The Comprehensive Dialysis Study (CDS):
A USRDS Special Study

Nancy G. Kutner,* Kirsten L. Johansen,†‡ George A. Kaysen,§ Sarah Pederson,‖
Shu-Cheng Chen,‖ Lawrence Y. Agodoa,** Paul W. Eggers,** and Glenn M. Chertow†¶

*Rehabilitation/Quality of Life Special Studies Center and †Nutrition Special Studies Center, United States Renal Data System (USRDS), Emory University, Atlanta, Georgia; ‡University of California San Francisco, San Francisco, California; §University of California Davis, Davis, California; ¶Stanford University, Palo Alto, California; ‖Coordinating Center, USRDS, University of Minnesota, Minneapolis, Minnesota; **National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland

Background and objectives: The Comprehensive Dialysis Study (CDS) aimed to understand factors contributing to physical, functional, and nutritional health status among patients starting dialysis.

Design, setting, participants, & measurements: A phone interview survey was conducted with patients from a geographically stratified national random sample of dialysis units, and quarterly serum samples were obtained for patients at a preidentified subset of units. The phone survey collected standardized measures of physical activity, employment and disability status, perceived health and well-being, and dietary intake. Serum samples were obtained to measure prealbumin, albumin, creatinine, normalized protein catabolic rate, and C-reactive protein. To comply with restrictions required under the Health Insurance Portability and Accountability Act (HIPAA), dialysis unit personnel could not participate in any research-related activities.

Results: Overall participation rate was 18.5%. One thousand six hundred forty-six patients affiliated with 295 dialysis units completed the phone survey; 361 patients affiliated with 68 dialysis units also completed a dietary intake survey, with 269 providing serum samples. Despite challenges in the design and implementation of CDS, the population was diverse and results should be generalizable.

Conclusions: Constraints within the dialysis industry and HIPAA requirements render the assembly of nationally representative cohorts extremely difficult. Nevertheless, the CDS represents the largest cohort of incident dialysis patients containing detailed information on self-reported physical activity and dietary intake and is one of few cohorts simultaneously measuring laboratory proxies of nutrition and inflammatory status. Data from CDS can be used to inform the design of interventions addressing several conditions that affect longevity and health status in ESRD.

units throughout the United States. The CMS identified patients who initiated maintenance dialysis from September 2005 to June 2007 at selected units, and the USRDS CC attempted to contact these patients to enlist their participation in the study.

The CDS design called for all study participants to provide baseline responses to a patient questionnaire (PQ) in a phone interview. Patients from a preidentified subset of dialysis units were also asked to provide data for a nutrition substudy by completing the Block 2000-Brief Food Questionnaire (FQ) (1), and quarterly serum samples were requested from dialysis units for these patients. Patients in the substudy were also recontacted and asked to complete both the PQ and FQ at 12 mo after baseline data collection. The original design called for recontacting all CDS participants at 12 mo after baseline, but resource constraints necessitated limiting the longitudinal component of the study to the nutrition substudy participants.

Institutional review boards (IRBs) at the Nutrition SSC (UCSF and UC Davis), the Rehabilitation/QOL SSC (Emory University), and the USRDS CC (University of Minnesota) approved the study. Informed consent was obtained. Patient anonymity is ensured at the USRDS CC by assigning a universal USRDS identifier to all data obtained for specific patients.

Dialysis Unit Sampling

The sampling frame of dialysis units was based on the April 7, 2005 version of the Dialysis Facility Compare database. Data on number of annual incident maintenance dialysis patients for each dialysis unit were obtained from the most recently available (2003) CMS ESRD Facility Survey. The two databases were merged to provide a size measure (annual incident dialysis patients) for each dialysis unit. The size measure allowed projection of incident patients in the sample and probability proportional to estimated size sampling for some units. A size measure based on the number of dialysis stations was computed for 315 units, 75% of which were certified after 2003 and therefore not included in the 2003 Facility Survey. The following types of facilities were eliminated from the sampling frame: Veterans Administration, children-only, outside the 50 states and District of Columbia, and transplant-only. The final sampling frame contained 4410 facilities.

The sampling frame was sorted by ESRD network and then by adjacent states within the network. Furthermore, the sampling frame was sorted within state in a serpentine manner by the size measure. A sample of 335 dialysis units was selected using equal probability systematic random sampling. Using systematic random sampling in conjunction with the sorted facility sampling frame yielded implicit geographical stratification (network and state within network) for the sample facilities.

The sample of 335 facilities was compared with the population of 4410 facilities on the size measure, number of dialysis stations, facility type (based on digits 3 and 4 of the 6-digit provider ID), chain or not, offer hemodialysis or not, offer peritoneal dialysis or not, and network. The facility sample matched the population closely on all of these characteristics.

Because it was not feasible to collect serum samples from patients affiliated with all of the 335 sample dialysis units, a subsample of 73 units was selected for the nutrition substudy. Systematic probability proportional to estimated size (using the size measure above) sampling was used to maximize the number of patients in the units selected for the nutrition substudy. The sort variable within state was changed from the size measure to zip code, because the size measure was used for the unequal probability sampling. Thus, both the dialysis unit sample identified for the nutrition substudy (n = 73) and the remaining units selected for the CDS (n = 262) were implicitly stratified by geography.

The USRDS CC notified the medical directors and administrators of dialysis units that were randomly selected for the study. The Nutrition SSC contacted units selected for the nutrition substudy to review procedures for lab draws. Of the 335 selected dialysis units, patients from 295 (88%) agreed to participate (Figure 1).

Patient Identification and Enrollment

Adult patients starting dialysis at CDS-selected facilities were identified by the CMS Standard Information Management System, and patient lists were shared with the USRDS CC on a monthly basis. From October 2005 through February 22, 2006, the USRDS CC made initial contact with potential participants by mailing to the patient’s dialysis facility (1) information addressed to the patient describing the study, (2) a consent form, and (3) a participation availability form. This mailing was sent to the patient’s home for those on home dialysis. Patients who completed and returned a signed consent form then received a phone call from DataBanque Research Services (Pittsburgh, PA), a data management service organization contracted by the CDS, to schedule a phone interview.

From February 23, 2006 through August 2007, IRB approval was sequentially obtained for several modifications to the patient contact protocol in an effort to enhance response rates. These modifications included: (1) mailings directly to the patient’s home address for all potential participants (2) follow-up phone calls by DataBanque to patients who had not returned a consent form to further explain the study, (3) a participation incentive of $25, and (4) the opportunity to complete a verbal consent at the time of the DataBanque phone call. Patients selected for the nutrition substudy continued to provide written consent to document their permission for provision of quarterly serum specimens.

Data Collection

Patient-Reported Data

The contents of the PQ are summarized in Table 1. Data were requested in phone interviews that were conducted in English or Spanish by trained interviewers located at DataBanque. Interviewer training was provided by CDS investigators. Interview responses were entered into a Computer-Assisted Telephone Interviewing (CATI) system. The data collection procedure was used for PQ at baseline (all participants), follow-up PQ at 12 mo (nutrition substudy participants), and FQ at baseline and 12 mo (nutrition substudy participants).

DataBanque developed a custom software program that incorporated data from the PQ and FQ into a CATI system. The system allowed for scheduling of interviews, callbacks, recording of survey answers, and reporting. The software allowed the interviewer to select the participant from the USRDS list and record that participant’s responses. An opening and closing script was developed and approved by the CDS Steering Committee.

Lists of patients to be contacted were provided monthly to DataBanque from the CC. Patients with a complete address were mailed an introductory letter thanking them for agreeing to participate in the study and re-explaining the process for participation. For those in the nutrition substudy, DataBanque mailed a copy of the FQ and a form showing representative serving sizes to help inform participants’ responses at the time of the DataBanque call.

DataBanque conducted the PQ interview (and FQ when applicable) on the day of the first call attempt, if the patient agreed. If the patient was unable to speak with the interviewer at that time, an appointment was scheduled to conduct the interview in the future. DataBanque attempted to schedule appointments no more than 14 d in advance unless a different appointment day was requested by the participant.
For patients with scheduled interviews, DataBanque called to remind participants of their scheduled appointments on the day before the interview. When patients could not complete the entire interview at one time, DataBanque allowed them to start an interview and schedule a time when they could complete the remainder. On the day of the appointment, an interviewer phoned the participant at the scheduled time, conducted the survey(s), and recorded responses in the software. A Spanish-speaking interviewer was on staff to accommodate Spanish-speaking participants. If an attempt was made to phone a patient and the patient was unreachable, the patient remained “active” on the DataBanque call list for up to 50 call attempts. There were no time restrictions placed on reaching 50 attempts. DataBanque attempted to call the patient every other day for the entire time the patient was on the active call list, varying the call times so that an attempt was made to reach the patient in the morning, afternoon, and evening on each successive attempt. Every effort was made to accommodate patients’ schedules, and many requested to be called weeks later. If DataBanque learned that the patient was in the hospital, an attempt was made to reach the individual 1 wk later, and the individual was put back on the active list at that time and remained so up until the maximum number of call attempts was reached. For follow-up interviews with patients in the nutrition substudy, it was DataBanque’s goal to interview patients 12 mo after the baseline interview ± 2 wk, unless the patient was unable to conduct the interview during that timeframe, in which case an attempt was made to interview patients as close as possible to their anniversary date. Thirty percent of interviews were recorded by DataBanque for quality assurance purposes (with approval by patients), so that a quality rating could be applied to each interviewer. Weekly status and quality reports were provided to the USRDS CC.

Table 1. Data collection items: CDS Patient Questionnaire

<table>
<thead>
<tr>
<th>Variables</th>
<th>Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sociodemographics</td>
<td>Age, gender, race, ethnicity, educational status, household status, smoking status</td>
</tr>
<tr>
<td>Treatment characteristics</td>
<td>Dialysis modality, home/clinic treatment, no. of HD days/wk, no. of hours/HD treatment, HD start time, type of current HD access</td>
</tr>
<tr>
<td>Patient-reported health status</td>
<td>SF-12, v0.2 (3)</td>
</tr>
<tr>
<td>Quality of life</td>
<td>KDQOL-36 symptoms/problems (4)</td>
</tr>
<tr>
<td></td>
<td>KDQOL-36 effects of kidney disease (4)</td>
</tr>
<tr>
<td></td>
<td>KDQOL-36 burden of kidney disease (4)</td>
</tr>
<tr>
<td></td>
<td>KDQOL-SF cognitive function (4)</td>
</tr>
<tr>
<td></td>
<td>Other symptoms/problems: Muscle weakness</td>
</tr>
<tr>
<td></td>
<td>Headaches</td>
</tr>
<tr>
<td></td>
<td>Postdialysis fatigue (HD patients) (5)</td>
</tr>
<tr>
<td></td>
<td>Restless legs (6)</td>
</tr>
<tr>
<td></td>
<td>Daytime sleepiness (7)</td>
</tr>
<tr>
<td></td>
<td>Nocturnal sleep quantity and complaints (8,9)</td>
</tr>
<tr>
<td>Satisfaction with care</td>
<td>KDQOL-SF satisfaction with care (4)</td>
</tr>
<tr>
<td>Functional status/physical activity</td>
<td>Satisfaction with physician interaction (10)</td>
</tr>
<tr>
<td>Usual activity/exercise</td>
<td>Human Activity Profile (11,12)</td>
</tr>
<tr>
<td>Depressed mood</td>
<td>Physical Activity Score (13,14)</td>
</tr>
<tr>
<td>Employment status</td>
<td>PHQ-2: two-item depression screener (15)</td>
</tr>
<tr>
<td>Pre-ESRD treatment</td>
<td>Working for pay: year prior and currently</td>
</tr>
<tr>
<td></td>
<td>Able to work</td>
</tr>
<tr>
<td></td>
<td>Social Security disability status</td>
</tr>
<tr>
<td>Transplant discussed</td>
<td>Transplant discussed since dialysis started</td>
</tr>
<tr>
<td></td>
<td>Cholesterol-lowering medications, blood pressure-lowering medications, modality choice</td>
</tr>
<tr>
<td>HD, hemodialysis; KDQOL, Kidney Disease Quality of Life; SF, short form; PHQ, Patient Health Questionnaire.</td>
<td></td>
</tr>
</tbody>
</table>
age call time was 30 to 35 min for the PQ and 55 to 60 min for the PQ and FQ. The average number of call attempts per patient over the complete period of data collection was approximately 4.5.

**Serum Samples**

Under the direction of the Nutrition SSC, serum specimens were obtained in conjunction with routine monthly laboratory studies from consenting patients at facilities in the nutrition substudy, for the purpose of conducting assays and creating a repository. At the time of the first blood draw, it was expected that one 3-ml tube of EDTA blood would be obtained for isolation of plasma and for separation of white blood cells for DNA analysis and repository storage. It was expected that 15 ml of blood would be drawn for serum isolation at study onset and then every 3 mos for 1 yr, specifically to measure prealbumin, albumin, C-reactive protein (CRP), and α-1 acid glycoprotein (α-1 AG)—clinical measures linked to nutritional status and inflammation but (with the exception of albumin) not routinely measured in dialysis practice. After receiving signed consents from patients at facilities selected for the nutrition substudy, the CC informed the Nutrition SSC lab at UC Davis. The lab then contacted the nurse manager and/or administrator of the participating facilities. If the facility agreed to participate, blood tubes, directions, and packaging materials were provided to the facilities. EDTA-treated blood was centrifuged according to routine dialysis unit procedures and then shipped on dry ice via overnight delivery to the UC Davis laboratory without further processing. If a patient moved to another facility, that facility was contacted and asked to participate. Shipping materials were sent either to the secondary facility or to the patient to deliver to the secondary facility for processing.

**Data Management and Analysis**

**Patient-Reported Data**

A complete download of the PQ database was sent by DataBanque monthly to the USRDS CC. DataBanque transmitted the PQ data to NutritionQuest (Berkeley CA), which later reported batched patient-specific dietary intake and nutrient data to the USRDS CC. Range and logical data checks were performed for all items in the PQ. Summary scores were constructed for variables that form scales.

**Laboratory Samples**

Serum was separated by centrifugation upon receipt, and up to five 500-μl aliquots were separated and stored over liquid nitrogen to create a repository. One aliquot was then analyzed for CRP, prealbumin, albumin, and α-1 AG by rate nephelometry using a Beckmann Array nephelometer. All nephelometric measurements were made in duplicate in each of two optical systems. The average of these values was used for calculations. The EDTA-treated sample was centrifuged at 2000 RPM for 10 min. The plasma was removed, aliquoted, and frozen. Theuffy coat was removed and washed, and the red blood cells were lysed with a solution of 10 mM Tris HCL (pH 7.6), 5 mM MgCl2, and 10mM NaCl. The white cell pellet was washed and then aliquoted and frozen.

**Results**

**Participation Rates and Patient Characteristics**

From September 1, 2005 to June 1, 2007, the Standard Information Management System identified 11,292 incident maintenance dialysis patients from the 335 sampled dialysis units who potentially met the criteria specified for the CDS. Reasons for patients subsequently being determined not eligible for the CDS included death, prior transplantation or imminent transplant scheduled, return of kidney function, incarceration, not English- or Spanish-speaking, hearing or speech impaired, cognitively impaired, in hospice, and moving out of the country. In addition, during the early phase of data collection, it was not possible to contact patients affiliated with facilities that were closed as a result of Hurricane Katrina or facilities that did not deliver CDS materials to patients. Because a phone interview was the method of data collection, individuals were also determined to be not eligible for the CDS if DataBanque encountered an incorrect or disconnected phone number and no additional contact information could be obtained. Individuals for whom the phone was always busy or no answer was obtained despite multiple call attempts were classified as unknown eligibility.

The overall survey response rate (number of interviewed patients divided by the total number of eligible patients) was 18.5%. Because the CDS implemented enhanced efforts to contact and encourage enrollment of potential participants in the time period February 23, 2006 through August, 2007, survey response rates were estimated separately for this time period and for the initial time period of October 2005 to February 22, 2006. The estimated response rate for the second phase of the study was 20.5%, and the estimated response rate for the initial phase of the study was 12.4%.

A total of 1646 eligible patients participated in the CDS at baseline. Baseline CDS participants were cared for at 295 dialysis units, located across all 18 Networks and in all 50 states and the District of Columbia, except Alaska and Vermont. At the time of the construction of the facility sampling frame, three fourths of these 295 units were chain affiliated. The majority (90%) of units with which patients were affiliated were free-standing, nonhospital based.

There were 361 patients at 68 dialysis units who participated in the nutrition substudy at baseline. Of the 68 dialysis units, 66 (97%) were free-standing, nonhospital based. Of the 361 baseline nutrition substudy participants, serum samples were obtained for 259 patients (76%).

Sociodemographic information for CDS participants is shown in Table 2, along with information for all U.S. incident dialysis patients over the same time frame. CDS participants were more likely to be aged 45 to 64 and less likely to be aged 75 and older than were the overall population of dialysis patients who started treatment in 2005 (as expected, on the basis of exclusion criteria), but the proportion of patients aged 18 to 44 and the proportion aged 65 to 74 closely matched the overall population of patients incident in 2005.

The proportions of participants on peritoneal dialysis (PD) in the CDS (about 10%) and in the nutrition substudy (about 14%) were higher than in the overall population who started dialysis in 2005 (7%). We had anticipated an enriched fraction of patients on PD owing to selection strategies favoring larger facilities for the nutrition substudy.

The average length of time that CDS participants had been on dialysis at their baseline interview was approximately 4 mo (median = 122 d; mean = 129 d). The range was 78 d to 357 d (only two patients had been on dialysis longer than 1 yr at the time of the baseline interview). Interviews were conducted in
Spanish with 119 CDS participants (approximately 7% of the patients interviewed).

Compared with DMMS Wave 2 patients, fewer CDS participants had less than a high school education (22.7%), and more CDS participants had some college or higher education (44.6%). The DMMS Wave 2 study provides the only national source of educational status information for patients beginning dialysis in the United States (2). Among patients included in the DMMS Wave 2, one third (31.6%) had less than a high school education, one third (35.1%) were high school graduates, and one third (33.3%) had some college or higher education.

For the 269 nutrition subset patients who provided serum samples, mean serum albumin concentration was 3.47 ± 0.47 g/dl, mean prealbumin was 31.5 ± 9.6 mg/dl, and mean α1-AG was 115 ± 36 mg/dl. CRP was not normally distributed. The median value of CRP was 7.4 mg/L, with an interquartile range of 3.96 to 12.7 mg/dl (range 1 to 278 mg/L).

**Discussion**

**Challenges in Implementation**

Several alternative study designs were considered before the CDS was implemented. The initial model proposed for the CDS was the DMMS Wave 2. However, since 1996–1997 when Wave 2 was performed, issues of confidentiality and privacy protection described under the Health Insurance Portability and Accountability Act (HIPAA) and the research community’s response have changed the face of clinical research. In the context of dialysis research, consolidation of the industry (such that two for-profit providers are now responsible for more than two-thirds of patients), mandates to adhere to widely accepted clinical practice guidelines (e.g., Kidney Disease Outcomes Quality Initiative) and increasing financial pressures have created additional obstacles to the collection of data for research purposes. Moreover, the Office of Human Research Protections of the Department of Health and Human Services has rigidly defined “engagement” in research, such that without requisite research training and distinct institutional review board approval, dialysis unit personnel are no longer allowed to participate in research. The CDS Steering Committee considered training data collectors who would go to dialysis facilities but eventually concluded that this approach would not be economically feasible if dialysis facilities were to be sampled nationally.

To retain the geographic diversity and scientific integrity of a randomly selected national sample of dialysis facilities, the Steering Committee elected to collect data from patients via a telephone interview rather than face-to-face. A data management service organization was enlisted to contact patients by phone and interview them. The CATI system developed by the organization and used for patient interviews was efficient, and the organization readily accommodated requests and recommendations from the CDS Steering Committee.

The first phase of the CDS involved facility sampling, with the goal of achieving a probability sample of dialysis units that was nationally representative and that would yield a sufficient number of incident patients to produce an adequate number of study respondents after allowing for anticipated noneligibility and nonparticipation. The number of incident patients identified as potential participants from sample dialysis units per month by CMS was consistent with the number that had been projected. Making contact with patients was problematic, however, especially in locations where dialysis unit personnel did not routinely facilitate distribution of study information to patients and when patients had nonworking mailing addresses and/or phone numbers. In part by design and in part owing to the lower than expected participation rate, the CDS sample cannot be considered truly representative of the U.S. dialysis population. Within the CDS, there were fewer persons with cognitive impairment and diminished health literacy than within the incident dialysis population at large. Outcomes
studies derived from the CDS cohort will need to take these and other issues related to participation in research into consideration.

The CDS experience provided valuable insights about the challenges of conducting USRDS special studies. Although national probability samples of facilities and patients allow estimation of population parameters, data collection from such samples can be very expensive and problematic. Geographic areas for a study can be purposely selected to maximize response from varied groups within the population of interest (e.g., urban and rural), or concentrations of racial/ethnic minorities. This may be a preferable model for future USRDS special studies. Dialysis facility cooperation remains crucial in this model, with active dialogue among researchers and dialysis providers. All stakeholders should value the study outcomes and collaborate in the interpretation and presentation of findings. In March 2006, participants in a dialysis facility research leadership meeting agreed that clinical and translational research should be a core value for providers as well as investigators. The CDS investigators and the USRDS strongly endorse support for this objective from the outset in future studies.

The CDS contains detailed information on health-related quality of life, self-reported physical activity, and macronutrient intake and represents one of few cohorts in which laboratory proxies of nutrition and inflammatory status were measured simultaneously. With linkage to the USRDS database, associations among these domains and survival, hospitalization, subsequent use of other medical services, and associated costs can be investigated. The CDS therefore provides a unique resource from which to inform the design of interventions addressing several related conditions that affect longevity and health status in ESRD.

Appendix: CDS Steering Committee
Nancy G. Kutner (Principal Investigator) and Donna Brogan (Co-principal Investigator), USRDS Rehabilitation/Quality of Life Special Studies Center; Glenn M. Chertow (Principal Investigator), George A. Kaysen (Co-principal Investigator), Kirsten L. Johansen (Co-Investigator), and D. Jordi Goldstein, USRDS Nutrition Special Studies Center; Lawrence Hunsicker, USRDS Economics Special Studies Center; Allan J. Collins, Shu-Cheng Chen, and Sarah Pederson, USRDS Coordinating Center; Doug Marsh and Vickie Peters, ESRD Network Coordinating Center; Sondra Frenz, DataBanque Research Services; Lawrence Y. Agodoa and Paul W. Eggers, NIDDK; and Pam Frederick and Condict Martak, CMS.

Acknowledgements
The CDS investigators gratefully acknowledge the dialysis units that assisted with the study and the time and interest of individual patients who participated. The interpretation and reporting of the data presented here are the responsibility of the authors and in no way should be seen as an official policy or interpretation of the United States government.

Disclosures
None.

References
1. Block Dietary Data Systems. www.nutritionquest.com