Appendix A: Products & Services

Products and services provided by the USRDS to support the work of the renal community are detailed in table b.a. The entire ADR is available at www.USRDS.org, with powerpoint slides of all figures and excel files of the data behind the graphs; included as well are pdf files of the researcher's guide. The site's render system allows users to create customized data tables and regional maps. Data on website use are presented in figure b.1.

Data Requests

Making information on ESRD available to the renal community is a primary objective of the USRDS, and we are committed to the timely fulfillment of data requests. In many cases requests can be answered through data published in the ADR or elsewhere. Requests for data not available in material published by the USRDS, but that require two hours or less of staff time, are fulfilled by the Coordinating Center without charge, usually within one week. More complex requests — requiring more than two hours of staff time — as well as requests for Standard Analysis Files and custom files, must be accompanied by a written proposal (see details below), and will be completed only upon written approval by the NIDDK Project Officer.

Research Files

The Coordinating Center maintains a set of Standard Analysis Files (SAFS) to meet diverse research needs and provide easy access to data used in the ADR. The SAFS were introduced in 1994, as the NIDDK began awarding new grants focusing on research using the USRDS data. The result has been an annual increase in the number of files provided by the USRDS.

Prior to 1994, all researcher files were created for specific projects. Since the introduction of the SAFS, however, custom files are generally limited to cases in which a researcher provides a patient finder file to be matched with the USRDS database. For more information on merged data requests, please contact the Coordinating Center at USRDS@USRDS.org.

The Core SAF dataset contains basic patient data, and is needed to use any of the other SAFS. Included are each patient's demographic information, payor and treatment history, limited transplant data, provider data, and data from many of the USRDS Special Studies. Approximately half of the researchers using the USRDS SAFS need only this data set. The Transplant data set contains detailed transplant and transplant follow-up data collected by CMS and UNOS. Data on hospital inpatient stays are found on the Hospital data set. All Medicare billing data are available by individual year (see Table A.3).

Standard Analysis Files

SAF use is governed by the USRDS policy on data release for investigator-initiated research, found later in these appendices. Research proposals must be approved by a USRDS Project Officer, and researchers must sign the USRDS "Agreement for Release of Data," on the same page. File prices are listed in Table A.3.

Most SAFs provide patient-specific data. All patient identifiers are removed or encrypted, but data confidentiality remains a serious concern. The USRDS "Agreement for Release of Data" describes restrictions on SAF use and disposition. SAFs include an encrypted ID number to allow patient data from multiple SAFs to be merged.

Core SAF Dataset

The Core SAFs contain the most frequently used data and are needed for use of the Transplant and Hospital datasets, or any data based on Medicare claims. Included files are as follows (also listed in Table b.b).

- **Patient** Contains one record per patient in the USRDS database, and gives basic demographic and ESRD-related data.
- **Residence** A longitudinal record, to ZIP code, of residence.
- **Payor History** Contains a new record for each patient at each change in insurance payor.
- **Treatment History/Modality Sequence** Contains a new record for each patient at each change in modality or dialysis provider.
- **Medical Evidence** Contains full data from the 1995 and 2005 versions of the CMS Medical Evidence form. In April 1995 a new version of the form went into use, with data on comorbidity, employment status, lab values at initiation, and Hispanic ethnicity.
- **Transplant** Contains basic data for all transplants (reported by CMS and UNOS), including the date of graft failure (detailed transplant data are contained on a separate transplant data set).
- **Transplant Wait List** Beginning with 2001 data (used in the 2002 ADR), this file has been updated to include basic patient demographic data and, from UNOS, all unique wait-list periods for each dialysis patient.
- **Facility** Conducted annually, the CMS End-Stage Renal Disease Facility Survey is the source of data for the Facility SAF. Geographic variables that could identify facilities are deleted. The survey period is January 1 through December 31.
- **Facility Cost Reports** CMS hospital and independent facility cost reports for 1989–1995 and 1989–1993, respectively, are available as SAFs. All geographic variables are deleted to ensure confidentiality. The files may be linked to the Facility SAF using the USRDS provider ID, though analyses at less than a regional or network level are not possible. Because these files are rarely used, additional data will be added only if there is sufficient demand.
- **Dialyzers** The Case Mix Severity, Case Mix Adequacy, and DMMS Special Studies collected information on patient dialyzers. SAFs for these studies describe the dialyzer through a code, which must be matched to information in the Dialyzer file to find the manufacturer and model along with characteristics such as membrane type and clearance. We believe that these data, from published sources available at the time of the study, accurately represent the dialyzer characteristics, but they should be used with caution.

Data from Special Studies

Topics for USRDS Special Studies are approved by the NIDDK, with recommendations from CMS, the Scientific Advisory Committee, the ESRD networks, and the Renal Community Council. Design and sampling plans are developed, samples are selected, and data collection forms and instructions are drafted, tested, and finalized. The main studies to date are summarized below, and are detailed in the Researcher's Guide.

Dialysis Morbidity & Mortality Study (DMMS) The DMMS was a USRDS Special Study in which data on demographics, comorbidity, laboratory values, treatment, socioeconomic factors, and insurance were collected, using dialysis records, for a random sample of U.S. patients. Waves 1, 3, and 4 are historical prospective studies on a total of 16,812 participants in which data were collected for patients on in-center hemodialysis on December 31, 1993. Data were abstracted from medical records, and patients were followed to the earliest of data abstraction, death, transplant, change in modality, or transfer to another facility. Wave 2 is

a prospective study of incident hemodialysis and peritoneal dialysis patients for 1996 and early 1997 and included 4,024 participants. Case Mix Adequacy Study of Dialysis: The objectives of this USRDS Special Study were to establish the relationship between the dose of delivered dialysis therapy and mortality, determine the strength of this relationship when data are adjusted for comorbidity, assess how this relationship changes with dialysis dose, assess how this relationship is affected by dialyzer reuse, and examine the impact of different dialysis membranes on patient morbidity and mortality.

The study consisted of two groups: an incident sample of ESRD patients who began hemodialysis in 1990, and a prevalent sample of hemodialysis patients whose ESRD began prior to 1990. A total of 7,096 patients from 523 dialysis units were included, with approximately 3,300 patients having both the pre- and post-BUN values needed to calculate delivered dialysis dose. Ninety-four percent of these cases were matched to the USRDS database. The ESRD networks collected these data in conjunction with their Medical Case Review data abstraction.

Case Mix Severity Study For this USRDS Special Study, data were collected on 5,255 patients incident in 1986–87 at 328 dialysis units nationwide. Objectives were to estimate the correlation of comorbidity and other factors existing at the onset of ESRD to mortality and hospitalization rates, while adjusting for age, gender, race, and primary diagnosis; evaluate possible associations of these factors with reported causes of death; assess the distribution of comorbidity and other factors among patients on different modalities; and compare relative mortality rates by treatment modality, adjusting for comorbid conditions and other factors.

Pediatric Growth & Development The objectives of the USRDS Pediatric Growth and Development Study were to establish a baseline for assessing the relation of patient growth and sexual maturation to modality, and establish a prototype for the ongoing collection of pediatric data. All patients prevalent in 1990 and born after December 31, 1970, were included in the study, a total of 3,067 patients at 548 units.

CAPD & Peritonitis Study The USRDS CAPD and Peritonitis Study examined the relation of peritonitis episodes in CAPD patients to connection device technology and other factors. The study population included all patients newly starting CAPD in the first six months of 1989, a maximum of 14 patients per dialysis unit. All units providing CAPD training participated in the study. The sample contains data on 3,385 patients from 706 units.

Transplant SAF Dataset

Due to changes in data collection sources over the years, data related to transplants are now presented in eight separate SAFs. The first two are included on the Core SAF, and the remaining six are included inthe Transplant data set.

TX includes minimum details on all transplants from all sources

- TXWAIT contains one record for each patient in the USRDS database per wait list event
- TXHCFA includes transplant information collected by CMS's PMMIS system prior to 1994
- TXUNOS includes transplant information collected since 1987 by UNOS, currently the main source of transplant data for the USRDS
- TXIRUNOS includes information on immunosuppressive drugs collected by UNOS at the time of transplantation events
- TXFUHCFA includes transplant follow-up reports collected by CMS prior to 1994; reports are completed at discharge, six months, each year post-transplant, and at graft failure
- TXFUUNOS includes transplant follow-up reports collected by UNOS since 1988
- TXIFUNOS includes information on immunosuppressive drugs, collected by UNOS at follow-up visits

Reference tables in Transplant Sections E and F are produced primarily from the CMS and UNOS transplant files.

In July of 1994, CMS and the Health Resources Services Administration (HRSA) consolidated transplant data into a single collection by UNOS under its HRSA contract. Expanded transplant data are shared among HRSA, CMS, and the NIH, and are thus available to the USRDS. This has resulted in the addition of data on a substantial number of non-Medicare transplant patients, including children.

CMS and UNOS transplant files overlap for 1988–1993, and some Medical Evidence (ME) forms and institutional claims records indicate transplants not included in either file. To resolve conflicts among the sources and create the transplant SAF, all UNOS transplants are first accepted into the file, with all pre-1988 CMS transplants accepted next. CMS transplants from 1988–1993 are then accepted if there is no transplant in the file for that patient within 30 days of the CMS transplant (it is common for dates between sources to differ by one day). Finally, transplants indicated on the ME form are accepted if no transplant is listed for the patient within 30 days of the Medical Evidence transplant date.

Hospital SAF Dataset
Hospitalization inpatient data are a subset of the data in the Institutional Claims file. No payment or cost variables are included on this dataset, which is for researchers who need data on hospital inpatient stays and on diagnoses and procedures for those stays, but who do not need payment data.

Comprehensive Dialysis Study Dataset
This dataset contains information from the Comprehensive Dialysis Study (CDS), a USRDS special data collection study to assess rehabilitation/quality of life and nutrition issues in incident dialysis patients. The study was conducted between 2005 and 2008. All 1,677 participants answered questions on physical activity level, health-related quality of life, and work/disability status during the first six months of after the initiation of ESRD therapy. In a subset of 400 participants, dietary intake and nutritional status were also assessed.

Dialysis Morbidity & Mortality Claims Dataset
This dataset contains Medicare claims for participants in the Dialysis Morbidity and Mortality Studies. Data are followed to the currently reported claims year.

Case Mix Adequacy Claims Dataset
This dataset contains Medicare claims for particiants in the Case Mix Adequacy Special Study. Medicare payment data for these patients are followed to the currently reported claims year.

Medicare Payment Data
Medicare payment data on institutional claims are available for pre-1989 through 2007, while data on physician/supplier claims are available for 1991–2007. The 2008 claims will be available, along with other updated USRDS SAFs, by the end of 2010.

Institutional claims consist of all inpatient/outpatient claims (inpatient, outpatient, skilled nursing facility, home health agency, and hospice), including outpatient dialysis claims. Physician/supplier claims account for 80 percent of claims but only 20 percent of dollars. The structure and content of the two types of claims differ, as do the files derived from them. Institutional claims are provided in two types of files: the Institutional Claims file, indicating claim type, dollar amounts, DRG code, type of dialysis involved (if any), and dates of service; and the Institutional Claims Detail file, containing details such as diagnosis and procedure codes. Many analyses require only the Institutional Claims files. Physician/supplier claims are contained in one type of file with one record for each claim line-item. The file includes dollar amounts, dates of service, diagnosis and procedure codes, and type and place of service.

Clinical Performance Measures Survey
The Clinical Performance Measures (CPM) data is a CMS project developed to collect information on the quality of care provided to the dialysis population. The data originates from yearly surveys of approximately 10,000 dialysis patients completed by the primary care facilities, and was formerly known as the ESRD Core Indicators Project. This project results in a rich source of detailed information, useful in analyses of healthcare delivery in a sample of the dialysis population.

To further expand the value and use of the CPM data, we have linked patient data from the USRDS SAFs, enabling complete claims extraction from the SAFs for all identified patients. The resulting claims history has been combined with the CPM data to form a complete mini-set of the USRDS data products with supporting files. This enables researchers to add patient-level laboratory and dialysis prescription detail to a broad range of healthcare service event data over many years.

The USRDS Coordinating Center has made the CPM data available as SAFs. The dataset contains CPM data collected in surveys from 1994–2008. A listing of available files and the corresponding costs can be found in Table b.e, or you may contact the USRDS Coordinating Center for further information. For a detailed explanation of why there are no 2009 CPM form data available, please view the CPM 2010 Researcher's Guide on the USRDS website.

CKD 5 Percent General Medicare Payment Data
The CKD cohort datasets are built from the 5 percent general Medicare Claims SAFs, and contain a patient master file, a payor sequence file, and a set of comorbidity files. We no longer produce datasets for diabetes and CHF based on the 5 percent Medicare claims.

Separately, a 5 percent general Medicare Hospital SAF (inpatient, outpatient, skilled nursing facility, home health, hospice, Part B, and durable medical equipment) for the CKD cohort is also available for 1992–2008; 2009 claims will be available by the end of 2011. Data are derived from the IP claims SAF files. No payment or cost variables are included, so these data are for researchers who need data on hospital inpatient stays and on diagnoses and procedures for those stays, but do not need payment data.

Pre-ESRD Medicare claims
The pre-ESRD claims (also known as the back-casted claims) are a collection of Medicare institutional and physician/supplier billing records incurred prior to the onset of ESRD. Included in these claims are any and all claims available from Medicare for incident patients during their incident year and the two prior calendar years.

The USRDS has made the pre-ESRD data available as SAFs. This dataset includes Medicare claims of ESRD patients from incident years 1995–2008 with 2009 data available by the end of 2010. The structure of the claims file is identical to the ESRD claims files and organized by calendar year. In addition, a pre-ESRD payor sequence is provided so researchers can determine Medicare enrollment for the periods prior to first ESRD service date. A listing of available files and the corresponding costs can be found in Table b.e.

Part D Data
Section 101 of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003 amended Title XVIII of the Social Security Act by establishing the Voluntary Prescription Drug Benefit Program (Part D). Effective January 1, 2006, Part D is an optional prescription drug benefit for individuals who are entitled to Medicare benefits under Part A or enrolled in Medicare benefits under Part B. The data from the first few months of 2006, when the benefit was very new, may be incomplete, and should be interpreted with caution.

The Part D data is obtained from CMS annually, with finder files provided by the USRDS. The Part D data are divided into two separate files: an annual enrollment file containing monthly indicators of enrollment in Part D, and a prescription drug event file (PDE) containing details of prescriptions filled by Part D beneficiaries.

Since the Part D benefit is voluntary, not all Medicare beneficiaries are enrolled. The annual enrollment file contains 12 monthly indicators that detail whether the beneficiary is enrolled in Part D, and if so, the type of plan. There are also monthly indicators for dual eligibility (Medicare and Medicaid), the retiree drug Subsidy, and the low income subsidy (LIS).

Linkages to the USRDS Database
The USRDS does provide the service of linking population cohorts to the USRDS dataset to determine ESRD status and outcomes for epidemiological research. Please contact the USRDS Coordinating Center for more information on the application process and the costs for this service.

File Media & Formats
SAFs are provided on DVDs as SAS files, and can be used by SAS on any 486 or Pentium PC with a DVD reader. The SAS format is widely used, easily transported, and largely self-documenting. SAS is a commercially available data management and statistical analysis software system that runs on most computers, and is almost universally available on university computer systems. The SAFs take full advantage of the program's ability to incorporate detailed documentation into the file. Researchers needing another format or medium must arrange for the conversion.

Costs
File prices cover file reproduction, documentation, administrative costs, and costs of technical support. Prices are subject to change.

Documentation
The Researcher's Guide to the USRDS database provides most of the SAF documentation. It includes a codebook of variables and a chapter on using the SAFs in SAS. Copies of data collection forms used by CMS, UNOS, and the USRDS Special Studies as well as the entire guide may be downloaded from the USRDS website (PDF copy included on Core SAF).

Data Use Acknowledgement
Publications using USRDS data should include an acknowledgment and this notice: The data reported here have been supplied by the United States Renal Data System (USRDS). The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy or interpretation of the U.S. government.

Data Release Policy
Since the SAFs and custom data files contain confidential, patient-specific data, their release requires the approval process described here. Investigators may contact the USRDS Project Officer (PO) at the NIDDK to discuss requests before preparing a proposal. To request and use USRDS data files, investigators must provide the PO with a detailed description of the proposed investigation (see Table b.d). The summary must include goals, background data, an in-depth description of study design and methodology, and resources available for completing the project, and may be the description from a grant proposal or other application. The project must comply with the Privacy Act of 1974, and the summary should provide enough information to enable assessment of compliance. Guidelines for Privacy Act adherence are found in the "Agreement for Release of Data," later in the appendices. With your completed research proposal, please include a signed agreement for release of information from each investigator and analyst who will use the data files.

Investigators must also indicate needed USRDS SAFs by name. If these files cannot meet requirements of the proposed research, investigators must specify precisely which data elements are needed, and budget for a substantially higher cost.

The investigator and the Coordinating Center (CC) will resolve any technical questions. The investigator will arrange payment with the CC, and payment must be received before the files will be released. Checks must be made payable to the Minneapolis Medical Research Foundation.

The NIH will review the project for technical merit and for conformity with the Privacy Act. The PO will notify the investigator(s) in writing of the outcome, and if the project is not approved will discuss reasons for the decision. The PO will send a copy of the approval letters to the CC. When payment for the files has been received by the CC, the CC will prepare the files and documentation and send them to the investigator.

Any reports or articles resulting from use of USRDS data must be submitted to the PO prior to submission for publication to assure adherence to the Privacy Act. The PO must respond within 30 days. If a report or article is determined not to adhere to the Act, it shall not be published until compliance is achieved. Assessment of compliance will not depend on the opinions and conclusions expressed by the investigators, nor will the PO's approval indicate government endorsement of the investigator's opinions and conclusions.

All publications using released data must contain the standard acknowledgement and disclaimer presented above. Investigators are requested to send copies of all final publications resulting from this research to both the PO and the CC.

Caveats

This policy establishes conditions and procedures for the release of data from the USRDS, and is intended to ensure that data are made available to investigators in the pursuit of legitimate biomedical, cost-effectiveness, or other economic research.

The USRDS will not release data that identify individual patients, providers, or facilities. Since it might be possible, however, to infer identity from SAF data, these data are considered confidential. The USRDS "Agreement for Release of Data" contains a number of general and specific restrictions on the use of USRDS data, and investigators are expected to abide by these restrictions. If individually identifiable data are needed, the request should be submitted directly to CMS. Use of these data to identify and/or contact patients, facilities, or providers is prohibited by USRDS policy and by the Privacy Act of 1974.

The USRDS CC will provide data in one DVD. Analytical services other than review of the proposal and preparation of the data file will not be provided under the USRDS contract, though CC personnel may participate in analyses funded by other sources.

A.1 USRDS Products & Services

Reports & guides

Annual Data Reports Available from the National Kidney and Urologic Disease Information Clearinghouse, 3 Information Way, Bethesda, MD 20892-3560; 301.654.4415, nkudic@info.niddk.nih.gov. ADR material is also published in the American Journal of Kidney Diseases.

Annual Data Report CD Contains the text and graphics of the ADR, data tables, PowerPoint slides, and the Researcher's Guide.

Researcher's Guide to the USRDS database

Provides a detailed description of the USRDS database and of the USRDS Standard Analysis Files; the basic reference for researchers who use USRDS data files.

www.usrds.org

Contains PDF files of the chapters, reference tables, and the Researcher's Guide; PowerPoint slides of atlas figures and USRDS conference presentations; Excel files of the data tables; notices regarding current news and analyses; links to related Internet sites; and email addresses for contacting the USRDS.

RenDER

The USRDS Renal Data Extraction and Referencing (RenDER) System is a querying application that allows users to create data tables and interactive maps . It can be accessed at www.usrds.org/odr/xrender_home.asp following a short registration; a tutorial is also available on this site to help new users.

Requests for data

Data requests: two-hour Questions and data requests that are not answered directly by the ADR can be addressed to the Coordinating Center; those that require less than two hours of staff time to fulfill will be processed without charge.

Data requests: more than two hours Questions and data requests that require over two hours of staff time must be submitted in writing and approved by the NIDDK Project Officer. Fulfillment of these requests is subject to staff availability, and costs are assessed on a case-by-case basis.

Standard Analysis Files SAFs provide patient-specific data from the USRDS to support ESRD research. A standard price list has been established for the files (Table A.3), and users must sign a Data Release Agreement with the NIDDK.

Merged data files Merged files can be created by the Coordinating Center for approved research projects. An hourly rate of $119.57 will be assessed for time spent on the request, and users must sign a data release agreement with the NIDDK. Contact the USRDS Coordinating Cener for more information.

Publications & presentations

Most USRDS research studies result in published papers or presentations at national meetings. Figures from abstracts and presentations can be found on the website, while published abstracts and papers can be found in the relevant journals.

Contact information

USRDS Coordinating Center
914 South 8th Street, Suite S2.100
Minneapolis, MN 55404
612.347.7776 or 1.888.99USRDS
Fax 612.347.5878
usrds@usrds.org

Data file contacts Shu-Cheng Chen, MS; schen@usrds.org
Beth Forrest, BBA; bforrest@usrds.org

A.2 Contents of the USRDS Core Standard Analysis Files

Patient (PATIENTS) one record for each ESRD patient. This is the patient master file and it contains patient demographic information, ESRD first service date, cause of renal failure, and more. Most other files will need to be linked to this file using the encrypted patient ID.

Residence (RESIDENC) for each patient, one record for each period in a different residence. This file is suitable for geographic variation analyses.

Treatment History (RXHIST) one record for each period a patient is on one modality. This file is suitable for analyses pertaining to treatment modalities and patterns.

Payor History (PAYHIST) one record for each period a patient is covered by one payor; each patient can have many records. This file would be used in analyses with respect to the impact of insurance payors on clinical outcomes.

Medical Evidence (MEDEVID95) one record for each 2728 form filed (1995 version). ESRD first service date, initial treatment modality, comorbid conditions, patient status at start of ESRD.

Medical Evidence (MEDEVID05) one record for each 2728 form filed (2005 version). ESRD first service date, initial treatment modality, comorbid conditions, patient status at start of ESRD, pre-ESRD care, and vascular access information.

Transplant (TX) one record for each transplant event; patients can have multiple events. This file would be used in transplant and transplant outcome analyses.

Transplant Waiting List (WAITLIST_KI, WAITLIST_KP) one or more records for each patient ever on list. This file would be used in analyses involving the the comparison of transplanted patients to dialysis patients who are transplant candidates. Patient selection to wait list.

Transplant Waiting Sequence (WAITSEQ_KI, WAITSEQ_KP) one or more records for each patient ever on list. Comparison of transplanted patients to dialysis patients who are transplant candidates. Patient selection to wait list.

Dialysis Morbidity and Mortality Special Study (DMMS) Wave 1: 5,670 patients; Wave 2: 4,024 patients; Wave 3–4: 11,142 patients. Comorbid conditions, adequacy of dialysis, dialysis prescription and other treatment parameters, laboratory test values, nutrition, vascular access.

Case Mix Adequacy Special Study (ADEQUACY) 7,096 patients. Comorbid conditions, adequacy of dialysis, dialysis prescription and other treatment parameters, laboratory values.

Case Mix Severity Special Study (CASEMIXS) 5,255 patients. Comorbid conditions, adequacy of dialysis, dialysis prescription and other treatment parameters, laboratory values.

Pediatric Growth and Development Special Study (PEDGROW) 3,067 patients. Growth, development, and other issues relating to pediatric ESRD patients.

CAPD Peritonitis Special Study (CAPD) 3,385 patients. CAPD and peritonitis.

Facility (FACILITY) one record for each year facility has operated. This data would be used to merge with the treatment history, transplant, or annual summary SAFs for analyses involving provider characteristics by encrypted ID.

Facility Cost Reports (FCOSHOS) one record per facility per year (1989–1995). This file contains the costs of staffing and dialysis facilities.

Dialyzers (DIALYZER) information on dialyzer characteristics; to be matched to patient dialyzer information in other SAFs. This file would be used in analyses involving the relation of dialyzer characteristics to patient outcomes.

Claim Codes (CLMCODES) one record for each diagnosis, procedure, or HCPCS code appearing in claims files. A starting point for analyses that will use diagnosis and procedure codes.

Formats all USRDS-defined SAS formats used by SAFs. Format library used to format values of categorical variables.

A.3 Prices for the USRDS Standard Analysis Files

Standard Analysis Files

Core dataset	$1,275	Needed in order to use the other files.
Transplant dataset	$500	Detailed transplant data from CMS and UNOS.
Hospital dataset	$500	Derived from the institutional claims; contains diagnosis and surgical procedure codes for each stay but does not include the cost data from the institutional claims records.
CDS survey dataset	$750	Survey information and laboratory values from the Comprehensive Dialysis Survey
DMMS claims	$500	Contains all of the Institutional and Physician/Supplier claims data for the patients in the USRDS Dialysis Morbidity and Mortality (DMMS) Special Study. Survey data are included in the Core dataset.
Case Mix Adequacy claims	$125	Contains all institutional and physician/supplier claims data for patients in the USRDS Case Mix Adequacy Special Study. Survey data are included in the Core dataset.

Pre-ESRD claims available for 1993 to 2010; price ranges from $200 to $800 per year and claim type. Prices subject to change.

ESRD Medicare Payment Data

	Institutional	Physician/ supplier	Part D
pre-1989	$250		
1989	$250		
1990	$250		
1991	$375	$500	
1992	$375	$500	
1993	$375	$500	
1994	$375	$625	
1995	$500	$625	
1996	$500	$750	
1997	$500	$875	
1998	$500	$875	
1999	$500	$875	
2000	$750	$875	
2001	$875	$875	
2002	$875	$1,000	
2003	$1,000	$1,125	
2004	$1,125	$1,125	
2005	$1,250	$1,250	
2006	$1,250	$1,250	$750
2007	$1,750	$1,375	$1,000
2008	$1,875	$1,500	$1,000
2009	$2,000	$1,625	$1,250
2010	$2,000	$1,750	$1,250

A.4 Prices for the CKD 5 Percent Medicare Sample Standard Analysis Files

Patient cohort finder $750 / Hospital file $250

	Institutional	Physician/ supplier		Institutional	Physician/ supplier	Part D
1992	$375	$375	2002	$500	$500	
1993	$375	$375	2003	$500	$500	
1994	$375	$375	2004	$500	$500	
1995	$375	$375	2005	$625	$625	
1996	$375	$500	2006	$750	$625	$350
1997	$375	$500	2007	$875	$625	$500
1998	$375	$500	2008	$1,000	$750	$500
1999	$500	$500	2009	$1,125	$875	$750
2000	$500	$500	2010	$1,125	$875	$750
2001	$500	$500				

A.5 Prices for the ESRD CPM/USRDS Files

ESRD CPM Survey data

Includes 1994–2008 hemodialysis survey years and 1995–2008 peritoneal dialysis survey years $1,250

ESRD CPM/SAF linked files

Core files	$500
Hospital	$200
Transplant	$200

ESRD CPM Medicare participant institutional & physician/supplier claims

are available for the years pre-1989 through 2011; $100–300 per year

A.6 Outline for Research Proposals Using USRDS Data

A data request applies only to the project stated in the proposal; a new proposal must be submitted for each additional use of the data

I Research topic title and submission date.

II Background information.

III Study design: objectives, hypothesis(es), analytical methods.

IV Data being requested: 1) List of Standard Analysis Files needed (if multiple years, please specify), or data fields needed in custom data file. 2) Description of data security: responsible party, computer access, etc. 3) Time frame for the project. 4) Statement that data will be returned to the USRDS or destroyed at the end of the project.

V To address patient privacy issues, to be consistent with HIPAA policies, and to insure that researchers are adhering to local privacy standards as well as to USRDS and CMS privacy policies, the USRDS now requires IRB approval for all research proposals. IRB approval is not required from those requesting aggregate data.

VI Outline of estimated costs of requested data; source of funding.

VII Agreement for Release of Data, signed by all researchers.

VIII For Principal Investigator and co-authors, **required**:
Name
Affiliation
Business address
Business phone & fax
Email address

Submit to

Paul Eggers, PhD
NIDDK
6707 Democracy Blvd, Room 615
Bethesda, MD 20892-5458
Phone 301.594.8305
Fax 301.480.3510
eggersp@extra.niddk.nih.gov

United States Renal Data System (USRDS)
Agreement for Release of Data

Project title _____

In this agreement, "Requestor Organization" means _____

A. The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), through the United States Renal Data System (USRDS) Coordinating Center (CC), will provide the Requestor with CDs, DVDs, or other media type containing the data extracted from the USRDS research database(the "Data"), which constitutes a Limited Data Set within the meaning of the HIPAA privacy regulations.

B. The sole purpose of providing the Data is the conduct of legitimate and approved biomedical, cost-effectiveness, and/or other economic research by the Requestor.

C. The Requestor shall not use the Data to identify individuals on the file.

D. The Requestor shall not combine or link the Data provided with any other collection or source of information that may contain information specific to individuals on the file, except where written authorization has been obtained through the approval process.

E. The Requestor shall not use the Data for purposes that are not related to biomedical research, cost-effectiveness, economic and/or other epidemiological research. Purposes for which the Data may not be used include, but are not limited to,
   - the identification and targeting of under- or over-served health service markets primarily for commercial benefit
   - the obtaining of information about providers or facilities for commercial benefit
   - insurance purposes such as redlining areas deemed to offer bad health insurance risks
   - adverse selection (e.g., identifying patients with high risk diagnoses)

   Any use of the Data for research not in the original proposal must be approved by the USRDS Project Officer (PO).

F. The Requestor shall not publish or otherwise disclose the Data in the file to any person or organization unless the Data have been aggregated (that is, combined into groupings of Data such that the Data are no longer specific to any individuals within each grouping), and no cells (aggregates of Data) contain information on fewer than ten individuals or fewer than five providers or facilities. The Requestor shall not publish or otherwise disclose Data that identify individual providers or facilities, or from which such identities could be inferred. However, the Requestor may release Data to a contractor for purposes of data processing or storage if (1) the Requestor specified in the research plan submitted to the USRDS Project Officer that Data would be released to the particular contractor, or the Requestor has obtained written authorization from the PO to release the Data to such contractor, and (2) the contractor has signed a data release agreement with the PO.

G. A copy of any aggregation of Data intended for publication shall be submitted to the PO for review for compliance with the confidentiality provisions of this agreement prior to submission for publication and, if not approved, shall not be published until compliance is achieved. The PO must respond within 30 days.

H. Appropriate administrative, technical, procedural, and physical safeguards shall be established by the Requestor to protect the confidentiality of the data and to prevent unauthorized access to it. The safeguards shall provide a level of security outlined in OMB Circular No. A-130, Appendix III — Security of Federal Automated Information System, which sets forth guidelines for security plans for automated information systems in Federal agencies.

I. No copies or derivatives shall be made of the data in this file except as necessary for the purpose authorized in this agreement. The Requestor shall keep an accurate written account of all such copies and derivative files, which will be furnished upon request to the PO. The USRDS data files covered in this data use agreement may be retained by the Requestor until the date specified by the PO in the approval letter, at which time Requestor may request renewal of this data use agreement to extend the retention period to comply with legal or institutional recorDKeeping requirements or to maintain the integrity of the research or research publications. If at any time during the data retention period the DUA between USRDS and CMS is canceled, the Requestor will be contacted to destroy the files in their possession. At the completion of the activities in the research plan, the file(s) and any derivative files and copies shall be destroyed. At that time the Requestor will inform the USRDS and the PO in writing that the files have been destroyed.

J. For the purpose of inspecting security procedures and arrangements, authorized representatives of the NIDDK and/or of CMS will, upon request, be granted access to premises where the Data are kept.

K. The following USRDS data file(s) is/are covered under this Agreement.

Name of Data file(s) requested (e.g. Core, Institutional claims, etc)	Year(s) if applicable
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____

REQUESTOR SIGNATURE: _____

_____
Authorized signatory (name, title & date)

_____
Requestor address

_____
Requestor telephone number

READ AND ACKNOWLEDGED:

_____	_____
Investigator/Analyst signature	Print Investigator/Analyst name & date
_____	_____
Investigator/Analyst signature	Print Investigator/Analyst name & date
_____	_____
Investigator/Analyst signature	Print Investigator/Analyst name & date

USRDS Project Officer - Lawrence Y. C. Agodoa, MD, NIDDK, NIH or Paul W. Eggers, PhD, NIDDK, NIH

_____
USRDS Project Officer signature & date

June 2012

United States Renal Data System (USRDS)
Merged Dataset Agreement for Release of Data

Project title _____

In this agreement, "Requestor Organization" means _____

A. The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), through the United States Renal Data System (USRDS) Coordinating Center (CC), will provide the Requestor CDs, DVDs or other media type containing data extracted from the USRDS research database. Prior to receiving USRDS data, the Requestor will provide USRDS with a list of personally identifiable information (PII) so USRDS can report which of the Requestor's subjects are in the USRDS end-stage renal disease (ESRD) data.

B. The sole purpose of providing the Data is the conduct of legitimate and approved biomedical, cost-effectiveness, and/or other economic research by the Requestor.

C. USRDS shall not use or disclose the Requestor's data for any purpose other than to create the data extracted from the USRDS database. In the event that the Requestor's data is used or disclosed for any purpose other than that covered by this agreement, USRDS will notify the Requestor immediately and agree to work with Requestor to address the use or disclosure. The USRDS will destroy the Requestor's data set one year after the linkage is complete unless otherwise specified by the Requestor in the research proposal.

D. The Requestor shall not combine or link the data provided with any other collection or source of information that may contain information specific to individuals on the file, except where a waiver of authorization has been approved by the Requestor's IRB/Privacy Board.

E. The Requestor shall not use the Data for purposes that are not related to biomedical research, cost-effectiveness, economic and/or other epidemiological research. Purposes for which the Data may not be used include, but are not limited to,
   - the identification and targeting of under- or over-served health service markets primarily for commercial benefit
   - the obtaining of information about providers or facilities for commercial benefit
   - insurance purposes such as redlining areas deemed to offer bad health insurance risks
   - adverse selection (e.g., identifying patients with high risk diagnoses)

   Any use of the Data for research not in the original proposal must be approved by the USRDS Project Officer (PO).

F. The Requestor shall not publish or otherwise disclose the data in the file to any person or organization unless the data have been aggregated (that is, combined into groupings of data such that the data are no longer specific to any individuals within each grouping), and no cells (aggregates of data) contain information on fewer than ten individuals or fewer than five providers or facilities. The Requestor shall not publish or otherwise disclose data that identify individual providers or facilities, or from which such identities could be inferred. However, the Requestor may release data to a contractor for purposes of data processing or storage if (1) the Requestor specified in the research plan submitted to the USRDS Project Officer that data would be released to the particular contractor, or the Requestor has obtained written authorization from the PO to release the data to such contractor, and (2) the contractor has signed a data release agreement with the PO.

G. A copy of any aggregation of data intended for publication shall be submitted to the PO for review for compliance with the confidentiality provisions of this agreement prior to submission for publication and, if not approved, shall not be published until compliance is achieved. The PO must respond within 30 days.

H. Appropriate administrative, technical, procedural, and physical safeguards shall be established by the Requestor to protect the confidentiality of the data and to prevent unauthorized access to it. The safeguards shall provide a level of security outlined in OMB Circular No. A-130, Appendix III — Security of Federal Automated Information System, which sets forth guidelines for security plans for automated information systems in Federal agencies.

I. No copies or derivatives shall be made of the data in this file except as necessary for the purpose authorized in this agreement. The Requestor shall keep an accurate written account of all such copies and derivative files, which will be furnished upon request to the PO. The USRDS data files covered in this data use agreement may be retained by the Requestor until the date specified by the PO in the approval letter, at which time Requestor may request renewal of this data use agreement to extend the retention period to comply with legal or institutional recorDKeeping requirements or to maintain the integrity of the research or research publications. If at any time during the data retention period the DUA between USRDS and CMS is canceled, the Requestor will be contacted to destroy the files in their possession. At the completion of the activities in the research plan, the file(s) and any derivative files and copies shall be destroyed. At that time the Requestor will inform the USRDS and the PO in writing that the files have been destroyed.

J. For the purpose of inspecting security procedures and arrangements, authorized representatives of the NIDDK and/or of CMS will, upon request, be granted access to premises where the Data are kept.

K. The following USRDS data file(s) is/are covered under this Agreement.

Name of Data file(s) requested (e.g. Core, Institutional claims, etc)	Year(s) if applicable

REQUESTOR SIGNATURE: _____

_____
Authorized signatory (name, title & date)

_____
Requestor address

_____
Requestor telephone number

READ AND ACKNOWLEDGED:

_____	_____
Requestor/investigator signature	Print name & date
_____	_____
Requestor/investigator signature	Print name & date
_____	_____
Requestor/investigator signature	Print name & date

USRDS Project Officer - Lawrence Y. C. Agodoa, MD, NIDDK, NIH or Paul W. Eggers, PhD, NIDDK, NIH

_____
USRDS Project Officer signature & date

June 2012

Appendix B: Medicare Claims

Introduction

The USRDS Coordinating Center created files from CMS billing data to incorporate into the USRDS database. These files include claims for some patients who are not included in the SAF.PATIENTS file and claims for some patients before the start of ESRD. These cases can be identified and handled by merging the claims files with SAF.PATIENTS, as discussed under the heading "Patients and Time Periods Included."

CMS Data Sources

Medicare claims are of two types: physician/supplier claims for all of Medicare Part B, and institutional claims primarily for Part A. Some Part B claims, however, are institutional claims, notably those for outpatient dialysis. The structure and content of the two types of claims are different, as are the files derived from them.

The institutional claims files are obtained from the CMS SAFs, and the physician/supplier data from the 100% National Claims History nearline file. Information on outpatient dialysis and hospital inpatient stays not included in the CMS SAFs is obtained from PMMIS/REBUS. Together, these sources provide data on all types of Medicare bills. The following CMS SAFs are used:

- Inpatient
- Outpatient
- Skilled Nursing Facility
- Home Health Agency
- Hospice

For institutional and physician/supplier claims files, data for a year is frozen at the end of the following June, so claims submitted after June of the year following the year of service are not included. All data are resolved to final bills, with duplicates and correction transactions resolved into a single final bill for the service in question.

For 1977 through 1990 the PMMIS/REBUS system provides an alternate source of data on hospital inpatient stays and outpatient dialysis, but it includes no charge or payment data. The inpatient data include diagnosis and procedure codes, and outpatient data include summaries of dialysis claims by calendar quarter and provider. This is the only source for data from before 1989, the year in which the CMS SAFs start. Starting with 1991, data from PMMIS/REBUS is used only when a matching hospital stay or dialysis record is not in the CMS SAFs. SAF data are given preference because of their greater detail. However, because these files contain no data for claims processed by CMS after the June following the year of service, some claims are missed. PMMIS/REBUS data are included in the Institutional Claims and Institutional Claim Details Files and can

be distinguished by the value of the HCFASAF variable (M or Q). CMS SAFs and Part B physician/supplier data both begin in 1991, and extend through the last date shown in Table A.3. Data for a given year usually become available in August or September of the following year, and are based on claims processed through June.

Bills submitted or finalized after the cutoff date are included in the SAF for the following year. When analyzing claims, it is important to realize that all claims contained in the SAFs for a given year may not have been incurred in that year, while some claims incurred in a given year may appear in the SAFs for the following year. Because the service dates of the claim correspond to the actual dates of service, they should be used to determine inclusion in analyses, not the calendar year of the SAFs. As the reporting window is 18 months for January claims and only six months for December claims, data are likely to be less complete as a year progresses.

Patients and Time Periods Included

The Medicare claims files contain data for some patients not included in the SAF.PATIENTS file. When the USRDS database is updated, all claims for all patients who show an indication of having ESRD are retrieved from the CMS database. Some patients are then filtered out, and not included in SAF.PATIENTS or the USRDS analyses. This procedure allows the USRDS Coordinating Center itself to exclude data, rather than request them anew from CMS should they be needed later. Patients may be filtered out because of problems with the data, as when two patients have the same Medicare ID or Social Security number, or a patient's listed birth date comes after the death date. In other cases, too little information is available to establish the presence of ESRD or a date of first ESRD service. Sometimes a person filtered out one year passes the filters the next year because data problems are resolved or new data confirm that the patient has ESRD.

Researchers need to decide whether to include the claims for these patients in their analyses. The claims can be excluded by merging the claims file with SAF.PATIENTS by USRDS_ID and selecting only patients who appear in SAF.PATIENTS.

The USRDS database also includes pre-ESRD claims for patients who were entitled to Medicare due to age or disability before they developed ESRD. Because these data are not available for all patients, and because it is likely that patients entitled to Medicare before ESRD are systematically different from those not entitled, analyses of these data must be designed with care.

To obtain claims from the ESRD period only, merge the claims file with PATIENTS to identify the first service date, and select only those claims occurring on or after this date. It is up to researchers to determine how or whether to include claims that straddle the first service date.

Basic File Structure

Institutional claims are for hospital inpatient stays, hospital outpatient services, most dialysis, skilled nursing facilities, home health agencies, and hospices. Dollar amounts are available in the Institutional Claims file. The Institutional Claims Details file contains diagnostic and procedural codes that can occur a number of times for each claim. For many analyses this file is not needed.

Physician/supplier claims are bills covering physician services and medical supplies. They account for approximately 80% of the claims but only 20% of the dollars. One diagnostic and one procedural code can occur on each physician/supplier claim, which is essentially a line-item record. One visit to a physician can generate multiple claims records.

While there are only minor differences in the structure of the data included in the five institutional claim types (hospital inpatient, hospital and freestanding outpatient, hospice, home health agency, skilled nursing facility), the structure of the physician/supplier claims is substantially different from that of the institutional claims.

Institutional claims are submitted on Part A claim forms, which have a large header portion followed by variable length trailers. Possible trailer fields include diagnoses, procedures, and revenue centers. Physician/supplier claims have a simpler header portion and fewer trailer fields, including the revenue center with a CMS Common Procedure Code Standard (HCPCS) procedural code. Unlike the International Classification of Diseases, Ninth Edition, Clinical Modification (ICD-9-CM) procedural codes on the institutional claims, which primarily record invasive surgical procedures, HCPCS codes record all procedures performed by physicians (e.g., patient histories) and all supplies, ranging from Band-Aids to dialysis machines.

Institutional Claims

Effective on January 1, 2008, CMS required an important change in billing requirements for ESRD facilities. This change in turn had a substantial effect on the content of the institutional claims files that the USRDS receives from CMS for ESRD patients. The USRDS Institutional files for years prior to 2008 will not change, but USRDS Institutional files for calendar year 2008 and later will undergo some changes in order to accommodate the changes in the Institutional files obtained from CMS. In this guide, descriptions of Institutional files for calendar years before 2008 will be followed by a description of any differences in the 2008 files compared to prior years.

The CMS mandated changes in billing requirements for ESRD facilities for 2008 essentially require ESRD facilities to report each separately billable service (e.g., dialysis, ESA administration) as a separate revenue center line item with a date of service for each service. In the past, ESRD facilities were able to bill an entire month of each separately billable service as a single line item (i.e., 1 line item for dialysis, 1 for EPO, 1 for IV Iron, etc.). Since January 1, 2008 ESRD facilities are required to report each service separately, so a typical month of dialysis would require 13 separate line items (as opposed to 1 line item in previous years), 13 separate line items for EPO administrations (instead of 1), and so forth. The practical result of this change (which was actually phased in over the entire 2007 calendar year) is an enormous increase in Revenue Center Detail records, from 76 million in 2006 to over 175 million in 2008. Meanwhile, the number of other detail records showed the usual yearly increases (from 39 million to 43 million over the same time period). The increased detail in ESRD billing may be useful for studying patterns of utilization, so the USRDS will make these additional Revenue Line Items available to researchers. The Revenue Center details will no longer be included in the Institutional details file, but rather in a separate Revenue Center detail SAF. CMS also altered the method for ESRD facilities reporting of EPO administration and dosage. These changes will be described in the section on EPO variables.

A "claim" file and a "claim detail" file, and starting with calendar year 2008, a "revenue detail" file, are created from the institutional files. The details can be linked back to the claims. The Institutional Claims File has one record per claim, with a claim generally representing a single instance of service, such as a hospital inpatient stay, an outpatient surgery, or a month of dialysis. Dollar values for total charges and payment amounts are stored in the claim file, which also shows the type and number of dialysis sessions included in the claim. Data in the Institutional Claims File allow researchers to determine dialysis treatment modality over time, compute hospitalization rates, and determine aggregate costs by time period and type of cost. These data are sufficient for many research studies and most USRDS products. Analyses of particular diagnoses, procedures, or revenue centers require the claims detail files. Tables B.1 and B.2 show the variables in the Institutional Claims, Claims Detail, and Revenue Detail files. The claims are uniquely identified by a compound key consisting of four variables: USRDS_ID, CLM_FROM, HCFASAF, and SEQ_KEYC; which are used to uniquely link claim records among the Institutional Claims, Claims Detail, and Revenue Detail files. The records in all institutional files are sorted by this compound key. The derivation of the dialysis and EPO variables on the Institutional Claims record is described below under Revenue Center Details.

Medicare Payment Variables

CLM_TOT is the total amount billed for the claim, while CLM_AMT is the amount actually paid by Medicare. For inpatient and skilled nursing facility claims, the cost also includes an amount for the CMS pass-through payments for items such as indirect medical education, capital, and kidney acquisition for transplants. To obtain this pass-through payment amount, multiply the per diem amount (PER_DIEM) by the count of

covered days (CVR_DCNT). In addition to these overall amounts, the billed amount for dialysis and for EPO are provided by the variables DIALCASH and EPOCASH.

Dialysis Variables

The variable RXCAT indicates the type of dialysis, if any, included in a claim. RXCAT is derived from DIALREVC and DIALCRC, which come from the Revenue Center and Claim Related Condition details, as described below. DIALSESS is the UNITS value from the Revenue Center detail which indicates dialysis. For in-center hemodialysis, this generally indicates a plausible value for the number of dialysis sessions. For other types of dialysis, particularly CAPD and CCPD, this may indicate the number of days. DIALCASH is REV_CH from the Revenue Center detail and is the provider's billed charge rather than the Medicare payment. The Revenue Center and Claim Related Condition details, which indicate dialysis, are not included in the Institutional Claim Details file prior to 2001, unless the claim has multiple details which indicate dialysis. All dialysis revenue line items are included in the revenue center details file starting with calendar year 2001.

EPO Variables

Summary variables are provided for EPO treatments covered by a claim. EPO treatments are identified by Revenue Center codes 0634 and 0635 on a Revenue Center detail. For claims prior to 2008, the variable for number of EPO administrations (EPOADMIN) is the UNITS variable from the Revenue Center detail, while the variable for EPO payments (EPOCASH) is the REV_CH variable. If the claim has multiple Revenue Center details indicating EPO, the EPOADMIN and EPOCASH are summed over these details. The Revenue Center details from which these variables come are retained in the Institutional Claim Details SAF starting with CY2001. The variables for the dose of EPO (EPODOSE) come from a Claim Related Value detail with code '68', and the variable for hematocrit (HCRIT) comes from a Claim Related Value detail with code '48'. The Claim Related Value details from which these variables come are not retained in the Institutional Claim Details SAF. CMS mandated line-item billing for ESRD facilities effective January 1, 2008. Facilities are required to report each EPO administration as a separate line item, and to report the EPO dose administered in the UNITS field as a multiple of 100 (e.g., a dose of 5,000 units would be reported as 50 in the UNITS field. Value code 68, the total EPO dose, is no longer required. Also starting with the 2008 claims, a summarization of dARBepoeitin (DPO) claims, identified by HCPCS codes, is also included on the claim.

Institutional Claim Details

The Institutional Claim Details file includes a variety of details about each claim. The records in this file can be linked back to the corresponding claim in the claims file. There may be none, one, or many records for each type of detail for a particular claim.

- ICD-9-CM diagnosis codes
- ICD-9-CM procedure codes
- CMS revenue center codes (line item, for years prior to 2008)
- HCPCS procedure codes (with line item)
- CMS claim related condition codes
- CMS claim related value codes

Table B.2 shows the variables appearing in the Institutional Claim Details and Revenue Center Detail files. There can be any number of Institutional Details records for each Institutional Claims record. The claim detail files are sorted by the same four-part compound key as the Claims file, so that this key can be used to link the files. The multi-file structure is a solution to the problem of a number of important data items that appear none, one, or many times on a given claim. Hospital inpatient stay claims, for example, always have DRG codes, but other types of institutional claims never have this code. All claims should have at least one ICD-9-CM diagnosis code, but they may have up to ten. A hospital inpatient claim probably uses one or more ICD-9-CM surgical procedure codes if the stay involved surgery, but may also have revenue center details which specify procedures using HCPCS and/or revenue center codes, and an outpatient claim is more likely to specify procedures using revenue center codes with HCPCS codes. Using a master and detail files creates a simple structure easily manipulated in SAS. Examples of useful value code details (CDTYPE=V) are height and weight, which are required elements for ESA and dialysis claims beginning with the 2007 claim

set. Code A8 indicates that the value of the UNITS variable is patient post-dialysis weight (in kilograms), while code A9 indicates that the UNITS variable holds the patient height (in centimeters).

Revenue Center Details

The Revenue Center details are the source of a number of important variables. The Revenue Center details correspond to the Revenue Center "trailers" on the CMS SAF records. A record "trailer" is a section of the file record that can appear a variable number of times; the number of occurrences is indicated by an additional variable resulting in records that vary in length depending upon the amount of data present. The CMS SAF records have nine types of trailers, making the record structure quite complex. The Revenue Center details (or record trailers) provide data about the breakdown of the total charges into charges from "each cost center for which a separate charge is billed (type of accommodation or ancillary)." A cost center is a division or unit within a hospital (e.g., radiology, emergency room, pathology). Each Revenue Center detail contains a variable for the amount charged (REV_CH in the Institutional Claims Detail file), and one detail, while Revenue Center code '0001,' is the sum of all of the REV_CH for all other Revenue Center details for that claim. To test the consistency of the Revenue Center Details and the CLM_TOT variable, The Coordinating Center examined the original CMS SAF records for 10,000 inpatient and 10,000 outpatient claims. In all cases the '0001' Revenue Center amount was the sum of the other Revenue Center amounts. In about 3% of the inpatient records, however, CLM_TOT was greater than this sum. Beginning in late 2000, CMS began providing a field in the revenue trailers called revenue center payment amount, which corresponds to the payment amount for each revenue center trailer for all outpatient claims. This field is included as an additional variable (REVPMT) in the Institutional Detail SAF, starting with calendar year 2001. This value allows researchers to more accurately determine the payment amount for individual types of Revenue Center services (such as Laboratory service, EPO, and dialysis) billed on outpatient claims. One caveat for using this variable is that the REVPMT summed overall Revenue Center Details for a given claim will not always agree with the CLM_AMT variable contained in the Institutional Claim SAF for that claim. Our analysis showed that the sum of REVPMT over all outpatient claims exceeded the CLM_AMT for all outpatient claims by approximately 3%. The SAS format $REVCEN gives labels for the Revenue Center codes (the CODE variable on records with CDTYPE = 'R'). The Revenue Center details are the source for the dialysis and EPO variables on the Institutional Claims file. Codes 0800-0809 and 0820-0889 indicate the type of dialysis (DIALREVC). UNITS provides the number of dialysis sessions (DIALSESS), and REV_CH provides the dialysis charges (DIALCASH).

DIALCASH should be treated with caution because its use may be inconsistently defined; it is not clear if the value is the charged amount or the CMS allowed charge, and definition of the value may vary from institution to institution. When a claim has only one dialysis Revenue Center code, as is usually the case for years prior to 2001, a Revenue Center detail record is not produced because the relevant data items are recorded on the Institutional Claim record. If a claim has multiple Revenue Center details indicating dialysis, the dialysis variables are derived from the first Revenue Center code encountered, giving precedence to the more specific codes. In this case, a detail record is created for each Revenue Center detail on the claim so users have the opportunity to interpret the multiple details. Other Institutional Revenue Center details are of lesser interest unless a HCPCS code is included indicating a more specific service. A code showing that a claim is for laboratory services, for example, frequently includes a HCPCS code indicating the specific test performed. Revenue Center Detail records are included regardless of the presence or absence of a HCPCS code. Before calendar year 2008, Revenue Center Detail records are included with other Claim Detail records, and after 2008 are placed in a separate file. For many analyses, the Revenue Center Detail records may not be required.

Institutional Claims Detail File Variables

CDTYPE, Code

CDTYPE indicates the type of code contained in the CODE variable. Both variables are present on every record, while the remaining variables are not present for some CDTYPES.

The SAS format $CDTYPEI indicates the meaning of each CDTYPE.

UNITS

Use of the UNITS variable varies with CDTYPE. When CDTYPE = "P" (ICD-9-CM Surgical Procedures), UNITS is a value created by the USRDS to indicate when the surgical procedure was performed, and time is expressed as the number of days from the date given by CLM_FROM, with CLM_FROM counted as 1. A value of 1 for UNITS indicates that the procedure was performed on the date given by CLM_FROM, and 2 indicates the day after CLM_FROM. When CDTYPE = "R," UNITS is described in the CMS file documentation as "a quantitative measure (unit) of services provided to a beneficiary associated with accommodation and ancillary revenue centers" described on an institutional claim. Depending on the type of service, units are measured by number of covered days in a particular accommodation, emergency room visits, clinic visits, dialysis treatments (sessions or days), outpatient therapy visits, and outpatient clinical diagnostic laboratory tests. The revenue center code or the HCPCS code indicates the type of service. Because the meaning of UNITS varies greatly, the variable must be used with caution. When using this variable, tabulate the distribution of values over the records being analyzed to ensure that the values look correct. When CDTYPE = "I", UNITS has a value of 1 or 0, where 1 indicates that this was the primary diagnosis for this claim and 0 indicates that it was a secondary diagnosis. The claim details are not necessarily sorted with the primary diagnosis first.

REV_CH

REV_CH occurs only on Revenue Center details (CDTYPE= "R") and indicates "the total charges (covered and non-covered) for all accommodations and services (related to the revenue code) for a billing period before reduction for the deductible and COInsurance amounts and before an adjustment for the cost of services provided." REV_CH corresponds in concept to the CLM_TOT variable on the Institutional Claims file, as discussed above under Revenue Center Details.

HCPCS

The CMS Common Procedure Coding Standard (HCPCS) "is a collection of codes that represent procedures, supplies, products, and services which may be provided to Medicare beneficiaries and to individuals enrolled in private health insurance programs." The HCPCS code occurs only on Revenue Center (CDTYPE = R) details but may not be present on all such records. HCPCS are an extension of the American Medical Association CPT-4 codes. Codes for certain pharmaceuticals, laboratory procedures, durable medical equipment, and radiology procedures are added to the CPT-4 codes to form HCPCS.

URR_CD

Starting in 1998, CMS began requiring the reporting of Urea Reduction Ratios (URRs) on outpatient hemodialysis claims. The URR is reported as a range that reflects the results for the month being billed. This information appears as a formatted value in the Revenue Center Details for hemodialysis claims.

Physician/Supplier File

In the physician/supplier file, a claim does not necessarily correspond to a logical instance of service, but is more likely to represent all services provided to a patient during the provider's billing period. Because procedures and costs are specified at the claim line item level, the file is constructed as a line item file, with one record per claim line item. For physician claims, the from/through dates can be used to identify a series of line items associated with a single visit. For supplier claims, however, the instance of service is more difficult to define. Bills for home dialysis dialysate, for example, specify the quantity and delivery date of the dialysate but not the time period over which it is to be dispensed.

Table B.3 shows the variables in the Physician/Supplier Claims files. It contains two file types, identified by the value of the CDTYPE variables. CDTYPE = B indicates a physician/supplier line item, which has data for all the other variables. The DIAG variable indicates the diagnosis code associated with this line item. CDTYPE = I indicates that this record contains only a diagnosis code (DIAG); in this case the diagnosis code is associated with all line items on the claim. Payment variables for these records should be missing.

Variables MOD1-MOD4 are included to further identify the type of service billed on the line item. They are used in conjunction with the HCPCS/CPT code on the line item, and their meaning can be found in the Current Procedural Terminology codebook and the HCPCS Level II codebook.

The physician/supplier specialty code (SPCLTY) can be useful for untangling the bills for a specific surgical procedure. The principal surgeon, physician surgical assistants, and anesthesiologist use the HCPCS referring to the major invasive surgery to bill for that surgery. The code for nephrologists is 39.

The place of service variable (PLCSRV) indicates where the service was rendered. It can be used to distinguish between inpatient and outpatient services and between home dialysis and in-unit dialysis supplies. The value 6 refers to an ESRD treatment center.

The CMS service code variable (HCSRVC) can be used to distinguish between the principal surgeon and assistants. The value for immunosuppressive drugs is G, for renal supplier in the home L, for monthly capitation payment (dialysis) M, and for kidney donor N.

Three cost fields appear on each physician/supplier line item: submitted charges (SBMTCH); allowed charges (ALOWCH), which are the lower of prevailing, customary, or actual as determined by CMS; and the payment amount (PMTAMT), the amount paid to the provider and/or beneficiary after deductible and co-insurance amounts have been paid for the services included as a line item on a physician/supplier claim.

B.1 Variables in the Institutional Claims SAF File

Name	Type	Length	Format	Comment
CLM_AMT	Num	8	BEST	Medicare payments
CLM_FROM	Num	8	MMDDYY	From date of service
CLM_THRU	Num	8	MMDDYY	Service through date
CLM_TOT	Num	8	BEST	Total charges
CVR_DCNT	Num	8	BEST	For Inpatient and SNF claims, the Medicare covered day count. See PER_DIEM.
DIALCASH	Num	8	BEST	Claim amounts for dialysis
DIALCRC	Char	5	$DIALCRC	Claim related condition for dialysis. Right digit of the primary claim related condition code indicates dialysis. See text for determining dialysis modality
DIALREVC	Char	5	$DIALRVC	Revenue center code for dialysis. Right two characters of the primary Revenue Center code indicate dialysis. See text for determining dialysis modality
DIALSESS	Num	8	BEST	Dialysis treatments based on the UNITS variable for the Revenue Center code indicated by DIALREVC. If multiple occurrences of Revenue Center code, UNITS are summed ACROSS occurrences
DISCSTAT	Char	2	$DRG_DES	Discharge status
DPOADMIN*	Num	8		Number of DPO administrations
DPOCASH*	Num	8		Total Charge for DPO on this claim
DPODOSE*	Num	8		
DRG_CD	Char	3		Diagnosis Related Group Code. Inpatient and SNF claims only. DRG 302 indicates kidney transplant
EPOADMIN	Num	8	BEST	Number of EPO administrations
EPOCASH	Num	8	BEST	Total Charge for EPO on this claim
EPODOSE	Num	8	BEST	Total EPO dosage (units) for this claim
HCFASAF	Char	1	$HCFASAF	CMS SAF source of this bill. Format: D Dialsis, H-Home health, I Inpatient. N Skilled nursing facility, O Outpatient, P Physician/supplier, S Hospice
HCRIT	Num	8	BEST	Hematocrit reported on ESA claim
HGB*	Num	8		Hemoglobin reported on ESA claim
PER_DIEM	Num	8	BEST	For inpatient and SNF claims, the HCFA pass-through payments. Hospital is reimbursed separately for PER_ DIEM times CVR_DCNT.
PRM_PYR	Char	1	$PRPAYR	Primary payer for this bill
PROVUSRD	Num	8	BEST	USRDS provider ID number for dialysis providers (HCFA SAF = D), blank for others
RXCAT	Char	1	$RXCATIC	Dialysis treatment modality
SEQ_KEYC	Char	2		Sequence # to ensure unique key
USRDS_ID	Num	8	BEST	Patient: Patient ID Used To Cross Reference To Other USRDS SAF Files

*For calendar years greater than 2007 only.

B.2 Variables in the Institutional Claims Detail SAF File

Name	Type	Length	Format	Comment
USRDS_ID	Num	8		USRDS_ID
CLM_FROM	Num	8	Date7.	Claim from date
HCFASAF	Char	1	$HCFASAF.	SAF source of bill
SEQ_KEYC	Char	2		Sequence number to ensure unique key
CODE	Char	5		Diagnosis, procedure, or revenue code
CDTYPE	Char	1	$CDTYPEI.	Defines type for variable 'CODE'
HCPCS*	Char	5		HCPCS code
REV_CH*	Num	8		Revenue center total charge
UNITS	Num	8		Varies by detail type. See documentation
URR_CD*	Char	8	$URRFMT.	Urea reduction ratio for reported hemo bills
REVPMT	Num	8		Line item payment amount
REV_DT**	Num	8	Date7.	Date of service for line item

*For Revenue Center Detail file only, starting with calendar year 2008; prior to 2008 these variables are included in the details files, but are always missing values for all details except Revenue Centers.

**Revenue Center Detail file, new starting with calendar year 2008.

B.3 Variables in the Physician/Supplier File

Name	Type	Length	Format	Comment
ALOWCH	Num	8		Allowed charges
CDTYPE	Char	1	$HCCDTYP	Line Item Type
CLM_FROM	Num	8	MMDDYY	From Date of Service
CLM_THRU	Num	8	MMDDYY	Thru Date of Service
DIAG	Char	5		ICD-9-CM Diagnostic Code
HCFASAF	Char	1	$HCFASAF	HCFA SAF Source of this Bill
HCPCS*	Char	5		HCPCS code
HCSRVC	Char	1	HCFASVC	HCFA Service Code
MOD1	Char	2		HCPCS/CPT 1st Modifier
MOD2	Char	2		HCPCS/CPT 2nd Modifier
MOD3	Char	2		HCPCS/CPT 3rd Modifier
MOD4	Char	2		HCPCS/CPT 4th Modifier
PLCSRV	Char	2	$PLACESV	Place of Service
PMTAMT	Num	8		Claim Payment Amount
PYRCOD	Char	1	$PRPAYR	Primary Payer Code
SBMTCH	Num	8		Submitted Charges
SPCLTY	Char	2	$PROVSP	Provider Specialty Code
SRVCCT	Num	8		Total Number Line Item Services
USRDS_ID	Num	8	BEST	USRDS_ID

Appendix C: Statistical Methods

Methods for Event Rate Calculation

Transplant rates and rates of ESRD incidence, prevalence, and mortality are often estimated overall or by groups. Some are based on population size (incidence and prevalence) or follow-up time (mortality). Rates can be direct estimates based on the data observed (observed rate), they can be estimated based on statistical models, or they can be adjusted for patient case-mix.

Observed Event Rate

The observed event rate calculation is straightforward. It is the number of events divided by the population size or the total follow-up time and sometime it need to multiple a number based the size of the rate or the unit people are used to use. For example, if state A had 1600 incident ESRD patients in 2004 and the state population size was 6,400,000 people, the incidence rate of state A in 2004 is:

r = 1600/6,400,000 x 1,000,000 = 250 per million people

The rate is multiplied by 1,000,000 because otherwise the rate is a very small number. If the total follow-up time of these ESRD patients in 2004 is 1100 patient-years, and 150 patient die in the incident year, the 2004 death rate of the incident patients in state A is:

r = 150/1100 x 1000 = 136.4 per thousand patient-years.

The standard error of the rate is usually estimated by $\sqrt{(r/d)}$, where r is the rate and d is the denominator in the rate calculation.

Model-Based Event Rate

Using the observed rate as the estimate of the real rate usually works well for the population level or for big groups. When it is necessary to calculate event rate for subgroups and some subgroups are small, the observed rate may not be an accurate estimate of the real rate, and a statistical model will be necessary. As an example, we can calculate the death rate of each age (every 5 years), race (white, black, Asian, Native American, and other), and gender combination group. We usually do not have enough data to calculate the death rate of the (age<5, female, Native American) group. If the Poisson model is the right model to fit the number of deaths, we can fit a Poisson model with the number of deaths as the response variable, age, race, and gender as independent variables, and the natural logarithm of follow-up time as the offset, i.e. $y \sim \text{Poisson}(t \cdot \exp[x\beta])$, where y is the number of deaths, x is the corresponding values of age, race, gender, and maybe some interactions, as well as 1, which corresponds to the intercept, and β is the corresponding coefficients. Then the death rate of a combination group will be $\exp[x\beta]$, where the x corresponds to the combination group. For a rate calculation, a logistic regression model, a Poisson regression model, and a Cox proportional hazard regression model can be used. The closer the model is to the saturated model, the better the rate estimate conditional on overfitting. In the ESRD Annual Data Report, both a Poisson model and a Cox model were used with/without random effects.

Adjusted Event Rates

When comparing the event rate of two groups with different patients, the event rate comparison may not be meaningful. For example, we can compare the death rates of Groups A and B, with Group A having far older patients than Group B. Group A has a higher death rate than Group B, perhaps because Group A made patients have a higher risk, or simply because the older patients have a higher risk than the younger patients. To figure out if Group A made people have a higher risk we need to adjust the patient age when calculating the rates. When comparing rates adjusted for a particular factor, any remaining observed differences between groups cannot be attributed to confounding by that factor. The two main adjustment techniques are the direct method and the indirect method; only the direct method allows the rates to be compared (Fleiss 1981).

Direct Adjustment

If each group has many categories (for example, combination groups defined by age, race, and gender), the direct adjusted rate is derived by applying the observed category-specific rates in the group to a single standard, or reference, population. This weighted average of the observed category-specific rates, with the weights taken from a standard reference population, provides for each group a single summary rate that reflects the numbers of events that would be expected if the group had identical distribution of the characteristic of interest as the reference. This makes the comparison valid, but the values of the adjusted rates are not meaningful. Adjusted rates and their explanation are reference dependent. Because the reference population must have the same categories as all of the groups, it must be chosen with caution.

The disadvantages of this method are:

- If one category in a group is small the corresponding category-specific rate will be unstable; this may make the adjusted rate for this group unstable as well.
- If one category is empty, the adjusted rate cannot be calculated for that group.

Model-Based Adjustment

Because of the disadvantages of the direct adjustment method, a model-based adjustment method is necessary. Using an appropriate model to calculate category-specific event rates as described in the Model-Based Estimates above, we can apply the direct adjustment method based on the model-based category-specific rates, instead of observed category-specific rate [Liu, 2006].

Survival Analysis

The most commonly used methods for survival analyses are the Kaplan-Meier method, log-rank test, and the Cox proportional regression model.

The Kaplan-Meier method is used to estimate the survival probability over time. The plots of the survival estimates are used to intuitively compare groups on patient survival. The log-rank test is a method for testing the hypothesis that there is no difference in survival probabilities over a time among groups. The Cox regression model is the most widely used method for survival analysis. It can be used to compare risk of mortality or other events among groups and to find risk factors. Details for these methods can be found in Kalbfleisch and Prentice 1980. More advanced methods can be found in Therneau and Grambsch 2000.

Bibliography

Fleiss JL. Statistics Methods for Rates and Proportions, New York: John Wiley and Sons, 1981.

Liu J, Louis TA, Pan W, Ma JZ, Collins AJ. State-level adjusted ESRD incident rates: use of observed vs model-predicted category-specific rates. Kidney Int. 2006;69(8):1459-1463.

Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data. New York: John Wiley and Sons, 1980.

Therneau TM, Grambsch, PM. Modeling Survival Data: Extending the Cox Model. New York: Springer, 2000.