

**2014 RESEARCHER'S GUIDE TO THE
USRDS DATABASE**

2014 Researcher's Guide to the USRDS Database

Table of Contents

Preface	4
Acknowledgments	4
USRDS Coordinating Center.....	4
NIDDK Project Officers.....	4
Suggested Citation.....	4
Introduction.....	4
USRDS website: www.usrds.org.....	4
RenDER.....	4
What's New in 2014	5
Variable Changes.....	5
Core SAF	5
Hospitalization and Institutional Claims SAF	5
Format and Label Changes.....	5
Modified Formats.....	5
File Structure Changes.....	6
Transplant Datasets.....	6
Institutional Physician/Supplier and Hospitalization Claims	6
Method Changes	6
Patient Match Crosswalk File.....	6
First Service Date and Lost to Follow-up	6
Getting Started.....	7
Basic SAS Use	7
SAS Formats for USRDS Data.....	8
Comment Lines.....	8
The SAF Directory.....	8
SAS Format Examples.....	9
Example 1: Incident Cohort	9
Example 2: Incident Patients Distribution by Demographic Data	9
Example 3: Point Prevalent Cohort	9
Example 4: 2005 Incident Patient Survival Rates (Kaplan-Meier).....	10
Example 5: Merge With Medical Evidence file (CMS 2728)	10
Example 6: 2008 Incident Patient Survival Rates (Kaplan-Meier).....	11
Example 7: Waiting List Access Rate.....	11
Example 8: Create a patient cohort of 2008 Medicare Incident Patients	12
Example 9: Total Admission Rate.....	12
Example 10: First Admission Rate	14
Methods & Databases	16
Section 1: ESRD Data and the USRDS Database System	16
Data Sources.....	16
Consolidated Renal Operations in a Web-enabled Network (CROWN).....	16
CMS ESRD Medical Evidence Report Form (CMS 2728).....	16
2005 Revision of the Medical Evidence Report Form (CMS 2728).....	17
CMS ESRD Death Notification Form (CMS 2746)	17
CMS Medicare Enrollment Database (EDB)	18
CMS Paid Claims Records	18
Organ Procurement and Transplantation Network (OPTN) Database.....	18
CMS ESRD Standard Analytical Files (SAF).....	18
CMS 5% General Medicare SAF	18

Clinical Performance Measures (CPM) Project	18
CMS Dialysis Facility Compare Data.....	19
CMS Annual Facility Survey (AFS)	19
CDC National Surveillance Data	19
Dialysis Morbidity and Mortality Study (DMMS)	19
Case Mix Adequacy Study (CMAS)	19
Case Mix Severity Study.....	20
Pediatric Growth and Development	20
Continuous Ambulatory Peritoneal Dialysis (CAPD) and Peritonitis Study.....	20
United States Census	20
USRDS Database System	20
Section 2: ESRD Patients.....	21
Identifying ESRD Patients.....	21
First ESRD Service Date.....	21
Death Date Determination.....	21
Transplant Dates	21
Graft Failure.....	22
Medicare and Non-Medicare Patients	22
Section 3: Treatment History	22
Treatment Modality Categories.....	23
How Treatment Modality is Determined.....	23
The 60-Day Collapsing Rule.....	24
Integration of the CROWNWeb and CMS Claims Databases	25
Determining Lost-to-Follow-up Periods	25
Section 4: Payer History	25
ESRD Payer Sequence File	25
Pre-ESRD Payer Sequence File	25
Data Sources.....	26
Payer Categories	26
How the Payer Category is Determined.....	26
Gaps in the Payer History	27
Section 5: Medicare Claims	27
Introduction	27
CMS Data Sources.....	27
Patients and Time Periods Included.....	28
Basic File Structure	29
Institutional Claims	30
Institutional Claims File Variables.....	30
Medicare Payment Variables	30
Dialysis Variables.....	30
EPO Variables	30
Institutional Claim Details.....	31
Revenue Center Details.....	31
Institutional Claim Details and Revenue Center Details File Variables	32
CDTYPE and CODE Variables.....	32
UNITS	32
REV_CH	33
HCPCS	33
URR_CD.....	33
Physician/Supplier File.....	33
Section 6: Transplant Process and Outcomes.....	34
Data Sources.....	34
Reconciliation of CMS and OPTN Events	35
Kidney and Pancreas Waiting list	35
Transplant SAF	36

Section 7: Morbidity and Hospitalization	37
Data Sources	37
Contents of the Hospital Dataset.....	37
Preparing Inpatient Claims Data for Use	37
Section 8: Survival and Mortality	38
Data Sources.....	38
Section 9: Providers	38
Data Sources.....	38
Provider Numbers Assigned by the USRDS	38
Profit Status	39
Chain Affiliation.....	39
Freestanding vs. Hospital-based Facility	39
Section 10: CKD Cohort (5% CKD Cohort Dataset)	39
Data Sources.....	39
Patient Master File	39
Payer Sequence File.....	40
Co-Morbid Files	40
Section 11: Comprehensive Dialysis Study SAF	41
Section 12: Medicare Prescription Drug	41
Data Sources.....	41
Part D Enrollment SAF	41
Prescription Drug Event SAF	42

Preface

The Researcher's Guide is intended for investigators within and outside of the USRDS Coordinating Center who wish to undertake research projects using data from the USRDS database. This guide places particular emphasis on the USRDS Standard Analysis File (SAF) dataset, the primary means by which USRDS data are available for use. The Researcher's Guide includes information needed to help researchers select and use the appropriate SAFs for the intended project.

Acknowledgments

Most original data for the USRDS are provided by the Centers for Medicare & Medicaid Services through the inter-agency agreement between the NIDDK/DKUHD of the National Institutes of Health, and the Bureau of Data Management and Strategy of the Centers for Medicare & Medicaid Services. The United States Renal Data System offers its grateful appreciation to everyone who contributed to the USRDS database and this guide, especially all the ESRD professionals and Networks on whose reporting of data this enterprise ultimately rests.

USRDS Coordinating Center

- Director: Rajiv Saran, MBBS, DTCD, MD, MRCP, MS
- Co-Deputy Director: Yi Li, PhD
- Co-Deputy Director: Bruce Robinson, MD, MS, FACP

NIDDK Project Officers

- Kevin C. Abbott, MD, MPH — Director, Kidney and Urology Epidemiology, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institutes of Health (NIH)
- Lawrence Y. C. Agodoa, MD — Director, End-stage Renal Disease Program, Division of Kidney, Urologic, and Hematologic Diseases (DKUHD), NIH
- Paul W. Eggers, PhD — NIDDK, NIH

Suggested Citation

The suggested citation for this guide is:

U.S. Renal Data System. 2014 Researcher's Guide to the USRDS Database. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2015.

Publications based upon USRDS data must include the following disclaimer: "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."

Introduction

USRDS website: www.usrds.org

The USRDS website provides users with access to the Annual Data Report (ADR), which is available in HTML and downloadable PDF, Excel, and PowerPoint files. ADR chapters and volumes are available in PDF; Reference Tables and the data underlying the figures and tables for each chapter are available in Excel; and, all chapter figures and tables are available in PowerPoint slides. Because of its size, downloading some portions of the ADR may require user patience.

RenDER

The USRDS Renal Data Extraction and Referencing (RenDER) System is an online data querying application accessible through the USRDS website, allowing access to a wealth of information regarding end-stage renal disease (ESRD) in the United States. It quickly returns an accurate table of data or an interactive map based on the user's query specifications. Tables can be copied into a spreadsheet application on the user's computer for further manipulation and investigation, and map images can be copied or saved to local applications. A database file download of the mapped data, which can be opened or imported by most spreadsheet applications, is also available.

The RenDER System allows easy access to some of the most frequently requested data. While the ADR thoroughly covers many ESRD statistics, it cannot reasonably contain the more detailed tables often requested by researchers. RenDER allows users to drill down into the data behind many of the tables published in the ADR, allowing cross interactions among various demographic fields. To access the RenDER tutorial, visit www.usrds.org/render/xrender_home.asp.

What's New in 2014

Changes to the 2014 USRDS Standard Analysis Files (SAF) are listed below, and are explained in further detail in the What's New 2014 text file included with the Core SAF.

Variable Changes

Several additions, deletions, and modifications were made to variables in the following 2014 files.

CORE SAF

This section describes variable changes in the Core SAF.

PATIENTS Dataset

Eight variables have been added to the PATIENTS file: birth date (BORN); date patient was first added to the kidney transplant waiting list, first removed from the list, and reason for removal if applicable (CAN_FIRST_LISTING_DT, CAN_REM_DT, CAN_REM_CD); first treatment modality date and date and type of first dialysis treatment (DIALTYP, FIRSTDIAL, FIRST_MODALITY); and Hispanic ethnicity (HISPANIC). The PDIS variable in the PATIENTS file now includes data from both the 1995 and 2005 Medical Evidence forms. In previous SAFs, the PDIS variable only included data from the 2005 Medical Evidence form.

MEDEVID, MEDEVID95, and MEDEVID05 Datasets

Four variables, DISGRPC (primary disease group), MEFDATE, CTYEAR and MEFYEAR have been added to the Medical Evidence files.

Part D Data

In Part D enrollment data, PDENROL2012, the variable indicating beneficiary enrollment reference year (BENE_ENROLLMT_REF_YR) has been added.

Transplant Dataset

A variable indicating whether the donated kidney came from an expanded criteria donor (DON_EXPAND_DON_KI) was added to the TX file.

HOSPITALIZATION AND INSTITUTIONAL CLAIMS SAF

Three variables that identify physicians (ATTENDING_PHYS, OPERATING_PHYS, OTHER_PHYS) have been added to the following files: 2012 institutional claims file INC2012; HOSP3; Case Mix Adequacy Study (CMAS) HOSP2; and Dialysis Morbidity and Mortality Study (DMMS) HOSP2. PRIMDIAG in previous year's HOSP3, CMAS HOSP2, and DMMS HOSP2 is the variable

indicating if the first diagnosis was primary. In this version, PRIMDIAG is the first ICD-9 diagnosis code. In addition, HSDIAG26 has been added to the HOSP3, CMAS HOSP2, and DMMS HOSP2 dataset to allow for all 26 ICD-9 diagnosis codes that can now appear on the claims.

PHYSICIAN/SUPPLIER CLAIMS SAF

Three variables have been added to the Physician/Supplier (PS) Claims SAF, CMAS PSCLAIM and DMMS PSCLAIM: CLMTYP, PHYSUSRDS_RFRG and PHYSUSRDS_PRFRMG. Additionally, the CDTYPE variable was removed from these files.

PHYSUSRDS_RFRG and PHYSUSRDS_PRFRMG have been added as physician identifiers and are created by the USRDS based on the physician's National Provider Identifier (NPI). CLMTYP indicates whether the record comes from durable medical equipment (DME) or PS claims data.

Format and Label Changes

A few cosmetic changes have been made to the labeling for the variables in the Medical Evidence files, namely, MEDEVID, MEDEVID95 and MEDEVID05. The Medical Evidence CMS 2728 form question number now appears at the beginning of the label, rather than the end. The word "Discontinued" appears in the label for those items from the 1995 version of the form that were discontinued with the current, 2005 version.

Additionally, \$PD96DET_ALL, which is a more complete format than what was used previously, has been attached to the PDIS variable in the PATIENTS, MEDEVID95, MEDEVID05 and MEDEVID files within the Core SAF. The LABMETH format was attached to the LABMETHOD variable in the MEDEVID05 and MEDEVID files within the Core Dataset. The \$PBCLMTYP format was attached to the CLMTYP variable in the ESRD and pre-ESRD Physician/Supplier claims files.

MODIFIED FORMATS

The existing \$ADRINDT, LD_REL, \$NETFMT and \$STATFIP formats have been modified, details of which are included in the What's New 2014 text file included in the Core SAF. A new entry, "B = Discontinued dialysis", was added to the existing \$RCATDT and \$RCATGP formats. This format is attached to DIALTYP and FIRST_MODALITY variables in the PATIENTS file and RXDETAIL RXGROUP variables in the RXHIST and RXHIST60 files within the Core SAF.

File Structure Changes

TRANSPLANT DATASETS

The TXUNOS_KI_PRE_JUL04 and TXUNOS_KI_POST_JUL04 files now include only kidney (KI) data. In prior SAFs, the TXUNOS_KI_PRE_JUL04 file included both kidney and kidney-pancreas (KP) data. TXUNOS_KP includes all years' kidney-pancreas data. More than 200 text and lab value status indicator variables were removed from these files. The complete list is found in the What's New 2014 text file included with the Core SAF.

INSTITUTIONAL PHYSICIAN/SUPPLIER AND HOSPITALIZATION CLAIMS

The 2014 Physician/Supplier (PS) files were created by merging the claim-level data from PS and durable medical equipment (DME) claims with the line-item data from the PS and DME claims, respectively, for each record, where applicable. In prior PS files, there was a record for each line-item (as there is currently), but each claim was also included as a separate record. REMIS summarized files are no longer included in ESRD claim data from 2012 and pre-ESRD claim data from 2010. See "Section 5: Medicare Claims."

Method Changes

PATIENT MATCH CROSSWALK FILE

The method for identifying patients changed somewhat for this year, which has resulted in cases

where patients who were identified as being separate individuals in the prior SAFs are now thought to be the same individual. In those cases, a single USRDS_ID is now used for all records. A list of these USRDS_IDs can be found in the USRDSID_MULTIPLE_XREF file included in the Core SAF. This file may be useful for prior users of the USRDS SAFs.

FIRST SERVICE DATE AND LOST TO FOLLOW-UP

Changes were made to the method of determining first service date and identifying those that were lost-to-follow-up. These changes are discussed in greater detail in "Section 3: Treatment History", and mainly affect the following files within the Core SAF: the RXHIST and RXHIST60 files, and the FIRST_SE variable in the PATIENTS file.

Some patients who existed in the prior year's PATIENTS file were not identified in the new patient matching algorithm due to inconsistent dates or incomplete identifiers. These patients are included in this year's PATIENTS file with ADRIND coded as "0". In addition, the patients found in the data system but with first service date and death date on same day, or no first service date are also coded as "ADRIND=0".

Figure 1. Structure of the USRDS Database

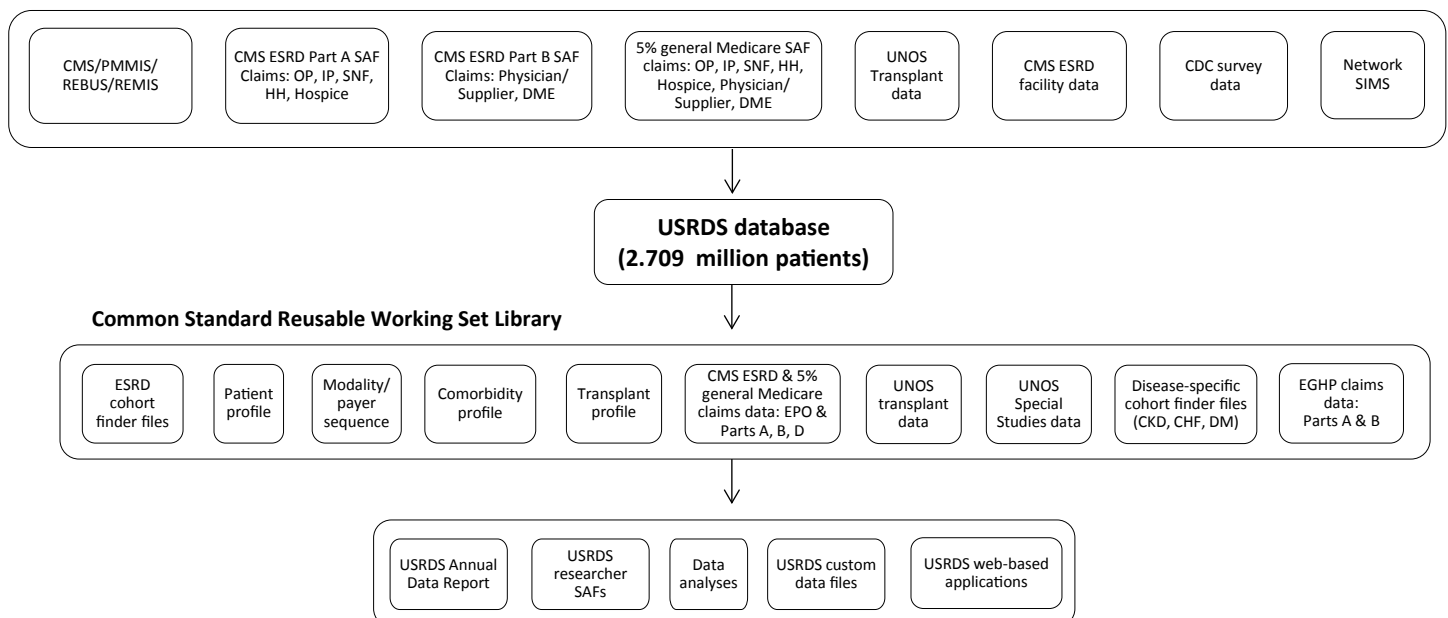


Table 1. Year-to-Year Comparison of Number of Records in USRDS Standard Analysis Files (SAFs)

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
PATIENTS	1,600,693	1,698,706	1,801,298	1,910,161	2,024,425	2,138,876	2,260,986	2,377,166	2,497,498	2,709,247
RESIDENC	2,388,928	2,587,007	2,791,903	2,968,884	3,189,860	3,404,854	3,699,752	3,879,561	4,175,840	4,875,215
MEDEVID†	896,447	1,164,686	1,147,657	1,265,566	1,385,589	1,505,285	1,630,270	1,753,496	1,862,922	2,105,748
RXHIST	10,733,455	11,658,208	12,533,612	13,532,743	11,626,297	11,741,390	12,392,072	13,038,727	13,696,997	13,003,645
FACILITY	60,138	64,870	69,764	74,831	80,071	85,572	91,332	97,201	103,210	115,973
TX	272,277	289,533	308,002	324,476	343,051	360,297	378,536	395,347	413,493	451,499
TXWAIT*	347,564	378,262	405,165	437,256	474,348	513,028	546,239	579,225	618,511	690,217

†MEDEVID is the total number of observations between Medevid95 and Medevid05.
*TXWAIT is the total number of observations between Waitlist_ki and Waitlist_kp.

Getting Started

Work on a typical study consists of determining the study variables, selecting the variables from the datasets in which they are stored, merging the selected variables into one or more datasets for analysis, and finally performing the statistical analyses of the data. This section offers dataset examples of this process to first-time users of the USRDS SAFs.

All USRDS data are stored in SAS datasets that were created in the Windows environment and can be used only on the Windows operating system. Using the datasets on another computer platform requires moving and converting the data to SAS datasets for that platform. Using another data analysis system also requires a conversion to a format compatible with that system.

We provide examples using basic SAS code. New SAS users should take classes, consult colleagues, or otherwise become familiar with the SAS system. Regardless of SAS experience, all users must know:

- The location of the SAS dataset
- The location of the SAS catalog of FORMATS

In the examples, the datasets and the FORMAT catalog are assumed to be in the same directory, namely, C:\SAF. Two SAS statements are needed to point to this information:

```
LIBNAME saf "c:\saf"; * Directory location of the data;
LIBNAME LIBRARY "c:\saf"; * Directory location of format catalog;
```

Always assume that these two statements are required in all code. (See Appendix C: Data Formatting for additional information.) The datasets and the format catalog may be in different directories.

Basic SAS Use

The SAS CONTENTS procedure generates a list of all variables in the dataset and a label associated with each. The information in this label is usually an adequate indication of whether the variable will be in use; however, PROC CONTENTS is always the best method for obtaining the latest variable list, as last minute updates may not be reflected in the printed documentation (see Appendix B: Data File Descriptions).

Note: In the following program examples, SAS commands, keywords, and procedures are in uppercase text. Datasets, comments, and variables in which SAS will perform operations, are in lowercase text.

To determine the contents of the PATIENTS dataset, or any dataset, use the following code (remember the two LIBNAME requirement).

```
PROC CONTENTS DATA=saf.patients;
    TITLE1 'DATASET: saf.patients';
RUN;
```

To see what the data look like, use the SAS procedure PRINT to list all observations of the dataset:

```
PROC PRINT data=saf.patients;
    TITLE1 'DATASET: saf.patients';
RUN;
```

A dataset can have millions of observations. To print the first 500, use the following code:

```
PROC PRINT DATA=saf.patients (OBS=500);
    TITLE1 'DATASET: saf.patients';
RUN;
```

To print a group of observations other than the first N observations, use the following code; this example prints observations 1500–2000, inclusive:

```
PROC PRINT DATA=saf.patients (FIRSTOBS=1500
OBS=2000);
    TITLE1 'DATASET: saf.patients';
RUN;
```

A dataset can contain hundreds of variables. Print selected variables using the VAR statement:

```
PROC PRINT DATA=saf.patients (OBS=500);
    VAR usrds_id sex race incyear;
    TITLE1 'DATASET: saf.patients';
RUN;
```

There are two parts to the “DATA=” expression. The first is the directory location, defined by the LIBNAME, and the second is the dataset name. The first part (directory location) implies that the dataset is permanent. Later examples do not include this part, and the datasets you create will disappear when you leave SAS. To permanently save a dataset, the first part of the expression must be included.

SAS Formats for USRDS Data

Information in a dataset may be coded. Thus, the variable GENDER may have the values F, M, or U, where “F” is the coded value for female, “M” is the value for male, and “U” is the value for unknown. Using one character instead of several saves disk storage, but because the coded values are not always easy for users to understand, a format is assigned to translate them. Many variables have been assigned formats by the USRDS.

In each program, SAS must be told where these formats are. Assume that the format catalog is in the directory C:\SAF, then the following SAS LIBNAME makes the formats accessible to your SAS programs:

```
LIBNAME LIBRARY "c:\saf"; * Directory location of format
catalog;
```

The SAS keyword LIBRARY must be used because it tells SAS to look for formats in the directory, C:\SAF. Other methods can accomplish this, but not as easily. To bypass the use of formats, use the following two SAS statements before running any SAS procedures:

```
LIBNAME LIBRARY;
OPTIONS NOFMterr;
```

Bypassing the use of formats allows you to see raw data instead of the formatted values, which may be useful when you need to write SAS IF statements to control the flow of your program in a SAS data step. (See Appendix C: Data Formatting for a tabular list of the formats with their coded values.)

Comment Lines

Comment lines in the SAS code look like this:

```
/* Comment line */
* Comment line;
```

These refer to any descriptive comment. The use of comments is optional, but is strongly recommended.

The SAF Directory

Throughout this section SAF is used as the permanent SAS LIBNAME. It is assumed that all of the USRDS SAF datasets and catalogs have been placed in this directory. If the datasets have been placed in the WINDOWS directory C:\SAF, then the following SAS LIBNAME would point to the SAF datasets:

```
LIBNAME saf "c:\saf";
```

Note: The datasets may be loaded into any directory, with the directory in the LIBNAME changed accordingly.

SAS Format Examples**EXAMPLE 1: INCIDENT COHORT**

-- Create a patient cohort of incident patients from 2000 to 2010 to use in the examples that follow. --;

```
DATA inc2000_2010 (KEEP=usrds_id first_se incyear first_modality);
    SET saf.patients;
    IF 2000<=incyear<=2010;
RUN;
PROC FREQ DATA=inc2000_2010;
    TABLE incyear;
    TITLE 'Frequency Distribution of Incident Patients 2000 – 2010 by Incident Year';
RUN;
PROC FREQ DATA=inc2000_2010;
    TABLE incyear * first_modality;
    TITLE 'Frequency Distribution of Incident Patients 2000 – 2010 by Incident Year and Modality';
RUN;
```

EXAMPLE 2: INCIDENT PATIENTS DISTRIBUTION BY DEMOGRAPHIC DATA

```
PROC FREQ DATA=saf.patients;
    TABLE incagec sex race disgrp;
    WHERE (incyear = 2010);
    TITLE 'Frequency Distribution of 2010 Incident Patients by Demographic Data';
RUN;
```

EXAMPLE 3: POINT PREVALENT COHORT

-- Create a patient cohort of point prevalent dialysis patients who were alive on January 1, 2010. --;

```
DATA pre_2010 (KEEP=usrds_id rxgroup);
    SET saf.rxhist60;
    BY usrds_id begdate;
    IF (begdate <= MDY(1,1,2010)) AND
        ((enddate = .) OR (enddate >= MDY(1,1,2010))) AND
        (rxgroup NOT IN ("B", "D", "X", "Z", "T")) THEN
    OUTPUT;
RUN;
PROC FREQ DATA=pre_2010;
    TABLE rxgroup;
    TITLE 'Frequency Distribution of 1-1-2010 Point Prevalent Patients by Modality';
RUN;
```

EXAMPLE 4: 2005 INCIDENT PATIENT SURVIVAL RATES (KAPLAN-MEIER)

-- Calculate survival rates of 2005 incident dialysis patients using Kaplan-Meier method, censoring at transplant. --;

```
DATA inc_2005_s;
    MERGE inc2000_2010 (IN = x1)
        saf.patients (IN = x2 KEEP=usrds_id died tx1date);
    BY usrds_id;
    IF x1 AND x2;
    * Extract only 2005 incident dialysis patients.;
    IF (incyear = 2005) AND (first_modality ^= "T");
    * Calculate the survival time (in months) of each incident patient.;
    t = (MIN(died, tx1date, MDY(12,31,2010)) - first_se + 1) / 30.4375;
    IF (t < 0) THEN t = 0;
    * Determine whether the patient is censored.;
    c = (MIN(died, tx1date, MDY(12,31,2010)) = died) and (MIN(died, tx1date, MDY(12,31,2010)) <> tx1date);
RUN;
PROC LIFETEST DATA=inc_2005_s METHOD=KM NOTABLE PLOTS=(s) OUTSURV=surv2005;
    TIME t*c(0);
    TITLE '2005 Incident Dialysis Patients Survival Rates';
RUN;
```

EXAMPLE 5: MERGE WITH MEDICAL EVIDENCE FILE (CMS 2728)

-- Demonstrate a way to extract comorbid conditions for a pre-defined study cohort from the Medical Evidence file. --;

```
DATA medevid;
    SET saf.medevid (KEEP=usrds_id cancer cararr carfail cva diabins diabprim dysrhyt
        hyper ihd mi pulmon pvasc);
    BY usrds_id;
    IF (FIRST.usrds_id);
RUN;
DATA me_2728;
    MERGE medevid (IN=x1)
        saf.patients (IN=x2 KEEP=usrds_id died tx1date first_se);
    BY usrds_id;
    IF x1 AND x2;
RUN;
DATA inc2008_me;
    MERGE inc2000_2010 (IN=x1 WHERE=(incyear = 2008))
        me_2728 (IN=x2);
    BY usrds_id;
    IF x1 AND x2;
RUN;
```

EXAMPLE 6: 2008 INCIDENT PATIENT SURVIVAL RATES (KAPLAN-MEIER)

-- Show a survival rate calculation stratified by patient comorbid condition, not censored at transplant. --;

```
DATA inc_2008_s;
    SET inc2008_me;
    BY usrds_id;
    IF (UPCASE(cancer) IN ("1""2""Y""N"));
    IF (UPCASE(cancer) IN ("1""Y")) THEN can = 1;
    ELSE can = 0;
    * Calculate the survival time (in months) of each incident patient.;
    t = (MIN(died, MDY(12,31,2010)) - first_se + 1) / 30.4375;
    IF (t < 0) THEN t = 0;
    * Determine whether the patient is censored.;
    c = (MIN(died, MDY(12,31,2010)) = died);
RUN;
PROC LIFETEST DATA=inc_2008_s METHOD=KM NOTABLE PLOTS=(s) OUTSURV=surv2008;
    TIME t*c(0);
    STRATA can;
    TITLE '2008 Incident ESRD Patients Survival Rates by Co-Morbidity';
RUN;
```

EXAMPLE 7: WAITING LIST ACCESS RATE

-- Determine waiting list access rate to December 31, 2010, of 2008 incident dialysis patients who were registered in the transplant waiting list. --;

```
DATA txwait;
    SET saf.waitseq_ki saf.waitseq_kp;
RUN;
PROC SORT data=txwait;
    By usrds_id;
RUN;
DATA inc_2008_w;
    MERGE inc2008_me (IN=x1) txwait (IN=x2 KEEP=usrds_id begin);
    BY usrds_id;
    IF x1;
    * Extract only 2008 incident dialysis patients.;
    IF (first_modality ^= "T");
    * Make sure all dialysis patients who were not put on the waiting list;
    * are censored at the end of follow-up period.;
    IF x1 AND ^x2 THEN begin = MDY(1,1,2011);
    * Calculate the waiting list access time (in month) of each incident patient.;
    t = (MIN(begin, died, MDY(12,31,2010)) - first_se + 1) / 30.4375;
    IF (t < 0) THEN t = 0;
    * Determine whether the patient is censored.;
    c = (MIN(begin, died, MDY(12,31,2010)) = begin);
RUN;
PROC LIFETEST DATA=inc_2008_w METHOD=KM NOTABLE PLOTS=(s) OUTSURV=wait2008;
    TIME t*c(0);
    TITLE '2008 Incident Dialysis Patients Waiting List Access Rates';
RUN;
```

EXAMPLE 8: CREATE A PATIENT COHORT OF 2008 MEDICARE INCIDENT PATIENTS

-- Create the 2008 incident patient file. --;

```
DATA inc_2008;
    SET saf.patients (WHERE=(incyear=2008));
RUN;
```

-- Find Medicare payer status at first service date. --;

```
DATA first_payer;
    SET saf.payhist;
    BY usrds_id begdate;
    IF first.usrds_id;
RUN;
```

-- Combine payer information with original incident cohort and limit to patients with Medicare payers at first service date. --;

```
PROC SORT DATA=inc_2008;
    BY usrds_id;
RUN;
DATA medicare_inc_2008;
    MERGE inc_2008 (IN=x1)
    first_payer (IN=x2 KEEP=usrds_id mcare payer dualelig);
    BY usrds_id;
    IF x1 AND mcare = "Y";
RUN;
```

EXAMPLE 9: TOTAL ADMISSION RATE

-- Create the patient driver for hospitalization rates, using previously created file. --;

```
DATA driver;
    SET medicare_inc_2008 (WHERE=(payer='MPAB'));
    sfu = first_se;
    efu = MIN(died, first_se+365);
    IF (died NE .) AND (died < sfu) THEN DELETE;
    FORMAT sfu efu MMDDYY10.;
RUN;
PROC SORT DATA=driver;
    BY usrds_id;
RUN;
```

-- Link to hospitalization file; gather hospitalizations within period of interest (includes only patients with hospitalizations). --;

```
PROC SQL;
    CREATE TABLE hospdat AS SELECT * FROM driver a, saf.inc2012 (KEEP=usrds_id clm_from clm_thru) b WHERE
    a.usrds_id=b.usrds_id AND a.sfu<=clm_thru AND a.efu>=b.clm_from;
QUIT;
PROC SORT DATA=hospdat;
    BY usrds_id clm_from clm_thru;
RUN;
```

DATA rate;

```

SET hospdat;
BY usrds_id;
RETAIN n_hos n_adm exptime rt_adm;
IF FIRST.usrds_id THEN
DO; /* INITIALIZE RETAIN VARIABLES FOR EACH USRDS_ID */
    n_hos=0;
    n_adm=0;
    exptime=0;
    rt_adm=0;
END;
* FOR HOSPITALIZATIONS OVERLAPPING STUDY START ONLY COUNT;
* HOSPITAL DAYS, NOT AS ADMISSION;
IF (clm_from<sfu<=clm_thru) THEN n_hos=n_hos+(MIN(clm_thru,efu)-sfu+1);
* IF WITHIN STUDY PERIOD COUNT DAYS (UP TO STUDY END) AND;
* ADMISSIONS;
ELSE IF (sfu<=clm_from<=efu) THEN
DO;
    n_hos=n_hos+(MIN(clm_thru,efu)-clm_from +1);
    n_adm=n_adm+1;
END;
IF LAST.usrds_id THEN /* Output one record per USRDS_ID */
DO;
    exptime=(efu-sfu-n_hos+1)/365;
    IF exptime>0 then rt_adm=n_adm*1000/exptime;
    ELSE IF (exptime<0) THEN exptime=0;
OUTPUT;
END;

```

RUN;

-- Combine result with original incident sample to include patients without hospitalizations. --;

DATA rateall;

```

MERGE driver (IN=x1) rate (IN=x2);
BY usrds_id;
IF x2=0 THEN
DO;
    n_hos=0;
    n_adm=0;
    rt_adm=0;
    exptime=(efu-sfu+1)/365;
END;
IF x1 THEN OUTPUT;

```

RUN;

-- Calculate mean admissions per 1,000 patient-years by sex. --;

```
PROC TABULATE DATA=rateall;
  VAR rt_adm;
  CLASS sex;
  WEIGHT exptime;
  TABLE sex=""*rt_adm=""*mean*f=8.1;
  KEYLABEL mean="";
  TITLE1 'Total Admission rates per 1,000 patient years by sex';
RUN;
```

EXAMPLE 10: FIRST ADMISSION RATE

-- Using hospitalization data from above, calculate first hospitalization rates. --;

```
DATA frate;
  SET hospdat;
  BY usrds_id;
  RETAIN n_adm exptime rt_adm flag;
  IF FIRST.usrds_id THEN
  DO; /* INITIALIZE RETAIN VARIABLES FOR EACH USRDS_ID */
    n_adm=0;
    exptime=0;
    rt_adm=0;
    flag=0;
  END;
  * FOR HOSPITALIZATIONS OVERLAPPING STUDY START SET EXPOSURE;
  * TIME TO ZERO TO EXCLUDE FROM RATE;
  IF (clm_from<sfu<=clm_thru) THEN
  DO;
    exptime=0;
    flag=1;
  END;
  * IF WITHIN STUDY PERIOD COUNT ADMISSION AND CALCULATE;
  * TIME TO ZERO TO EXCLUDE FROM RATE;
  ELSE IF (sfu<=clm_from<=efu) AND flag=0 THEN
  DO;
    exptime=(clm_from-sfu+1)/365;
    n_adm=n_adm+1;
    flag=1;
  END;
  IF LAST.usrds_id THEN /* OUTPUT ONE RECORD PER USRDS_ID */
  DO;
    IF exptime>0 then rt_fadm=n_adm*1000/exptime;
    OUTPUT;
  END;
RUN;
```


-- Combine result with original incident sample to include patients without hospitalizations. --;

```
DATA frateall;
  MERGE driver (IN=x1) frate (IN=x2);
  BY usrds_id;
  IF x2=0 THEN
  DO;
    n_adm=0;
    rt_fadm=0;
    exptime=(efu-sfu+1)/365;
  END;
  IF x1 THEN OUTPUT;
RUN;
```

-- Calculate mean admissions per 1,000 patient-years by sex. --;

```
PROC TABULATE DATA=frateall;
  VAR rt_fadm;
  CLASS sex;
  WEIGHT exptime;
  TABLE sex=***rt_fadm=***mean*f=8.1;
  KEYLABEL mean=***;
  TITLE1 'First Admission rates per 1,000 patient years';
RUN;
```

Methods & Databases

Section 1: ESRD Data and the USRDS Database System

The main objective of the USRDS Coordinating Center (CC) is to use all relevant ESRD data to create an integrated and consistent database system for outcomes research. The USRDS ESRD database includes ESRD patient demographic and diagnosis data, biochemical values, dialysis claims, and information on treatment history, hospitalization events, and physician/supplier services.

Data Sources

The data used by the USRDS CC originates from the Centers for Medicare & Medicare Services (CMS), Organ Procurement and Transplantation Network (OPTN), the Centers for Disease Control (CDC), the ESRD Networks, the USRDS Special Studies and the U.S. Census.

CONSOLIDATED RENAL OPERATIONS IN A WEB-ENABLED NETWORK (CROWN)

The major source of ESRD patient information for the USRDS is currently the Centers for Medicare & Medicare Services (CMS) Consolidated Renal Operations in a Web-enabled Network (CROWN) data system. This database contains demographic, diagnostic, and treatment history information for all Medicare beneficiaries with ESRD. Data for non-Medicare patients have also been included since 1995, when ESRD Medical Evidence Report forms (CMS 2728) became mandatory for all ESRD patients.

The original CMS ESRD database was called the Program Management and Medical Information System (PMMIS); this was replaced by the Renal Beneficiary and Utilization System (REBUS) in 1995. Having advanced its database technology, CMS migrated the REBUS database into an Oracle relational database system, called the Renal Management Information System (REMIS), in the fall of 2003. In 2003, the Standard Information Management System (SIMS) database of the ESRD Networks was also established; SIMS included information to track patient movement in and out of ESRD facilities, and their transitions from one treatment modality to another. The REMIS database originally included all patients who were alive and had ESRD as of January 1, 1995, or were incident after this date, an approach

adopted from the procedure used to create the SIMS database. However, because the REMIS system as it existed did not include legacy patients for longitudinal studies, CMS expanded it in the fall of 2004 to include all ESRD patients. Together, REMIS and SIMS comprised the CROWN system. In May 2012, Internet-based access to the data system, CROWNWeb, was rolled out nationally. It replaced the functionality of SIMS, interfaces with REMIS, and also provides new data to support calculation of clinical measures.

CMS regularly updates the REMIS/CROWNWeb database, using the Medicare Enrollment Database (EDB), Medicare inpatient and outpatient claims, the OPTN transplant database, ESRD Medical Evidence Report forms (CMS 2728), and ESRD Death Notification forms (CMS 2746). CMS has established data integrity rules to ensure accurate identification of patients in the CMS databases. Each ESRD patient (new or existing) is identified with a unique patient identification number common to both databases, guaranteeing that data for all patients are consistently managed over time.

Working solely with data from the Medical Evidence Report, the CC could establish the first ESRD service data for these patients, but could not consistently generate a detailed treatment history. The integration of the CROWNWeb event data into the CC database, however, allows for the examination of issues that arise in the non-Medicare ESRD population, such as the large and growing number of lost-to-follow-up patients, and for gathering data on patients for whom no data were previously available on initial modality or death.

CMS ESRD MEDICAL EVIDENCE REPORT FORM (CMS 2728)

The CMS ESRD Medical Evidence Report form (CMS 2728) is used to register patients at the onset of ESRD, and must be submitted by dialysis or transplant providers within 45 days of treatment initiation. It establishes Medicare eligibility for individuals who previously were not Medicare beneficiaries, reclassifies previously eligible Medicare beneficiaries as ESRD patients, and provides demographic and diagnostic information for all new ESRD patients regardless of Medicare entitlement. The CMS, USRDS, and renal research communities rely on the form to ascertain patient demographics, primary diagnosis, comorbidities, and biochemical test results at the time of ESRD initiation.

Prior to 1995, dialysis units and transplant centers were required to file the Medical Evidence Report form only for Medicare-eligible patients. Since the 1995 revision, however, providers are required to complete the form for all new ESRD patients regardless of Medicare eligibility status. The 1995 revised form included new fields for comorbid conditions, employment status, expanded race categories, ethnicity, and biochemical data at ESRD onset. The third major revision of the Medical Evidence form, in May 2005, remedied several shortcomings of the 1995 form and its earlier versions and is described in more detail in the next section. Key additions target pre-ESRD care and vascular access use, and additional new fields collect information on glycosylated hemoglobin (HgbA1c) and lipid testing, on the frequency of hemodialysis (HD) sessions, and on whether patients are informed of transplant options. The Medical Evidence form is the only source of information about the cause of a patient's ESRD. Because the list of diseases has been revised, the USRDS stores the codes from each version so that detail is not lost through conversion of one set of codes to the other.

Only one Medical Evidence form (CMS 2728) is expected for each ESRD patient for the entire ESRD treatment period; however, multiple forms may be filed for patients whose insurance eligibility changes due to therapy changes. For example, transplant patients with functioning grafts after 3 years lose Medicare benefits if ESRD was the sole qualification for Medicare eligibility. If such a patient experiences graft failure and returns to dialysis, a second Medical Evidence Report must be filed to reestablish Medicare eligibility. Dialysis patients who discontinue dialysis for more than 12 months also lose Medicare ESRD benefits. If such a patient returns to dialysis or undergoes kidney transplant, a second Medical Evidence form must be filed to reestablish Medicare eligibility.

2005 REVISION OF THE MEDICAL EVIDENCE REPORT FORM (CMS 2728)

The revision of the Medical Evidence Report form (CMS 2728) introduced in May 2005 includes new data collection methods and new variables. It allows users to specify whether the Medicare registration is initial (new ESRD patient), re-entitlement (reinstating Medicare entitlement after a lapse due to no claims being filed for 12 or more months or a functioning graft for 36 or more months), or supplemental (updating missing or incorrect information). This clarifies the intended

use of the form without recourse to the "First Regular Dialysis Start Date," and helps chronicle the historical sequence of multiple forms for the same patient.

Multiple races can be specified. CMS provides a single variable containing a concatenated string representing all selected race codes with binary digits (0s and 1s). This string must be decoded to determine patient race (or races). Similar formatting requirements apply to information on comorbid conditions, medical coverage, and reasons for not informing patients of transplant options. Because the required programming is substantial, we include, for each of these 4 data categories, the original variable with the concatenated string and a new variable with the decoded values (i.e., information noted on form CMS 2728). However, the decoded multiple race values must be presented as "Other" or "Multiple race" to maintain consistency with the legacy race information.

Data fields for nephrologist care, dietitian care, and access type were also added, with their respective time intervals relative to ESRD onset. Data on the laboratory values hematocrit, creatinine clearance, BUN, and urea clearance are no longer collected. Added laboratory values are HbA1c and lipid profiles (TC, LDL, HDL cholesterol, and TG). Additional fields relate to whether patients were informed of transplant options, and if not, why not, and donor type. Comorbid conditions and primary diagnoses were also added. Both the 2005 and 1995 versions of the CMS 2728 form are provided in the USRDS Core SAF dataset and are available in the USRDS Researchers' Guide, Appendix D: Data Collection Forms on the USRDS website: www.usrds.org/research.aspx.

CMS ESRD DEATH NOTIFICATION FORM (CMS 2746)

The ESRD Death Notification form (CMS 2746) is used to report the death of ESRD patients. According to CMS policy, this form must be submitted by dialysis or transplant providers within 30 days of a patient's death, and provides the date and causes of death (primary and secondary), reasons for discontinuation of renal replacement therapy, if applicable, and evidence of hospice care prior to death. It is the primary source of death information for CMS and the USRDS, identifying more than 99 percent of deaths. The USRDS also utilizes several supplemental data sources for ascertaining date and cause of death.

CMS MEDICARE ENROLLMENT DATABASE (EDB)

The CMS Enrollment Database (EDB) is the designated repository of all Medicare beneficiary enrollment and entitlement data, including current and historical information on beneficiary residence, Medicare as Secondary Payer (MSP) status, and Health Insurance Claim/Beneficiary Identification Code (HIC/BIC) cross-referencing.

CMS PAID CLAIMS RECORDS

Inpatient transplant and outpatient dialysis claims records are sometimes used to identify new ESRD patients for whom no Medical Evidence Report form has been filed. These patients are most likely to be non-Medicare patients or beneficiaries already receiving Medicare because of age or disability. They will eventually be entered into the CROWNWeb database, and hence the USRDS database, through the claims records.

For patients without Medical Evidence Reports, these claims are the only reliable information from which to determine first ESRD service dates. These paid claims records, however, only supplement and do not replace other sources of information on incidence and prevalence. Bills for some Medicare-eligible patients may not be submitted to or paid by Medicare. These patients are MSP patients covered by private insurance, HMOs, Medicaid, or the Department of Veterans Affairs (DVA).

ORGAN PROCUREMENT AND TRANSPLANTATION NETWORK (OPTN) DATABASE

CMS began collecting data on all Medicare kidney transplants in the early 1980s. In 1984, the National Organ Transplant Act called for the creation of an Organ Procurement and Transplant Network (OPTN) to collect data and maintain a registry for organ matching and transplantation. The United Network for Organ Sharing (UNOS) was awarded the OPTN contract in 1988 to provide a national system for allocating donor organs and to maintain a centralized data depository for organ transplants. OPTN also began collecting data on all transplants. The OPTN and CMS collection efforts were consolidated in 1994, and only the OPTN continued to collect data on transplant donors and recipients.

CMS ESRD STANDARD ANALYTICAL FILES (SAF)

The CMS SAF contains data from final action claims submitted by Medicare beneficiaries, in which all adjustments have been resolved. For Part A

institutional claims, the USRDS uses the following 100% SAF claims:

- Inpatient
- Outpatient
- Skilled nursing facility (SNF)
- Home health agency (HHA)
- Hospice

For Part-B physician/supplier 100% SAF claims:

- Physician/supplier
- Durable medical equipment (DME)

CMS SAFs are updated each quarter through June of the next year, when the annual files are finalized. Datasets for the current year are created six months into the year and updated quarterly until finalized at 18 months, after which files are frozen and will not include late arriving claims. Annual files are thus approximately 98% complete. The USRDS 2014 SAF includes all claims up to December 31, 2012. Patient-specific demographic and diagnosis information, however, includes data as recent as June 2014.

CMS 5% GENERAL MEDICARE SAF

The 5% general Medicare SAF has the same structure and data elements as the ESRD 100% SAF, but they were extracted from the general Medicare billing database as a random sample of 5% of the entire Medicare population. Because the sample is selected using the last two digits of patient Social Security Numbers, one should expect to see the same group of Medicare beneficiaries represented in the 5% SAFs each year, with exceptions for death, disenrollment, and new Medicare entitlements.

The USRDS CC uses these files to conduct studies on Healthy People 2020 objectives, comparing preventive care and other non-ESRD disease treatments in general Medicare and ESRD patients. In addition, these files are used to create the CKD cohort dataset.

CLINICAL PERFORMANCE MEASURES (CPM) PROJECT

CMS developed the ESRD CPM (formerly the ESRD Core Indicators Project) to collect information on the quality of care provided to dialysis patients. The data originate from surveys completed by primary care facilities, and focus on dialysis adequacy measures, anemia management, and vascular access use. Additional clinical parameters such as albumin are also available. These data were collected annually

beginning in 1994, using a random sample of patients aged 18 years and older, alive and on dialysis at the end of each calendar year; on average, about 8,500 in-center hemodialysis patients and 1,500 peritoneal dialysis patients were surveyed each year. Data collection for all pediatric patients aged 12 to 17 years began in 2000, and in 2002 was expanded to all in-center hemodialysis patients aged younger than 18 years. Starting in 2005, all peritoneal dialysis (PD) patients in the U.S. were sampled for the pediatric PD data collection. In anticipation of the national release of the CROWNWeb system and its supporting performance measures reports, CMS concluded its CPM project in 2009, making 2008 its final survey year. The USRDS CC, in collaboration with CMS, provides CPM/USRDS merged data to the general research community.

CMS DIALYSIS FACILITY COMPARE DATA

The USRDS uses the CMS Dialysis Facility Compare data to define chain and ownership information for each renal facility. Before the 2003 ADR, similar data were extracted from the Independent Renal Facility Cost Report (CMS 265-94).

CMS ANNUAL FACILITY SURVEY (AFS)

In addition to the CMS ESRD databases, independent ESRD patient counts are available from the CMS Annual Facility Survey (AFS; CMS 2744), which all Medicare-certified dialysis units and transplant centers are required to complete at the end of each calendar year. The AFS reports counts of patients being treated at the end of the year, new ESRD patients starting during the year, and patients who died during the year. Counts of Medicare and non-Medicare end-of-year patients are included. While AFS files do not carry patient-specific demographic and diagnosis information, they do provide independent patient counts used to complement the CMS patient-specific records.

CDC NATIONAL SURVEILLANCE DATA

From 1993–1997 and 1999–2002, the CDC used its survey National Surveillance of Dialysis-Associated Diseases in the United States to collect information from dialysis facilities on patient and staff counts, membrane types, reuse practices, water treatment methods, therapy types, vascular access use, antibiotic use, hepatitis vaccination and conversion rates (for both staff and patients), and the incidence of HIV, AIDS, and tuberculosis. None of the information is

patient-specific. Because the CDC terminated this program in 2003, the last surveillance report is for 2002 data. The CDC did not conduct a survey in 1998.

DIALYSIS MORBIDITY AND MORTALITY STUDY (DMMS)

The DMMS was an observational study that collected data on demographics, comorbidity, laboratory values, treatment, socioeconomic factors, and insurance for a random sample of U.S. dialysis patients, using dialysis records. Data were collected on 6,000 ESRD patients in each of Waves I, III, and IV, and 4,500 patients in Wave II, a total of 22,500 patients over three years. Waves I, III, and IV are each historical prospective studies in which data were collected for patients receiving in-center hemodialysis on December 31, 1993. Data were abstracted from patient medical records, and each patient was followed from December 31, 1993, through the earliest of data abstraction, death, transplant, change in modality, or transfer to another facility. Wave II is a true prospective study of incident hemodialysis and peritoneal dialysis patients for 1996 and some incident patients entering the ESRD program in the first part of the 1997 calendar year.

CASE MIX ADEQUACY STUDY (CMAS)

The objectives of the USRDS Case Mix Adequacy Study of Dialysis 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 at different dialysis doses
- Assess how this relationship is affected by dialyzer reuse
- Assess the impact of different dialysis membranes on patient morbidity and mortality

The study consisted of two groups of patients: an incident sample of ESRD patients who began hemodialysis during 1990, and a prevalent sample of hemodialysis patients with ESRD onset before 1990. A total of 7,096 patients from 523 dialysis units were included, the pre- and post-BUN values needed to calculate delivered dialysis dose were present for approximately 3,300 patients. 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

The objectives of this study were to:

- Estimate the correlation of comorbidity and other factors present at ESRD onset with subsequent mortality and hospitalization rates, adjusting for age, sex, 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 using different treatment modalities
- Compare relative mortality rates by treatment modality, adjusting for selected comorbid conditions and other factors

Data were collected for 5,255 incident patients in 1986 and 1987 at 328 dialysis units nationwide.

PEDIATRIC GROWTH AND DEVELOPMENT

The objectives of the USRDS Pediatric ESRD Growth and Development Study were to:

- Establish a baseline for assessing the relation of pediatric ESRD patient growth and sexual maturation to modality
- 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 dialysis units.

CONTINUOUS AMBULATORY PERITONEAL DIALYSIS (CAPD) AND PERITONITIS STUDY

The USRDS CAPD and Peritonitis Rates 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, up to a maximum of 14 patients per dialysis unit. All units providing CAPD training participated in the study. The sample includes 3,385 patients from 706 units.

UNITED STATES CENSUS

The U.S. population data are from the 2000 and 2010 U.S. Census, and also incorporate CDC postcensal and intercensal population estimates. The data and methods for these estimates are available at http://www.cdc.gov/nchs/nvss/bridged_race.htm. Both intercensal and postcensal estimate datasets are available at http://www.cdc.gov/nchs/nvss/bridged_race/data_documentation.

htm. USRDS summarizes the data with different race categories at state and national level.

USRDS DATABASE SYSTEM

The USRDS CC has developed a centralized ESRD patient database by integrating data from the above data sources and establishing methods to identify patients with ESRD. We use this database to update and maintain data on demographics, clinical measurements, biochemical lab test results, renal replacement therapy, treatment history, and all medical service events reported in the Medicare claims database. Through this patient-oriented database we attempt to define each individual ESRD patient through multiple stages of data cleaning, conversion, validation, and consolidation. Establishment of a universal patient identification system was critical to ensure that unique patient identification numbers are assigned to each ESRD patient at the time of ESRD initiation, and to accurately track counts and rates of incident and prevalent cohorts over time.

In 1994, the USRDS CC introduced the Standard Analysis Files (SAFs)—not to be confused with the CMS Standard Analytical Files for Part A and Part B claims data—to help meet the ESRD data needs of a wide variety of research studies. These SAFs were subsequently enhanced to include not only ESRD clinical and claims data from CMS, but also transplant and waiting list data from OPTN. All SAFs are indexed or sorted by the unique USRDS-specified patient identification number, and patient identifiers (name, address, SSN, HIC/BIC, etc.) are removed to protect patient confidentiality.

The CC also uses the USRDS ESRD database to generate datasets for the tables, graphs, and maps in the USRDS ADR.

The USRDS ESRD database is updated quarterly with data obtained from the various data sources. The CC generally receives CMS SAF claims data, Facility Survey data, patient lists and patient events, Medical Evidence Report, and Death Notification data, on a quarterly basis. We receive OPTN transplant and waiting list data monthly. The EDB is currently updated annually, although we will be moving toward more regular updates. These multiple updates allow the CC to assess growth of the ESRD population, the demographic distribution of ESRD patients, and changes in the percent of patients diagnosed with major diseases secondary to ESRD.

Section 2: ESRD Patients

ESRD is defined as chronic renal failure requiring renal replacement treatment—dialysis or transplant—to sustain life. It is not the same as acute renal failure, from which patients are expected to recover within weeks or months. A Medical Evidence Report must be completed immediately by renal providers for all ESRD patients to register them in the CMS ESRD database and to apply for Medicare eligibility if they were not previously eligible.

To establish the incident and prevalent cohorts by year, modality, primary cause of renal failure, and other factors, patient demographic and clinical information are required as well as treatment history data. This information can be obtained from the USRDS Core SAF dataset from the following files: PATIENTS, MEDEVID, RXHIST, RXHIST60, and PAYHIST.

Identifying ESRD Patients

A person is identified as having ESRD when a physician certifies the disease on the Medical Evidence Report form (CMS 2728), or when other evidence of chronic dialysis or kidney transplant exists. Patients with acute kidney failure who are on dialysis for days or weeks, but who then recover kidney function, are excluded from the database if their Medical Evidence forms have not been submitted. Patients who die soon after kidney failure without receiving dialysis are sometimes missed.

First ESRD Service Date

The first ESRD service date (FSD) is the single most important data element in the USRDS database, and each patient must, at a minimum, have a valid FSD. This date is used to determine each new patient's incident year and the first year in which the patient is counted as prevalent. The date 90 days after the FSD is used as the starting point for many patient survival outcomes analyses. This 90-day rule allows each new ESRD patient to obtain Medicare services despite potential delays in completing the Medicare eligibility application process, and it provides an adequate time period for patients to arrive at a stable and suitable dialytic treatment modality.

In most cases, the first service date is derived by identifying the earliest date of various potential indicators:

- The start of dialysis for chronic kidney failure as reported on the Medical Evidence form
- The first CROWNWeb/SIMS event

- A kidney transplant as reported on a CMS or OPTN transplant form, a Medical Evidence form, or a hospital inpatient claim, or
- The first Medicare dialysis claim

There are two exceptions to the first ESRD service date determination:

- If the CROWNWeb/SIMS event and Medical Evidence form agree (within 30 days of each other) and are more than 90 days after the first Medicare dialysis claim, and, if in addition, there is no transplant event between the first dialysis claim and the earlier of the CROWNWeb/SIMS event date and Medical Evidence form date, then first service date is defined as the earlier of the CROWNWeb/SIMS event date and Medical Evidence form date.
- If the Medical Evidence form date is one year earlier than the first CROWNWeb/SIMS event date, and if the first claim date or first transplant date agrees with the first CROWNWeb/SIMS event date, then the CROWNWeb/SIMS first event date is used as the first service date.

Death Date Determination

After the ESRD first service date, the date of death is the most critical piece of information in the ESRD database. Death dates are obtained from several sources, including the CMS Medicare Enrollment Database, CMS forms 2746 and 2728, the OPTN transplant follow-up form, CROWNWeb database, the Social Security Death Master File, and inpatient claims. Because multiple sources report death information for the same patient, one patient may have several reported dates. For these patients, we primarily use the median of the various death dates reported. However, in the small number of cases where there are only two death dates and they are more than 70 days apart, we use instead the most recent of the two dates.

Transplant Dates

The CMS and OPTN transplant data files overlap for 1988–1993, and transplants are also identified from Medical Evidence forms that indicate transplant as the initial modality, from CROWNWeb transplant events, and from institutional inpatient claims. To resolve any conflicts among these three sources, the USRDS has adopted the following procedure:

- Before 1988, all transplant events found in CMS PMMIS/REBUS/REMIS Transplant files are used.

- After 1994, all transplant events found in OPTN files are used.
- Between 1988 and 1993, all transplant events found in OPTN files are used, and additional transplant events from the CMS PMMIS/REBUS/REMIS Transplant file are used only if they occur at least 30 days on either side of a previously accepted transplant event.
- Additionally, transplant events associated with reported incident transplant patients from the Medical Evidence Report are used if they occur at least 30 days on either side of a previously accepted transplant event. Transplant events found in CMS Inpatient claims records are also included as transplants found in the CROWNWeb patient events data.

Each transplant event found in the Transplant file of the USRDS Core SAF dataset is thus a unique event derived from the OPTN database, the CMS Transplant database, Medical Evidence Report records, CROWNWeb patient events, or institutional claims files.

Graft Failure

The CC assumes that a graft failure date reported in the OPTN transplant follow-up file or the REMIS identification file is correct unless death or a new transplant occurs before this date. However, a graft failure date may be unrecorded in both files. In this case, the CC derives the graft failure date from the following sources:

- Date of death
- Date of subsequent transplant
- Date of return to regular dialysis, indicated by a continuous period of dialysis billing records covering a minimum of 60 days with at least 22 reported dialysis treatments
- Date of return to dialysis reported on the Medical Evidence Report, or the date of graft nephrectomy from the OPTN transplant follow-up record or a Medicare claim
- If no failure date is available, then the earliest of the above dates is used as the graft failure date

Medicare and Non-Medicare Patients

Beneficiaries are enrolled in Medicare based on criteria defined in Title XVIII of the Social Security Act of 1965 and in subsequent amendments to the Act. A person in one of these four categories is eligible to apply for

Medicare entitlement:

- Aged 65 years and older
- Disabled
- Enrolled in the ESRD program
- Railroad Retirement Board (RRB)

Most ESRD patients are eligible to apply for Medicare as their primary insurance payer. Some, however, are not immediately eligible for Medicare primary payer coverage because of their employment status and pre-existing primary insurance payers. These are covered by payers such as Employer Group Health Plans (EGHPs), the Department of Veteran Affairs, and private insurers. Typically, these patients wait 30–33 months before becoming eligible for Medicare as primary payer, and are not included in the EDB database during the waiting period. Many of these patients, particularly those who are new since 1995, have FSDs established by Medical Evidence Reports, even though they have no dialysis claims or hospitalization events in the CMS claims database. Treatment history events can be tracked by the CROWNWeb patient events file which follows changes in modality, transplants, transfers between facilities and death. In the REMIS database, all non-Medicare ESRD patients are designated “ZZ” in the two-character Beneficiary Identification Code field. All ESRD patients, regardless of their Medicare Eligibility status, are in the CROWNWeb system.

The USRDS recognizes that these non-Medicare (“ZZ”) patients are ESRD patients and should therefore be included in patient counts for incidence, prevalence, and treatment modality, as well as mortality and transplant rate calculations. Calculations of hospitalization statistics, however, should not include these patients because of the small number of claims available in the first 30 to 33 months after first ESRD service.

Section 3: Treatment History

The USRDS uses all available data to create a treatment history for each patient in the database, including all modality events, their duration, and the renal providers involved in each patient’s care. This history defines incident and prevalent cohorts and determines censoring points and outcomes for observational studies. The CROWNWeb event database is the primary source of the modality sequence file, and the dialysis claims are used as a way of confirming placements and identifying problem cases. As described in the previous section, we use all

available sources to determine first service dates, deaths, transplants, and transplant failures.

Treatment Modality Categories

Table 2 lists modality categories used by the USRDS. They can be described as detailed or general. For most analyses, the general categories, which combine detailed modality categories, are sufficient. The recovered renal function (RRF) modality/event can be established only if it occurs within the first 180 days of first service date, and the RRF period persists for at least 90 days. The RRF modality/event is similar to the lost-to-follow-up event in that patients with an RRF event are not included in the prevalent populations for outcomes analyses. However, as with lost-to-follow-up events, these patients remain in the modality sequence so that subsequent renal episodes can be closely tracked in a timely manner. A new modality/event — discontinued dialysis — was introduced in this year's SAF from the CROWNWeb Event file.

Type	Detailed	General
1	Center HD	HD
2	Center Self HD	HD
3	Home HD	HD
4	HD Training	HD
5	CAPD	PD
6	CAPD Training	PD
7	CCPD	PD
8	CCPD Training	PD
9	Other PD	PD
A	Uncertain Dialysis	Unknown Dialysis
B	Discontinued Dialysis	Discontinued Dialysis
T	Functioning Tx	Transplant
X	Lost to follow-up	Lost to follow-up
D	Death	Death
Z	Recovered Function	Recovered Function

How Treatment Modality is Determined

The treatment history files in the USRDS Core SAF dataset record the sequence of modalities for each patient. The variables included in these files are presented in Table 3. Each record in the file indicates a period of therapy with a given modality, and any change in provider or detailed modality results in a new record. Several data sources are used to determine the treatment history, including the Medical Evidence file, the CROWNWeb Event file, the OPTN Transplant events file, and Medicare claims files.

File Name	Variable label	Variable name
RXHIST (Detailed treatment history)	USRDS_ID	USRDS patient identification
	BEGDATE	Beginning date of a modality period
	ENDDATE	Ending date of a modality period
	BEGDAY	Start day of modality period (First Service Date = 1)
	ENDDAY	End day of modality period (First Service Date = 1)
	RXDETAIL	Detailed treatment modality for period
	RXGROUP	Grouped treatment modality for period
PROVUSRD	USRDS assigned facility identification	
RXHIST60 (Condensed treatment history)	USRDS_ID	USRDS patient identification
	BEGDATE	Beginning date of a modality period
	ENDDATE	Ending date of a modality period
	BEGDAY	Start day of modality period (First service date = 1)
	ENDDAY	End day of modality period (First service date = 1)
	RXGROUP	Grouped treatment modality for period

In constructing this treatment history file (RXHIST), these conventions are followed:

- The sequence always begins with the first ESRD service date and ends with the earliest of date of death or end of the period for which the data are complete. For the category death, the end dates are missing from a patient's last record. Other categories always have an end date.
- Death is treated as the last event of the day, and is the absolute end point of the treatment history.
- If two dialysis billing periods overlap, the earlier is terminated at the start of the later. If a dialysis billing period is entirely contained within another billing period, the contained period is ignored.

- A functioning transplant is considered treatment, not recovery from ESRD.
- If a dialysis billing period overlaps a transplant date, the dialysis period is closed at the transplant date.
- If a graft failure is recorded in the database, but there are no subsequent records of dialysis or another transplant, a period of “unknown dialysis” is inserted. If no dialysis billing records or re-transplant appears within one year, the patient is designated as lost-to-follow-up at the end of that year, and this status continues until dialysis or new transplant records appear.
- Once established, a modality is assumed to continue until a change in provider or detailed modality occurs. For dialysis patients, if no new dialysis billing data appear for one year, the patient is reclassified as lost-to-follow-up at the end of that one-year period.
- A functioning graft is assumed to continue until an indication of graft failure or death appears, or evidence of regular or maintenance dialysis.

The 60-Day Collapsing Rule

The CC uses the convention that a dialysis modality must continue for at least 60 days to be considered stable. A transplant is considered a stable modality regardless of duration.

Because the dialysis treatment history is derived from the CROWNWeb events file, the Medical Evidence Report, OPTN transplant events, and Medicare billing records, it includes intermixed and overlapping details that are not required or are unsuitable for most analyses. A long period of CAPD, for example, may be interrupted by a short period of inpatient hemodialysis treatment, or a patient may go on vacation and receive dialysis from a different provider. By applying the 60-day rule, we collapse modality periods of less than 60 days, and periods that differ only by provider, giving a less complex treatment history for analyses that require less detail than is available in the detailed treatment history. For maximum flexibility, we provide two treatment history files, one with full detail (RXHIST), and one applying the 60-day collapsing rule (RXHIST60). Tables 4 and 5 show an example of how these two files differ. The 60-day file is constructed from the detailed file as follows:

- Consecutive records with the same modality group are collapsed into a single record. This removes changes of provider only and changes between detailed modalities within the same general modality group.

- Remaining modality periods, except functioning transplant < 60 days are recoded as uncertain dialysis, and consecutive dialysis records labeled uncertain are combined.
- If a modality is interrupted by a different modality that lasts < 60 days (e.g., a period of peritoneal dialysis between two longer periods of hemodialysis), the short modality period is ignored and the longer modality extends over the entire period.

Table 4. RXHIST

USRDS ID	BEGDATE	ENDDATE	RXDETAIL	PROVUSRD
900000001	1/1/91	1/31/91	Center HD	5
900000001	2/1/91	5/31/91	Center HD	7
900000001	6/1/91	6/31/91	CAPD training	14
900000001	7/1/91	1/31/92	CAPD	14
900000001	2/1/92	9/30/92	Center HD	5
900000001	10/1/92	10/15/92	CAPD	14
900000001	10/17/92	5/31/93	Center HD	5
900000001	6/1/93	7/15/93	Transplant	23
900000001	7/16/93	2/12/95	Center HD	5
900000001	2/13/95	1/31/96	Center self HD	5
900000001	2/1/96	2/29/96	HD training	5
900000001	3/1/96	2/12/97	Home HD	5
900000001	2/13/97	.	Death	

Table 5. RXHIST60

USRDS ID	BEGDATE	ENDDATE	RXDETAIL
900000001	1/1/91	5/31/91	Hemodialysis ^a
900000001	6/1/91	1/31/92	CAPD ^b
900000001	2/1/92	5/31/93	Hemodialysis ^c
900000001	6/1/93	7/15/93	Transplant ^d
900000001	7/16/93	2/12/97	Hemodialysis ^e
900000001	2/13/97	.	Death

^aProvider change only.
^bSame modality category.
^cCAPD record only 15 days.
^dThe 60-day rule does not apply to transplants.
^eSame modality category.

Integration of the CROWNWeb and CMS Claims Databases

For patients who either do not appear in the CROWNWeb events file or for whom the only event is “New ESRD Patient,” and patients who have transfer-out gaps, the Medicare dialysis claims file is used. For “Transfer Out” and “Transfer Out for a Transplant” events with large gaps (seven days or more), claims falling in gaps are included, with the exception that no claims data are included if the “Transfer Out for a Transplant” event has a corresponding transplant/transplant failure event that occurred within (before or after) 30 days. Claims data are also included for the periods after “Transplant Failure” events and “Discontinued Dialysis” modality if the periods are longer than seven days.

Because the claims data capture the modality “Center Self Hemodialysis” more accurately than the CROWNWeb data, this claims-based designation overrides other dialysis modalities from CROWNWeb. Any CROWNWeb dialysis event that falls into a “Center Self Hemodialysis” period as determined by claims is recoded as “Center Self Hemodialysis.”

Some events that do not make sense are removed. These include events that occur before a patient’s first service date, those falling between “Transplant” and “Transplant Failure,” and “Transfer Out for A Transplant” events that occur 60 days or less after the corresponding “Transplant.”

Determining Lost-to-Follow-up Periods

Gaps frequently exist in the CROWNWeb and billing data upon which modality periods are based. The USRDS assumes that a modality continues until death or the next modality-determining event. A patient with a functioning transplant is assumed to maintain it unless a new CROWNWeb event, claim event, or death date is encountered in the data. A dialysis modality, in contrast, is assumed to continue for only 365 days from the date of the last claim, in the absence of a death date or dialysis claims. After this period the patient is declared lost-to-follow-up, until the occurrence of a new CROWNWeb event, dialysis claim, or transplant event.

Patients are considered lost-to-follow-up beginning 365 days after a “Transplant Failure” event or “Discontinued Dialysis” modality. Patients for whom the only event is a first service date, and who do not exist in any other files were also treated as

lost-to-follow-up, beginning one year after the first service date. A number of events can result in a lack of dialysis data and eventual reclassification of a patient as lost-to-follow-up:

- The patient may have recovered renal function (RRF) and many no longer have ESRD. For a valid patient classification, this event must occur within 180 days of the first service date, and the RRF period must persist for at least 90 days.
- The patient may no longer reside in the U.S.
- The patient’s death may not have been reported to the Social Security Administration or to CMS.

Section 4: Payer History

ESRD Payer Sequence File

The ESRD payer history file is similar to the treatment history file. CMS payer information is used to create a continuous sequential history of payers for each patient in the ESRD database, beginning with the first ESRD service date. Each patient’s FSD in the payer history file is the same date reported in the treatment history file. Data from the Medicare Enrollment Database and dialysis claims information are used to categorize payer status as Medicare primary payer (MPP), Medicare secondary payer (MSP), or non-Medicare. The claims database contains data only for MPP and MSP patients, so claims based analyses, such as analyses of cost and hospitalization, will not include non-Medicare patients. Non-Medicare patients therefore must be identified and excluded when determining numbers of patients or patient years at risk for analyses of cost per patient or hospitalization rates. In addition, as it is impossible to determine the complete hospitalization history or complete cost of care for ESRD patients with MSP coverage, such analyses should also exclude patients during the periods when they have this coverage. The payer history file can be used to make these exclusions.

Pre-ESRD Payer Sequence File

The pre-ESRD payer sequence file can be used to determine payer status for the periods prior to first ESRD service date. This payer sequence file is similar to the standard ESRD payer service file except that the pre-ESRD payer sequence file begins at the first evidence of Medicare enrollment from the Enrollment Database, rather than first ESRD service date, as is the case with the

ESRD payer sequence. The pre-ESRD payer sequence ends the day before the first ESRD service date.

Data Sources

The payer for any given time period is determined by evaluating several data sources. The CMS Enrollment Database (EDB) is used to determine Part A, Part B, Group Health Organization, MSP Primary Payer, Third Party Part A, and Third Party Part B payers. The CMS claims billing files provide dates of regular maintenance dialysis, used as an indicator of Medicare as primary payer. The PATIENTS file in the USRDS Core SAF dataset contains dates of death, used to help establish the end point of the payer sequence.

Payer Categories

Table 6 shows the payer categories used by the USRDS and indicates whether a patient is considered a Medicare patient (yes or no), and whether the patient has dual Medicare/Medicaid eligibility for that payer time period (yes or no).

Medicare Code	Medicare/Medicaid Description	Medicare Patient	Medicare/Medicaid Dual Eligibility
MPAB	Medicare Primary, Part A and Part B	Y	Y or N
MPO	Medicare Primary, Other	Y	Y or N
MSP-EGHP	Medicare as Secondary Payer with EGHP	Y	N
MSP-nonEGHP	Medicare as Secondary Payer with no EGHP	Y	N
HMO	Group Health Org	Y	N
WAIT	90 Day Waiting Period	Y or N	Y or N
OTH	Other/Unknown	N	N

How the Payer Category is Determined

The payer history file in the USRDS Core SAF dataset records the sequence of payers for each patient. Table 7 shows the variables included in this file. Each record in the file represents a time period covered by a particular payer. Any change in payer, Medicare status, or dual eligibility status results in a new record.

Note that the "WAIT" status can be either Medicare or non-Medicare, depending on the Medicare status of the sequence following the "WAIT" period.

Variable label	Variable Description
USRDS_ID	USRDS Patient Identification
BEGDATE	Beginning date of a payer period
ENDDATE	Ending date of a payer period
PAYER	Payer for the period
MCARE	Payer includes Medicare (Y/N)
DUALELIG	Payer includes Medicare and Medicaid (Y/N)

Because several data sources are used to determine payer history, more than one payer may be reported for any given time period. In constructing the history, the following conventions are followed:

- The sequence begins with the first ESRD service date. If the patient has died, the end date for the last patient record is the date of death. If the patient has not died, the end date is missing from the patient's last record.
- The expected ESRD entitlement date is defined as the remainder of the month of incidence, plus two full calendar months. If the payer on the first service date is unknown, the patient is assumed to be in the waiting period prior to entitlement, and the first payer is assigned to the value "WAIT." The "WAIT" period lasts until the expected entitlement date, or until another payer is identified. The Medicare status of the "WAIT" period is assumed to be the same as the Medicare status of the following period.
- If the reported end date for a payer is missing in the data source, the payer is assumed to continue sharing responsibility through the date of death or through the period of available payer information, whichever comes first.
- If more than one payer is identified for a time period, precedence for assigning the payer for the sequence is as follows: HMO, Medicare Primary if indicated by the claims billing file, MSP, then Medicare Primary if indicated by the Enrollment Database.
- If payer and dual eligibility status are the same in consecutive payer sequences, the sequences are collapsed into one time period, starting with the beginning date of the first and ending with the end date of the last.
- Unlike the treatment history file, the payer history file does not require a payer to continue for any specific time period to be considered stable, and all changes in payer are reported. At their discretion,

users may apply additional rules to manage the payer history sequence.

Gaps in the Payer History

Some patient payer histories include gaps with Other/Unknown as payer between sequences with identified payers. This might occur for several reasons, similar to the explanations for gaps in the treatment history:

- Some patients recover enough renal function to discontinue dialysis and the USRDS database may contain no payer information for non-dialysis time periods.
- Payer information for successful transplant patients may not be reported in the data sources because Medicare eligibility terminated after three years with a functioning graft.
- A patient may leave the country and become lost-to-follow-up.
- A patient may die and the death data may not reach the USRDS.
- CMS may not have collected payer information.
- Reporting, data entry, and clerical errors may obscure the record of FSD and the corresponding payer.
- Errors in patient identification may cause data for a single patient to be split between two patients or to be associated with the wrong patient.

If a patient is alive but the last reported payer in the data sources ends on a date before the reported data extraction date, an additional payer sequence is created with Other/Unknown payer to extend the patient's payer history. Thus, only patients who have died have end dates in their last payer sequences, and the last end date is the date of death.

Some patients have no identified payers. Their payer histories are represented in the file as one payer sequence, starting at first ESRD service date, with missing end date and Other/Unknown payer.

Section 5: Medicare Claims

Introduction

The USRDS Coordinating Center has created files from CMS billing data and incorporated them 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 in the section titled "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 Standard Analytical Files (SAFs), and the physician/supplier data are from the 100% National Claims History Nearline file. For ESRD claim data prior to 2012 and pre-ESRD claim data prior to 2010, information on outpatient dialysis and hospital inpatient stays not included in the CMS SAFs are obtained from PMMIS/REBUS/REMIS. 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 given year is frozen at the end of June of the subsequent year, 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 ESRD claim data prior to 2012 and pre-ESRD claim data prior to 2010, the PMMIS/REBUS/REMIS 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/REMIS were used only when a matching hospital stay or dialysis record is not in the CMS SAFs; PMMIS/REBUS/REMIS data are distinguished by the value of the HCFASAF variable (M or Q). SAF data were given preference because of their greater detail and are now the only source. ESRD claim data from 2012 and

pre-ESRD claim data from 2010 no longer include data from REMIS summarized files.

CMS Institutional SAFs and Physician Supplier and DME (Part B) claims are available on a quarterly basis and are finalized and closed for a particular service year 18 months following the start of that service year. For example, the CMS 2014 claims SAFs are finalized and closed in June 2015. This means that for a claim incurred on January 1, 2014, this system allows for an 18-month window for processing; claims incurred in December 2014 have a 6-month window for processing to complete.

The service dates of the claim correspond to the actual dates of service and should be used to determine inclusion in analyses. The calendar year of the SAFs should not be used to determine inclusion in analyses. A small number of claims included in a given year's file may have service dates do not correspond to that claim year.

Patients and Time Periods Included

The claims database contains data only for MPP and MSP patients, so claims based analyses, such as analyses of cost and hospitalization, will not include non-Medicare patients. Non-Medicare patients therefore must be identified and excluded when determining numbers of patients or patient years at risk for analyses of cost per patient or hospitalization rates. In addition, as it is impossible to determine the complete hospitalization history or complete cost of care for ESRD patients with MSP coverage, such analyses should also exclude patients during the periods when they have this coverage. The payer history file can be used to make these exclusions.

In addition, 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 are 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, such 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.

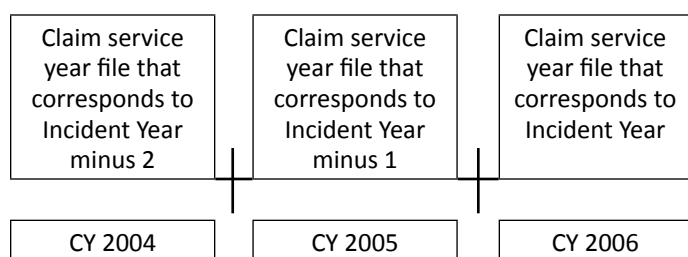
The USRDS database also includes pre-ESRD claims for patients who were entitled to Medicare due to age or disability before they developed ESRD. Because these data are not available for all patients, and because it is likely that patients entitled to Medicare before ESRD are systematically different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

Pre-ESRD Medicare Institutional claims and Physician/Supplier claims are available starting from the 1993 claim service year. The structure of the pre-ESRD claims files is identical to the ESRD claims files (see the "Basic File Structure" section below). Each pre-ESRD claim service year file contains claims for multiple incident years. For example, the claims labeled 2004 contain the pre-ESRD claims for patient incidents in 2004, 2005 and 2006. The claims are NOT constrained to an exact two year period. Patients who are treated for ESRD for the first time are known as incident ESRD patients. Patients who are incident for ESRD at the end of a calendar year may have claims extending back for nearly three years prior to their incident date.

To obtain a patient's complete pre-ESRD records, three pre-ESRD claim files should be used. These are the claim service year file that corresponds to the calendar year of incidence for that patient, the claim service year file that corresponds to the incident calendar year minus 1, and the claim service year file that corresponds to the incident calendar year minus 2. For example, to obtain pre-ESRD Physician/Supplier records for a patient incident in 2006, PS2006, PS2005 and PS2004 should all be used (see Figure 2).

Figure 2. Example of Pre-ESRD Physician/Supplier Record for an Incident ESRD Patient in 2006



It is important to note that some overlap exists between the pre-ESRD claims sets prior to 2010 and the ESRD claims sets prior to 2012, as both claims sets include complete claims information for the calendar year of incidence. Care must be taken to avoid double counting of claims during this time frame. Our recommendation is to search the pre-ESRD claims up to, but not including first ESRD service date, and the ESRD claims starting at first ESRD service date. Pre-ESRD files from claim service year 2010 and on contain only pre-ESRD claims. ESRD files from claim service year 2012 and on contain only ESRD claims.

Basic File Structure

The claims are organized by claim service year. The majority of records in a particular claim service year file are claims with a service date within that claim service year. For example, in PS2004 (Physician/Supplier records incurred in claim service year 2004) most of the records have a claim from date within calendar year 2004.

Institutional claims are for hospital inpatient stays, hospital outpatient services, and care provided at dialysis facilities, 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. From 2008, claim details are separated into Institutional Claim Details file and Revenue Center Details file.

Physician/supplier claims are bills covering physician, laboratory, and radiology services, as well as 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 record, which is a line-item of the entire claim. One visit to a physician can generate multiple line-item 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.

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 are changed in order to accommodate the changes in the Institutional files obtained from CMS. 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., one line item for dialysis, one for erythropoietin (EPO), one 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 for each session (as opposed to one line item in previous years), 13 separate line items for EPO administrations (instead of one), 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. Complete listings of the variables in the Institutional Claims, Claim Details, and Revenue Center Details files are available in Appendix B: Data File Descriptions. 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.

Institutional 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, which are described in the “Institutional Claim Details” and “Revenue Center Details” sections. Complete listings of the variables appearing in the Institutional Claims files are available in Appendix B: Data File Descriptions. The detailed information for some Institutional Claims file variables are described in the next section.

Institutional Claims File Variables

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. The UNITS value from the Revenue Center detail that indicates dialysis is used to calculate the number of dialysis sessions. For in-center hemodialysis, UNITS generally indicates a plausible value for the number of dialysis sessions. For other types of dialysis, particularly CAPD and CCPD, this variable may indicate the number of days. An adjustment is made for CAPD and CCPD to calculate “hemodialysis equivalent” sessions for these modalities. DIALSESS is the resulting number of total dialysis sessions. 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. All dialysis revenue line items are included in the Institutional Claim details file between 2001 and 2007, and are included in the Revenue Center details file starting with calendar year 2008.

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 are retained in the Institutional Claim Details SAF starting with CY2001. The variable for the dose of EPO (EPODOSE), prior to 2008 either came from a Claim Related Value detail with code “68”, or have to be calculated by dividing the total payments variable (RPRVPMT) by price per unit for HCPCS codes “Jo885”, and “Jo886”, and, with the 2012 file, code “Q4o8i”, which began appearing on bills after 2007. 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 now 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. EPODOSE is created as a composite variable which checks for both the relevant HCPCS and the CRV details and considers the year in which the bill was incurred to properly handle the UNITS variable. Also starting with the 2008 claims, a summarization of darbepoetin (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)
- CMS claim related condition codes
- CMS claim related value codes

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 26. 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). Prior to 2012, for CODETYPE = “V”, only the records with CODE equals “A8”, “A9”, “37”, and “48” were kept in Institutional Claim details. From 2012 and on, all records for CODETYPE“V” are kept in Institutional Claim details. Complete listings of the variables appearing in the Institutional Claim Details files are available in Appendix B: Data File Descriptions.

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. Complete listings of the variables appearing in the Revenue Center Details files are available in Appendix B: Data File Descriptions. The detailed information for some Revenue Center Details file variables are described in the next section.

Institutional Claim Details and Revenue Center Details File Variables

CDTYPE and CODE Variables

The variable CDTYPE indicates the type of code contained in the CODE variable (e.g., a procedure code, a revenue center code, etc.). For instance, a code of "P" in the CDTYPE variable indicates that the value of the CODE variable is an ICD-9 procedure code. Both variables are present on every record, but the variables UNITS, REV_CH, HCPCS, and URR_CD 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". From 2008, CDTYPE = "R" records are all in Revenue Center details file. 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. Starting with calendar year 2010, the HCPCS code may be accompanied by up to five HCPCS modifier codes (MOD1-MOD5). These modifiers are used to report information related to the services described by the HCPCS code. For example, beginning July 1, 2010, CMS requires the use of the modifiers V5 (catheter), V6 (graft), and V7 (fistula) to report the type of vascular access in use, V8 (infection) and V9 (no infection) to report absence or presence of access infection on the last dialysis line item of each dialysis bill. Another example would be modifiers for IV (JA) and subcutaneous (JB) administration of ESA.

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 payments 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.

Prior to 2012 file, claim-level and line-item data were treated as distinct and were stacked together. From 2012 file and on, claim-level and line-item data are treated as related and line-item data that are matched to a claim were linked to their "parent" claim.

From 2012 file and on, CLMTYP variable, which indicates whether the record comes from physician/supplier claim or durable medical equipment claim, is added to Physician/Supplier Claims files. CLMTYP = "P" indicates a physician/supplier claim; CLMTYP = "D" indicates a durable medical equipment claim. Claims may or may not have supporting line item data; notably, payment variables for records with no line item data will be missing.

From 2012 file and on, CDTYPE variable is dropped. Prior to 2012 file, 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.

The DIAG variable indicates the diagnosis code associated with the claim. 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 (PLCSRVC) 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 payment 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 (a complete listing of the variables in the Physician/Supplier Claims files is available in Appendix B: Data File Descriptions).

After the implementation of the bundled payment system in 2011, approximately 15% fewer physician/supplier claims were submitted to CMS. We believe this is because the implementation of bundled care changed the way in which healthcare providers applied for Medicare reimbursement for ESRD patient care. Specifically, a given patient’s dialysis facility now submits claims for any bundled care, even if it is performed by an outside physician. In turn, the facility reimburses the provider for any bundled care

provided to the ESRD patient. Therefore, when looking at temporal trends in payment data, for example, researchers will see a decline in payments after January 2011. In addition, when looking at trends that cross the bundle implementation date, it is necessary to restrict payment analyses to care that falls outside of the bundle. Researchers may use the list of bundle-covered medication and laboratory procedures at http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ESRDpayment/Consolidated_Billing.html.

Section 6: Transplant Process and Outcomes

Transplant patients constitute a unique subset of the ESRD population and are often studied separately from dialysis patients. Researchers may wish to simply count the number of transplant events that meet certain criteria, or calculate transplant event rates and survival probabilities. Using the USRDS transplant data, researchers can obtain information on transplant events, such as donor and recipient characteristics, and on patient- and graft-related outcomes.

Data Sources

Basic transplant variables are contained in the Transplant file in the Core SAF. Transplant event data are combined from CMS and OPTN databases. Before 1988, CMS was the primary source for all transplant event data. Between 1988 and 1993, both CMS and OPTN collected information regarding transplant events. Since 1994, OPTN has been the primary source for all transplant event data (see Figure 3).

Figure 3. Source of Transplant Events in the USRDS Population

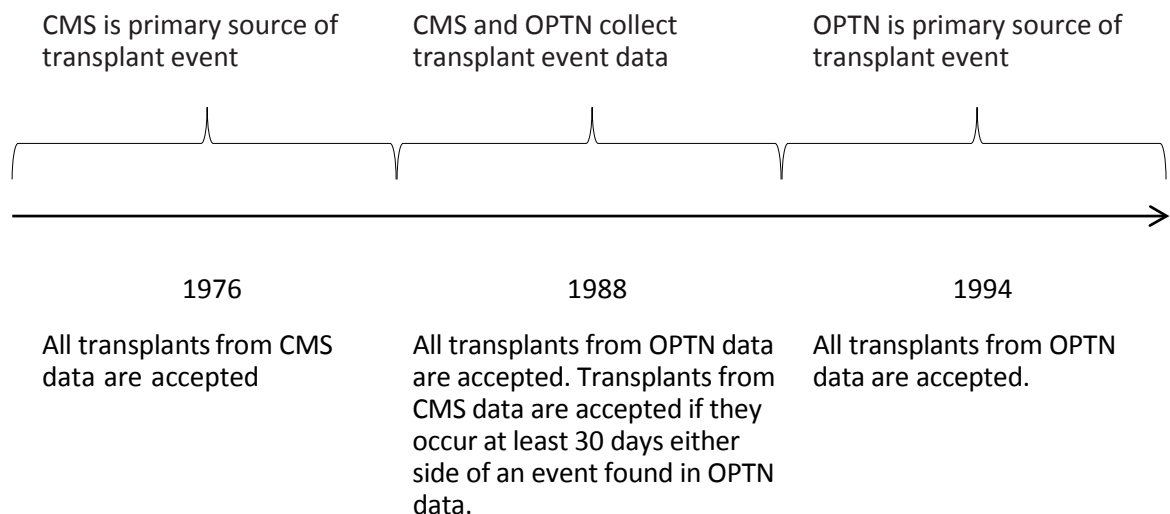


Table 8. USRDS Transplant Files				
File	Data Source	Comments	Years	Dataset
TX	Constructed by USRDS CC from OPTN, CMS PMMIS/REBUS, Transplant, Medical Evidence files	Identifying data on all known transplants	All 3/9/1963-9/3/2014	Core
WAITLIST_KI	OPTN Transplant Waiting List file	Contains data of patients on kidney waiting list per listing event	Listing date 3/1/1970	Core
WAITSEQ_KI	OPTN Transplant Waiting List file	Contains condensed, center-specific, kidney waiting list dates	Listing date 3/1/1970	Core
WAITLIST_KP	OPTN Transplant Waiting List file	Contains data of patients on kidney-pancreas waiting list per listing event	Listing date 9/2/1986	Core
WAITSEQ_KP	OPTN Transplant Waiting List file	Contains condensed, center-specific, kidney-pancreas waiting list dates	Listing date 9/2/1986	Core
TXHCFA	CMS PMMIS/REBUS Transplant file	Details from CMS	1976-93	Transplant
TXUNOS_KI_PRE_JUL04	OPTN candidates, recipients, donors, histocompatibility files	Contains kidney recipients transplant details	Transplant date 10/1/1987-6/29/2004	Transplant
TXUNOS_KI_POST_JUL04	OPTN candidates, recipients, donors, histocompatibility files	Contains kidney recipients transplant details	Transplant date 6/30/2004+	Transplant
TXUNOS_KP	OPTN candidates, recipients, donors, histocompatibility files	Contains kidney-pancreas recipients transplant details	Transplant date 10/15/1987+	Transplant
TXFUHCFA	CMS PMMIS/REBUS transplant follow-up file	Follow-up details, CMS	1988+	Transplant
TXFUUNOS_KI	OPTN kidney recipients follow-up file	Contains kidney recipients transplant follow-up details	Transplant date 1/1/1988+	Transplant
TXFUUNOS_KP	OPTN kidney-pancreas recipients follow-up file	Contains kidney-pancreas recipients transplant follow-up details	Transplant date 1/5/1988+	Transplant
TXIRUNOS	OPTN immunosuppression at time of transplant	Immunosuppression details for kidney and kidney-pancreas recipients at the time of transplant	Transplant date 1/1/1988+	Transplant
TXIFUNOS	OPTN immunosuppression at follow-up	Immunosuppression details for kidney and kidney-pancreas recipients at follow-up	Transplant date 1/1/1988+	Transplant

Reconciliation of CMS and OPTN Events

The USRDS has implemented a decision algorithm, described in detail in “Section 2: ESRD Patients”, which reconciles identical transplant events when multiple sources contain conflicting information. Each transplant event found in the Transplant file (TX) of the Core SAF dataset is thus a unique event that the USRDS believes occurred after reviewing both OPTN and CMS REMIS Transplant and Medical Evidence record files. Table 8 details the various USRDS SAFs related to transplant. The Transplant file (TX) in the Core SAF contains transplant dates, basic descriptive data, and causes of graft failure, when known.

Kidney and Pancreas Waiting list

The Core SAF also contains four files with information on the kidney and simultaneous kidney-pancreas waiting list files from OPTN.

Most investigators are simply interested in the dates of listing at specific transplant centers. The WAITSEQ_KI and WAITSEQ_KP files contain entry and removal date sequences per patient per transplant center for the kidney alone and simultaneous kidney-pancreas waiting lists, respectively. These records are collapsed from the raw OPTN waiting list files WAITLIST_KI and WAITLIST_KP. Both the raw OPTN waiting list files and the sequence files contain two patient identifiers,

PID and USRDS_ID. ESRD patients included in the USRDS patient profile have a USRDS_ID.

Because the sequence files are collapsed versions of the waiting list files, investigators should use caution when linking back to the raw waiting list files to obtain information such as Panel Reactive Antibody (PRA) values. Investigators will need to obtain all raw waiting list records and then make decisions as to which record contains the information most relevant to their needs. Some specific assumptions the CC used to collapse waiting list records are as follows:

- Candidates listed in error were removed (REMCODE=10).
- Patients with missing listing dates (EDATE=.) were removed.
- Patients with listing and removal dates on the same date were removed.
- Transplant dates were taken from the USRDS Transplant file and not from the OPTN Waiting List file.
- The transplant dates on the USRDS Transplant file are cleaned and reconciled with other sources of data, so these dates were used to ensure consistency. In the event that a patient not known to the USRDS received a transplant, the OPTN transplant date found in the raw waiting list file was used.
- Known transplant dates are used to truncate waiting periods at all centers at which a patient is currently listed. For example, if a patient underwent transplant but was not removed from a center's waiting list, a removal date would be imputed on the transplant date. This is true for all centers at which a patient is listed.
- Patients with inactive waiting periods were removed; the waiting list sequence lists only active periods.
- Overlapping waiting periods at the same center were collapsed.

Transplant SAF

To obtain additional data regarding transplant events, users need the Transplant SAF, which contains detailed information, in separate files, from CMS and OPTN (UNOS). The Transplant file in the Core SAF contains reconciled transplant event data from all available sources. As some researchers may find it beneficial to see the transplant data obtained from each source, data from each source are included.

Two transplant files contain data collected by CMS and OPTN (UNOS) on transplant follow-up visits. The follow-up records in TXFUHCFA and TXFUUNOS overlap in time, specifically 1988 to 1993, and contain information collected during patient follow-up visits, which typically occur at 6 months, 1 year, and yearly thereafter. The TXIRUNOS and TXIFUNOS files contain information on immunosuppression treatment. Data in these files are from the Kidney Transplant Recipient Registration Worksheet and Kidney Transplant Recipient Follow-Up Worksheet. TXIRUNOS contains data on treatments at the time of transplant, and TXIFUNOS contains data on treatments updated at each follow-up visit, if available. The TXUNOS_KI_PRE_JULO4 and TXUNOS_KI_POST_JULO4 files now include only kidney (KI) data. In prior SAFs, the TXUNOS_KI_PRE_JULO4 file included both kidney and kidney-pancreas (KP) data. TXUNOS_KP includes all years' kidney-pancreas data.

Each file on the Transplant dataset should be considered a separate analytical file. Attempting to combine information from various files involves reconciling information across files.

Starting in 2003, OPTN began to use the bitmask technique to maintain multiple values (i.e. selections/choices) within one variable. This method applies to all multiple selection questions. Previously, OPTN used as many variables as needed to collect all possible answers for such questions. For example, for the question "Treatment: Other therapies," OPTN used three variables, PHOTOPH, PLASMA and LYMPHOID with answers "Yes/No." After 2003, OPTN combined these variables into a single variable, THERAPIES, for the same question, using the bitmask technique. The basic code values are as follows: 1 = "Photopheresis"; 2 = "Plasmapheresis"; 4 = "Total Lymphoid Irradiation (TLI)", while 3 represents the sum of values 1 and 2; and 5 indicates that both 1 and 4 are selected. To assist researchers in decrypting these multiple-selection/value variables, USRDS has identified all possible additive combinations of these bitmask-value variables and made them available as SAS formats.

Appendix D: Data Collection Forms lists the CMS and OPTN data collection forms, which are available at www.usrds.org. As a cross-reference between the data file and the form, the SAS variable labels (shown in Appendix B: Data File Descriptions) indicate, whenever possible, the question number from the form. The label PM7694, for example, indicates that the variable can be

found on the PMMIS from 1976 to 1994. If the question number on the form did not change over time, this number is also indicated in the label. PM819421b indicates question 21b on the PMMIS form collected between 1981 and 1994. Table 9 provides details regarding these labels.

File	Variable Label Prefix	Definition
TXHCFA	PM7681	From CMS form, 1976-1981
	PM8194	From CMS form, 1981-1994
	PM7694	Appears on all CMS forms, 1976-1994
	PM819421b	CMS form, 1981-1994, question 21b
TXUNOS*	TCR	Transplant candidate registration
	CDR	Deceased donor registration
	LDR	Living donor registration
	DHS	Donor histocompatibility
	RHS	Recipient histocompatibility
	KIR	Kidney transplant recipient registration
	KPR	Kidney-pancreas transplant recipient registration
TXFUHCFA	TFU	From CMS form before 1994
TXFUUOS**	KIF	Kidney transplant recipient follow-up
	KPF	Kidney-pancreas transplant recipient follow-up
* Including TXUNOS_KI_PRE_JUL04; TXUNOS_KI_POST_JULY04, and TXUNOS_KP.		
** Including TXFUUNOS_KI and TXFUIUNOS_KP.		

Most descriptive data on transplant events can be found in the Core SAF Transplant file, which can be used to construct counts of various transplant-related events. Transplant rates can be constructed by combining transplant data with a patient treatment history file, RXHIST or RXHIST60; rates in the ADR are typically calculated using RXHIST60, the treatment history file with the 60-day rule built in. Transplant survival, both graft and patient, can be calculated using the transplant failure date found in the Transplant file and the date of death found in the PATIENT file.

Section 7: Morbidity and Hospitalization

Morbidity associated with ESRD can be determined from information on hospitalizations and acute events, documented in Medicare claims files through

ICD-9-CM (International Classification of Diseases, Ninth Revision, Clinical Modification) and CPT (Current Procedural Terminology) codes. The USRDS provides data on inpatient admissions, discharges, and diagnosis and procedure codes, which may be used to produce different types of hospitalization rates, such as total admission rates and hospital day rates.

Data Sources

Inpatient hospitalization data, a subset of the data in the Institutional Claims file, are included in the HOSP SAFs. These data originate from the Part A institutional inpatient claims data and are supplemented by the REMIS inpatient data, as hospitalizations may appear in both sources or in one but not the other. The patient treatment history file (RXHIST or RXHIST60) provides the patient modality treatment history, which may be used to construct hospitalization rate data, while demographic data are obtained from the PATIENTS file.

Contents of the Hospital Dataset

A complete list of variables in the HOSP dataset can be found in Appendix B: Data File Descriptions. The following list outlines some of the key variables:

- HCFASAF indicates the data source of the claim.
- CLM_FROM provides the from date of service, indicating the admission date.
- CLM_THRU provides the through date of service, indicating the date of discharge.
- HSDIAG₁-HSDIAG₁₀ provide up to 10 ICD-9-CM diagnosis codes.
- HSSURG₁-HSSURG₁₀ provide up to 10 ICD-9-CM procedure codes.
- DRG_CD provides diagnostic-related group (DRG) codes, which categorize inpatient stays by ICD-9-CM diagnosis and procedure codes.
- PRIMDIAG: equals “yes” when HSDIAG₁ provides the principal diagnosis code.

Preparing Inpatient Claims Data for Use

Preparing data for analyses requires that they be cleaned, with overlapping hospitalizations for the same patient combined. The principal diagnosis and procedure codes from the first of two overlapping hospitalizations should be retained, with the combined hospitalization extending from the first admission date to the last discharge date.

To create a dataset to calculate rates for period prevalent patient cohorts, attach data for hospitalizations occurring during the selected year by patient USRDS_ID to a period prevalent cohort file for the year. In this way, all patients in the file, including those with multiple hospitalizations and those with none, are included in the rate calculation. The 90-day rule defines each patient start date as day 91 of ESRD. Use of this rule helps to avoid incomplete hospitalization data from center hemodialysis patients aged younger than 65 years and not disabled, who cannot bill Medicare until 90 days after first ESRD service date. To calculate rates by patient characteristics, demographic data may be attached to the dataset by patient USRDS_ID from the PATIENTS file.

Because hospitalization data are incomplete for non-Medicare patients and for patients classified as Medicare secondary payer (MSP), these patients should generally be excluded, thereby limiting their contribution to patient-years at risk. The payer history file in the Core SAF can be used to do this. Another method of MSP patient exclusion involves excluding dialysis patients who do not reach a certain level of Medicare paid dialysis bills. For instance, dialysis patient start dates (January 1 of the year for prevalent patients and day 91 of ESRD for incident patients) can be required to fall between start and end dates based on Medicare paid dialysis claims as follows:

- Claims start date: the first day of the first month in which there is at least \$675 of Medicare paid dialysis claims.
- Claims end date: the end of a three-month period in which there is less than \$675 of paid claims in each month.

If a patient's start date does not fall between the claims start and end dates, that patient is excluded from the analysis for that year.

Section 8: Survival and Mortality

Survival and mortality analyses are often used to compare outcomes among treatment modalities, age groups, or races, or to evaluate outcome trends over time. Primary analyses include unadjusted survival probabilities using the Kaplan-Meier method, adjusted survival probabilities using the Cox regression model, unadjusted death rates, and adjusted death rates.

Data Sources

Survival and mortality analyses often require patient

demographic information such as age, sex, race, primary cause of ESRD, death date, and cause of death, and treatment modality history data, which can be obtained from the USRDS Core SAF files PATIENTS and RXHIST or RXHIST6o.

Section 9: Providers

The Facility SAF file is constructed from data supplied by the CMS Annual Facility Survey, the CDC National Surveillance of Dialysis-Associated Diseases, the CMS Dialysis Facility Compare and the CMS CROWNWeb. Data are at the facility level and there are no patient level data. Over time, facilities may be purchased and sold. Thus, a facility may have the same provider number for several years, then be purchased by a different owner and receive a new provider number. The physical facility and location may be the same, and most staff may remain, but the provider number has changed.

The Facility SAF file contains over 100 provider characteristics, including:

- Transplant counts
- Self-dialysis information
- HD, IPD, CAPD, and CCPD patient counts
- Number of in-center new dialysis patients at a facility
- Profit status, and hospital-based or freestanding status
- Chain affiliation
- Reuse practices

Data Sources

The CMS Annual Facility Survey is obtained as part of the CROWN data that the USRDS receives. The CDC National Surveillance of Dialysis-Associated Diseases survey was obtained from the CDC annually through 2002. As the last CDC National Surveillance survey occurred in 2002, all data for CDC variables are missing after 2002. A list of the CMS and CDC surveys can be found in Appendix D: Data Collection Forms. Those forms are also available at www.usrds.org. The CMS Dialysis Facility Compare data are sent quarterly from CMS.

Provider Numbers Assigned by the USRDS

Each provider has been assigned an identification number by the CMS. However, in an effort to ensure confidentiality and conceal provider identity, the USRDS assigned an anonymous number to each provider, which are not the original provider numbers

assigned by the CMS. Each provider maintains its USRDS-assigned facility ID throughout the SAFs to allow linkage.

Researchers who wish to conduct patient-level studies incorporating the Facility SAF data can link patient information from the Detailed Treatment History SAF (RXHIST), the Condensed Treatment History SAF (RXHIST6o), or other SAFs, to the Facility SAF via the USRDS provider number and year.

Profit Status

The profit variable, NU_P_NP, has the following values, “For-profit”, “Non-profit”, and “Unknown” (case sensitive).

The For-profit group consists of facilities categorized on the CMS survey as individual-profit, partnership-profit, corporation-profit, or other-profit. Non-profit facilities include individual-non-profit, partnership-nonprofit, corporate-nonprofit, other-nonprofit, state-government non-federal, county-government non-federal, city government non-federal, city/county-government non-federal, hospital district/authority government non-federal, other-government non-federal, Veterans Administration Government Federal, Public Health Service Government Federal, military-government federal, or other-government federal. Those 18 categories are designated by the variable TYPOWNER.

However, the 2002 CMS survey dropped this variable and it is missing in subsequent survey periods. Profit status is now determined solely from the CMS Dialysis Facility Compare data.

Chain Affiliation

Originally, the USRDS definition of a chain was 20 or more facilities in 2 or more states owned by the same corporation. The number of chains can vary from year to year depending on changes of facility ownership and the opening of new facilities. Because of the many changes occurring with regard to facility ownership, the original chain definition was modified. Starting with the 2005 facility survey, chains were further classified into Large Dialysis Organizations (LDOs) and Small Dialysis Organizations (SDOs). SDOs are defined as those dialysis organizations that operate at least 20 but not more than 199 facilities. LDOs are those organizations with 200 or more facilities. The chain variable, CHAIN_ID, contains a text string identifying the chain ownership of the facility. A blank

value for CHAIN_ID indicates that the facility has no chain affiliation. Facilities with no chain affiliation are divided into hospital-based facilities and independent facilities.

Freestanding vs. Hospital-based Facility

A hospital-based dialysis facility is associated with a hospital; it can be located in the hospital or at a remote location. Freestanding dialysis facilities operate independently of a hospital.

A study analyzing freestanding vs. hospital-based units will generally include three groups: freestanding, hospital-based, and unknown. The variable NU_HBFS designates this status. A code of 1 represents hospital-based unit, and 2 represents free-standing. The unknown category will consist of any provider in the researcher's study not found in the Facility database.

Section 10: CKD Cohort (5% CKD Cohort Dataset)

Traditionally, the USRDS has focused on the final stages of kidney disease, due to the comprehensive data available from Medicare on all patients with ESRD. However, there is no similar national reporting system for patients with kidney disease prior to end stage. For several years, the USRDS has been reporting on chronic kidney disease in Volume 1 of the USRDS ADR and making standard analytic files available using a cohort of patients diagnosed with CKD from the Medicare 5 percent database.

Data Sources

To create the CKD cohort dataset by year, the CC uses 5% General Medicare Claims data (1992–2012) and patient demographic information. The CKD cohort dataset contains three basic data components: Patient Master File, Payer Sequence File, and a series of Co-Morbid Files. To illustrate the methodology used to ascertain patients with the CKD and the underlying data structure and file organization, below is a description of how the CKD cohort dataset was created using the 5% General Medicare Claims data.

Patient Master File

For patients in this file, at least one CKD ICD-9 diagnosis code (Table 9.1) was identified in the 5% IP, OP, HH, HS, SNF, and PB SAFs, with one record per patient. The two key variables CKD_xx (1-year entry period CKD indicator) and CKD_xxyy (2-year entry period CKD indicator) are defined as follows:

CKD_xx = Y if a patient in year xx (e.g., CKD_95 or CKD_02):

- Was Part A and Part B entitled and not enrolled in an HMO for the entire year (i.e., Payer_Seq_File)
- Had any one CKD ICD-9 diagnosis code from IP or HH or SNF, or any two CKD ICD-9 diagnosis code combinations from PB or OP with different claim dates

CKD_xx = N otherwise

CKD_xxyy = Y if a patient in a two-year entry period (e.g., CKD_9596, CKD_9900, or CKD_0102):

- Was Part A and Part B entitled and not enrolled in an HMO for the entire two-year entry period (i.e., Payer_Seq_File)
- Was alive and did not develop ESRD as of December 31 of the two-year entry period
- Had any one CKD ICD-9 diagnosis code from IP or HH or SNF, or any two CKD ICD-9 diagnosis code combinations from PB or OP with different claim dates identified at any time during the two-year entry period (either in one of the two years or across the two years)

CKD_xxyy = N otherwise

Patients with CKD_95 = Y are a subset of patients in the 1995 Co-Morbid file (Co_Morbid_95) in the Co-Morbid Files dataset; patients with CKD_9900 = Y are a subset of the patients from the 1999 and 2000 Co-Morbid files (Co_Morbid_99 and Co_Morbid_00) in the Co-Morbid Files dataset respectively.

Some patients in the Patient Master File may have all CKD_xx = N and all CKD_xxyy = N. This implies that these patients had at least one CKD ICD-9 diagnosis code identified in the 1992 to 2010 5% IP, OP, HH, HS, SNF, and PB SAFs but did not meet the one-year entry period and two-year entry period CKD eligibility criteria described above.

Each patient is identified by two patient identification numbers. One is FIVEP_ID, a unique patient ID in the 5% Medicare sample population and the primary linking key within the USRDS CKD Cohort Finder dataset. Each year, new FIVEP_ID numbers are generated only for new patients added to the 5% sample. All existing patients (repeated patients) are referenced with their previously assigned numbers. A second patient identification number, USRDS_ID, is the unique ID in the USRDS database system for identifying ESRD patients. Only patients with ESRD are assigned a valid USRDS_ID number. FIVEP_ID and USRDS_ID are not related.

Payer Sequence File

This file was created from the 5% Medicare denominator files and the IP, OP, HH, HS, SNF and PB SAFs. It contains Medicare coverage information for 5% patients with at least 1 CKD ICD-9 diagnosis code identified through the 5% IP, OP, HH, HS, SNF, and PB SAFs. A patient might have one or many sequence records in which each record indicates different insurance coverage within a well-defined time period. There are five unique insurance types:

- A: Entitled with Part A only
- B: Entitled with Part B only
- AB: Entitled with Part A and Part B
- H: Enrolled in a Medicare HMO health plan
- N: Non-Medicare

The first record contains the earliest available date for coverage information in the 5% Medicare denominator files. The last record contains the last known available date (i.e., the most recent) for coverage information in the 5% Medicare denominator files. FIVEP_ID is the primary patient identification number.

Co-Morbid Files

The Co-Morbid files are constructed from the 5% Medicare IP, OP, HH, HS, SNF, and PB SAFs and named individually by the respective calendar years. Each file contains all patients whose CKD diagnosis was identified from the 5% Medicare IP, OP, HH, HS, SNF, and PB SAFs within that year (i.e., Co_Morbid_95 contains all patients whose CKD disease events were identified from the 1995 5% Medicare Claims SAFs). However, patients were not necessarily required to be entitled to Parts A and B and not enrolled in an HMO for that entire year. A patient might have one or many records within a calendar year in which each record depicts one unique CKD disease event. Records are considered duplicates if they are extracted from the same type of claim file with the same ICD-9 diagnosis code on the same date. These files do not include duplicates.

For example, each record in Co-Morbid_95 consists of one CKD code per claim file per service date per unique patient in 1995. The combination of claim type, CKD code, and service date for a patient establishes uniqueness in each Co-Morbid file. FIVEP_ID is the primary patient identification number.

Description	ICD-9 Diagnosis Codes
CKD	016.0, 095.4 189.0, 189.9 223.0, 236.91, 250.4, 271.4, 274.1, 283.11 403.x0, 403.x1, 404.x0, 404.x1, 404.x2, 404.x3, 440.1, 442.1, 447.3, 572.4, 580-588, 591 642.1, 646.2 753.12-753.17, 753.19, 753.2, 794.4

Section 11: Comprehensive Dialysis Study SAF

The Comprehensive Dialysis Study (CDS) collected data on patient demographics, contact information, treatment, laboratory values, quality of life (QOL) survey interviews, and nutrition survey interviews for U.S. dialysis patients who started treatment during 2005–2007 at 335 randomly selected dialysis facilities. Patients were selected via a monthly REMIS/SIMS database custom extract process using the following criteria:

- Incident dialysis patients (age \geq 18) who had survived for at least two months
- No prior transplant
- Dialyzed at one of the pre-determined 335 facilities

Over a period of two years, 11,292 patients were selected. Of these, 1,678 consented and participated in the surveys and/or lab assays. The CDS SAF contains the data for these patients. Of the 1,678 patients, 1,279 participated only in the QOL survey; 364 participated in the QOL and nutrition surveys and/or lab assays (227 with QOL, lab, and nutrition; 4 with QOL and lab; and 133 with QOL and nutrition); 35 did not participate in either survey but provided lab data.

The CDS SAF dataset consists of:

- CDS Patient file. Contains 1 record for each of the 1,678 patients. A patient record includes a unique identifier USRDS_ID (unique identifier used to cross reference to other USRDS SAF files), demographic data, and a patient category indicator (PAT_CAT_BL), which indicates baseline survey and lab participation status (QOL-only, QOL w/lab data, QOL and nutrition w/o lab data, Lab data w/o survey).
- CDS QOL Baseline file. This file contains the Baseline QOL survey answer data and some derivative score data from QOL only and QOL/Nutrition participants.

- CDS Food Baseline file. This file contains the Baseline Nutrition survey answer data and data generated by “Block Dietary Data System” on QOL/Nutrition participants.
- CDS Lab file. This file contains up to five sets of lab data from a subset of QOL/Nutrition participants.
- CDS Dictionary. This file contains each CDS SAF file contents, and variable formats.

Section 12: Medicare Prescription Drug

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, while the benefit was very new, may not be complete, and should be interpreted with caution.

Data Sources

The ESRD and pre-ESRD Part D data are obtained from ESRD and pre-ESRD CMS annually with USRDS provided Finder Files. 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.

Part D Enrollment SAF

Since the Part D benefit is voluntary, not all Medicare beneficiaries are enrolled in Part D. The annual enrollment file contains 12 monthly indicators that detail whether the beneficiary is enrolled in Part D, and if so, what type of plan the beneficiary is enrolled in. There are also monthly indicators for dual eligibility (Medicare and Medicaid), monthly indicators for Retiree Drug Subsidy, and low income subsidy (LIS). A complete listing of the variables included in the annual enrollment file is available in Appendix B: Data File Descriptions.

The variables PTD_CNTRCTID_01- PTD_CNTRCTID_12 indicate Part D enrollment as follows:

- H = Managed care organizations other than Regional Preferred Provider Organization

- R = Regional Preferred Provider Organization
- S = Stand-alone Prescription Drug Plan (PDP)
- E = Employer sponsored
- O = Not Medicare enrolled
- X = Medicare enrolled, but no Part D enrollment record
- N = Not Part D enrolled

Values of H, R, S, and E for these variables are generally considered to indicate Part D enrollment.

Low income subsidy status can be determined from the variables CST_SHR_GRP_CD_01- CST_SHR_GRP_CD_12. If beneficiary is Medicare enrolled and Part D enrolled:

- 01 = 100% premium subsidy and no copayment
- 02 = 100% premium subsidy and low copayment
- 03 = 100% premium subsidy and high copayment
- 04 = LIS, 100% premium-subsidy and high copayment
- 05 = LIS, 100% premium-subsidy and 15% copayment
- 06 = LIS, 75% premium-subsidy and 15% copayment
- 07 = LIS, 50% premium-subsidy and 15% copayment
- 08 = LIS, 25% premium-subsidy and 15% copayment
- 09 = No premium subsidy or cost sharing

If beneficiary is Medicare enrolled and not Part D enrolled:

- 10 = Not enrolled in Part D, but employer is entitled for RDS subsidy
- 11 = Creditable coverage but no RDS
- 12 = Not Part D enrolled, no creditable coverage, no RDS
- 13 = None of the above conditions have been met

Prescription Drug Event SAF

This SAF contains details on prescription drug utilization, including brand name, generic name, dosage form, drug strength, quantity dispensed, date of service, and total prescription cost. The variables in this SAF are listed in Appendix B: Data File Descriptions. More detailed information on the Part D benefit can be found at: <http://www.cms.gov/PrescriptionDrugCovGenIn/>.