hornbrook responded that they merely looked at dispensing, not why people took Riluzole. It is possible that there is off-label dispensing, even inside the HMOs.

Open Discussion Session

Wendy E. Kaye, PhD
Senior Epidemiologist
McKing Consulting Corporation

During this session, Dr. Kaye led participants in an open discussion regarding the issues related to developing a national surveillance system and registries for ALS and MS.

Discussion

- Dr. Hornbrook wondered whether anyone planned to work on negative prediction models and criteria for defining when a person does not have ALS.

- Dr. Sorenson responded that this was a very active interest for the Mayo Clinic. With their data from Olmstead County, they can generate very good predictive models on who does and who does not have ALS. They are very interested in determining who they are missing and who is appearing in the database who does not have ALS.

- Mr. Culpepper responded that the MS Center of Excellence is attempting to do this with the MS side as well. The intent of their original strategy was to focus on those who were in the system, but who were not MS.

- Dr. Tyry indicated that in NARCOMS there is a very small percentage of people who have reported being misdiagnosed. It might be helpful to examine the characteristics of those individuals as well.

- Dr. Sorenson responded that that was yet another challenge. That represents a population of patients for whom one neurologist may say that have the disease, while another may say they do not. That is a different scenario from people who are miscoded or appear with another diagnosis. Determining what to do when there is a disagreement between clinicians is an even greater challenge.

- Dr. Van Den Eeden replied that partially it would depend upon when they wanted to say someone was a case or not a case.

- Dr. Kasarskis recapped what he believed he understood, in no particular order, from the morning's presentations: It seems clear that one dataset clearly does not serve all masters. While life would be simpler with a biomarker or a set of biomarkers, at least from the ALS side, they do not have this. Thus, the goodness of anyone's datasets rests on how good one's diagnostician is. Several people mentioned that they had an adjudicator (e.g., an expert). Bertrand Russell said an expert is a person that other experts say is an expert. From his own experience reviewing charts from many places around the country in the context of the VA registry, Dr. Kasarskis found that there were some people who were clearly not experts and others who clearly had seen a lot of ALS, zeroed right into the heart of the matter, and laid out a perfect history. Yet, someone must try to reduce that to a single code in a single dataset, which is an extremely difficult and perhaps impossible task. With

no offense to anyone, he did not believe they would really further anybody's agenda by trying to invent new terminology, especially that equivocated on the same words. *El Escorial* is really not a perfect system, but to some degree, they have already co-opted the terms "possible," "definite," and "probable." To come up with another definition in another study is not going to allow any sort of cross-comparison whatsoever. Instead, it will cloud the water and people will never move beyond the debate about what they are really talking about in terms of a dataset. He also wondered what the goal was for the entire exercise. If they wanted a dataset of people for whom there are very few false positives, meaning really true cases, then that would feed into a clinical study. If they wanted to determine a spectrum of motor neuron diseases to find out who was in the population, that would be a different set of questions that would allow "fuzziness" in the dataset. There was no right or wrong answer, he just wondered why they were collecting the data and what they were trying to monitor. He thought that the European approach was to call it "motor neuron diseases" of which classic ALS is a subset.

- Regarding the definitions, Dr. Kaye agreed with Dr. Kasarskis about the *El Escorial* criteria. However, surveillance also has definitions of "definite," "probable," and "unlikely." They do need to be careful about what they say and perhaps should included a descriptor in front of it, but as far as these projects are concerned, those terms are going to be used. For surveillance purposes, they are definitely ALS, probably ALS, or they are not likely to be ALS. There also must be an understanding that people can shift categories as more years of data become available. Regarding the other motor neuron diseases, CDC / ATSDR did ask for all of those disease categories because they thought it would be a global process. Unfortunately, so far it appears that a lot of the other MND codes are being used inappropriately. In South Carolina, 50% of the cases were coded as other motor neuron diseases because they are using the codes as descriptors of symptoms and not diseases. However, they are not doing that with ALS.

- If they were to do this again, Dr. Sorenson wondered whether they would include the ICD-9 code 335.2. A small minority of those cases have ALS.

- Ms. Royer said from what they have found so far, the majority of the other and unspecified motor neuron disease codes (335.29, 335.2) are not turning out to be ALS cases when they are the only code that identified that case. There also appears to be very few instances when pseudobulbar palsy or progressive muscular atrophy is the only code that identifies an individual for inclusion in the cohort, and upon chart review it appears to be a potential ALS case. It appears that pseudobulbar palsy and progressive muscular atrophy are being used to describe symptoms of other diseases.

- Dr. Sorenson pointed out that they could enrich their database by culling those out of the population.

- Dr. Kaye added that there could also be other descriptors that could better sort those out ahead of time.

- Mr. Culpepper noted that another issue was that every dataset is unique in terms of the types of data that are available. Therefore, what can be done in one dataset to rule out cases is not necessarily going to be available in another dataset. For each of the data sources they will have to develop specific algorithms that are pertinent to that data source, with which everyone is comfortable, and that are giving them a reasonable estimate of who the real cases are. If they can do that, they can prioritize the cases as definite, probable, and possible or something along those lines based upon the strength of evidence within that

35

given data source. Once the matching is done, they will be able to examine the ratings across datasets to help further define what is definite, probable, and possible.

- Dr. Benatar pointed out that within a dataset, defining cases of ALS rests on the fundamental premise that there is agreement about when to abstract a chart, which is the gold standard. Thus, there probably should be some discussion about whether they will rely on El Escorial criteria or whether other criteria will be used. Perhaps they will want to have several levels of confidence of diagnosis, including El Escorial and then some gestalt impression based on several features that are included. If they do not have agreement on what rules should be used when abstracting data to make those determinations, all of the other algorithms that are tested against any gold standard will not be meaningful.

- Dr. Kaye referred people to the Excel data sheet the projects are to send back to CDC / ATSDR, noting that there are two columns: one for the neurologist and one for El Escorial criteria. Those are different.

- Dr. Benatar said that the neurologist is still making the determination about El Escorial criteria.

- Dr. Kaye responded that they are not because they do not have the data that would feed into the El Escorial criteria, so they cannot use it at all.

- Dr. Stricker said that of the 400 charts he has reviewed, the El Escorial criteria are used in about 12 of those.

- Dr. Benatar said that meant he was making a gestalt judgment for the other field, but the question regarded what that gestalt judgment was based on and whether it was the same as the gestalt judgment of other investigators.

- Dr. Sorenson pointed out that as the algorithm becomes more complicated, they start missing data as it builds. A simple, straightforward program is needed. Perhaps less than 5% of neurologists who see patients who have been diagnosed with ALS for a second opinion change the diagnosis. He would also argue that if any neurologist gave a patient a diagnosis of ALS that would be good enough.

- Dr. Benatar clarified that what he was saying was that they need to have consensus on the subject. Perhaps someone could propose some rules for them all to consider.

- Dr. Newman said he thought in essence that was what was happening during this discussion in slow motion. That is, a number of programs are underway simultaneously which are examining large databases in different ways and attempting to determine the most efficient way "to be wrong the least". While the point was well-taken with respect to terminology, given the opportunity to review everything in the chart, investigators may be able to make a determination. However, they must be able to look at 500,000 pieces of Medicare data and 1.4 minutes later be able to say what the set is that has a high likelihood of being ALS. He thought it was Don Moulder who said that it was rare for somebody who was experienced to look at ALS and think it could be anything else. The simplicity of that is in great contrast to the tortured efforts of El Escorial, as well as the El Escorial revised criteria in which many people are still dying without meeting "definite." There are going to be false positives and false negatives, but if a neurologist says someone has ALS, that is one of the best predictors.

- Dr. Benatar expressed concern about the emergency department physician who states that there is a history of ALS.

- Dr. Newman responded that the emergency department physician is likely to have more than one data point in that same database.

- Dr. Sorenson pointed out that some of the centers were grappling with the fact that they have isolated data points and they have to make a decision about whether a patient is in or out.

- Dr. Hornbrook's sense was that they would have to make do with all of the data they could possibly get. He will have some cases for which he has only one or two data points. That is going to be a class of cases that will be labeled accordingly. But, that's not all of his cases. A lot of cases will give them two or more years of a progression of symptoms, treatments, and co-morbidities that combined will give them the gestalt and perhaps even more than that.

- Dr. Benatar agreed that some cases would be easy, while others would be more difficult. All he was suggesting was that an agreed upon set of principles should be used for the hard cases. Otherwise, they would not be speaking the same language. When there is clear evidence of progression, they have ALS, they are on Riluzole, et cetera, then they can be included as definite ALS. He was talking about the more difficult cases in which a judgment must be made because there is a lack of data or there is just one data point.

- From the health services research perspective, Dr. Hornbrook said he had sometimes had few data points because the patient had tremendous barriers to access. If the person has very good access, they present all of the time for their migraines, foot aches, muscle problems, joint problems, et cetera and if there is only one ALS mention he is a lot more suspicious. He does not believe that the abstraction data set is preferred for neurologist adjudication, which is why they are staying with the full medical record context in their adjudication approach, because they want all of the environmental information.

- Dr. Benatar suggested that they should have agreement upon whether they should or should not be using a data abstraction form. That seemed to be the extent of the disagreement.

- Dr. Kaye responded that the projects were somewhat different from each other for a reason. CDC / ATSDR does understand that there will be some differences because many of the HMOs have electronic records and, therefore, do not have the issue of having to do data abstraction versus a more traditional setting where data abstraction must be done.

- Dr. Benatar suggested that the minimum requirement should be that each site should write down the rules they are using.

- Dr. Kaye indicated that they reviewed a number of the cases in South Carolina, which showed their rationale for why cases were included in one category or another and it is more of a gestalt.

- Dr. Sorenson asked Ms. Royer what portion of the patients they uncovered who had ALS from a single emergency department visit for pneumonia, which had no documentation of anything else from these isolated data points.

- Ms. Royer responded that she did not remember the number right off, but sometimes the only chart she had available was one emergency department record from which to try to confirm a case.

- Dr. Sorenson said his gestalt was that those numbers are fairly low, and they are struggling with a lot of time and effort over a very small fragment of patients and whether it is going to make or break the database. It is almost an insignificant group of patients.

- Dr. Benatar pointed out that if investigators had a death certificate and some other source to make the diagnosis of ALS, they could not then go back to say that being identified in two data sources meant they have ALS. That just becomes circular.

- Dr. Sorenson responded that they went through the VA and to the Medicare database, finding that in the vast majority of cases it is obvious—they have ALS or they do not. Only in a small number of cases did they have to struggle to determine whether a patient had ALS or not.

- Dr. Stickler said that of the 135 records he reviewed in South Carolina with an ALS or late ALS code, 51 did not have enough information and 36 had enough to be probable or possible, but not to be definite. As he has reviewed these, he has been making notes in the text box about why he thought someone could have ALS and is hoping to go back to create parameters.

- Ms. Royer added there were 10 cases out of the 135 reviewed to date where the diagnosis of ALS appeared in the chart, but there was not enough clinical data to confirm a case.

- Dr. Sorenson replied that they were struggling with a really small number. Just like anything else, to increase diagnostic sensitivity and specificity, as they move toward 100%, their costs are going to skyrocket. At some point, they have to live with what is financially possible and how much sloppiness they can live with. A huge amount of effort can be put into tracking down a handful of cases, or they can focus on what is reasonable. There is no way in a national registry that charts can be perused to this level to decide if a case should be in or out.

- Dr. Benatar pointed out that that was why this was done at the feasibility stage, in order to know the reliability of the data sources.

- Dr. Cwik said it seemed to her that everybody should be using the same definitions and the same criteria. If not, they would not be able to compare the datasets or pool the data. This would destroy the registry.

- Dr. Strickland agreed with Dr. Sorenson's comments about what a small proportion of ALS patients this is going to be, but at the same time he had 3.2 million patients at Kaiser.

- Dr. LaRocca found it surprising that in a project like this there was not more agreement on definitions. Epidemiology is really all about approximations. They really are spending a great deal of time on trying to achieve what would seem to be perfection. Instead, they need to devote more time to looking at the sources of error and defining what the parameters of the errors are in the various estimates because they really do not have the time or resources to achieve perfection or anything near it. What they really need to know is

how good their approximations are. They also must keep in mind that they are trying to achieve a reasonable approximation, at a reasonable cost, within a reasonable period of time.

- Dr. Newman asked how Dr. LaRocca would suggest getting an idea of how wrong they are. As they heard earlier, the reasons they are doing this is to know what the range of incidence is in order to know when there is a cluster. Without knowing how much "slop" is built in, it was not clear they would be able to answer that question.

- Dr. LaRocca responded that a good example might be the study that Mr. Culpepper showed them in which they developed an algorithm and then tested it on a sub-set of the available dataset. That is probably the most cost-effective way to examine the level of errors, as long as the sub-set is reasonably representative of the larger population.

- Dr. Kaye indicated that Dr. Randy Durbin, who is a statistical consultant, will be working on some methodology.   Basically, through capture-recapture statistics, he will study where the overlap is, which will allow for making best estimates on what has been found. Dr. Strickland did this with the Parkinson's data in Nebraska. It is based on fish and wildlife, but specific applications have been designed for use in public health. That is one thing they will be doing with the combined datasets that might help answer some of those questions. They must keep in mind that surveillance is never perfect and that they will always be missing somebody. It would be nice to know some characteristics of the kinds of people they are most likely to be missing.

- Mr. Culpepper pointed out that the other issue pertained to how they would use that information. If the datasets are categorized as definite, probable, and possible, for some questions they may only want to focus on the definites. If examining raw incidence, they may want to include everybody in order to get a range of what the real values are. Even if there is some disagreement or question about some of the cases, once they are identified, there is no reason to discard them per se.

- Dr. Benatar suggested three questions to ask during data abstraction to make a determination: 1) Is there a chart diagnosis of ALS? That weeds out the question of miscoding, so a Parkinson's patient can be quickly dispensed with; 2) Has a diagnosis of ALS been made by a neurologist?; 3) Is there an El Escorial diagnosis of ALS at the various levels of certainty.

- Dr. Kaye responded that all of those questions are in the Excel data: 1) "What was the code in the chart?"; 2) What were the El Escorial criteria; and 3) For each diagnosis it asks whether it was made by a neurologist, family medicine practitioner, et cetera. Procedure codes were originally considered as well, but that went beyond the scope.

- Dr. Benatar said the chart was unlikely to have a code, which was why the question should be, "Is there a chart diagnosis of ALS?" It was not clear to him how the Excel could be completed if the abstraction was first.

- Dr. Kaye replied that the Excel is the sum of what is found.

- Dr. Benatar stressed that they would have to generate that automatically from the abstractions because otherwise they would have to go back and abstract again.

- Dr. Kaye acknowledged that they must ensure that there is no disconnect between the abstraction forms versus the Excel. She thought it was possibly just a terminology issue.

- Dr. Tyry expressed an interest in going back to the databases to look for people who reported being misdiagnosed. Most of the data they have is self-report, so she was unclear how reliable that would be, but self-reports are considered as cases. They do ask whether they have a doctor's diagnosis.

- Dr. LaRocca said they had done some verification and self-report was not that unreliable.

- Mr. Culpepper said they see the same thing in their survey data and do ask whether there is a physician diagnosis.

- Dr. Kasarskis said that in the VA registry, they apply a screener asking, "Has any health care provider told you that you have ALS or any motor neuron disease?"

- Dr. Sorenson noted that the MS group was much further along than ALS. He did not believe ALS should hold themselves to a different standard from what MS has done. He thought the same rules should be applied to ALS as had been applied to MS.

- Dr. Teter indicated that they are also tracking number of relapses, how often relapse occurs, et cetera. She would not doubt the neurologists' diagnoses.

- Dr. Kaye pointed out that even though the MS registries were at two extremes with one registry being almost entirely clinical diagnoses and the other entirely self-report, they were getting similar results.

- Dr. Hornbrook thought that end-of-life care required documentation from hospice.

- Dr. Kaye clarified that for Medicare eligibility is determined prior to admittance to long-term care. They do not have to put a diagnosis code in the field in order to get paid. In some states, there was 100% ICD coding in long-term care while in other states there was no coding. Long-term facilities are not required to indicate a reason for eligibility.

- Dr. Benatar wondered if everyone was in agreement that someone should be included if their hospice record indicated that they had ALS.

- Mr. Culpepper thought that would be okay, but that it should be documented in some other way also.

- Dr. Hornbrook pointed out that another issue pertained to who was filling out the codes and why. A clerk following the coding rules was different from trying to drive the care system with the correct data.

- Dr. Sorenson said he stood by accepting a diagnosis from a neurologist.

- Dr. Thurman pointed out that, with respect to self-report, one issue of concern was that even if asked whether the doctor who made their diagnosis was a neurologist, some people may not know what kind of doctor they saw.

- Dr. Sorenson stressed that doctors are afraid to make the ALS diagnosis.

- Mr. Culpepper said that in the VA system, most people develop ALS while active. If they present to VA after a definitive diagnosis is made, the investigators can work backward and forward from that point to validate whether the diagnosis is a true case.

- Dr. Kasarskis agreed with Dr. Sorenson that rarely would a primary care physician diagnose ALS. Every one of these cases has two or three other diagnoses and physicians (e.g., primary, ENT, orthopedist, rehab) along the way before the "light bulb goes on." His prediction was that there would be a lot of ICD-9 codes before an actual ALS diagnosis.

- Dr. Teter said that while they see cases in the MS clinic, people go back to primary care for treatment. They suspect that a portion of registrants lost to follow-up may be treated in primary care.

- Dr. Tyry said they do not know who their responders treating physicians are, but they do ask people if their care is taking place in a specialized MS center or clinic. About half say "no." Not everybody sees a neurologist every six months.

- Dr. Sorenson thought the same would be true in ALS. Probably about half present to a clinic and the other half are seen locally by their family physician and / or neurologist. Probably all were diagnosed by a neurologist.

- Dr. Hornbrook said they see a pattern of patients with early symptoms. It is likely that family physicians try many things until a family grows increasingly angry. Then they go to the neurologist. He has seen patients go outside Kaiser to a university and then send Kaiser the bills.

- Dr. Kaye inquired as to how many of the investigators had had an opportunity to examine any of the pharmacy data.

- Dr. Hornbrook stressed that pharmaceuticals can tell them a great deal about what a physician believes their patient to have.

- Mr. Culpepper said that within the VA, there are safeguards. All DMT are non-formulary, but usually require a sign-off. They did not use Mitoxantrone as a criterion for DMT usage because it is used for other things.

- Dr. Kasarskis pointed out that the absence of prescribing something did not mean anything.

- Dr. Kaye wondered if service-connection for MS or ALS was a reliable indicator by itself.

- Dr. Kasarskis thought that there should be greater caution with respect to the service-connection label. It could be granted on exam by a nurse practitioner or a retired physician versus an A-team making this designation. The encounter is somewhat different in the VA in that people are presenting to be certified as service-connected in order to receive benefits. It is a legal encounter rather than a health care encounter. Merely to say someone is service-connected is not enough to determine whether someone is a case. The first criterion of being diagnosed by a neurologist and / or being on DMT are fine, but he did not endorse making a decision based simply on service-connection.

- In thinking about the long-term, Dr. Kaye requested discussion about the grayer area of cases in disease progression and how that might affect whether someone is included in "definite" or "possible."

- Mr. Culpepper replied that one problem with the MS data in the VA is that they do not have this information in the extant databases. They have diagnosis codes, procedure codes, pharmacy use, and service-connected status, but not sub-type. They have to use survey and direct data collection efforts to garner that information. They do not even have date of onset or date of diagnosis to be able to calculate disease duration and use that as a proxy, as poor as it may be.

- Dr. Kaye clarified that she was referring more to not becoming known to the system until a certain stage of illness, meaning that they were missing people early on because they are not identified this way. For the VA data, for example, someone had to be included every year as having MS, but someone may receive a diagnosis and then not have another one for 15 years.

- Mr. Culpepper responded that when they applied the algorithm, it did not matter what year someone was prescribed DMT or what year they may have received a service-connected disability. The investigators wanted to know whether someone had a primary diagnosis for MS once per year. If somebody was diagnosed with MS in 1999, did not have another diagnosis for MS for two years, and then suddenly had a diagnosis and a DMT, they would be retrospectively pulled into the cohort back to 1998 even if that was prior to a definitive diagnosis. This is not known for everyone currently, except for the survey respondents.

- Dr. Teter said that they have found diagnoses of clinically definite MS for people who are in their 50s or 60s who may have had MS for 20-25 years, but perhaps not diagnosed because it was relapsing-remitting. They may have gone to someone for the symptoms or may have ignored them. The point being that no matter what the investigators do, they will be underestimating how many people actually have MS.

- Dr. Tyry noted that neurologists may not correctly document the type of MS. Based on the validation study, it seemed that the patients were better able to report that than the neurologists in some cases. That was one reason they did not use the categories of "primary progressive" and "secondary progressive." Instead, they describe what happens in the disease history in terms of relapse, worsening symptoms, et cetera.

- Dr. Sorenson said the same was true for ALS. Referring to a manuscript in which Dr. Kasarskis was involved, which examined the correlation between neurologists and their El Escorial diagnostic criteria and found that there was only a 60% concordance rate between the neurologists and how they classified people as possible, probable, and definite ALS.

- Dr. Kaye requested that panelists discuss the lack of overlap in a lot of the databases.

- Dr. Benatar indicated that Emory has about 8-9 data sources, which seem largely independent although there is some overlap. There is some overlap found by going into community neurologists' offices, but they are largely independent. If one of the goals of the registry is not only to have accurately diagnosed cases, but also all of the cases, it makes him nervous about what will need to be done to get to all of those cases.

- Dr. Kaye asked whether that argued for a self-identification component with the stipulations that had already been made. That is, someone could self-identify and state whether and by whom they were diagnosed.

- Dr. Benatar thought there was already a marker of self-identification in those who register with ALSA. That is a relatively small number in Georgia. Emory had more people in their ALS clinic, so he worried about whether self-identification would be sufficient.

- Ms. Usher said she was surprised at the numbers who had registered with ALSA, although it was not clear to her why it was so low. There were 189 based on her review of ALSA's records. She thought self-report was fine, but it will potentially miss people.

- Dr. Benatar wondered about the potential for selection bias based on how the investigators choose which charts to abstract, and whether there was something self-fulfilling about where they were headed if they were doing most chart abstractions from tertiary referral centers that were highly specialized in ALS care. Those people identified at Emory may not have the same profile as those people who are diagnosed elsewhere. They will get at that to some extent by moving around different institutions throughout the US, and that the HMO data will address that to a large extent, but it is still troubling.

- Dr. Kaye responded that South Carolina is abstracting nearly 100% regardless of where people are.

- Dr. Benatar noted that South Carolina is having the problem of incomplete data to be able to decide whether someone has ALS. In a tertiary referral center there is a bias in the population, but there is a more reliable diagnosis. In other facilities, perhaps the diagnosis is less reliable. This may be a tension that cannot be resolved.

- Dr. Hornbrook pointed out that the problem with proving the negative is that it is extremely expensive to draw enough charts to prove the null hypothesis that there are no cases. At the same time, they would have the opportunity to examine the negative prediction to look at various patterns of near case criteria: no mention of the right drug, no mention of the right disease, but a symptom complex that would lead one to suspect cases. This might result in a subset of cases which would be most rich for finding missed cases.

- By asking about Rilutek® use, Dr. Kasarskis noted that they had not heard anything about utilization of medical resources and how that changes over time. That is essentially a surrogate marker for the veracity of the ALS diagnosis. When people with ALS are first diagnosed, they usually do not have power wheelchairs, walkers, et cetera. However, as time goes on they do. Bob Miller's group had at least an abstract showing the cost of ALS annually by ALS functional rating scale. That definition included lost wages and lost benefits for the patient and caregiver. While Dr. Kasarskis realized that this was detailed economic data, certainly some of these commonly used assistive devices appearing in a new person could help verify or disprove some of the codes.

- Dr. Kaye said that one thing that is in a holding pattern awaiting finances, CMS is willing to give ATSDR every encounter for the individual once that individual has been identified regardless of whether MS or ALS was mentioned within one of the 10 ICD-9 codes. This will enable them to see patterns of usage.

- Dr. Kasarskis responded that this will not work for MS because the drugs actually work.

- With respect to what information associations have, Mr. Gibson noted that if they were to do this study differently, they would have more of a comprehensive look. From what he gathered from the three ALS studies, they went to the chapter that dealt with the clinical person. Because of government silos, there was not enough time for collaboration to have a full ascertainment of the resources available. Within each chapter there are various silos. For example, there is a silo of advocates from the State of Georgia who are not part of the chapter and do not need resources for whatever reasons.

- Dr. Kaye noted that one issue in particular with ALSA was that they had to go to every chapter; whereas, with the National MS Society, they could go to one location.

- Ms. Kennedy requested clarification of what information they were requesting from ALSA.

- Dr. Kaye responding that they were asking for any information they could get (e.g., the number of people with that diagnosis in a state, names, dates of birth, et cetera) so that they can be matched with the other datasets and the final dataset can be de-duplicated as much as possible, with the understanding that there may still be duplicates.

- Dr. Van Den Eeden added that some people will steal someone's health card, so the data gets in for multiple patients under the same health record number. It takes a lot of detective work. They are now requiring people to show their ID when checking in to see the doctor.

- Mr. Culpepper indicated that they have not done any matching in the VA. Everything is connected by Social Security Number. They have found some errors and some transposition takes place. Surprisingly in that large a system, this does not occur with great frequency. The issue will really come into play when attempting to link across systems.

- Dr. Gunter said that New Mexico is involved in the development of an information exchange across organizations, so there are a number of published algorithms. There are various types of software that can take different constellations of variables and use them for matching. She wondered what South Carolina used to do their linkages.

- Ms. Royer responded that they use a statistical algorithm.

- Dr. Durbin indicated that CDC's Division of Cancer Prevention and Control (DCPC) had done a remarkably good job of matching, and in the last four years they have developed some in-house software that they share freely: The Link Plus product does the probabilistic matching. It is an easy to use stand alone application which can detect duplicates or link files. Although originally designed to be used by cancer registries, it can be used with any type of data. It has a large amount of flexibility in that it will allow weights to be assigned to the algorithms thought to be most appropriate for the specific data and data types. He encouraged everyone to think about what was available that had already been paid for by taxpayer dollars.

- Concern was expressed about the emphasis on European style names. A lot of people in Southern California are Asian. Asian women customarily change their names upon getting married, although they will sometimes follow the American custom. Sometimes the Chinese characters that make up an Asian name will be put together as two syllables. Sometimes they will be a first name and a middle name. Frequently, the Latinos will use the maternity along with paternity. This is very complicated.

- Dr. Teter indicated that the way their algorithm is set up, they can do matches up to five or six points, but then the last name does not match. This requires further discussion. Their analysts suggested adding more levels because it has more quality if it contains the last name. Someone who scores a three, four, or five regardless of whether the last name is included, it has to be more definite. Scoring one point each on sex, state, city, and date of birth will result in a score of 4 but then the last name does not match up. With a group of 5,000 people with a score of 4, New York's argument was that they need more layers in order to ascertain whether a 4 is a match with the last name.

- Dr. Tyry suggested trying the program to which Dr. Durbin referred and weight the last name more heavily rather than creating more levels.

- Dr. Kaye clarified that the score was supposed to be done after the link. The idea was that a human matched them up and the score was assigned afterward. A human should not be assigning people with different last names as a match to start with unless they had data to show that one was a maiden name.

- Dr. Teter indicated that their system is automated. It reads everything, matches according to the algorithm, and assigns a score to it. At this time while preliminary testing on matching is being done by last name, she is not sure whether the matches are correct.

- Dr. Kaye pointed out that New York was using this differently than it was intended. That is, it was to determine how good the match was after it was made, versus using it to make the match.

- Dr. Tyry indicated that NARCOMS did the first name, last name, and date of birth one at a time. Once those three are down, everything else seems to fall into place. There are very few other combinations that will result in additional sure matches.

- Dr. Teter said that New York had thought about starting with last name, but MS being a disease that affects people primarily in their 20s, there is a high likelihood of women getting married. Thus, matching on last name to begin with would miss a population.

- Dr. Gibson inquired as to what the plans were to do the updates of the ATSDR studies that address ALS and MS.

- Dr. Kaye responded that the written study information has been updated with all of the data for both the MS and ALS projects. A new fact sheet was needed, which she spent a lot of time updating, but then she and Dr. Muravov were told that the correct format had not been used. They then spent an inordinate amount of time converting it into a different format with a lot less information in it. They will get it cleared through the system so that it the updated fact sheet can be posted.

Closing Remarks

Wendy E. Kaye, PhD
Senior Epidemiologist
McKing Consulting Corporation

In conclusion, Dr. Kaye indicated that this group would likely not meet again until there were data in from the projects. The plan was for ATSDR to conduct some analyses of the combined datasets, with some analyses done on the individual site levels. Additional experts in the areas of MS and ALS will be brought in to discuss the issues, with perhaps more concrete information pertaining to whether they should proceed on the same path or if changes would need to be made. Those will be the suggestions that ATSDR is seeking in order to move forward with an ALS registry and MS surveillance. Based on when ATSDR expects to receive data from all of the pilot projects, most likely the next meeting will be around May 2009. In order for ATSDR to plan around other annual meetings / conferences, she requested that panelists email her with specific dates of ALS or MS conferences that people might be attending. Dr. Kaye thanked everyone for their attendance and all of their work on this project. With no further business posed, she officially adjourned the meeting.

Participants

Susan Andrade, ScD
Meyers Primary Care Institute
Univ. of Massachusetts Medical School
Worcester,

Khemissa Bejaoui, PhD
Research Scientist
Marshfield Clinic Research Foundation
Marshfield, WI

Michael Benatar, MD
Neurologist
Emory Department of Neurology
Atlanta, GA

Lucie Bruijn, PhD
Science Director and Vice President
Amyotrophic Lateral Sclerosis Association
Calabasas Hills, CA

Kimberly Cantor
Manager , Federal Government Relations
National Multiple Sclerosis Society
Washington, DC

William Culpepper II, MA
Baltimore Veterans Affairs Medical Center
Baltimore, MD

Valerie Cwik, MD
Medical Director
Muscular Dystrophy Association
Tucson, AZ

Randy Durbin, MSPH, DO,
Senior Statistician
McKing Consulting Corporation
Atlanta, GA

Stevan Gibson
Vice President,
Government Relations & Public Affairs
Amyotrophic Lateral Sclerosis Association
Washington, DC

Maggie Gunter, PhD
President/ Executive Director
Lovelace Clinic Foundation
Albuquerque, NM

Carolyn Harris
Division of Health Studies
Agency for Toxic Substances and Disease Registry
Atlanta, GA

Stuart N. Hoffman, DO
Department of Neurology
Geisinger Medical Center
Danville, PA

Mark C. Hornbrook, PhD
Chief Scientist, The Center for Health Research
Northwest/Hawai'i/Southeast
Kaiser Permanente Northwest
Portland, OR

Edward Kasarskis, MD, PhD
Co-Principal Investigator, VA Neurology Service
VA ALS Registry
Lexington, KY

Wendy Kaye, PhD
Senior Epidemiologist
McKing Consulting Corporation
Atlanta, GA

Annie Kennedy
Vice President – Advocacy
Muscular Dystrophy Association
National Advocacy Office
Washington, DC

Nicolas LaRocca, PhD
Vice President, Health Care Delivery and Policy Research
National Multiple Sclerosis Society
New York, NY

Jane Larkindale
Muscular Dystrophy Association
Tucson, AZ

Jay Mandrekar, PhD
Statistician
Mayo Clinic
Rochester, MN

Sharon Matland, RN, MBA
Vice President, Patient Services
Amyotrophic Lateral Sclerosis Association
Calabasas Hills, CA

Oleg Muravov, MD, PhD
Medical Epidemiologist
Principle Investigator ALS and MS Pilot Projects
Agency for Toxic Substances and Disease Registry
Atlanta, GA

Lorene Nelson, PhD
Associate Profession & Chief
Division of Epidemiology
Department of Health Research & Policy
Standford University School of Medicine
Stanford, CA

Daniel S. Newman, MD
Director, Harry J. Hoenselaar ALS Clinic
Henry Ford Hospital
Detroit, MI

Rebecca Pentz, PhD
Professor of Research Ethics
Emory Winship Cancer Institute
Emory University
Atlanta, GA

Julie Royer, MSPH
Statistician, Office of Research and Statistics
South Carolina Budget and Control Board
Columbia, SC

Judy Smith
Agency for Toxic Substances and Disease Registry
Atlanta, GA

Eric Sorenson, MD
Neurologist
Mayo Clinic
Rochester, MN

Anne Sowell, PhD
Agency for Toxic Substances and Disease Registry
 Atlanta, GA

David Stickler, MD
Assistant Professor of Neurology
Director, ALS Clinic
Medical University of South Carolina
Charleston, SC

Daniel Strickland, PhD
Epidemiologist, KaiserResearch & Evaluation
Kaiser Permanente Southern California
Pasadena, CA,

Barbara Teter, PhD, MPH
Director, Research and Development
Jacobs Neurological Institute
NYS Multiple Sclerosis Consortium
Buffalo, NY

David Thurman, MD
Centers for Disease Control and Prevention
Atlanta, GA

Tuula Tyry, PhD, MAED
Project Manager, NARCOMS
Barrow Neurological Institute
St. Joseph's Hospital and Medical Center
Phoenix, AZ

Sharon Usher, RN
Research Nurse, Emory Department of Neurology
Atlanta, GA

Stephen Van Den Eeden, PhD
Senior Epidemiologist, Division of Research
Kaiser Permanente Northern California
Oakland, CA

Priscilla Velentgas, PhD
Department of Ambulatory Care & Prevention
Harvard Pilgrim Health Care
Harvard Medical School
Boston, MA

Patrick Wildman
Director, Communications and Public Policy
Amyotrophic Lateral Sclerosis Association
Washington, DC

G. David Willamson, PhD
Director, Division of Health Studies
Agency for Toxic Substances and Disease Registry
Atlanta, GA

Dhelia Williamson, PhD
Centers for Disease Control and Prvention
Atlanta, GA

Jenny Wu, MS
Agency for Toxic Substances and Disease Registry
Atlanta, GA