The National Alzheimer’s Coordinating Center (NACC) Database: The Uniform Data Set

Duane L. Beekly, BS,* Erin M. Ramos, PhD, MPH,* William W. Lee, PhD,* Woodrow D. Deitrich, BS,* Mary E. Jacka,* Joylee Wu, MS,* Janene L. Hubbard, BS,* Thomas D. Koepsell, MD, MPH,* John C. Morris, MD,† and Walter A. Kukull, PhD*;
The NIA Alzheimer’s Disease Centers

Abstract: The National Alzheimer’s Coordinating Center (NACC) is responsible for developing and maintaining a database of participant information collected from the 29 Alzheimer’s Disease Centers (ADCs) funded by the National Institute on Aging (NIA). The NIA appointed the ADC Clinical Task Force to determine and define an expanded, standardized clinical data set, called the Uniform Data Set (UDS). The goal of the UDS is to provide ADC researchers a standard set of assessment procedures, collected longitudinally, to better characterize ADC participants with mild Alzheimer disease and mild cognitive impairment in comparison with nondemented controls. NACC implemented the UDS (September 2005) by developing data collection forms for initial and follow-up visits based on Clinical Task Force definitions, a relational database, and a data submission system accessible by all ADCs. The NIA requires ADCs to submit UDS data to NACC for all their Clinical Core participants. Thus, the NACC web site (https://www.alz.washington.edu) was enhanced to provide efficient and secure access data submission and retrieval systems.

Key Words: Alzheimer disease, database, dementia, NACC, UDS, data submission

*Alzheimer Dis Assoc Disord 2007;21:249–258*

The first repository for Alzheimer’s Disease Center (ADC)-wide data collection was begun by Dr Dennis Evans at the Rush-Presbyterian-St Luke’s Medical Center (Alzheimer’s Disease Center Core) in 1997. This data collection effort, known as the Minimum Data Set (MDS), was limited to a brief set of data elements that were retrospectively collected on enrolled participants accumulated since each ADC began and served as an inventory of participant information.1 This MDS and the responsibility for managing it were transferred to National Alzheimer’s Coordinating Center (NACC) in 1999. NACC expanded this data set to include additional demographic, clinical, and specimen data elements.2 A neuropathology data set (NP) containing standardized neuropathology data and Center-related data sets, such as descriptions of neuropsychologic tests administered and neuropathology protocols used at each ADC were also incorporated into the NACC database.

The NACC database, including the MDS, NP, and Centers data sets, has been operational since 2000. As the MDS contained cross-sectional data of limited scope, it was determined that a more comprehensive database had to be developed to support a broader range of clinical research. In June 2002, the Clinical Task Force (CTF) was formed. The CTF, chaired by John C. Morris, consists of the 5 ADC-elected Clinical Core Steering Committee members (John C. Morris, Charles DeCarli, Norman Foster, Neill Graff-Radford, and Elaine Peskind) and 5 additional members appointed by the National Institute on Aging (NIA) (Helena Chui, Jeffrey Cummings, Steven Ferris, Douglas Galasko, and Sandra Weintraub). Their first charge was to expand the original MDS into a data set which would include more detailed and standardized clinical, diagnostic, and cognitive data on all participants.3 NACC’s tasks were to group the data elements selected by the CTF and create the Uniform Data Set (UDS) forms, design a database to store these forms, create a data submission system allowing the Centers to submit UDS data on a continuous basis, and create a data access system to allow researchers to review and analyze the data.

METHODS

Software and Hardware

NACC built the UDS database using the commercial database, Oracle.4 This is a relational database that allows storage of many types of data including numeric, text, and images. SAS software5 is used for the interface from the web site to the database, and many programs...
were written in SAS to create the UDS data submission and access systems. The database is stored and maintained on a Sun Sparc Sunfire 880 (Sun Microsystems, Inc, Santa Clara, CA) using the Solaris 9 operating System. The database can be accessed via secure Internet connection using a web browser.

NACC Web Site

The NACC web site (https://www.alz.washington.edu) is designed to serve ADC investigators, NACC personnel, NIA officials, and the general public. The web site is used by the data managers at each Center to submit, error check, and finalize data. Investigators use the web site to access data and retrieve data files or to identify resources available at ADCs for other research. The NIA uses the web site to access reports and to follow ADC enrollment and the data collection progress. The public may also use the web site to obtain links to caregiver information and to obtain limited access to the NACC database. Specific sections of the web site are password protected, and available only to specific classes of users; for example, NIA personnel may have a different level of access than an ADC data manager. User accounts and passwords can be obtained from NACC with the appropriate authorizations. All data sent to or from the NACC web site are encrypted with at least a 128-bit encryption suite. NACC has purchased a certificate from Verisign (VeriSign, Inc, Mountain View, CA), one of the best-known certificate sites.

UDS Form Design

NACC designed the UDS forms to be straightforward to use for both the clinicians and data managers. Forms are composed of individual data elements and grouped into packets that are administered using standardized instructions at a particular visit. Currently, there are 4 packet types: initial, follow-up, milestone, and status. Each packet comprises 1 or more forms. For example, the initial packet contains forms A1 to A5 (demographic and administrative data), B1 to B9 (clinical evaluation data), C1 (neuropsychologic data), D1 (diagnostic data), and E1 (imaging and specimen data inventory). Each form represents a group of data elements that are administered together. For example, form A1 is the demographic form that contains data elements such as sex, race, ethnicity, and marital status and form B1 is physical evaluation form that includes height, weight, and blood pressure.

In the future, additional packets will be developed to better capture features of other non-AD dementias including frontotemporal dementia, Lewy body disease, and vascular dementia. The database design enables the addition of more packets that NACC will develop as they are authorized and defined by the NIA, the CTF, and the ADC Directors.

Database Design

The database was designed to accomplish the following aims:
A. To separate the “clean” data packets from the packets with errors.
B. To allow submission of multiple types of packets including follow-up packets.
C. To provide an electronic data element dictionary (DED).
D. To incorporate and facilitate electronic error checking of data.
E. To integrate the new UDS with prior NACC databases.
F. To facilitate data quality control and data monitoring.
G. To implement data form version changes seamlessly through “version control” software.
H. To provide efficient and secure access to data.

To achieve these aims, NACC designed a database that is divided into 3 sections (Fig. 1). These sections are labeled as (1) Working database, (2) Current database,
and (3) Frozen databases. Forms are first uploaded to the Working database, where the data can be easily examined and error checked by both ADC and NACC personnel. Packets that have all required forms and have passed all required data quality checks are copied to the Current database. Approximately every 6 months, data in the Current database are frozen and certified. This process generates a Frozen database whose content is stable and suitable for research use. Frozen databases are kept in perpetuity so that published results may be recreated. This 3-tiered system allows for efficient data submission and integrated quality assurance (QA).

Data are stored in the relational database as Oracle tables. Each form comprises 1 table. Each table is related to the other tables by Center ID, Patient ID, and visit number. The Working database, Current database, and Frozen databases all have their own set of tables for the forms. Thus, form A1 table is named frmwa1, frmca1, and f1fa1 for the Working, Current, and Frozen database, respectively. In addition to the data tables, the UDS database was designed to have a number of “index” tables. These index tables make error checking, finalizing data, and accessing data easier and quicker. The index tables include:

(1) A table that specifies all patient IDs in the database (1 row per patient ID).
(2) A table that specifies all patient IDs and visits for the patient ID in the database (1 row per ID per visit).
(3) A table that specifies which forms comprise each packet.

The new UDS database has been integrated with the entire NACC database (Fig. 2). The MDS, NP, and UDS all use identical codes for ADC identification and for patient identification. Therefore, the data may be accessed separately for each set of data or seamlessly from all the combined data. NACC will continue to collect neuropathologic examination data on all autopsied participants whether they were in the UDS or MDS at their last clinical visit. An MDS Status Form has been developed and implemented which allows the Centers to explain what happened to IDs in the MDS that are considered lost to follow-up, declined further participation, or have been discontinued by the ADC before their enrollment in the UDS.

Version Control

As the forms and data elements collected can change over time, version control has been built into the NACC Database. Each form is submitted with a version data element. This version data element is used by the submission software to determine which data elements to read in and which error checks to execute. The electronic DED also has versions associated with each data element, as does the electronic error checking tables. These are described below.

Electronic DED

An electronic DED was developed that describes and defines all data elements. The electronic DED enables the addition of new data elements and modification of existing data elements. In the NACC database, there is one DED Oracle table for each type of form. For example, there is an Oracle table for UDS form A1 and another table for form B1. Within each table, there is 1 row for each data element, packet, and version. For example, a row may describe the data element “race,”

---

FIGURE 2. A conceptualization of the design of the relational NACC database. Sections of the overall database include the Minimum Data, the Neuropathology Data, the Centers Data, the Collaborative Data, and the Other AD Data, and can be accessed individually or in combination with other data sections.
packet I, version 1. Another row may describe data element race, packet F, version 1. The version may change so there could be a row describing data element race, packet I, version 2. Each row uniquely describes a data element for a specific version and packet. Some of the descriptions include the allowed numeric range of the data element, title to display on the screen for web data entry, and when the data element may be left blank. Organizing the database content in tables allows for ease of error checking and changing titles. Data elements may be added with new versions without impacting the system. All interfaces to the electronic DED are through the NACC web site NACC data managers may access and update all of the parameters for a chosen data element by just entering changes into the web page. This simplifies updating data elements.

**Electronic Error Checking**

Error checking is automated through the use of 2 error checking tables and these tables may be accessed through the NACC web site. A general table uses an error number to define errors. An example of a general error is: Item A = 3 but Item B is blank. There are currently 85 errors. The second table defines specific errors and is related with the first error table by the error number (or ID). For example, a specific error may be “RACE” = 3 (Hispanic) and “Hispanic Origin” is blank. Errors are listed by packet, form, and given a descriptive error number. Both general errors and specific errors may be added, modified, or deleted by NACC data personnel through the error checking control web page found on the NACC web site.

The SAS error checking programs read the error checking tables and apply them to the data a Center has submitted. A list of data elements with errors is generated and displayed on the web site. Also a printable list is created. The Center data manager can then fix the errors through web data entry or uploading another file.

**Data Submission System**

The data submission system was designed so that the ADCs could easily submit their data to NACC. Data must be submitted within 2 weeks of completion of a packet. Therefore, a system was developed to allow uploading of data or web data entry at any time. A user-friendly interface (Fig. 3) was developed that can be accessed from the NACC web site. ADC personnel can only upload, error check, and edit their own Center’s data. Data may be submitted to NACC by uploading files or by a web data entry system.

**File Upload**

Programs to upload files were written so that a Center data manager can transmit a file to NACC through the NACC web site. A file created locally can thus be uploaded to NACC for processing by the ADC’s data manager (Fig. 4). Types of data files that may be uploaded include fixed ASCII format, comma delimited, tab delimited, and SAS version 7, 8, or 9 files. Once uploaded, the file is preprocessed to ensure that the format is correct and that all required data elements have been included in the file. If the file passes the preprocessing checks, it is loaded into the Working database.

**Web Data Entry**

Data can also be entered and modified through the NACC web site. Data can only be modified or added directly into the Working database. Once data modification is complete, a patient ID and visit can be finalized into the Current database and thus the Current database is changed. Patient IDs can be added or deleted. Once a patient ID is added, visits for a patient may be added and then forms for visits may be added (Fig. 5). After a form has been added, new data can be entered or existing data can be modified. The NACC web data entry system allows Centers to enter and update their data through a condensed data interface or electronic forms. Data may be double entered to assure correctness in data entry.

**Error Checking**

Once the data are in the Working database, the data may be error checked (Fig. 6). If errors occur, they must be corrected either through the web data entry system or by uploading a revised file. In addition alerts occur as well. Alerts identify improbable (but possibly correct)

---

**FIGURE 3.** The NACC UDS submission system home page.
Data values that should be checked and confirmed before being considered final. These alerts may either be verified (ie, data are confirmed as correct) or corrected using the methods described above. Two types of error checking must be completed before a packet may be finalized: within form error checks that evaluate data elements within a form for errors, and cross form checks that evaluate data consistency across forms. All error checks are controlled through the electronic error checking system.

Finalizing Data

Once a packet is complete with all required forms present, no errors exist, and all alerts have been verified, the packet may be finalized into the Current database. The “finalize” function performs a final check on all packets. If a packet passes the check, it is copied to the Current database (Fig. 7). Packets within the Current database are considered to be error free and are archived in the Frozen database at the next freeze date.

Reports

Numerous reports are available to the Center submitting data. These include precertification reports, past certification reports, and Center reports. The precertification reports are exact replicas of the certification reports sent to the Center’s Director during the semiannual data freeze. These reports allow the data manager at the Center to see the progress made in submitting and finalizing data to the Current database. The past certification reports allow the Center to recreate past data freezes. Center reports are useful for quality control for the Center. Various reports are available that detail the percentage of missing data for specified forms and other QA concerns. A monthly report is sent to all Centers and the NIA detailing progress in the amount and quality of data collected in NACC’s UDS database. Among the purposes of the report are to encourage a steady flow of data to NACC, and to call attention to any problems in the levels of missing data.

Downloading Data

Data from the Working or Current database may be downloaded to a local PC in a variety of formats including Microsoft Excel, ASCII Fixed Field, and SAS version 7, 8, and 9 files. A Center may download only its own data. This allows Centers to enter data through the NACC submission system and then download and load data into their local databases.

Printing Forms

The “print form” option is a particularly useful feature of the data submission system and enables a Center to print a copy of the UDS data forms for a particular participant, regardless of whether the data were entered through the Web Data Entry System or was uploaded in a batch file. The forms are nearly identical to the actual paper UDS forms and can be used to help complete the data forms at subsequent follow-up visits. A patient ID and visit can be selected along with the form(s) to be printed. The forms are displayed on the screen as a web page and the local computer’s print command can be used to print the forms.
Certification

Packets must be certified before they are permanently stored in a Frozen database. All data in the Current database are frozen semiannually. Once frozen, certification reports are generated and sent to the Center Directors. If a Director agrees with the reports, she/he signs the report and returns it to NACC. The data remain frozen in the Frozen database, and the Current database is opened for more data entry. If the Director disagrees with the report, the Center has 2 weeks to correct and certify the data. If not certified, the data are removed from the Frozen database before this particular frozen database is made final.

QA

An internal NACC QA committee was appointed to monitor the quality of UDS data being submitted and each Center appointed a QA officer. In addition to specifying errors and alerts to be applied at data submission time, QA reports were created. These QA reports are used to: (1) monitor data quality across Centers, (2) detect unusual data patterns, and (3) create a monthly report for the NIA, CTF, and Center Directors.

In addition, NACC periodically requests each Center to send a copy of the paper forms for randomly selected patient IDs and visits. Each Center receives an express mail package from NACC and has 3 days to gather the requested data packets and return them to NACC. Once at NACC, members of the QA committee review the packets to be sure the UDS forms have been used and that data have not been transcribed from other Center forms. Report cards are created for each Center and sent to the Director, the NIA, and the CTF.

An subset of the packets received from the Centers are entered into a designated QA section of the NACC database. The data are compared with the corresponding data previously entered into the database by the Centers. An error rate is calculated per Center and per form. This helps determine whether Centers are having difficulties completing the forms.

Electronic Forms System

Some ADC clinicians prefer entering participant data directly into electronic data forms during the participant interview. Therefore, NACC developed an electronic forms system (E-Forms) so that clinicians could by-pass paper forms and enter data through the NACC web site directly. Forms displayed on web pages closely resemble the paper forms. An example of the electronic data form A1 (subject demographics) is presented in Figure 8 (owing to space constraints, only the top half of the form is shown). Data are processed in the same way as batch file uploads or the web data entry system. The electronic forms system was designed for clinicians to enter E-Forms without all the other functions needed by data managers. Functions such as downloading data and finalizing data forms have been omitted. Data managers can access these functions after data forms are completed by the clinicians.

Data Access System

NACC has developed a data access system so that UDS data may be accessed through the NACC web site. This system is divided into member and nonmember access. Personnel at the Centers, NACC, the NIA, and persons sponsored by a Center may become a member and have access to a complete UDS data set that contains...
combined data from all ADCs, not just from 1 Center. Nonmembers have limited access. A data access request form may be submitted through the web site for both members and nonmembers. This mechanism allows a user to describe the data requested which is then evaluated by NACC and the NACC Publication Committee. The NACC Publications Committee is comprised of 5 elected members who represent 5 different ADCs and was established to review concepts for publications based on data in the NACC database. If the request is approved, a data set or report is sent to the user. This entire process, from submission of the data request to generation of an appropriate data file takes approximately 2 weeks. However, if the user only needs a brief summary of available data, a variety of standardized reports are available and a generalized frequency report generator that allows the immediate creation of 1 way or 2 way tables from most UDS data elements (Fig. 9) and also combined UDS, MDS, and NP data elements.

**FIGURE 8.** An example of the electronic data form A1 (subject demographics) is presented in Figure 8 (owing to space constraints, only the top half of the form is shown).
The Uniform Data Set

Create 2-Way Tables for Initial Packet UDS Data Elements

Using Packets from March 1, 2007 Data Freeze

Documentation/Data Backup

Documentation is imperative for the operation of all databases and data submission systems. NACC has developed an extensive and complete set of user and system documentation. User documentation includes materials such as the data submission manual, DED, data templates, coding manuals, and answers to frequently asked questions. System documentation includes the database systems and operational procedures manuals. This documentation is available on the NACC web site (Fig. 10). Documentation is updated continuously and users are notified by e-mail when there has been an update.

All data are backed up to tape every night. A tape is carried off-site to protect against natural disasters. Complete monthly backups are performed and the tapes are kept perpetually. The Oracle database also has a rollback feature that allows for recovery of data in the case of loss.
Training/Feedback

Each section of the data submission system (upload files, web data entry, etc) has a customized help screen. If the online help is insufficient, individuals are encouraged to call a NACC phone line, which is continually monitored during business hours. Furthermore, NACC maintains a special e-mail account where questions and comments can be sent. This e-mail account is monitored daily and NACC staff attempts to respond to all questions within 48 hours. A data manager’s bulletin board is available on the NACC web site. Here, data managers can post questions, respond to comments from other data managers, or request specific feedback from NACC. In addition to the online help and NACC phone line, NACC personnel give help and training-related presentations at semiannual Center-wide meetings.

RESULTS

In the first year of operation of the UDS, more than 3400 initial packets have been submitted to NACC and certified. Table 1 stratifies these packets by race and sex. These packets represent standardized data collected by the Centers for a variety of participants. Table 2 shows these participants stratified by clinical diagnosis. The first data freeze was executed on July 1, 2006, and all Centers certified their data. These data are now ready to be analyzed. Numerous research ideas have been suggested and data analyses are underway. NACC encourages researchers across the country to access and use the NACC database. A data request system is available and investigators can request data sets in the file formats most convenient for them and their statistical software packages.

DISCUSSION

A major expansion of a standardized, longitudinal database system for clinical research on dementia has been completed and implemented. The ADCs are routinely using the UDS system to collect and submit data. NACC is continuing with its efforts to integrate the UDS with the MDS and NP. All of these data sets will be placed in Oracle for transparent access and use.

In the future, additional UDS packets will be created. A Spanish language packet and a follow-up telephone administration packet are under development and frontal temporal dementia and Lewy body disease packets are being considered. The UDS has generated many research ideas from the ADCs and from within NACC itself. The NACC database will be an important resource for ADC investigators for both cross-sectional and longitudinal research questions.

The NIA Alzheimer’s Disease Centers members are: Arizona ADC: Eric M. Reiman, MD, Banner Health Institute, Phoenix, AZ; Boston University: Neil Kowall, MD, Bedford VA Medical Center GRECC Program, Bedford, MA; Case Western Reserve University: Gary Landreth, PhD, University Hospitals of Cleveland, Cleveland, OH; Columbia University: Michael Shelanski, MD, PhD, Columbia University, New York, NY; Duke University Medical Center: Kathleen Welsh-Bohmer, PhD, Bryan ADRC, Durham, NC; Emory University: Allan I. Levey, MD, PhD, Emory University, Atlanta, GA; Florida ADC: Huntington Potter, PhD, Byrd Alzheimer’s Institute, Tampa, FL; Indiana University: Bernardino Ghetti, MD, Indiana University School of Medicine, Indianapolis, IN; Johns Hopkins University: Donald Price, MD, Johns Hopkins University School of Medicine, Baltimore, MD; Massachusetts General Hospital: Bradley Hyman, MD, PhD, Massachusetts General Hospital, Charlestown, MA; Mayo Clinic: Ronald C. Petersen, PhD, MD, Mayo Clinic, Rochester, MN; Mount Sinai School of Medicine: Mary Sano, PhD, Mount Sinai School of Medicine, New York, NY; New York University: Steven H. Ferris, PhD, Silberstein Aging and Dementia Research Center, New York, NY; Northwestern University Medical School, Chicago, IL; Oregon Health and Science University: Jeffrey Kaye, MD, Aging and Alzheimer Disease Center, Portland, OR; Rush University Medical Center: David A. Bennett, MD, Rush University Medical Center, Chicago, IL; Stanford University: Jerome Yesavage, MD, Stanford University, Palo Alto, CA; University of Alabama at Birmingham: Daniel Marson, JD, PhD, Sparks Research Center, Birmingham, AL; University of Arkansas for Medical Sciences: Cornelia Beck, RN, PhD, University of Arkansas for Medical Science, Little Rock, AR; University of California, Davis: Charles DeCarli, MD, University of California, Davis Medical Center, Sacramento, CA; University of California, Irvine: Carl Cotman, PhD, University of California, Irvine, CA; University of California Los Angeles: Jeffrey L. Cummings, MD, University of California Los Angeles, Los Angeles, CA; University of California, San Diego: Leon J. Thal, MD, University of California, San Diego School of Medicine, La Jolla, CA; University of Kentucky: William Markesbery, MD, University of

<table>
<thead>
<tr>
<th>Race</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>1326</td>
<td>1569</td>
<td>2895</td>
</tr>
<tr>
<td>Black/African American</td>
<td>93</td>
<td>273</td>
<td>366</td>
</tr>
<tr>
<td>American Indian/Alaska Native</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Native Hawaiian/Pacific Islander</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Asian</td>
<td>16</td>
<td>25</td>
<td>41</td>
</tr>
<tr>
<td>Other</td>
<td>27</td>
<td>59</td>
<td>86</td>
</tr>
<tr>
<td>Unknown</td>
<td>7</td>
<td>9</td>
<td>16</td>
</tr>
<tr>
<td>Total</td>
<td>1470</td>
<td>1940</td>
<td>3410</td>
</tr>
</tbody>
</table>

TABLE 2. Primary Diagnoses for the 3410 UDS Participants Stored and Maintained in the NACC Database

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal cognition</td>
<td>1322  (38.8)</td>
</tr>
<tr>
<td>Demented</td>
<td>1355  (39.7)</td>
</tr>
<tr>
<td>Mild cognitive impairment</td>
<td>618   (18.1)</td>
</tr>
<tr>
<td>Impaired, not mild cognitive impairment</td>
<td>115   (3.4)</td>
</tr>
</tbody>
</table>

© 2007 Lippincott Williams & Wilkins

257
Kentucky, Lexington, KY; University of Michigan: Sid Gilman, MD, FRCP, University of Michigan, Ann Arbor, MI; University of Pennsylvania: John Q. Trojanowski, MD, PhD, University of Pennsylvania, Philadelphia, PA; University of Pittsburgh: Steven T. DeKosky, MD, University of Pittsburgh, Pittsburgh, PA; University of Southern California: Helena Chui, MD, University of Southern California, Los Angeles, CA; University of Texas Southwestern: Roger Rosenberg, MD, University of Texas SW Medical Center, Dallas, TX; University of Washington: Murray Raskind, MD, VA Puget Sound Health Care System Mental Health Services, Seattle, WA; Washington University: John C. Morris, MD, Washington University School of Medicine, St Louis, MO.

REFERENCES