

2013 Researcher's Guide

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 Files (SAFs), the primary means by which USRDS data are available for use. The Researcher's Guide includes the information needed to select and use the appropriate SAFs for the intended project.

Acknowledgements

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.

Allan J. Collins, MD, FACP

Director, USRDS Coordinating Center; University of Minnesota, Hennepin County Medical Center

Robert N. Foley, MB, MSc

Deputy Director, USRDS Coordinating Center; University of Minnesota, Hennepin County Medical Center

Lawrence Y. C. Agodoa, MD

Co-Project Officer, USRDS; Director, End-stage Renal Disease Program, DKUHD

Paul W. Eggers, PhD

Co-project Officer, USRDS; Program Director, Kidney and Urology Epidemiology, NIDDK

Cheryl Arko, BA
Senior Systems Analyst;
Programmer

Beth Forrest, BBA
Data Systems Coordinator

Jiannong Liu, PhD
Senior Biostatistician

Delaney Berrini, BS
Publications Specialist

Susan Everson, PhD
Senior Editor; Graphic Designer

Tricia Roberts, MS
Biostatistician

Nan Booth, MSW, MPH
Medical Editor

Eric Frazier, BS
Information Systems; Programmer

Craig Solid, PhD
Senior Epidemiologist;
Biostatistician

Shu-Cheng Chen, MS, MPH
Director of Information Systems

David Gilbertson, PhD
Director of Analytical Services

Jon Snyder, PhD
Senior Epidemiologist

Edward Constantini, MA
Senior Editor

Suying Li, PhD
Biostatistician; Economist

Xinyue Wang, BA/BS
Systems Analyst; Programmer

James Ebben, BS
Manager, Data Systems

Shuling Li, PhD
Biostatistician

Suggested Citation

The suggested citation for this guide is: U.S. Renal Data System Coordinating Center, USRDS 2013 Researcher's Guide to the USRDS Database, 2013.

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

Introduction 5

Getting Started: Examples of datasets and steps to help start constructing analyses 7

Methods & Databases 17

- 17 Section 1 · ESRD Data Sources and the USRDS Database System: General descriptions of data available to the USRDS and the processes involved in their creation
- 22 Section 2 · ESRD Patients: Methods used to define ESRD patients
- 24 Section 3 · Treatment History: Methods used to create a treatment history for each patient
- 28 Section 4 · Payer History: Methods for analyzing insurance payer information
- 30 Section 5 · Transplant Process and Outcomes: Methods for analyzing transplant patient data
- 34 Section 6 · Morbidity and Hospitalization: Methods for analyzing morbidity and hospitalization data
- 36 Section 7 · Survival and Mortality: Methods for analyzing survival and mortality data
- 37 Section 8 · Providers: Description of the Facility database, which contains data on US dialysis facilities
- 39 Section 9 · CKD Cohort: Description of the CKD cohort finder files by year, using the 5% General Medicare Claims data and ESRD patient demographic information
- 41 Section 10 · Comprehensive Dialysis Study:
- 41 Section 11 · Medicare Prescription Drug:

Appendices 47

- 47 A: USRDS Products and Services
- 61 B: Medicare Claims
- 71 C: Statistical Methods used by the USRDS
- 73 D: Data File Descriptions, with data dictionaries
- 253 E: Data Formatting
- 357 F: Data Forms: Lists of data forms available at www.USRDS.org: CMS, UNOS, and CDC ESRD data forms; Special Study ESRD data forms; DMMS Waves I, II, III, IV, and FACS study forms.
- 361 G: Glossary

Introduction

History of the USRDS

The first USRDS contract was awarded in 1988 to the Urban Institute in Washington, DC, which worked in conjunction with investigators at the University of Michigan at Ann Arbor. Under a new contract created in 1999, the USRDS was divided into a Coordinating Center (CC) and four Special Studies Centers (SSCs), each of which continues to operate under the direction of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) at the National Institutes of Health (NIH).

The CC and Cardiovascular SSC contracts were awarded in 1999 to the Minneapolis Medical Research Foundation, and are directed by Allan Collins, MD and Charles Herzog, MD, respectively. The Economic SSC was awarded to the University of Iowa under the direction of Lawrence Hunsicker, MD. In 2000, the Nutrition SSC was awarded to the University of California at San Francisco, directed by Glenn Chertow, MD, and the Rehabilitation SSC to Emory University, directed by Nancy Kutner, PhD.

In early 2007, the CC and the SSCs entered a new seven-year contract period. The Cardiovascular, Nutrition, and Quality of Life/Rehabilitation SSCs remain as before, and the activities of the Economic SSC are included with those of the CC.

USRDS website: www.usrds.org

The USRDS website provides users with access to PDF files of the printed Annual Data Report (ADR), Excel files of the Reference Tables and the data underlying the graphs and state maps, and PowerPoint slides of USRDS presentations and ADR figures. Because of the size and complexity of the ADR files, downloading time for large portions of the book may be substantial.

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 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. For more information, visit http://USRDS.org/render/xrender_home.asp to access the RENDER tutorial.

What's New in 2013

Data File Descriptions

On page 77 in the Core CD 1 file under TX: Transplant, RSEX has a new comment. It was changed from *Recipient Gender* to *Recipient Gender (from profile)*.

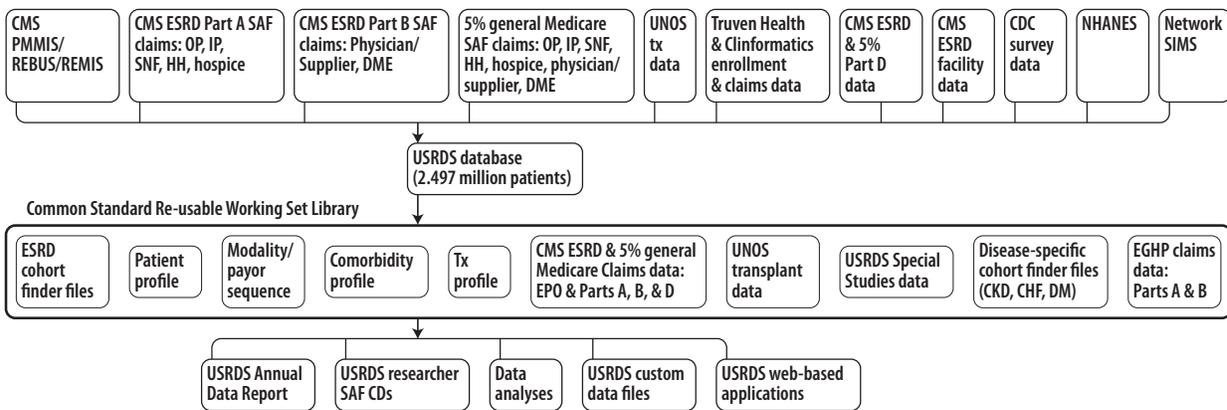
There is a new file in DMMS (pages 220-222). HOSP1 and HOSP2 replace former years' HOSP file. HOSP1 has the same variables as former years' HOSP file. HOSP2 has the same variables as former years' HOSP3 file in the Hospital CD.

The file Medicare 5% Sample CKD Patient Hospital Stay Summary File CKD_92_to_YR_hosp_clm in Medicare 5% Sample CKD Patient Hospital Stay Summary Dataset has been split into two files: CKD_92_to_09_hosp_clm and CKD_post09_hosp_clm (page 249). The difference between the two files is CKD_92_to_09_hosp_clm contains up to 10 ICD9 diagnosis codes and up to 10 ICD9 procedure codes. CKD_post09_hosp_clm incorporates CMS institutional claims file's new format containing up to 25 ICD9 diagnosis codes and up to 25 ICD9 procedure codes.

Data Formatting

Two new formats were added: MEDIATN (page 311) and RACEETH (page 329). In addition, two new entries were added to existing format REMCD (page 331).

1 Structure of the USRDS Database



2 Year-to-Year Comparison of Number of Records in USRDS Standard Analysis Files (SAF)

SAF	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
PATIENTS	1,516,251	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
RESIDENC	2,215,953	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
MEDEVID	786,297	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
RXHIST	10,630,350	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
FACILITY	55,547	60,138	64,870	69,764	74,831	80,071	85,572	91,332	97,201	103,210
TX	256,315	272,277	289,533	308,002	324,476	343,051	360,297	378,536	395,347	413,493
TXWAIT	313,446 [*]	347,564 [*]	378,262 [*]	405,165 [*]	437,256	474,348	513,028	546,239	579,225	618,511

^{*}TXWAIT is the total number of observations between Waitlist_ki and Waitlist_kp.

[†]MEDEVID is the total number of observations between Medevide95 and Medevide05.

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 SAF`.

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 requires a conversion to a format compatible with that system.

The examples use 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 `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 of 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 `Data Descriptions`.)

Note: In the following program examples, `SAS` keywords 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 requirement of the two `LIBNAMES`.)

```
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 thousands 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 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 strongly recommended.

The SAF Directory

Throughout this section SAF has been 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.

```
LIBNAME core_cd 'C:\USRDS\CORE_CD\SAF';  
LIBNAME hosp_cd 'C:\USRDS\HOSP_CD\SAF';  
LIBNAME in_cd 'C:\USRDS\IN\SAF';  
LIBNAME cpm_cd 'C:\USRDS\CPM\SAF';  
LIBNAME library 'C:\USRDS\CORE_CD\SAF';
```

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 esrddate inc_year rxgroup);  
    SET core_cd.rxhist60;  
        BY usrds_id begdate;  
  
    FORMAT esrddate MMDDYY10;  
  
    IF (FIRST.usrds_id) AND  
        (MDY(1,1,2000) <= begdate <= MDY(12,31,2010));  
  
    esrddate = begdate;  
    inc_year = YEAR(esrddate);  
  
RUN;  
  
PROC FREQ DATA=inc2000_2010;  
TABLE inc_year;  
TITLE 'Frequency Distribution of Incident Patients 2000 – 2010 by Incident Year';  
RUN;  
  
PROC FREQ DATA=inc2000_2010;  
TABLE inc_year * rxgroup;  
TITLE 'Frequency Distribution of Incident Patients 2000 - 2010 by Incident Year and Modali-
```

```
ty';  
RUN;
```

Example 2: Incident Patients Distribution by Demographic Data

```
PROC FREQ DATA=core_cd.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 DIALY-  
SIS PATIENTS WHO WERE ALIVE ON JANUARY 1, 2010.  
DATA pre_2010 (KEEP=usrds_id rxgroup);  
SET core_cd.rxhist60;  
BY usrds_id begdate;  
  
IF (begdate <= MDY(1,1,2010)) AND  
((enddate = .) OR (enddate >= MDY(1,1,2010))) AND  
(rxgroup NOT IN ('D' 'X' 'Z')) 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: 1995 Incident Patient Survival Rates (Kaplan Meier)

```
CALCULATE 5-YEAR SURVIVAL RATES OF 2005 INCIDENT DIALYSIS PATIENTS USING KAPLAN MEIER METHOD.  
DATA inc_2005_s;  
MERGE inc2000_2010 (IN = x1)  
core_cd.patients (IN = x2 KEEP=usrds_id died tx1date);  
BY usrds_id;  
  
IF x1 AND x2;  
  
* Extract only 2005 incident dialysis patients.;  
IF (inc_year = 2005) AND (rxgroup ^= 'T');  
  
* Calculate the survival time (in month) of each incident patient.;  
t = (MIN(died, tx1date, MDY(12,31,2010)) - esrddate + 1) / 30.4375;  
IF (t < 0) THEN t = 0;  
  
* Determine whether the patient is censored.;  
c = (MIN(died, tx1date, MDY(12,31,2010)) = died);  
RUN;  
  
PROC LIFETEST DATA=inc_2005_s METHOD=KM NOTABLE PLOTS=(s) OUTSURV=surv2005;  
TIME t*c(0);  
TITLE '2005 Incident Dialysis Patients 5-Year 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 core_cd.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)
          core_cd.patients (IN=x2 KEEP=usrds_id died tx1date);
    BY usrds_id;

    IF x1 AND x2;
RUN;

DATA inc2008_me;
    MERGE inc2000_2010 (IN=x1 WHERE=(inc_year = 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.

```
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 month) of each incident patient.;
    t = (MIN(died, MDY(12,31,2010)) - esrddate + 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 Dialysis Patients 2-Year Survival Rates by Co-Morbidity';

RUN;

Example 7: Waitlist Access Rate

DETERMINE WAITLIST ACCESS RATE TO DECEMBER 31, 2010, OF 2008 INCIDENT DIALYSIS PATIENTS WHO WERE REGISTERED IN THE TRANSPLANT WAITLIST.

```
DATA txwait;
    set core_cd.waitseq_ki core_cd.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 (rxgroup ^= 'T');

    * Make sure all dialysis patients who were not put on the waitlist;
    * get to be censored at the end of follow-up period.;
    IF x1 AND ^x2 THEN begin = MDY(1,1,2011);

    * Calculate the waitlist access time (in month) of each incident patient.;
    t = (MIN(begin, died, MDY(12,31,2010)) - esrddate + 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 Waitlist Access Rates';

RUN;
```

Example 8: Total Admission Rate

CREATE THE PATIENT DRIVER FOR HOSPITALIZATION RATES, OR USE PREVIOUSLY CREATED FILE.

```
DATA driver;
    SET core_cd.patients (WHERE=(incyear=2008));

    sfu = first_se + 91;
    efu = MIN(died, first_se+455);

    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, hosp_cd.hosp (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);
        n_adm=n_adm+1;
    END;

    IF LAST.usrds_id THEN          /* OUTPUT ONE RECORD PER USRDS_ID */
    DO;
        exptime=(efu-sfu-n_hos)/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 1000 patient-years by gender.

```

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';

RUN;

```

Example 9: 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
    * EXPOSURE TIME;
    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 1000 PATIENT-YEARS BY GENDER.

```

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;

```

Example 10: Create a patient cohort of 2008 Medicare Incident patients

CREATE THE 2008 INCIDENT PATIENT FILE.

```

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

```

FIND MEDICARE PAYER STATUS AT FIRST SERVICE DATE.

```

DATA first_payer;
    SET core_cd.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;

```


Methods & Databases

Section 1: ESRD Data and the USRDS Database System

The main objective of the USRDS CC is to use all relevant ESRD data to create an integrated and consistent database system for outcomes research. The CC 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 CMS, UNOS, the CDC, the ESRD Networks, and the USRDS Special Studies.

PMMIS/REBUS/REMIS DATABASE SYSTEM

The major source of ESRD patient information for the USRDS is the CMS Renal Beneficiary and Utilization System (REBUS), which was adopted in 1995 as the On-Line Transaction Processing (OLTP) system from its predecessor, the Program Management and Medical Information System (PMMIS) database. The PMMIS/REBUS database contains demographic, diagnosis, and treatment history information for all Medicare beneficiaries with ESRD. The database has been expanded to include non-Medicare patients, as discussed below.

Having advanced its database technology, CMS transformed the REBUS database into an Oracle relational database system, called the Renal Management Information System (REMIS), in the fall of 2003. This database included all patients who were alive and had ESRD as of January 1, 1995, or were incident after this date. This approach was adopted from the procedure used to create the Networks' Standard Information Management System (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.

CMS regularly updates the PMMIS/REBUS/REMIS database, using the Medicare Enrollment Database (EDB), Medicare inpatient and outpatient claims, the UNOS transplant database, the ESRD Medical Evidence Report (CMS-2728), the ESRD Death Notification (CMS-2746) obtained from renal providers, and the ESRD Networks' SIMS database. CMS has established data integrity rules to ensure accurate identification of patients in the SIMS and CMS databases. Each ESRD patient (new and old) is identified with a unique patient identification number common to both databases, guaranteeing that data for all patients are consistently managed over time.

STANDARD INFORMATION MANAGEMENT SYSTEM (SIMS) DATABASE

The USRDS CC continues to collaborate with CMS and the ESRD Networks to address data-tracking issues related to non-Medicare ESRD patients. 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 SIMS 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 the gathering of data on patients for whom no data were previously available on initial modality or death.

CMS MEDICAL EVIDENCE REPORT (FORM CMS-2728)

The CMS Medical Evidence Report is completed by the renal provider for each new ESRD patient, and is sent to CMS through the ESRD Networks. 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.

Before 1995, dialysis units and transplant centers were required to file the Medical Evidence Report only for Medicare-eligible patients. The form was revised in 1995, and providers were subsequently required to complete it for all new ESRD patients regardless of Medicare eligibility status. The 1995 revised form also included new fields for comorbid conditions, employment status, expanded race categories, ethnicity, and biochemical data at ESRD onset.

Only one Medical Evidence Report 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 Report must be filed to reestablish Medicare eligibility.

REVISED MEDICAL EVIDENCE REPORT (FORM CMS-2728)

Another revision of the Medical Evidence Report was introduced in May 2005. It 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, and 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 form are provided in the USRDS Core SAF dataset and are available for download in PDF format from the USRDS website: <http://www.USRDS.org/reference.htm>.

CMS ESRD DEATH NOTIFICATION (FORM CMS-2746)

Like the Medical Evidence Report, the Death Notification form is data rich, and CMS requires renal providers to complete it. Providers usually have 45 days to report ESRD death events to their respective ESRD Networks, including information about place, time, and cause of death. Data are thus available to the USRDS CC for research on cause-specific mortality outcomes.

CMS MEDICARE ENROLLMENT DATABASE (EDB)

The CMS Enrollment Database 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 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 PMMIS/REBUS/REMIS 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).

UNOS TRANSPLANT DATABASE

CMS began collecting data on all Medicare kidney transplants in the early 1980s. In 1988, UNOS was created to provide a national system for allocating donor organs and to maintain a centralized data depository for organ transplants. UNOS also began collecting data on all transplants. These two collection efforts were consolidated in 1994, and UNOS became the sole source of data on transplant donors and recipients.

The CMS and UNOS transplant data files overlap for 1988–1993, and some patients with Medical Evidence Reports indicating transplant as the initial modality are not included in either file. To resolve the 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 UNOS files are used.
- Between 1988 and 1993, all transplant events found in UNOS 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.
- Each transplant event found in the Transplant file of the USRDS SAF Core dataset is thus a unique event derived from the UNOS, CMS PMMIS/REBUS/REMIS Transplant, and Medical Evidence record files.

CMS ESRD STANDARD ANALYTICAL FILES (SAFS)

The CMS SAFS contain 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. Data-sets 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 2013 ADR includes all claims up to December 31, 2011. Patient-specific demographic and diagnosis information, however, includes data as recent as October 2012.

CMS 5% GENERAL MEDICARE SAFS

The 5% general Medicare SAFS have the same structure and data elements as the ESRD 100% SAFS, 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 CKD cohort dataset.

CLINICAL PERFORMANCE MEASURES PROJECT (CPM)

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 have been collected annually since 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 8500 in-center hemodialysis patients and 1500 peritoneal dialysis patients are 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 PD patients in the US have been sampled for the pediatric PD data collection. The USRDS CC, in collaboration with CMS, provides CPM/USRDS merged data to the general research community.

In anticipation of the national release of the CROWN web system and its supporting performance measures reports, CMS concluded its CPM project in 2009, making 2008 its final survey year.

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, which all Medicare-approved 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 1993-2002, the CDC used its 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. Also, 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 US dialysis patients, using dialysis records. Data were collected on 6000 ESRD patients in each of Waves I, III, and IV, and 4500 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

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 7096 patients from 523 dialysis units were included, the pre- and post-BUN values needed to calculate delivered dialysis dose were present for approximately 3300 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, gender, race, and primary diagnosis
- evaluate possible associations of these factors with reported causes of death.
- assess the distribution of comorbidity and other factors among patients using different treatment modalities.
- compare relative mortality rates by treatment modality, adjusting for selected comorbid conditions and other factors.

Data were collected for 5255 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 3067 patients at 548 dialysis units.

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 3385 patients from 706 units.

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, 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 wait list data from UNOS. All SAFS are 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 regularly with data obtained from the various data sources. The CC generally receives CMS SAF claims data, Facility Survey data, CDC Survey data, and UNOS transplant and wait-list data once a year, and REMIS and EDB data more frequently. 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.

Data Sources

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 dataset with files “PATIENTS,” “MEDEVID,” “RXHIST,” “RXHIST60,” and “PAYHIST.”

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 most patient survival outcomes analyses. This rule allows each new ESRD patient to generate 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.

The FSD is derived by taking the earliest of:

- the date of the start of dialysis for chronic renal failure, as reported on the Medical Evidence report,
- the date of a kidney transplant, as reported on a CMS or UNOS transplant form, a Medical Evidence report, or a hospital inpatient claim, or
- the date of the first Medicare dialysis claim.

Most FSDs are derived from the Medical Evidence Report. In the absence of this form, the date of the first Medicare dialysis claim or transplantation usually supplies the FSD. In the few cases in which the date of the earliest dialysis claim is earlier than the first dialysis date reported on the Medical Evidence Report, the earliest claim date is used as the FSD.

Identifying ESRD Patients

A person is identified as having ESRD when a physician certifies the disease on the Medical Evidence (ME) 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 ME forms have not been submitted. Patients who die soon after kidney failure without receiving dialysis are sometimes missed.

The ESRD First 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 the incident year for each new patient 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 most survival analyses.

The FSD is derived by taking the earliest of dialysis start date, as reported on the ME form, for chronic kidney failure; the date of kidney transplant, as reported on a CMS or OPTN transplant form, an ME form, or a hospital inpatient claim; or the date of the first Medicare dialysis claim. Most FSDs are obtained from the ME form. In the absence of this form, the FSD usually comes from date of the first Medicare dialysis claim or transplant. In the few cases in which the date of the earliest dialysis claim precedes the first dialysis date reported on the ME form, the earliest claim date is used as the FSD. However, starting with the 2007 ADR, FSDs for patients entering the ESRD program after December 31, 1994, are defined solely by the regular dialysis start date or the preemptive transplant date, whichever is earliest, on the ME form. This new method of determining FSD has been introduced in this ADR to align more closely to methods used by CMS. After years of careful monitoring and repeated comparative analyses of the traditional USRDS method to the new ME method, the USRDS believes it is appropriate to apply the ME method to incident patients entering the ESRD program on or after January 1, 1995.

Incidence and Prevalence

Incidence is defined as the number of people in a population who are newly diagnosed with a disease in a given time period, typically a year. Prevalence is the number of people in a population who have the disease at a given point in time (point prevalence) or during a given time period (period prevalence). The USRDS generally reports point prevalence, used primarily throughout the ADR, as of December 31, and period prevalence for a calendar year. Annual period prevalent data thus include people who have the disease at the end of the year and those who had the disease during the year and died before the year's end. Please refer to Getting Started for further details about defining incident and prevalent patient cohorts using the USRDS SAFS.

The USRDS treats successful transplant as a therapy, not as recovery from ESRD. Patients who undergo transplant at the time of ESRD initiation are counted as incident patients, and those with functioning grafts as prevalent.

Because data are available only for patients whose ESRD therapy is reported to CMS, patients who die of ESRD before receiving treatment or whose therapy is not reported to CMS are not included in the database. The terms incidence and prevalence are thus qualified as incidence and prevalence of reported ESRD. Some ESRD registries, such as the European Dialysis and Transplantation Association, use the term "acceptance

into ESRD therapy.” The USRDS, however, believes that “incidence of reported ESRD therapy” is more precise, because “acceptance” implies that remaining patients are rejected, when in fact they may simply not be identified as ESRD cases or may not be reported to CMS.

Point prevalence is useful in public health research because it measures the current burden of the disease on the health care delivery system. Period prevalence is appropriate for cost analysis because it indicates the total disease burden over the course of the year. The USRDS CC focuses primarily on ESRD incidence as the most useful measure for medical and epidemiological research that examines disease causality and its effect on various subpopulations.

Medicare and Non-Medicare (‘zz’) 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. However, some patients are not immediately eligible for Medicare primary payer coverage because of their employment status and pre-existing primary insurance 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. Some of these patients, particularly those who are new since 1995, have FSDs established by Medical Evidence Reports, but have no dialysis claims or hospitalization events in the CMS claims database to establish treatment history events. In the PMMIS/REBUS database, these patients are designated ZZ, or non-Medicare (the PMMIS/REBUS group assigns ZZ in the 2-character Beneficiary Identification Code field to identify all non-Medicare ESRD patients). CMS does not generally include these patients in the datasets released to researchers.

The USRDS recognizes that ZZ patients are ESRD patients and should therefore be included in patient counts for incidence, prevalence, and treatment modality. However, calculations of standardized mortality ratios (SMRS), standardized hospitalization ratios (SHRS), and standardized transplantation ratios (STRS) should not include these patients because of the small number of claims available in the first 30 to 33 months after first ESRD service. Furthermore, linking ZZ patients to their ESRD Death Notification (CMS 2746) or to the UNOS transplant data may be impossible; determining comorbid conditions or Part A and Part B services may also be impossible. Due to the limited availability of event data, event rates that include these patients must be assessed with caution.

Section 3: Treatment History

The USRDS CC uses billing information from the CMS ESRD database to create a longitudinal history of ESRD treatment for each patient in the database. This history defines incident and prevalent cohorts and determines censoring points and outcomes for observational studies.

Data Sources

The FSD is established by evaluating information from several sources. REMIS includes first ESRD service date from the Medical Evidence Report. The CMS dialysis billing information can also establish FSD if data are missing from other files. The UNOS transplant dataset provides FSD information for incident transplant patients. Data combined from these sources establish the first ESRD service date. Section 2 describes the decision algorithm used.

Treatment Modality Categories

Table 3.1 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. A new modality/event (recovered renal function; RRF) was introduced in the 2007 ADR. RRF can be established only if it occurs within the first 180 days of FSD, and the RRF period persists for at least 90 days. The RRF modality (i.e., event) is similar to the lost-to-follow-up event in that patients with an RRF event will not be included in the prevalent populations for outcomes analyses. However, as with lost-to-follow-up events, these patients remain in the modality sequence so subsequent renal failure episodes can be closely tracked in a timely manner.

Table 3.1. Treatment Modality Categories

Type	Detailed	General
1	Center HD	HD
2	Center Self HD	HD
3	Home HD	HD
4	HD Training	HD
5	CAPD	CAPD
6	CAPD Training	CAPD
7	CCPD	CCPD
8	CCPD Training	CCPD
9	Other PD	Unknown Dialysis
A	Uncertain Dialysis	Unknown 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 dataset record the sequence of modalities for each patient. The variables included in these files are presented in Table 3.2. 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 SIMS Event file, the UNOS Transplant events file, and Medicare claims files.

Table 3.2 Treatment History SAF Variables

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 categories death, loss to follow-up, and functioning graft, the end date may be missing from a patient's last record. Dialysis 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 2 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 1 year, the patient is designated 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 1 year, the patient is reclassified as lost to follow-up at the end of that 1-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.

Graft Failure

The CC assumes that a graft failure date reported in the UNOS transplant follow-up file or the REBUS 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 UNOS 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.

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 Medical Evidence Report, UNOS 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). Table 3.3 shows 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 3.3. - Detailed vs. Condensed Treatment History 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	

RXHIST60				
USRDS ID	BEGDATE	ENDDATE	RXGROUP	
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.

Determining Lost-to-Follow-up Periods

The functioning transplant modality is assumed to continue until a graft failure occurs. A dialysis modality is assumed to continue for a maximum of 1 year in the absence of dialysis billing data or other confirmation of dialysis therapy. After 1 year with no dialysis billing data, the patient is classified as lost to follow-up until evidence of a new modality appears. Patients cannot be classified as lost to follow-up during the first 3 years of ESRD because Medicare may be secondary payer for up to 33 months. The dialysis billing data frequently contain gaps, as listed below.

- The first service date may be known from the Medical Evidence Report, but many ESRD patients are not Medicare eligible for 90 days, delaying billing data for modality determination.
- For patients who have medical insurance other than Medicare, Medicare is the secondary payer for up to 33 months. The first ESRD service date may be known from the Medical Evidence Report, but dialysis billing data are unavailable until Medicare becomes the primary payer.
- Some patients may recover enough renal function to discontinue dialysis, and the USRDS database contains no data for non-dialysis time periods.
- 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.
- Self-dialysis treatments are billed by 2 methods, 1 of which may not be available in the CMS PMMIS/REBUS/REMIS system, causing some home dialysis billing to be missed.
- Some dialysis bills for patients enrolled in Managed Care Organizations may be missed by the CMS billing system.

- Reporting, data entry, and clerical errors may obscure the first service date; for example, first-ever dialysis, not the start of maintenance dialysis, may be reported on the Medical Evidence Report.
- Errors in beneficiary identification may cause data for a single patient to be split between two patients or to be associated with the wrong patient.

Defining Home Hemodialysis Patients

The 2006 ADR instituted an improved method for defining home hemodialysis patients. Previously, modality event data were taken from the SIMS event file only in the case of a gap or missing data in the USRDS Modality Sequence (i.e., Treatment History), which was constructed primarily from Medicare billing data and the REMIS database. However, after applying a series of systematic validation rounds to the SIMS database and examining the root causes of persistent under-counting of home hemodialysis patients, the CC concluded that SIMS event data should play a more prominent role. Thus, for the 2006 ADR, the CC incorporated all home hemodialysis events from SIMS data, with the billing data and the REMIS database, into the standard process of creating the final USRDS Modality Sequence. Subsequently, counts of incident and prevalent home hemodialysis patients rose substantially for all years reported in the 2006 ADR. In the event of data discrepancies, researchers should use home hemodialysis counts from the 2006 or later ADR.

Section 4: Payer History

The 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 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 dataset contains dates of death, used to help establish the end point of the payer sequence.

Payer Categories

Table 4.1 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).

Table 4.1. - Payer Categories

<i>Medicare Code</i>	<i>Medicare/Medicaid Description</i>	<i>Medicare Patient</i>	<i>Mcare/Mcaid Dual Eligibility</i>
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	Medicare as Secondary Payer	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 SAF in the USRDS Core dataset records the sequence of payers for each patient. Table 4.2 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. This is a change from the payer history files for prior years, where each payer code corresponded to only one possible Medicare status.

Table 4.2 - Payer History SAF Variables

<i>Variable</i>	<i>Description</i>
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. Note this is a change from prior years, where MSP was given precedence over other payers.
- If payer and dual eligibility status are the same in consecutive payer sequences, the sequences are collapsed into 1 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 3 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 collect 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.

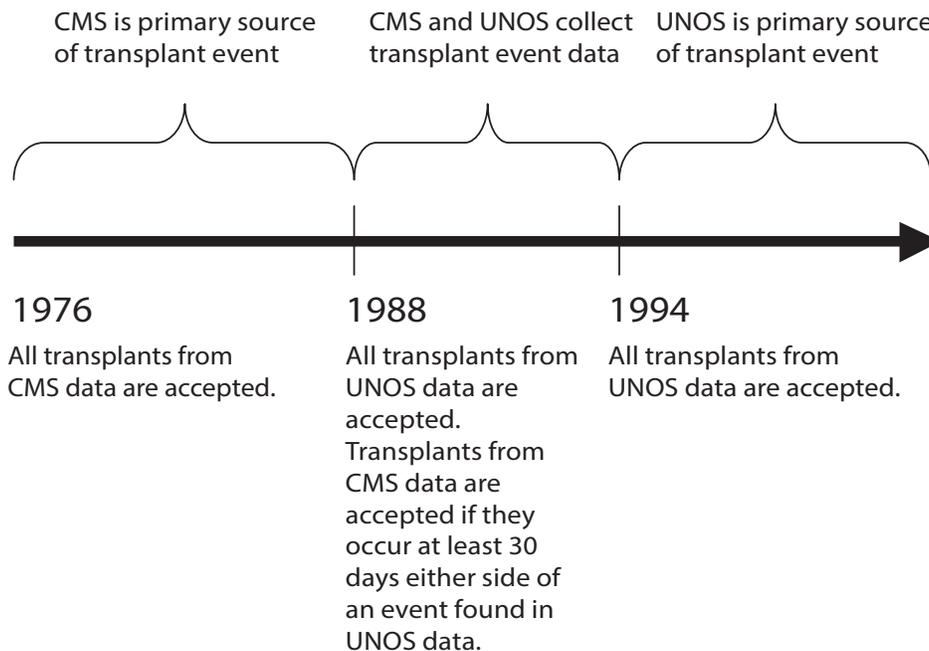
Section 5: Transplant Process and Outcomes

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 (TX) on the Core CD. Transplant event data are combined from various sources including CMS (PMMIS/REBUS/REMIS files) and UNOS. Before 1988, CMS was the primary source for all transplant event data. Between 1988 and 1993, both CMS and UNOS collected information regarding transplant events. Since 1994, UNOS has been the primary source for all transplant event data. (See figure 5.1.)

Figure 5.1 - Source of Transplant Events in the USRDS Population



Reconciliation of CMS and UNOS Events

The USRDS has implemented a decision algorithm that reconciles identical transplant events when multiple sources contain conflicting information.

- Before 1988, all transplant events found in CMS PMMIS/REBUS/REMIS transplant files are used.
- After 1994, all transplant events found in the UNOS files are used.
- Between 1988 and 1993, all transplant events found in the UNOS 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 the reported incident transplant patients in the Medical Evidence Report are used if they occur at least 30 days on either side of a previously accepted transplant event.
- Each transplant event found in the Transplant file (TX) of the Core dataset is thus a unique event that the USRDS believes occurred after reviewing both UNOS and CMS PMMIS/REBUS/REMIS Transplant and Medical Evidence record files.

Table 5.2 details the various USRDS SAFs related to transplant. The Transplant file (TX) on the Core SAF contains transplant dates, basic descriptive data, and causes of graft failure, when known.

Table 5.2 - USRDS Transplant Files

File	Data Source	Comments	Years	Dataset
TX	Constructed by USRDS CC from UNOS, CMS PMMIS/REBUS, Transplant, Medical Evidence files	Identifying data on all known transplants	All	Core
WAITLIST_KI	UNOS Transplant Waiting List file	Contains data of patients on kidney waiting list per listing event	Listing date 1980s -	Core
WAITSEQ_KI	UNOS Transplant Waiting List file	Contains condensed, center-specific, kidney waiting list dates	Listing date 1980s -	Core
WAITLIST_KP	UNOS Transplant Waiting List file	Contains data of patients on kidney-pancreas waiting list per listing event	Listing date 1990s -	Core
WAITSEQ_KP	UNOS Transplant Waiting List file	Contains condensed, center-specific, kidney-pancreas waiting list dates sequence	Listing date 1990s -	Core
TXHCFA	CMS PMMIS/REBUS Transplant file	Details from CMS	1976-93	Transplant
TXUNOS_KI_PRE_JUL04	UNOS candidates, recipients, donors, histocompatibility files	Contains kidney/kidney-pancreas recipients transplant details	Transplant date 1/1/1988-6/29/2004	Transplant
TXUNOS_KI_POST_JUL04	UNOS candidates, recipients, donors, histocompatibility files	Contains kidney recipients transplant details	Transplant date 2003 -	Transplant
TXUNOS_KP	UNOS candidates, recipients, donors, histocompatibility files	Contains kidney-pancreas recipients transplant details		Transplant
TXFUHCFA	CMS PMMIS/REBUS transplant follow-up file	Follow-up details, CMS	1988+	Transplant
TXFUUNOS_KI	UNOS kidney recipients follow-up file	Contains kidney recipients transplant follow-up details	Transplant date 1/1/1988 -	Transplant
TXFUUNOS_KP	UNOS kidney-pancreas recipients follow-up file	Contains kidney-pancreas recipients transplant follow-up details	Transplant date 1/1/1988 -	Transplant

Kidney and Pancreas Waiting list

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

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 UNOS waiting list files `WAITLIST_KI` and `WAITLIST_KP`. Both the raw UNOS 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 `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 UNOS 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 UNOS 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 UNOS. The Transplant file (`TX`) on 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 UNOS on transplant follow-up visits. The follow-up records in `TXFUHCFA` and `TXFUUNOS` overlap in time, especially 1988 to 1993, and contain information collected during patient follow-up visits, which typically occur at 6 months, 1 year, and yearly thereafter. Before 2003, UNOS used 1 file to store kidney and kidney/pancreas recipient registration worksheet data. Starting in 2003, UNOS separated these 2 recipient data groups. However, UNOS did not move the legacy kidney/pancreas recipient data into the respective new files. Users who need kidney/pancreas recipient longitudinal data should use variable `KPPROC` values in 300 level codes in `TXUNOS_KI_PRE_JULO4` to extract the legacy kidney/pancreas transplant records. The users can use `TRR_ID` from the extracted records merging records in `TXFUUNOS_KI` to obtain all legacy kidney/pancreas follow-up records.

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.

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, UNOS began to use the bitmask technique to maintain multiple values (i.e. selections/choices) within 1 variable. This method applies to all multiple selection questions. Previously, UNOS used as many variables as needed to collect all possible answers for such questions. For example, for the question “Treatment: Other therapies,” UNOS used 3 variables, “PHOTOPH,” “PLASMA” and “LYMPHOID” with answer “Yes/No.” After 2003, UNOS combined these variables into a single variable, “THERAPIES,” for the same question, using bitmask technique. The basic code values are value 1 = Photopheresis, value 2 = Plasmapheresis, value 4 = Total Lymphoid Irradiation (TLI). Value 3 represents the sum of values 1 and 2.; similarly value 5 implies that 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 formats for researcher convenience.

Appendix F, Data Forms, lists the CMS and UNOS 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 D, 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. Variables from the UNOS follow-up forms are labeled using section A-M, letters assigned by the CC because the form does not include question numbers. Table 5.3 gives details regarding these labels.

Users who need kidney/pancreas recipient longitudinal data should use variable KPPROC values in 300 level codes in TXUNOS_KI_PRE_JULO4 to extract the legacy kidney/pancreas transplant records. The users can use TRR_ID from the extracted records to merge the records in TXFUUNOS_KI to obtain all legacy kidney/pancreas follow-up records.

Table 5.3 - Variable Label Prefixes for Transplant SAFs

File	Variable Label Prefix	Definition
TXHCFA	TXPM7681	From CMS form, 1976-1981
	TXPM8194	From CMS form, 1981-1994
	TXPM7694	Appears on all CMS forms, 1976-1994
	TXPM819421b	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		From CMS form before 1994
TXFUUNOS**	KIF	Kidney transplant recipient follow-up
	KPF	Kidney-pancreas transplant recipient follow-up

* Include TXUNOS_KI_PRE_JULO4; TXUNOS_KI_POST_JULY04, and TXUNOS_KP.

** Include TXFUUNOS_KI and TXFUUNOS_KP.

Most descriptive data on transplant events can be found in the Core SAF’s TX 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 6: 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 first hospitalization, total admission, and hospital day rates.

Data Sources

Inpatient hospitalization data, a subset of the data in the Institutional Claims file, are included in the `HOSP` data file. 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 D of the 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 2 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 `UID` 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. Use data in which the 90-day rule defines patient start dates. The 90-day rule defines each patient start date as day 91 of ESRD. This avoids 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 `UID` 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 be excluded, thereby limiting their contribution to patient-years at risk. One method of MSP patient exclusion used in the ADR hospitalization rate calculation involves excluding dialysis patients who do not reach a certain level of Medicare paid dialysis bills. Dialysis patient start dates (January 1 of the year for prevalent patients and day 91 of ESRD for incident patients) must 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.

Hospitalization Rates

When we refer the hospitalization rate of a cohort, we usually mean the underlying hospitalization rate, which is not observable. But hospitalization rates can be estimated by observed rates or model-based rates. The calculation of the observed hospitalization rate is straightforward. It is the number of hospitalization admissions observed divided by the follow-up time. For the ESRD population or large groups, the observed hospitalization rate works well for estimating the underlying hospitalization rate. But for small groups, the observed hospitalization may not be a good estimator. It may be unstable. Model-based rate may be more appropriate.

Statistical modeling is used to produce stabilized rates for small groups. Current USRDS methodology for computing total admission and hospital day rates uses a generalized linear model with log link and Poisson distribution, details for model-based event rate calculations can be found in Appendix C, Statistical Methods.

Notice that the calculation of follow-up time may be affected by which patient group you are looking at. For example, when you calculate the hospitalization rate for the January 1 point prevalent dialysis patients in 2010, you will follow patients from January 1, 2010 to the earliest date of death, 3 days before kidney transplant, loss-to-follow-up, or December 31, 2010. If calculating hospitalization rates for hemodialysis patients, you also stop follow-up when patients are changed to PD. Also, the days that a patient was in-hospital should be excluded from the follow-up time of that patient because he or she was not at risk for another hospitalization in those days. When a patient is discharged from one hospital to another, we usually consider that as one hospitalization.

When hospitalization was used as an outcome, length of stay, or number of hospital days is also used as a measurement. Number of hospital days per patient year can be calculated the same way as the hospitalization rate. But, for hospital days per patient year calculation, the days that patient was in-hospital is counted for the follow-up time.

Unadjusted vs. Stabilized Hospitalization Rates

Hospitalization rates may be computed as raw unadjusted rates or as model-based rates. Calculation of raw rates is fairly straightforward, but caution is needed when interpreting raw rates because they may be unstable for small patient groups. In the case of unstable raw rates, a pooled rate (comprising data from several combined years) or a model-based rate may be more appropriate.

Statistical modeling is used to produce stabilized rates, preventing instability of raw rates in groups with small sample sizes and few patient years at risk. Current USRDS methodology for computing total admission and hospital day rates uses a generalized linear model with log link and Poisson distribution. Calculation of raw rates is described below, and methods of calculating raw and model-based rates are described in further detail in Appendix C, Statistical Methods.

Computation of Unadjusted Rates per Patient- Year

Unadjusted first admission, total admission, and hospital day rates can be computed for period prevalent patients for selected years. For each year, the unadjusted rates can be calculated as total events divided by total time at risk. These rates can also be pooled to increase stability, but follow-up is for single calendar year periods using cohorts of patients alive at the beginning of each year. The number of events and the number of years at risk for each event are computed separately for each year and summed over the pooled

years; rates are then computed by dividing the total admissions or days by the total time at risk. Rates may be expressed per patient-year or per 1000 patient-years at risk.

For patients in the all-dialysis, hemodialysis, and peritoneal dialysis categories, the period at risk for all hospitalization analyses is from January 1 or day 91 of ESRD until the earliest of death, 3 days before transplant, or December 31. The only modality change considered a censoring event is from dialysis to transplant. For dialysis patients in the all-ESRD category, in contrast, the analysis period for hospitalization is censored only at death or December 31; modality change is not a censoring event. For transplant patients in the all-ESRD and transplant categories, the analysis period is censored at the earliest of death, 3 years after the transplant date, or December 31. Censoring transplant patients at 3 years after the date of transplant is necessary because Medicare eligibility may cease and hospitalization data may be incomplete.

If a hospitalization begins before January 1 or day 91 of ESRD and continues into the analysis year, the time at risk for first admission begins the day of discharge from this bridge hospitalization. Patients with bridge hospitalizations that span the entire analysis period are excluded from the first admission rates.

Time at risk is calculated differently for length of stay and for total admissions. Because hospitalized patients remain at risk for additional hospital days, rates for hospital days include hospital days in the time at risk. But because currently hospitalized patients are not at risk for additional admissions, hospital days for each year are subtracted from the time at risk for total admissions. For hospitalizations in which admission occurs the same day as discharge, zero days are subtracted from the time at risk for total admissions. When bridge hospitalizations span the start of the analysis period, only the days within the period are subtracted from the time at risk for total admissions.

All admissions and hospital days that occur during the analysis period are included, respectively, in the total admissions and length of stay for each year. Admission for a hospitalization that occurs before and spans the start of the analysis period is excluded from the total admissions for that analysis period, and only the hospitalization days within the period are counted in the total days for length of stay rates. The minimum length of stay is 1 day; hospitalizations with admission and discharge on the same day or with discharge the day after admission are counted as 1 day.

Cause-specific Unadjusted Rates per Patient- Year

Cause-specific hospitalization rates may be computed by counting only hospitalizations with selected principal ICD-9-CM diagnosis codes. In the denominator of the rate, computation of time at risk is similar to computation described above for total admission rates. However, in the numerator, only hospitalizations with specified principal diagnosis codes are counted in total cause-specific admissions. For example, principal ICD-9-CM diagnosis codes of 480-486 and 487.0 can be used for pneumonia.

Unadjusted Rates per Patient

Rates of hospital days and total admissions may also be calculated per patient. For a selected year, the numerator remains the total number of events, and the denominator is the total number of period prevalent patients for the year. Rates per patient are susceptible to bias because they do not consider the actual length of time that a patient is at risk during the year. For example, a greater number of patients at risk for only a fraction of the year may produce smaller rates.

Section 7: 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. Results can be reported for both incident and prevalent cohorts for the following groups: all-ESRD patients, all dialysis patients, all hemodialysis patients, all peritoneal dialysis patients, all transplant patients, or all ESRD patients aged 65 years and over, etc. 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, etc.

Data Sources

These survival and mortality analyses require patient demographic information such as age, gender, race, primary cause of ESRD, death date, and cause of death, and treatment modality history data, which can be obtained from the USRDS CORE CD files PATIENTS and RXHIST or RXHIST60.

Unadjusted First-year Survival Probabilities: Incident Patients

First-year survival probability was estimated from the Kaplan-Meier method. Patients were followed from the date of ESRD onset to the earliest date of death, lost-to-follow-up, or 1 year. For dialysis patients, we also stop follow-up at transplant date, and also stop at modality change date for PD and HD patients. Same as hospitalization rates, the first-year survival probability from Kaplan-Meier may not be stable when the group is small. A Cox proportional hazard regression model might be used for subgroup first-year survival probability estimation. At times, the first-year survival probability was also calculated for those who survived the first 90 days after onset of ESRD. For this first-year survival calculation, patients were followed from day 91 to the earliest date of death, lost-to-follow-up, or 1 year plus 90 days. This is usually done for Medicare patients.

Adjusted First-year Survival Probabilities: Incident Patients

The adjusted first-year survival probability can be estimated for a specific cohort with a reference population based on a Cox proportional regression model with only variables adjusting for the model. The Cox model was fit for the specific cohort and the estimates of the baseline survival and coefficients were applied to the reference population to calculate expected first-year survival probability for each patient in the reference population. The average of the expected first-year survival probability over the reference population would be the estimate of the adjusted first-year survival probability.

Adjusted First-year Death Rates: Incident Hemodialysis Cohorts

Adjusted first-year death rates for incident cohorts can be estimated using a Cox model. These death rates are presented using aggregate categories for age, gender, race, and primary cause of ESRD (diabetes, hypertension, glomerulonephritis, and other). A death rate estimated for one of these variables is adjusted for the remaining three; overall death rates for all patients are adjusted for each of the four variables. The method for calculating these death rates is also described in Appendix C, Statistical Methods.

Unadjusted Death Rates: Period Prevalent Cohorts

Period prevalent cohorts include patients who are prevalent at the beginning of the period and those who become incident during the time period (e.g., 2010). For the 2010 period prevalent dialysis patients, follow-up starts on January 1, 2010, for prevalent patients and day 1 of dialysis for incident dialysis patients, and stops at transplant, death, or December 31. The observed death rate was calculated by dividing the number of deaths by total of follow-up time. For subgroup death rate calculation, a Poisson model might be used.

Adjusted Death Rates: Period Prevalent Cohorts

The adjusted death rate for a period prevalent cohort was calculated similarly to the adjusted first year survival probability, with the exception of the Poisson model and the follow-up time of patients in the reference groups.

Section 8: Providers

The Facility SAF file is constructed from data supplied by the CMS Annual Facility Survey, the CDC National Surveillance of Dialysis-Associated Diseases, and the CMS Dialysis Facility Compare database. Information is for dialysis facilities only and data are at the facility level; there are no individual patient-level data.

Construction of the Facility SAF is accomplished by merging the CMS Annual Survey data with the CMS Dialysis Facility Compare database by provider number and year. All provider numbers in the Facility SAF file are assigned by the USRDS in an effort to ensure dialysis facility confidentiality; they are not the original provider numbers assigned by CMS.

Over time, dialysis 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. Linking the 2 numbers is not possible.

Researchers who wish to conduct patient-level studies incorporating the Facility SAF can link patient information from the Detailed Treatment History SAF (RXHIST), the Condensed Treatment History SAF (RXHIST60), or other SAFs with associated dialysis providers, to the Facility SAF via the USRDS provider number and year.

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. CMS Dialysis Facility Compare data are sent quarterly from CMS. The CDC National Surveillance of Dialysis-Associated Diseases survey was requested from the CDC annually through 2002. A list of the CMS and CDC Surveys can be found in Appendix F, Data Forms. The forms are available at www.USRDS.org. As the last CDC National Surveillance survey occurred in 2002, all data for CDC variables are missing after this date.

Provider Numbers Assigned by the USRDS

Each provider has been assigned an identification number by CMS. To conceal provider identity, the USRDS converts this number to an anonymous number.

Profit Status

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

The for-profit group consists of the facilities categorized on the CMS survey. They include individual-profit, partnership-profit, corporation-profit, or other-profit. Non-profit facilities are listed as 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. These 18 categories are designated by the variable TYPOWNER. The 2002 CMS Survey (2004 ADR SAF) dropped this variable and it is missing in subsequent survey periods. Profit status is now determined solely from 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 (LDO) and Small Dialysis Organizations (SDO). 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. The grouping variable, NU_TYPE, combines all of the small Small Dialysis Organizations into a single group

labeled SDO. LDOS are not grouped. 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 include 3 groups: freestanding, hospital-based, and unknown. The variable NU_HBFS designates this status. The number 1 represents hospital-based 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 9: CKD Cohort (5% CKD Cohort Dataset)

Due to growing CKD concern that affects ESRD patients, the USRDS has constructed a set of CKD cohort dataset to help researchers gain a better understanding of causes and effects of this illness and their associations to morbidity and mortality in patients with ESRD.

Data Sources

To create the CKD cohort dataset by year, the CC uses 5% General Medicare Claims data (1992-2010) 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 1 CKD ICD-9 diagnosis code (Table 9.1) was identified in the 5% IP, OP, HH, HS, SNF, and PB SAFS, with 1 record per patient. Definitions of the two key variables CKD_xx (1-year entry period CKD indicator) and CKD_xxyy (2-year entry period CKD indicator) are defined below.

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 comorbid 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 1 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 1-year entry period and 2-year entry period CKD eligibility criteria described above.

Each patient is identified by 2 patient identification numbers. One is FIVEP_ID, a unique patient ID in the 5% 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 duplications. For example, each record in Co-Morbid_95 consists of 1 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.

Table 9.1 - CKD ICD-9 Diagnosis codes

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.XO, 403.X1, 404.XO, 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 10: Comprehensive Dialysis Study SAF Files

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 US dialysis patients who started treatment 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-determine 335 facilities

Over a period of 2 years, 11,292 patients were selected. Of these, 1677 consented and participated in the surveys or lab assays. The CDS SAF files contain the data for these patients. Of the 1677 patients, 1279 participated only in the QOL survey; 364 participated in the QOL and nutrition surveys (231 with and 133 without lab data); 34 did not participate in either survey but provided lab data.

SAF files for the CDS dataset consist of:

- CDS Patient file. Contains 1 record for each of the 1677 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 and nutrition w/lab data, QOL and nutrition w/o ab 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 5 sets of lab data from a subset of QOL/Nutrition participants.
- CDS Dictionary. This file contains each CDS SAF file contents, variable's format.

Section 11: 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 Part D data is obtained from CMS annually with USRDS provided Finder Files. The Part D data is 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). The variables included in the annual enrollment file are shown in table 11.1.

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 Table 11.2. More detailed information on the Part D benefit can be found at: <http://www.cms.gov/PrescriptionDrugCovGenIn/>.

11.1 Variables in the Part D Enrollment SAF

Name	Type	Length	Format	Comment
CRDTBL_CVRG_SW	Char	1		Creditable Coverage Switch
CST_SHR_GRP_CD_01	Char	2		Jan. Cost Share Group Code
CST_SHR_GRP_CD_02	Char	2		Feb. Cost Share Group Code
CST_SHR_GRP_CD_03	Char	2		Mar. Cost Share Group Code
CST_SHR_GRP_CD_04	Char	2		Apr. Cost Share Group Code
CST_SHR_GRP_CD_05	Char	2		May Cost Share Group Code
CST_SHR_GRP_CD_06	Char	2		Jun. Cost Share Group Code
CST_SHR_GRP_CD_07	Char	2		Jul. Cost Share Group Code
CST_SHR_GRP_CD_08	Char	2		Aug. Cost Share Group Code
CST_SHR_GRP_CD_09	Char	2		Sep. Cost Share Group Code
CST_SHR_GRP_CD_10	Char	2		Oct. Cost Share Group Code
CST_SHR_GRP_CD_11	Char	2		Nov. Cost Share Group Code
CST_SHR_GRP_CD_12	Char	2		Dec. Cost Share Group Code
DUAL_STUS_CD_01	Char	2		Jan. Dual Status Code
DUAL_STUS_CD_02	Char	2		Feb. Dual Status Code
DUAL_STUS_CD_03	Char	2		Mar. Dual Status Code
DUAL_STUS_CD_04	Char	2		Apr. Dual Status Code
DUAL_STUS_CD_05	Char	2		May Dual Status Code
DUAL_STUS_CD_06	Char	2		Jun. Dual Status Code
DUAL_STUS_CD_07	Char	2		Jul. Dual Status Code
DUAL_STUS_CD_08	Char	2		Aug. Dual Status Code
DUAL_STUS_CD_09	Char	2		Sep. Dual Status Code

11.1 Variables in the Part D Enrollment SAF (continued)

Name	Type	Length	Format	Comment
DUAL_STUS_CD_10	Char	2		Oct. Dual Status Code
DUAL_STUS_CD_11	Char	2		Nov. Dual Status Code
DUAL_STUS_CD_12	Char	2		Dec. Dual Status Code
PTD_CNTRCTID_01	Char	1		Jan. Contract Indicator
PTD_CNTRCTID_02	Char	1		Feb. Contract Indicator
PTD_CNTRCTID_03	Char	1		Mar. Contract Indicator
PTD_CNTRCTID_04	Char	1		Apr. Contract Indicator
PTD_CNTRCTID_05	Char	1		May Contract Indicator
PTD_CNTRCTID_06	Char	1		Jun. Contract Indicator
PTD_CNTRCTID_07	Char	1		Jul. Contract Indicator
PTD_CNTRCTID_08	Char	1		Aug. Contract Indicator
PTD_CNTRCTID_09	Char	1		Sep. Contract Indicator
PTD_CNTRCTID_10	Char	1		Oct. Contract Indicator
PTD_CNTRCTID_11	Char	1		Nov. Contract Indicator
PTD_CNTRCTID_12	Char	1		Dec. Contract Indicator
RDS_IND_01	Char	1		Jan. RDS Code - Retiree Drug Subsidy Code
RDS_IND_02	Char	1		Feb. RDS Code - Retiree Drug Subsidy Code
RDS_IND_03	Char	1		Mar. RDS Code - Retiree Drug Subsidy Code
RDS_IND_04	Char	1		Apr. RDS Code - Retiree Drug Subsidy Code
RDS_IND_05	Char	1		May RDS Code - Retiree Drug Subsidy Code
RDS_IND_06	Char	1		Jun. RDS Code - Retiree Drug Subsidy Code
RDS_IND_07	Char	1		Jul. RDS Code - Retiree Drug Subsidy Code
RDS_IND_08	Char	1		Aug. RDS Code - Retiree Drug Subsidy Code
RDS_IND_09	Char	1		Sep. RDS Code - Retiree Drug Subsidy Code
RDS_IND_10	Char	1		Oct. RDS Code - Retiree Drug Subsidy Code
RDS_IND_11	Char	1		Nov. RDS Code - Retiree Drug Subsidy Code
RDS_IND_12	Char	1		Dec. RDS Code - Retiree Drug Subsidy Code
DUAL_ELGBL_MOS_NUM	Char	2		Dual Eligible Months Number
PLAN_CVRG_MOS_NUM	Char	2		Plan Coverage Months Number
RDS_CVRG_MOS_NUM	Char	2		Retiree Drug Subsidy Coverage Months Number
USRDS_ID	Num	8		USRDS_ID

11.2

Variables in the Part D Event SAF

Name	Type	Length	Format	Comment
PDE_ID	Char	15		Encrypted 723 PDE ID
SRVC_DT	Num	4	Date9.	RX Service Date (DOS)
NDCNUM	Char	19		National Drug Code
BN	Char	30		Brand Name
GNN	Char	30		Generic Name - Short Version
GCDF	Char	2		Dosage Form Code
GCDF_DESC	Char	40		Dosage Form Code Description
STR	Char	10		Drug Strength Description
QTY_DSPNSD_NUM	Char	8		Quantity Dispensed
DAYS_SUPLY_NUM	Char	3		Days Supply
QUANTITY_LIMIT_YN	Char	2		Whether or Not the Drug has Quantity Limits
PRIOR_AUTHORIZATION_YN	Char	2		Whether or Not the Drug Requires Prior Authorization
PTNT_PAY_AMT	Num	8		Patient Pay Amount
TOT_RX_CST_AMT	Num	8		Gross Drug Cost
BENEFIT_PHASE	Char	2		The Benefit Phase of the Part D Event
CTSTRPHC_CVRG_CD	Char	1		Catastrophic Coverage Code
STEP	Char	2		Maximum Step Number
TIER_ID	Char	2		Medicare Part D Formulary Tier Identifier
USRDS_ID	Num	8		USRDS_ID

