

APPENDIX A
USRDS PRODUCTS AND SERVICES



INTRODUCTION

Table A.1 describes the products and services provided by the USRDS to support ESRD research and the work of the renal community.

The entire ADR is available at www.usrds.org. The site also offers color slides of figures, a PDF file of the *Researcher's Guide*, and USRDS contact information. Through this site, users can create customized data sets and regional maps. Data regarding site use are presented in Figure A.1.

DIALYSIS UNIT-SPECIFIC SMR/SHR REPORTS

From 1996 through 1999, the USRDS produced more than 2300 unit-specific reports each year, compiling information about the patients treated in each dialysis facility, and including standardized mortality ratios and standardized hospitalization ratios. These reports are now produced by the Kidney Epidemiology and Cost Center at the University of Michigan (www.med.umich.edu/kidney).

DATA REQUESTS

Making information on ESRD available to the renal community is a primary objective of the USRDS, and the CC is committed to the timely fulfillment of data requests. Many of these requests can be fulfilled using data published in the ADR or elsewhere. The CC fulfills data requests that require two hours or less of staff time without charge, usually within one week. More complex requests that require more than two hours of staff time, and requests for Standard Analysis Files and custom files, must be accompanied by a written proposal (see Table A.1), and are fulfilled only upon written approval by the NIDDK Project Officer.

DATA FILES AVAILABLE TO RESEARCHERS

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

Before 1994, all files provided to researchers were custom files created for a specific research project. 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.

The Core SAF CD-ROM contains basic patient data and is needed to use any other SAF. It includes each patient's demographic information, treatment history, limited transplant data, and all data from the USRDS Special Studies. Approximately half of the researchers using the USRDS SAFs need only the Core CD. Detailed transplant and transplant follow-up information is provided on a separate CD containing data collected by CMS and UNOS. Hospital inpatient data are provided on the Hospital CD, and Medicare payment data are available either in a full set or by individual year (Table A.2).

STANDARD ANALYSIS FILES (SAFS)

Use of the SAFs is governed by the USRDS Policy on Data Release for Investigator-Initiated Research, described on page 42. Proposals must be approved by the USRDS Project Officer, and the researcher must sign the USRDS Agreement for Release of Data, reproduced on page 44. SAF prices are listed in Table A.3.

Most SAFs provide patient-specific data. All patient identifiers (name, address, Social Security number, etc.) are removed from the files or encrypted, but data confidentiality is nevertheless a serious concern. The Agreement for Release of Data therefore includes restrictions on the use and disposition of the SAFs. The SAFs include encrypted ID numbers to allow patient data from multiple SAFs to be merged when needed.

CORE CD

The USRDS has carried out a number of Special Studies. Topics are approved by the NIDDK, with recommendations from CMS, the USRDS Scientific Advisory Committee, the ESRD networks, and the Renal Community Council. For each study, design and sampling plans were developed, samples were selected, and data collection forms and instructions drafted, tested, and finalized. The main studies are summarized below.

The Core CD contains the most frequently used SAFs, including those from the USRDS Special Studies, and is needed for use of the Transplant CD, the Hospital CD, or any CD based on Medicare claims data. Core CD files are the following:

Patient (file name: PATIENTS)

Contains one record per patient in the USRDS database, and gives basic demographic and ESRD-related data.

Residence (file name: RESIDENC)

Provides a longitudinal record to the ZIP code level of each patient's place of residence.

Treatment history (file names: RXHIST and RXHIST60)

Also referred to as the Modality Sequence file; contains a new record for each patient at each change in treatment modality or dialysis provider.

Payer History (file name: PAYHIST)

Contains detailed longitudinal insurance payer information for each unique patient from the date of ESRD initiation.

Medical Evidence (file names: MEDEVID95, MEDEVID05, & MEDEVID)

The MEDEVID95 and MEDEVID05 files contain data from the 1995 and 2005 versions of the CMS End-Stage Renal Disease Medical Evidence Report (CMS-2728), respectively. The MEDEVID file combines data from both form versions with additional processing in code value conversions and consistency, specifically targeting common and similar data elements to ensure data integrity. All unique data elements on the original Medical Evidence Form (1987 version) were appropriately incorporated into other files, such as the 'Patients' and 'RXHIST' files on the Core CD, and are not included in the MEDEVID file. The 1995 version included data on comorbidity, employment status, lab values at start

of dialysis, and Hispanic ethnicity. The 2005 version added more information on treatment before starting renal replacement therapy, additional comorbid conditions, and type of vascular access used at initiation.

Transplant (file name: TX)

Contains basic data for all transplants, including graft failure date (detailed transplant data are contained on a separate transplant CD).

Transplant waiting list (file names: WAITLIST_KI and WAITLIST_KP)

Since the creation of the 2001 Core CD, basic patient demographic data and data on all unique wait-list periods for each patient appear in the transplant wait-list files. These files include one or more records for each patient wait-listed for

kidney (KI) or simultaneous kidney-pancreas (KP) transplants from the UNOS transplant wait-list, with each record representing a unique wait-list period.

Transplant waiting list sequence (file names: WAITSEQ_KI and WAITSEQ_KP)

These condensed files list unique waiting periods per patient per transplant center. They are the condensed waiting list date sequence files for kidney (KI) and simultaneous kidney-pancreas (KP) transplants. Both files are center-specific and therefore contain beginning and ending dates for each patient at each transplant center at which they are registered on the waiting list. Investigators who wish to investigate the raw UNOS kidney and simultaneous kidney-pancreas waiting list files can use WAITLIST_KI and WAITLIST_KP. Refer to Section 5, Transplant Process and Outcomes, for a more detailed description of the waiting list sequence files.

Table A.1 USRDS Products and Services for ESRD Researchers and the General Renal Community. Products are provided without charge except as noted.

Reports and 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. Material from the ADR is also published in the American Journal of Kidney Disease.
ADR CD-ROM	Contains the text and graphics of the ADR, data tables, color PowerPoint® slides, and the <i>Researcher's Guide</i> .
<i>Researcher's Guide</i> to the USRDS Database	Provides a detailed description of the USRDS database and of the USRDS Standard Analysis Files, and is the basic reference for researchers who use USRDS data files.
www.usrds.org	Offers PDF files of ADR chapters, reference tables, and the <i>Researcher's Guide</i> ; 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 available to help new users.

Requests for data

Two-hour data requests	Questions and data requests not addressed directly by the ADR can be addressed to the Coordinating Center; those requiring less than two hours of staff time to fulfill will be processed without charge.
Extensive data requests ted	Questions and data requests that require more than 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 database to support ESRD research. A standard pricelist has been established for the files. Users must sign a Data Release Agreement with the NIDDK.
Custom data files	Custom files can be created by the Coordinating Center for projects requiring data not provided in the SAFs, at an hourly rate of \$103.97. Users must sign a data release agreement with the NIDDK.

Publications and presentations

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

Contact information

Data requests and publication orders	USRDS Coordinating Center, 914 South 8th Street, Suite S-206, Minneapolis, MN 55404. Phone 612.347.7776 or 1.888.99USRDS, Fax 612.347.5878. www.usrds.org
Data file contact	Shu-Cheng Chen, MS, schen@usrds.org

Dialysis Mortality and Morbidity Special Study Waves 1-4 (File Name: DMMS)

Each wave includes a data collection instrument for collecting core data, allowing collection of a consistent set of fundamental data for research questions that require a large sample size.

Facility (file name: FACILITY)

The CMS ESRD Annual Facility Survey is the source of data for the Facility SAF, which can be linked to the Facility Cost Report files using the USRDS provider ID. Because of this link, geographic variables that could be used to identify facilities have been deleted. The survey period is January 1 through December 31.

Facility Cost Reports (file name: FCOSIND)

The CMS hospital and independent facility cost reports for the years 1989–1995 and 1989–1993 are available as Standard Analysis Files. All geographic variables have been deleted to ensure confidentiality. The file may be linked with the Facility SAF using the USRDS provider ID; geographic analyses at less than a regional or ESRD network level, however, are not possible. Because use of these files has been minimal, data for additional years will be added only if there is sufficient demand.

Dialyzers (file name: DIALYZER)

The Case Mix Severity, Case Mix Adequacy, and DMMS Special Studies all collected information on the manufacturer that must be matched to information in the Dialyzer file to identify the manufacturer and model and characteristics such as membrane type and clearance. The data in this file come from published sources available at the time of the study. We believe these data accurately represent the dialyzer characteristics, but they should be used with caution.

TRANSPLANT CDS

Detailed transplant data are available on the Transplant CDs.

KIDNEY TRANSPLANT-CMS (FILE NAME: TXHCFA)

Includes transplant information collected by the CMS PMMIS system before 1994.

Kidney Transplant-UNOS (File Names: TXUNOS_KI and TXUNOS_KP)

Includes information on kidney (KI) and kidney-pancreas (KP) transplants collected since 1987 by UNOS, the main source of transplant data for the USRDS.

Immunosuppression at Registration-UNOS (File Name: TXIRUNOS)

Includes information on immunosuppressive drugs; collected by UNOS at the time of transplantation events.

Kidney Transplant Followups-CMS (File Name: TXFUHCFA)

Includes transplant followup reports collected by CMS before 1994; reports are completed at discharge, at six months, each year post-transplant, and at graft failure.

Kidney Transplant Followups-UNOS (File Names: TXFUUNOS_KI and TXFUUNOS_KP)

Includes followup reports on kidney (KI) and kidney-pancreas (KP) transplants collected by UNOS since 1988.

Immunosuppression at Followup-UNOS (File Name: TXIFUNOS)

Includes information on immunosuppressive drugs collected by UNOS at followup visits.

HOSPITAL CD

Hospitalization inpatient data from the USRDS database are a subset of the data in the institutional claims file. This CD includes data on hospital inpatient stays and on diagnoses and procedures during those stays. It includes no payment or cost variables.

DIALYSIS MORBIDITY AND MORTALITY CD

This CD contains files from the Dialysis Morbidity and Mortality Study and extracted data from all other SAFs for the patients in this study. All data on Medicare payments for these patients are followed to the currently reported claims year.

CASE MIX ADEQUACY CD

This CD contains the Case Mix Adequacy Special Study file and extracted data from all other SAFs for the patients in this study. All data on Medicare payments for these patients are followed to the currently reported claims year. Along with analyses related to the study itself, this file is useful for developing analyses that will later be run on the full Medicare payments files.

MEDICARE CLAIMS PAYMENT CDS

Medicare payment data on institutional claims are available for the years indicated in Table A.3. These data sets can be purchased by individual year.

Institutional claims consist of all Part A claims (inpatient, outpatient, skilled nursing facility, home health agency, and hospice) and some Part B claims, notably outpatient dialysis. All physician/supplier claims are Medicare Part B; these are about 80 percent of the claims but account for only 20 percent of the dollars.

The structure and content of the two types of claims are different, as are the files derived from them. The Institutional claims file indicates the type of claim, the dollar amounts, the DRG code, the type of dialysis (if any), and the dates of service. The institutional claims detail file contains details such as diagnosis and procedure codes. Many analyses require only the institutional claims file. Physician/supplier claims files have one record for each claim line-item, showing dollar amounts, dates of service, diagnosis and procedure codes, and type and place of service.

Table A.2 Contents of the USRDS Core Standard Analysis File CD-ROM

This file is needed to use any other Standard Analysis File. The data are provided on four CDs.

File Name	Unit of Observation	Uses
Patient (PATIENTS)	One record for each ESRD patient.	Incidence, prevalence, patient survival. Most other files must link to this file using the encrypted patient ID.
Residence (RESIDENC)	One record for each patient for each period in a different residence.	Regional analyses.
Treatment History (RXHIST)	One record for each period on one modality. A patient may have many records.	Modality distribution and treatment patterns.
Pay History (PAYHIST)	One record for each period with one payer. A patient have many records.	Effect of insurance payer on clinical outcomes.
Medical Evidence (MEDEVID95)	One record for each CMS-2728 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 CMS-2728 filed (2005 version).	ESRD first service date, initial treatment modality, comorbid conditions, patient status at start of ESRD.
Transplant (TX)	One record for each transplant event. A patient may have multiple transplants.	Transplant and transplant outcome analyses.
Transplant Waiting List (WAITLIST_KI, WAITLIST_KP)	One record per unique waitlist period per patient.	Raw waiting list data as reported by the OPTN.
Transplant Waiting List Sequence (WAITSEQ_KI, WAITSEQ_KP)	One record per continuous waitlist period per patient per transplant center.	Versions of the above files collapsed for easier use. Waiting list start and stop dates for each center at which a candidate is listed, with adjacent waiting periods that might appear in WAITLIST_KI and WAITLIST_KP collapsed.
Dialysis Mortality and Morbidity Special Study (DMMS)	Wave 1, 5670 patients. Wave 2, 4024 patients. Wave 3, 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)	7096 patients.	Comorbid conditions, adequacy of dialysis, dialysis prescription and other treatment parameters, laboratory values.
Case Mix Severity Special Study (CASEMIXS)	5255 patients.	Comorbid conditions, adequacy of dialysis, dialysis prescription and other treatment parameters, laboratory values.
Pediatric Growth Development Special Study (PEDGROW)	3067 patients.	Growth, development, and other issues relating to pediatric and ESRD patients.
CAPD Peritonitis Special Study (CAPD)	3385 patients.	CAPD and peritonitis.
Facility (FACILITY)	One record for each year facility has operated.	Merge with the treatment history, transplant, or annual summary SAFs for analyses involving provider characteristics by encrypted facility ID.
Facility Cost Reports (FCOSHOS)	One record per facility per year (1989–1995).	Costs and staffing of dialysis facilities.
Dialyzers (DIALYZER)	Information on dialyzer characteristics; match to patient dialyzer information in other files.	Relation of dialyzer characteristics to patient outcomes.
Claim Codes (CLMCODES)	One record for each diagnosis, procedure, or HCPCS code in claims files.	Frequency of occurrence of each code; a starting point for analyses using diagnosis and procedure codes.
FORMATS	All USRDS-defined SAS formats used by the SAFs.	Format library used to format values of categorical variables.

REQUIREMENTS FOR USING THE SAFS

- ◆ Computer: at a minimum, a 486 or Pentium PC. Smaller runs have been done on 486/100 PCs. The files can be converted to SAS transport format for use on any computer with access to SAS.
- ◆ CD-ROM drive: any PC with a CD-ROM drive should be able to read the SAF CDs.
- ◆ Disk storage: between 10 and 600 megabytes are needed to use the Core CD, depending on the files being used. The data on each CD require from 550 to 650 megabytes of disk storage. Additional space is needed for temporary work files and for files users create.
- ◆ Software: SAS. Files converted to SAS transport format can be used by SPSS.
- ◆ People with SAS experience: The SAF documentation provides basic information about loading the files into SAS and using them, but further work with the files requires SAS experience.

File media and formats

The SAFs are provided on CD-ROM disks as SAS files, and can be used directly by SAS on any 486 or Pentium PC with a CD-ROM reader.

The SAS format was chosen for the USRDS SAFs because it 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, from mainframes to PCs, and it is almost universally available on university computer systems. The USRDS SAFs take full advantage of the program's ability to incorporate a large amount of documentation into the file.

Researchers who require a different program format or a medium other than CD-ROM must arrange for the conversion themselves. The USRDS may be able to convert files to alternative formats or media, but at substantially greater cost.

COST

The price of the files covers the cost of reproducing and shipping the file and its documentation, administrative costs of handling the sales, and costs of technical support for researchers. Checks must be made payable to the Minneapolis Medical Research Foundation. Prices are subject to change.

DOCUMENTATION

This *Researcher's Guide* provides most of the documentation for the SAFs by including a codebook of variables in the files and a chapter on techniques for using the SAFs in SAS (Getting Started, page 7). The *Researcher's Guide* is available in PDF format at www.USRDS.org.

ACKNOWLEDGEMENT FOR USE OF USRDS DATA

Publications that use USRDS data should include an acknowledgment and the following 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 POLICIES AND PROCEDURES

Because the SAFs and custom data files contain confidential, patient-specific data, release of these files requires the approval process described here. Investigators may contact the USRDS Project Officer at the NIDDK to discuss their requests before preparing a written proposal. To request and use USRDS data files, investigators should do the following:

- ◆ Provide the USRDS Project Officer with a detailed description of the proposed investigation (outlined in Table A.4). The project summary must include goals, background data, an in-depth description of the study design and analytic methodology, and resources available for completing the project. It may be the project description from a grant or other funding application. The proposed project must comply with the Privacy Act of 1974, and the project summary should provide enough information to enable assessment of compliance. Guidelines for adherence to the Privacy Act appear in the USRDS "Agreement for Release of Data," page 44.
- ◆ Indicate which USRDS SAFs will be needed. If the USRDS SAFs cannot meet the requirements of the proposed research, the proposal must specify precisely which data elements are needed, and investigators must budget for a substantially higher cost.
- ◆ If the project is approved, return a signed copy of the USRDS "Agreement for Release of Data" to the Project Officer. The investigator and the Coordinating Center will resolve any technical questions and arrange for payment. Payment must be received before the files will be released. Checks must be payable to the Minneapolis Medical Research Foundation.

The NIH will review the project for technical merit and conformity with the Privacy Act. The Project Officer will notify the investigator(s) in writing of the outcome, and if the project is not approved will discuss reasons for the decision. The Project Officer will send the approval letter to the Coordinating Center. The process of reviewing the data request, generating the data file, and releasing the data takes the Coordinating Center about three months.

When the signed "Agreement for Release of Data" and payment for the files have been received by the Coordinating Center, the files and documentation will be prepared and sent to the investigator.

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

All publications using the released data must contain the standard acknowledgement and disclaimer presented above. The investigator is requested to send copies of all final publications resulting from this research to the Project Officer and to the USRDS Coordinating Center.

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. However, because inferring the identity of individual patients, providers, or facilities from the data in the SAFs might be possible, these data are considered confidential. The USRDS “Agreement for Release of Data” contains 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 the Centers for Medicare & Medicaid Services.

Use of these data to identify or contact patients, facilities, or providers in the files is prohibited both by USRDS policy and by the Privacy Act of 1974.

The USRDS Coordinating Center will provide data in any of the usual media (tape, disk, or hard copy). Analytical services, other than review of the proposal and preparation of the data file, are not provided under the USRDS contract, although Coordinating Center personnel may participate in analyses funded by other sources.

SAFs or other data files from USRDS Special Studies become available one year after the data have been collected, edited, and entered into the database.

Table A.4. Outline for Research Proposals Using USRDS Data

- I. Research topic title and submission date.
- II. Background information.
- III. Study design
 - A. Objectives
 - B. Hypotheses
 - C. Analytical methods
- IV. Requested data
 - A. List of Standard Analysis Files needed (specify if multiple years) or fields in custom data file.
 - B. Description of data security measures (responsible party, computer access, etc.).
 - C. Timeframe for the project.
 - D. 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 adherence to local privacy standards and USRDS and CMS privacy policies, the USRDS requires IRB approval for all research proposals except those requesting aggregate data.
- VI. Outline of estimated costs of requested data, funding source.
- VII. Agreement for Release of Data signed by all researchers.
- VIII. Information for Principal Investigator and co-investigators:
 - Name
 - Affiliation
 - Business address
 - Business phone and 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

Table A.3. Prices for USRDS Standard Analysis Files on CD-ROM
Checks must be payable to the Minneapolis Medical Research Foundation

<i>SAF CDs</i>	<i>Files</i>	<i>Price</i>	<i>Description</i>
Core	4	\$1000	Needed to use any other SAF CD.
Transplant	2	\$250	Detailed transplant data from CMS and UNOS.
Hospital	2	\$250	Derived from the institutional claims and institutional claims details CDs. Contains diagnosis and surgical procedure codes for each stay, but no cost data.
DMMS claims	4	\$500	All institutional and physician/supplier claims data for patients in the USRDS Dialysis Morbidity and Mortality Special Study. Special Study data collection form data are on the Core CD.
Case Mix Adequacy	1	\$125	Contains all institutional and physician/supplier claims data for patients in the USRDS Case Mix Adequacy Special Study, including Special Study data collection form data.
CPM/USRDS merged dataset	Variable		The CPM 2001 SAF CD set includes CPM data collections from the 1994 through 2000 surveys, with a claims history from the USRDS 2001 SAF CD. Additional years of merged CPM SAF CDs are also available. For more information, visit www.usrds.org/cpm.htm or call 1.888.99USRDS.

Medicare Claims CDs

Year	Institutional claims			Physician/supplier claims	
	Claim Files	Detail Files	Price	Files	Price
pre-1989*	1	1	\$250		
1989	1	1	\$250		
1990	1	1	\$250		
1991	1	2	\$375	4	\$500
1992	1	2	\$375	4	\$500
1993	1	2	\$375	4	\$500
1994	1	2	\$375	5	\$625
1995	1	3	\$500	5	\$625
1996	1	3	\$500	6	\$750
1997	1	3	\$500	7	\$875
1998	1	3	\$500	7	\$875
1999	1	3	\$500	7	\$875
2000	1	5	\$750	7	\$875
2001	1	5	\$750	7	\$875
2002	1	6	\$875	8	\$1000
2003	1	6	\$875	9	\$1125
2004	1	7	\$1000	9	\$1125
2005	1	8	\$1125	10	\$1250

*Pre-1989 files include only hospital inpatient stays and quarterly summaries of outpatient dialysis with no cost data.

†2006 claims will be available in early 2008; prices subject to change.

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

Project title _____

In this agreement, “Recipient” 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 Recipient with tapes, disks, and/or hard copies containing data extracted from the USRDS research database.
- 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 Recipient.
- C. The Recipient shall not use the data to identify individuals on the file.
- D. The Recipient 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 Recipient shall not use the data for purposes that are not related to biomedical research, cost-effectiveness, or other economic 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 Recipient 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 Recipient shall not publish or otherwise disclose data that identify individual providers or facilities, or from which such identities could be inferred. However, the Recipient may release data to a contractor for purposes of data processing or storage if (1) the Recipient specified in the research plan submitted to the USRDS Project Officer that data would be released to the particular contractor, or the Recipient 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 Recipient 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 Recipient 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 Recipient until _____. At the completion of the activities in the research plan, the file shall be returned to the USRDS CC at the Recipient’s expense, and any derivative files and copies shall be destroyed.
- J. For the purpose of inspecting security procedures and arrangements, authorized representatives of the PO and/or of CMS will, upon request, be granted access to premises where data in this file are kept.

Recipient typed name, title, & organization

Recipient telephone number

Recipient signature & date

Contractor typed name, title, & organization, as appropriate

Contractor telephone number

Contractor signature & date

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

USRDS Project Officer signature & date

May 2004

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 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 (page 40). 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 (SAF.INCLAIM, SAF.INDETAIL, and SAF.PSCLAIM) 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 variable 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/sup-

plier 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

A “claim” file and a “claim 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 and allowed charges 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 file.

Table B.1 shows the variables in the Institutional Claims file. The claims are uniquely identified by a compound key consisting of the first four variables listed in Table B.1. The records in the institutional claims 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 unless the claim has multiple details which indicate dialysis.

EPO VARIABLES

Summary variables are provided for EPO treatments covered by a claim. EPO treatments are identified by Revenue Center codes 0634 and 0635 and by HCPCS codes from Q9920 through Q9940 on a Revenue Center detail. 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 not retained in the Institutional Claim Details SAF.

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.

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)
- ◆ 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 file. There can be any number of Institutional Claim Details records for each Institutional Claims record. The claim detail file is sorted by the same four-part compound key as the Claims file, so that this key can be used to link the two files.

The two-file structure is a solution to the problem of a number of important data items that appear to be zero one or many times in 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 codes, and an outpatient claim is more likely to specify procedures using revenue center details with HCPCS codes. Using a master and a detail file creates a simple structure easily manipulated in SAS.

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 as a variable several 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, with 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 2001 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 over all 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, 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.

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 CDYPES. Table B.2 indicates the meaning of each CDTYPE and the SAS format that provides labels for the values of CODE for 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, pints of blood, 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

Table B.1. Variables in the Institutional Claims SAF File

Name	Type	Length	Format	Comment
USRDS_ID	Num	8		USRDS patient ID number.
CLM_FROM	Num	8	Date7.	From date of service.
HCFASAF	Char	1	\$HCFASAF.	CMS SAF source of this bill. Format: D Dialysis, H Home health, I Inpatient. N Skilled nursing facility, O Outpatient, P Physician/supplier, S Hospice
SEQ_KEYC	Char	2		Sequence number to ensure unique key
CLM_THRU	Num	8	Date7.	Service through date
PROVUSRD	Char	4		USRDS provider ID number for dialysis providers (HCFA SAF = D), blank for others
PRM_PYR	Char	1	\$PRPAYR.	Primary payer for this bill
CLM_TOT	Num	8		Total charges
CLM_AMT	Num	8		Medicare payments
DIALYSESS	Num	8		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
DIALREVC	Char	2		Revenue center code for dialysis. Right two characters of the primary Revenue Center code indicate dialysis. See text for determining dialysis modality
DIALCRC	Char	5		Claim related condition for dialysis. Right digit of the primary claim related condition code indicates dialysis. See text for determining dialysis modality
EPODOSE	Num	4		Dose of EPO from claim related value code 68
EPOADMIN	Num	4		Number of EPO administrations
EPOCASH	Num	4		Charges for EPO
HCRIT	Num	4		Hematocrit from claim related value code 48
DRG_CD	Char	3	\$DRGLAB.	Diagnosis Related Group Code. Inpatient and SNF claims only. DRG 302 indicates kidney transplant
DISCSTAT	Char	2		Destination on discharge from an inpatient or SNF stay. Files produced before 1998 provide this variable as an Institutional Claims Detail record with CDTYPE = S
PER_DIEM	Num	8		For inpatient and SNF claims, the HCFA pass-through payments. Hospital is reimbursed separately for PER_DIEM times CVR_DCNT.
CVR_DCNT	Num	8		For Inpatient and SNF claims, the Medicare covered day count. See PER_DIEM.
RXCAT	Char	1	\$RXCATIC.	Dialysis treatment modality

Table B.2. Variables in the Institutional Claims Detail SAF File

Name	Type	Length	Format	Comment
USRDS_ID	Num	8		USRDS_ID
CDTYPE	Char	1	\$CDTYPEL.	Defines type for CODE
CLM_FROM	Num	8	Date7.	Claim from date
CODE	Char	5		See documentation
HCFASAF	Char	1	\$HCFASAF.	SAF source of bill
HCPCS	Char	5		HCPCS code
REV_CH	Num	8		Revenue center total charge
SEQ_KEYC	Char	2		
UNITS	Num	8		
URR_CD	Char	8	\$URRFMT.	Urea reduction ratio for reported hemo bills
REVPMT	Num	8		Line item payment amount

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.

Table B.3. Variables in the Physician/Supplier File

Name	Type	Length	Format	Comment
USRDS_ID	Num	8		USRDS patient ID number
CLM_FROM	Date	8	Date7.	From date of service
HCFASAF	Char	1	\$HCFASAF.	CMS SAF source of this bill always = P
CLM_THRU	Date	8	Date7.	Thru date of service
CDTYPE	Char	1	\$HCCDTYP.	Line item type; B ¹ = Phy/supp line item, I ¹ = ICD-9-CM diagnosis
DIAG	Char	5	\$ICD9D.	ICD-9-CM diagnostic code
SBMTCH	Num	8		Submitted charges
ALLOWCH	Num	8		Allowed charges
PMTAMT	Num	8		Claim payment amount
HCPCS	Char	5	\$HCPCS.	HCPCS Procedure Code
MOD1	Char	2		First HCPCS modifier
MOD2	Char	2		HCPCS modifier
MOD3	Char	2		HCPCS modifier
MOD4	Char	2		HCPCS modifier
SPCLTY	Char	2	\$PROVSP.	Provider Specialty Code
PLCSRV	Char	2	\$PLACESV.	Place of Service
HCSRVC	Char	1	\$HCFSVC.	HCEA Service Code
SRVCCT	Num	8		Number of services
PYRCOD	Char	1	\$PRPAYR.	Primary Payer Code

APPENDIX C
STATISTICAL METHODS



METHODS FOR CALCULATING RATES

ESRD incidence rates, prevalence rates, death rates, and transplant rates are calculated for the total population or by groups such as year, Network, state, Health Service Area, age, gender, and primary cause of ESRD. Raw rates, model-based rates, and adjusted rates can be calculated using the following methods.

RAW RATES

Calculating raw rates is straightforward. Some rates are based on counts and others on total follow-up time. For example, incidence and prevalence rates are based on counts, but death rates are usually based on follow-up time. 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:

$$r = 1600/6,400,000 \times 1,000,000 = 250 \text{ per million people}$$

If the total follow-up time of these patients in 2004 is 1100 patient-years, and 150 patient die in the incident year, the incident year death rate of state A is:

$$r_d = 150/1100 \times 1000 = 136.4 \text{ per thousand patient-years}$$

Calculating the standard errors for the estimated rates depends on the calculation of rates. If rate r depends on counts, its standard deviation is $\sqrt{r(r-1)/n}$ where n is the denominator of r . When n is large and r is very small, the standard deviation is estimated by $\sqrt{r/n}$. If rate r depends on another factor only, the second method can be used to calculate the standard deviations. If the units of the rates change (the first method assumes $0 = r = 1$), the standard deviations change accordingly.

MODEL-BASED RATES

If the sizes or total follow-up times are very small for some groups the raw rates may not be stable, and statistical models are necessary. Those most often used for calculating rates are the Poisson model and the Cox proportional regression model. Some software packages give the corresponding standard deviations; if not, the delta method can be applied based on output. The ADR uses a generalized mixed model for calculating death rates, first hospitalization rates, and first transplant rates. The Poisson model can be used without random effects according to the data.

The model used in the ADR is a generalized mixed linear model with log links, Poisson distribution, age, gender, race, primary diagnosis, their two-way interactions as fixed effects, and their four-way interactions as random effects. The response variable is death count and the offset is the log of total follow-up time. This model can produce predicted death counts. Predicted death rates are predicted death counts divided by total follow-up times. Most software packages give standard deviations for predicted counts, but not for predicted rates. The delta method or some other technique is needed to obtain standard deviations for predicted death rates.

For the Cox regression model, which reduces to the Kaplan-Meier method if there are no covariates, the predicted cumulative hazard at time t is the death rate in time period $(0, t]$, but special attention must be paid to the unit of the death rate.

ADJUSTED RATES

Because of the make-up of patient populations, raw rates might not reflect between-group differences that might be confounded by the differences between underlying population distributions. Adjusted, or standardized, rates are needed for comparison. 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). Indirect adjusted rates are usually used as a standard for comparison with observed rates, similar to the standard mortality ratio.

Direct adjustment

If each group has many categories of interest, 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 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 meaningless. 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.

For example, to compare state-level incidence rates in 2004 under the assumption that race distributions in all states are the same, it is necessary to calculate the standardized incidence rates adjusted for race for each state. Because the race distributions are quite different in different states, the national population is a good choice for the reference. Using the national population at the end of 2004 as the standard population, and assuming the incidence rate of state A in 2004 is 173 per million people, the race-specific rates for state A and the national population distribution with respect to race are:

	Incidence Rate, State A	National Population, Percent
White	153	75.1
Black	250	12.3
Native American	303	0.9
Asian/Pacific	174	3.6
Other	220	8.0

The adjusted incidence rate of state A with the national population as the reference is:

$$(153 \times 75.1\%) + (250 \times 12.3\%) + (303 \times 0.9\%) + (174 \times 3.6\%) + (220 \times 8\%) = 158.73 \text{ per million people.}$$

Thus, if the population of state A had the same race proportions as the country as a whole, the incidence rate would be 158.73 instead of 173.

Adjusted-rate standard deviations can be calculated as follows: Calculate standard deviations for each category-specific rate in each group as described for raw rates, convert them to variances (the square of the standard deviation), then calculate the weighted sums of the variances as the variance of adjusted rates, where the weights are the squares of the weights in the calculation of adjusted rates. The square roots of the variances are the standard deviations of the adjusted rates.

The disadvantages of this method are:

- ◆ If one category in a group has a smaller population (the denominator of the rate) the corresponding category-specific rate will be unstable, making the adjusted rate for this group also unstable.
- ◆ If one category has no population in a group, the adjusted rate cannot be calculated for that group.
- ◆ There is no naïve method for calculating robust standard errors for adjusted rates.

Indirect adjustment

The indirect adjustment method is similar to the direct method. The indirect adjusted (standardized) rate for a group is derived by applying the category-specific event rates of the standard (reference) population to the group; it is the weighted average of the category-specific rates in the reference population with weights taken from the group. The indirect adjusted rate can be interpreted as the rate for the group if it had had the same category-specific rates as the reference population.

For example, if the population distribution with respect to race in state A and the national race-specific incidence rates per million population are as follows:

	National Incidence Rate	Population in State A, Percent
White	229	89.4
Black	970	3.5
Native American	862	1.1
Asian/Pacific	396	2.9
Other	358	3.0

The indirect adjusted rates for state A with national rates as the reference is

$$(229 \times 89.4\%) + (970 \times 3.5\%) + (862 \times 1.1\%) + (396 \times 2.9\%) + (358 \times 3\%) = 270.38 \text{ per million people.}$$

Choosing an appropriate reference allows calculation of indirect adjusted rates, but the rates are not comparable. For example, if Minnesota and Illinois had the same population distribution with respect to race, the indirect standardized rates, adjusted for race with the national population as the reference, would be the same no matter how different the real incidence rates between these two states.

MODEL-BASED ADJUSTMENT

Because of the disadvantages of the direct adjustment method, a model-based adjustment method is necessary. Sometimes models can be run and the average values of the covariates from the reference substituted. Unless the model is linear, however, averaging the covariates is not the same as averaging in the model, and the simple average might not provide accurate results. An appropriate method is to obtain the predicted death rates for each category in each group using a statistical model, then use the direct adjustment method to calculate adjusted rates based on predicted rates with the given reference population. If the model contains continuous covariates, individual predicted rates are needed.

For example, to calculate adjusted state-level one-year death rates for 2004 incident patients, adjusted for age, gender, race, and primary diagnosis, first calculate the predicted death rates for all intersection subgroups (categories) of age, gender, race, and primary diagnosis in each state (group), using a model. The Cox regression model can be used with state as a stratum variable because of the need to consider survival time and the questionable proportionality of the baseline hazard functions across states. The national 2004 ESRD incident patients are chosen as reference. The adjusted death rates can be calculated using the following procedures.

1. Fit the Cox regression model with survival time as response variable; age, gender, race, and primary diagnosis as explanatory variables; and state as the stratum variable.
2. Use the BASELINE statement in SAS Proc PHREG to calculate the baseline survival at day 365 for all states.
3. Calculate one death rate for each subgroup in each state at day 365 (one year) $r = \exp(x\beta) \log(S_0)$, using the estimated effect coefficient β and estimated survival of the state at day 365 S_0 , where x is the corresponding covariate value of the subgroup.
4. Calculate the weight for each subgroup based on the survival time in the first year for all 2004 incident patients in the country as a whole.
5. Calculate the adjusted death rate for each state based on the predicted rates from step 3, the weights from step 4, and the direct adjustment method.

Subgroup-specific death rates in steps 2 and 3 can also be calculated based on individual predicted deaths. The method for calculating adjusted survival is similar.

A disadvantage of this method is the lack of a closed form for calculating standard errors of the adjusted rates. The non-parametric bootstrap method can be used; it works well, but the computation is time consuming.

SURVIVAL ANALYSIS

The most commonly used methods for survival analyses are the Kaplan-Meier method for calculating survival probabilities, the log rank method for comparing survival between two or more groups, and the Cox proportional regression model for comparisons among groups and making inferences. The difficulty with survival analysis is defining the starting point (time zero) for survival time, especially in observational

studies. The time zero selection should be meaningful and should not cause selection bias; late or early start points may exclude some patients.

KAPLAN-MEIER

The Kaplan-Meier method (Kalbfleisch and Prentice 1980) is the most widely used non-parametric method for survival analysis. It is used for calculating survival functions and death rates. The plot of the Kaplan-Meier estimate of the survival function is a step-function, in which the estimated survival probabilities are constant between two successive death times which decrease at each death time. To find the survival probability at time t , find the closest death time $t_1 = t$; the survival probability at time t_1 is the estimate of survival probability at time t .

The standard error of the estimated survival probability at time t is given by the Greenwood formula. The confidence interval of the estimated survival probability at time t can be given by assuming that the estimated survival is normally distributed based on the standard error from the Greenwood formula.

LOG-RANK TEST

The log-rank test (Kalbfleisch and Prentice 1980) is a method for testing the hypothesis that there is no difference in survival probabilities over a time period among groups. The general log-rank test does not take into account the effect of other factors, such as different age groups, multiple clinical centers, etc. The stratified log-rank test (stratified by the factors of interest) will perform the log-rank test for each stratum and give the testing result by combining the information from all strata. This method can take account only of factors that can be classified. It cannot take account of continuous factors. If some strata have few patients, the result might not be stable. The modeling approach is more flexible.

COX REGRESSION MODEL

The Cox regression model (Kalbfleisch and Prentice 1980) is the most widely used semi-parametric method for survival analysis. It can be used to compare survival among groups and to find risk factors. The Cox model assumes:

$$h(t | X = x) = h_0(t) \exp(x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p)$$

where $h(t | X = x)$ is the hazard function and $X = x = (X_1, X_2, \dots, X_p)$, $h_0(t)$ is the baseline hazard; i.e. $h_0(t) = h(t | X = 0)$, $x = (x_1, x_2, \dots, x_p)$ are the potential risk factors, and $\beta_1, \beta_2, \dots, \beta_p$ are the corresponding effects. If β_i is significantly different from zero, X_i is considered a risk factor.

This model is also called the Cox proportional regression model because if two individuals have characteristic $x^1 = (x_1^1, x_2^1, \dots, x_p^1)$ and $x^2 = (x_1^2, x_2^2, \dots, x_p^2)$, respectively, then the ratio of their hazard (relative risk) is:

$$h(t | X = x^1) / h(t | X = x^2) = \exp[(x_1^1 - x_1^2)\beta_1 + (x_2^1 - x_2^2)\beta_2 + \dots + (x_p^1 - x_p^2)\beta_p]$$

which does not depend on time. The basic assumption of the Cox regression model is proportionality. There are some

methods to test this assumption (Therneau and Grambsch 2000). If the hazards are not proportional for a variable, the analysis can be stratified by that variable. The stratified Cox model assumes the other factors have the same effects in all strata, but different strata may have different (non-proportional) baseline hazards, and usually they are not proportional. In survival analysis of ESRD patients, hazards for diabetic and non-diabetic patients are usually not proportional.

BIBLIOGRAPHY

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

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.