

Chapter XIV

A Guide to the USRDS Database

The USRDS database is a comprehensive and integrated source of data on ESRD in the United States and is intended to be used for biomedical, cost effectiveness, and other economic research by investigators throughout the United States and abroad. The database has been designed to optimize its usefulness in supporting this research.

Chapter II, "Access to the USRDS Data," describes the various methods by which the data in the USRDS database are made available to researchers. It also gives an overview of the Standard Analysis Files (SAFs), data confidentiality restrictions, and procedures for their use.

This chapter describes the coverage and overall structure of the database, the content of the database, and key variables, including treatment modality. Particular attention is given to the SAFs. Chapter XV, "Analytical Methods," describes the analytical methods used for the analyses in this USRDS Annual Data Report (ADR). Chapter XV will give the researcher a fuller picture of the types of analyses which can be supported by the USRDS database. Appendix B provides copies of the data collection forms for the USRDS Special Studies for which SAFs are available.

Documentation of the USRDS database includes the chapters mentioned above of this *ADR*, the *Researcher's Guide to the USRDS Database*, and the *USRDS Operations and Forms Manual*. The *Researcher's Guide* is available from the USRDS CC, by contacting the database contact person in Table II-1 in Chapter 2.

The description of the database is also stored electronically in a data dictionary that is itself part of the database. The printed data dictionary comprises some 500 pages in printed form.

The database is stored on a Digital Equipment Corporation VAX 4000-300 computer and occupies about 8 gigabytes (billions of characters) of disk storage. It is stored and accessed using the Oracle relational database management system. SAS software running on the VAX or on networked PCs is used for most analyses. SAS ACCESS software is used to move data between Oracle and SAS. Depending on the volume of data, the SAFs are provided as SAS transport format data sets on 9 track 6250 bpi magnetic tape, or as ASCII files on PC diskettes.

Coverage of the Database

The data in the USRDS database cover the period from 1977 to the present, starting with patients alive in 1977. The USRDS patient database is updated once each year, generally during May and June, using data from the HCFA Program Management & Medical Information System (PMMIS). For instance, this 1994 Annual Data Report (ADR) is based on data updated through May 1993.

Because of delays in the submission and processing of Medicare bills, these data generally are not considered to be complete until 15 months have passed. Thus, although we have data through May 1993, the data are only considered complete through February, 1992, and this ADR reports results only through December 31, 1991. Analyses in this ADR which use data from 1992 are labeled as preliminary and should be treated with caution.

The database is estimated to include more than 93 percent of the ESRD population in the United States. Types of persons not completely covered include:

- Persons or dependents not eligible under the Social Security Act, such as farm workers who did not pay into the

Social Security system for the requisite number of quarters

- Persons under 65 who are not otherwise covered by Medicare (for instance, eligible due to disability) and who die before Medicare eligibility would begin (generally 60 to 90 days after starting chronic maintenance dialysis)
- Medicare ineligible persons who were treated by the Department of Veterans Affairs (DVA) hospitals before 1990. Only minimal data will be available for DVA patients treated only at DVA facilities

Data may also be incomplete for transplant patients under age 65 who have had a successful graft for more than three years, unless otherwise covered by Medicare. For example, a patient with a functioning graft could still be classified as disabled three years after a successful transplant and thus remain eligible for Medicare. Approximately one third of all first transplants are no longer Medicare eligible at 3.5 years after the transplant. Transplant patients who lose Medicare eligibility remain in the database, and Transplant Follow-up forms continue to be filed for many of them. In most cases, the USRDS database will contain the date of death for transplant patients even if it occurs after they have lost eligibility.

While these groups are generally not in the database, the database does include some data about some persons in these groups. Persons initially insured outside the Medicare system may enter the ESRD program, and therefore the USRDS database, after their other coverage terminates. Further, as noted above, certain data continue to be filed for many, if not most, transplant patients, even after three years with a successful graft. Many DVA patients have Medicare eligibility even though their primary treatment is with DVA. Most DVA patients alive in 1990 and thereafter are included in the database, but data may be limited to that collected on the Medical Evidence Form. Data about dialysis treatments or non-transplant hospitalizations performed in DVA facilities are not in the database, so that some longitudinal data may be limited for DVA patients.

The USRDS database includes data from the HCFA Annual Facility Survey (AFS). This survey reports the number of new dialysis patients (both Medicare and non-Medicare eligible) each year, and the number of patients at the end of each calendar year. The Annual Facility Survey portion of the database covers virtually all Medicare-approved institutional dialysis and transplant providers in the United States from 1980 through 1991.

This Facility Survey has been used to evaluate the completeness of the USRDS patient database. The patient database has more than 90 percent of the count of new patients reported by the AFS. When the USRDS counts are reduced by the proportion of patients reported as non-Medicare, then in some years the USRDS patient database count of new patients is very close to the estimate of Medicare patients from the AFS. This is discussed further in the *Researcher's Guide to the USRDS Database*.

Prior to 1990, USRDS coverage of DVA facilities was limited to those facilities which were Medicare-approved or which voluntarily completed the Annual Facility Survey. Starting with 1990, most DVA facilities completed the Annual Facility Survey. However, the 1991 Facility Survey still reported only about 80 percent of the number of patients counted by the DVA in its own ESRD patient census. Counts for subsequent years have not been evaluated.

The number and quality of data items available to the USRDS have changed over time. A new Chronic Renal Disease Medical Evidence Form (HCFA 2728) was introduced in 1981, which added the categories of Native American/Alaskan Native and Asian/Pacific Islander to the race classification. Compliance with completing the Medical Evidence Form improved greatly, and data items, such as primary disease causing ESRD, have a much smaller proportion of missing data starting about 1982. A new Transplant Form, introduced in 1983, provides much greater detail about transplants. A new Death Notification Form, with a revised and more detailed classification for cause of death, was introduced in 1990. More stringent HCFA reporting requirements for hospital billing were implemented in 1983 with the DRG system; this appears to have improved the quality of some aspects of the hospitalization data.

A new Chronic Renal Disease Medical Evidence Form (HCFA-2728) will go into use in January 1995 and will provide a number of new data items.

Data Sources

The USRDS database is constructed primarily from data collected by the Health Care Financing Administration (HCFA) in connection with HCFA's administration of the Medicare ESRD program. Much of the data are collected through the 18 ESRD Networks, which are responsible for quality assurance as well as for processing submitted forms such as the Annual Facility Survey and the Chronic Renal Disease Medical Evidence, Transplant, Transplant Follow-up, and Death Notification Forms. Data for USRDS special studies are also usually collected by the ESRD Networks. The data collection forms for selected special studies are included in Appendix B. Many of the HCFA data also come directly from the Medicare billing system.

The database also includes supplementary types of data tables drawn from a variety of sources. Text labels and descriptions for the various diagnostic, procedure, and geographic codes are used in producing reports. Cross reference tables are used to merge data using different geographic coding schemes. Census population estimates by year, age, race, and sex are used for computing rates. Data from the Bureau of the Census about characteristics of patient residence ZIP Codes are used to supplement the socioeconomic data contained in the patient data. Land area, population density, and latitude and longitude coordinates of ZIP Codes are used for various types of geographic analysis, such as computing distances between patient residences and ESRD providers.

Structure of the Database

Figure XIV-1 shows the overall structure of the database. It consists of three primary types of data: data about patients, data about medical events occurring to patients, and data about institutional providers of medical services. From these primary types of data, annual summaries are derived of the treatment history of each patient. These summaries are designed to facilitate many of the analytical methods used in ADR Reference Tables.

The database includes all of the data collected for the various USRDS special studies, as described in Chapter XIII, "USRDS Research Studies."

The USRDS is currently in the process of adding economic data to the database to support cost effectiveness and other economic research. This will result in adding a Medicare billing data file to the database and in incorporating economic data in the Treatment History (Modality Sequence) file and perhaps in other places. Figure XIV-1 also shows the SAFs which contain the data in each part of this structure. A list of the SAFs appears in Table XIV-1, which is repeated from Chapter II. The SAFs with patient data can be linked together using the USRDS encrypted patient ID number, which is contained in each of the patient data files. The SAFs are provided as SAS files, and SAS can readily perform this linking. The Treatment History records can be linked to the data for the ESRD facility providing treatment (in the Facility SAF) based on the USRDS encrypted provider ID number, which is contained on both the Treatment History and Facility SAFs.

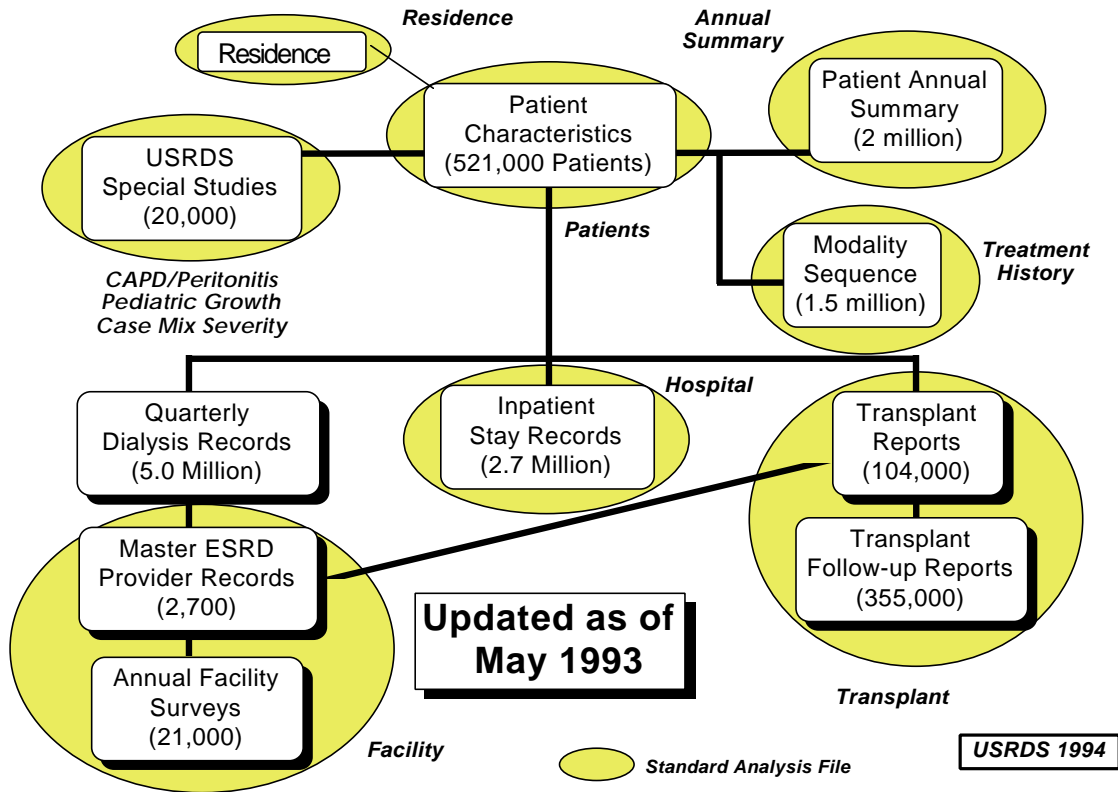


Figure XIV-1

Structure of the USRDS Database and Standard Analysis Files.

In order to protect patient confidentiality, the USRDS SAFs contain no patient identifiers which can be used directly to identify a patient, such as name, address, Social Security number or Medicare beneficiary ID. The USRDS patient and facility ID numbers are randomly assigned numbers which uniquely identify each patient and provider in the USRDS database and which can be used to link data from the database but which have no meaning outside of the database.

well as the Standard Analysis “Files.” In a relational database such as the USRDS database, the data are organized as two-dimensional tables in which the rows represent entities (such as patients, providers, or events) and the columns represent attributes of those entities. The tables can also be thought of as rectangular data files where the table rows are equivalent to the records and the table columns are equivalent to variables or fields on the file records.

In the discussion which follows, we will sometimes refer to “tables” in the database as

USRDS Standard Analysis Files

File Name	SAS File Names	Unit of Observation	Uses
Patient	PATIENTS	Patient	Incidence, prevalence, patient survival. Most other files will need to be linked to this file using the encrypted ID.
Residence	RESIDENC	For each patient, one record for each period in a different residence.	Regional analyses
Treatment History (Modality Sequence)	MODSEQ	Patient. One record for each spell a patient spends on one modality	Modality distribution and treatment patterns. Permits an easy determination of treatment modality at a point in time and changes in modality over time.
Hospital	HOSP	Patient hospital stay	Hospitalization analyses
Transplant	TX	Transplant. Can have multiple transplants for one patient.	Transplant and transplant outcome analyses. Includes file of denominators for rates of transplantation per patient year on dialysis.
Transplant Follow-up	TXFU	For each patient, one record for each follow-up form filed. Should be filed at discharge, 6 months, 12 months, then annually.	Immunosuppressive therapy, rehabilitation, rejection episodes.
Case Mix Severity (USRDS Special Study)	CASEMIXS	Patient. One record for each patient in the study.	Comorbid conditions, adequacy of dialysis, dialysis prescription and other treatment parameters, laboratory test values.
Pediatric Growth and Development (USRDS Special Study)	PEDGROW	One record for each patient in the study.	Growth, development, and other issues relating to pediatric ESRD Patients
CAPD Peritonitis (USRDS Special Study)	CAPD	One record for each patient in the study.	CAPD and peritonitis.
Case Mix Adequacy (USRDS Special Study)	ADEQUACY	One record for each patient in the study	Comorbid conditions, adequacy of dialysis, dialysis prescription and other treatment parameters, laboratory test values. Initial file of 4000 cases available 8/94. Full 6000 cases available 6/95.
Facility	FACILITY	One record for each year during which each facility was in operation	Merge with the treatment history, transplant, or annual summary SAFs for analyses involving provider characteristics.
Annual Summary	PATSUM	One record for each year during which a patient was alive at some time.	End of year point prevalence, modality distribution, denominators for rates of transplantation per patient year on dialysis.

Table XIV-1

Tables in a relational database can be joined together based on values of specified columns. For instance, most of the database tables discussed in this chapter can be joined together using the patient ID number. The facility tables can be joined to dialysis or transplant tables using the USRDS encrypted provider ID. In many cases, the SAFs correspond to database tables.

The remaining sections of this chapter discuss the various parts of the database in greater detail.

Data About Patients

The Patient SAF provides a roster of all patients in the database whose first service date is at least 15 months prior to the date the database was updated and who have no unresolved data problems which cause them to be excluded from analysis. The 15 month rule was discussed earlier under the topic, "Coverage of the Database." The data in the Patients SAF are included in the Transplant and Special Studies SAFs for the patients that appear in those files. In order to link patient characteristics with treatments in the Treatment History SAF or with the Hospital SAF, it is necessary to merge the Patients file with the other file, a task which is easily done in SAS.

The Patient file contains basic data about the characteristics of the patient. These data are stored in the ID, Medical Evidence, and Residence tables in the database. The file also contains some summary items about the patient's treatment, such as treatment modality at day 90, number of transplants, and characteristics of the first transplant.

The sociodemographic data on patients are limited to race, sex, and age. Data items such as income, marital status, and education are not available. The availability of some additional socioeconomic data for the general population by ZIP Code or County allows characteristics of residents of the ZIP Code or County area to be used as proxies for characteristics of the patient.

Files of characteristics of ZIP Codes and County are available from a number of commercial sources. The fact that ZIP Code boundaries change, sometimes extensively, over time must be taken into account when performing such analyses.

Race

Patient race is recorded using the following categories:

- Black
- White
- Native American/Alaskan Native
- Asian/Pacific Islander
- Other
- Unknown

The race of each patient is determined by combining the race code reported on the Chronic Renal Disease Medical Evidence Form (HCFA-2728) with the race code recorded in the HCFA Medicare Health Insurance Master File. The resulting race code has indicated "other" or "unknown" in less than 2 percent of all cases in recent years

The Native American/Alaskan Native and the Asian/Pacific Islander race categories were added to the Chronic Renal Disease Medical Evidence Form (HCFA-2728) in 1981. Counts of new patients for these categories are significantly under reported before 1982. In addition, starting in 1982, compliance with the Medical Evidence Form greatly improved. As a result of these changes, the proportion of new patients in the "other" and "unknown" race categories drops from about five percent before 1981 to about one percent from 1982 to 1986. The proportion in the "other" category increases after 1986 due to a large increase in the proportion of this category in the Medicare Health Insurance Master File, but the combined proportion of "other" and "unknown" remains under two percent. The discontinuity in the reporting of race before and after 1982 makes analyses by race difficult for earlier time periods.

Primary Disease Causing ESRD

Variables in the Patient File

Variable Group	Variable Name	Description
Identifier	USRDS_ID	USRDS encrypted patient ID number for use when merging with other files.
Demographics	BORN	Date of birth
	INCAGEC	Age at start of ESRD, in 5-year groups
	SEX	Sex
	RACE	Race
	DIED	Date of death
	CDEATH	Cause of death
	CDEATH89	Cause of death recoded to pre-1990 codes
ESRD Incidence	RXSTOP	Treatment stopped prior to death (1990 on)
	FIRST_SE	Date of first ESRD service
	INCYEAR	Year of first ESRD service
	PDIS	Primary disease causing ESRD
	DISGRP	Grouped cause of ESRD
	DISGRPC	Grouped cause of ESRD (USRDS standard 8 groups)
	NETWORK	ESRD Network at start of ESRD
	STATE	State of residence at incidence
	COUNTY	County of residence at incidence
	ZIPCODE	Zipcode of residence at incidence
ESRD treatment	RXDAY90	Treatment modality on day 90
	TOT_TX	Total transplants
	TXDATE	Date of first transplant
	DONOR_TY	Donor type of first transplant
	TXFAILED	Date of failure of first transplant

Table XIV-2

The primary disease causing renal failure is recorded on the Medical Evidence Form, which is completed as part of the process of certifying the ESRD patient's eligibility for Medicare benefits. Starting in 1982, primary disease is recorded for more than 90 percent of incident cases, but missing Medical Evidence Forms significantly reduce the completeness of this data item before 1982.

eight categories are presented in Sections A and B of the Reference Tables. Section A also presents a more limited set of data about more detailed disease categories.

Table XIV-3 shows the primary disease categories generally used in this report. The variable DISGRPC in the Patients SAF uses the eight broad categories shown, while the variable DISGRP uses the more detailed categories. Further data about primary disease appear in Chapter IV. Incidence and prevalence for the

**Incidence of treated ESRD by Primary Disease
Group for all Patients, 1988-1991**

Primary Disease Group	Total 1988-91 ¹	% of Total
All ESRD, (reference)	177,841	100.0
1 Diabetes	60,060	35.6
2 Hypertension	50,369	29.8
3 Glomerulonephritis	22,578	13.3
4 Cystic Kidney Diseases	5,398	3.2
5 Other Urologic Disease	9,199	5.5
Interstitial Nephritis	5,477	3.2
Obstructive Nephropathy	3,722	2.3
6 Other Known Cause	10,866	6
Collagen Vascular Diseases	3,782	2.2
Malignancies	2,249	1.3
Metabolic Diseases	873	0.5
Congenital/Other Hereditary Disease	1,342	0.7
Sickle Cell Disease	154	<0.1
AIDS-Related	586	0.3
Other ESRD	1,880	1.1
7 Cause Labeled Unknown	10,196	6.0
8 Missing Information	9,175	.

¹Divide total by 4 to determine average annual counts. Note these figures include patients in Puerto Rico and U.S. Territories.

Medicare Patients Only.

Table XIV-3

Coding for primary disease on the HCFA Chronic Renal Disease Medical Evidence Form (HCFA-2728) is based on a selected group of ICD-9-CM diagnostic codes. The list of allowable codes has been expanded in recent years to provide more detail. There are a number of inconsistencies, however, between the HCFA coding system and the full ICD-9-CM coding system. These problems and the USRDS approach to using the diagnostic codes are documented in a USRDS Technical Memo (Webb).

Diagnosis is also recorded for inpatient hospital stays. Up to five diagnostic codes are recorded for each stay, using the full ICD-9-CM coding scheme.

Date of First ESRD Service

The date of first ESRD service is used as the date of ESRD incidence, and most analyses depend in some way on this date. This date is derived from

the first dialysis or transplant date recorded on the Medical Evidence Form (HCFA-2728), the dates of transplants from the transplant form, the earliest Medicare dialysis bill date, and the Medicare eligibility date. Priority generally is given to the first dialysis date on the Medical Evidence Form (HCFA-2728). The USRDS Data Validation special study included a validation of the date of first ESRD service, as reported in Chapter IX of the 1992 ADR. For a sample of 734 patients from the USRDS database, the first service date from the database was compared with the date in the medical record. In 64 percent of the cases, the date matched exactly, and in 94.4 percent of the cases the dates were within 60 days of each other.

There are many cases where there is a period after the first service date for which we have no other data about the patient. The most common case is that of patients under 65 whose Medicare eligibility begins up to 90 days after the start of regular maintenance dialysis. However, there are cases where the delay is much longer. For

instance, if a patient is covered by the Department of Veterans Affairs (DVA), or by another insurance carrier for four years and then becomes eligible for Medicare and has a Medical Evidence Form filed, his first service date would be four years prior to his appearing in the Medicare file. Gaps may also be caused by errors in recording or coding the first service date.

The USRDS applies a set of quality control (QC) filters to the database to eliminate from analysis the cases which have obviously incomplete, inconsistent, or incorrect data. Many of the cases identified through these QC filters appear to not be true ESRD patients. These may be persons who experienced acute, rather than chronic, renal failure, or may be persons who are incorrectly classified as ESRD in the Medicare database. Overall, 4.2 percent of the cases in the database are removed by the QC filters, but for data since 1983 only one percent or fewer have been removed. **The QC filters have been applied to all the data used in the USRDS ADRs, including the Reference Tables.**

Death Date and Cause

The date of death is likely to be reliable because it comes from three sources: the Social Security system, the ESRD Death Notification Form (HCFA-2746), and hospital bill discharge data. The date of death is provided to the HCFA system from the Social Security system, where it is obtained from the individual's official Death Certificate. The HCFA ESRD Death Notification Form (HCFA-2746) is used primarily for other information about the beneficiary's death. The USRDS Data Validation special study has confirmed the accuracy of the death dates (see Chapter IX of the 1992 ADR). In that study, the date of death matched exactly in 83.2 percent of the cases and matched within 60 days in all cases.

The USRDS follows the convention of defining modality and prevalence based on a patient's

status at the end of the day after all changes occurring that day have taken place. Thus, patients who die on December 31 are excluded from the December 31 point prevalence counts but are included in the annual period prevalence counts.

One primary and two secondary causes of death can be recorded on the ESRD Death Notification Form. During 1990, a new version of this form was introduced which has an expanded coding scheme for cause of death. The new coding scheme uses codes which are mutually exclusive with the old codes so that the old codes could be derived from the new codes. The variable CDEATH on the Patients SAF contains the primary cause of death. The variable CDEATH89 recodes the data from the new form into the pre-1990 categories. With the new death codes, the variable RXSTOP must be used to determine whether the discontinuing of ESRD treatment was a cause of death.

Chapter XV includes a discussion of the effects of this change in the cause of death coding.

Residence and Geography

The database includes data on the place of residence of patients and the location of facilities. These locations are coded in terms of ESRD Network, State, County, and ZIP Code. See Figure III-3 for a map of the ESRD Networks. State and County are coded using the FIPS (Federal Information Processing Standards) codes.

The Patient SAF includes the residence at the time of first ESRD service. The Transplant SAF contains the residence at the time of the transplant. Since patients may move, a longitudinal Residence SAF is included with the Patient SAF package. Table XIV-4 shows the variables included in the Residence SAF.

Variables in the Residence File History File

Variable Group	Variable Name	Variable Description
Identifier	USRDS_ID	USRDS patient ID number for use when merging with other files.
Residence Period	BEGRESLAS	Begin of residence date for this patient.
Detailed Modality Period	BEGDATE	Begin of residence for this patient, as dates and as days from first service date.
	BEGDAY	
Location	ENDDATE	ESRD Network number
	NETWK	State FIPS code
	STATE	County FIPS code
	PROVIDER	ESRD Service provider during this period
Modality Period Under 60-day Rule	ZIPCODE	Postal ZIP code during this period in days
	ZIPCODES	
	BILLED	Number of billed dialysis days in this period
	RXDETAIL	Detailed treatment modality code
	RXGROUP	Grouped treatment modality code
Modality Period Under 60-day Rule	ENDDAT60	Begin and end date of this 60-day rule period, as dates and as days from first service date.
	ENDDAY60	
	RXGRP60	Grouped treatment modality code for this 60-day rule period

Table XIV-4

The Residence SAF contains a history of each patient's place of residence since the start of ESRD. Since the Residence data generally are based on only one observation per year, the dates of change of residence are imprecise. Below is a list of the variables on the Residence SAF.

The geographic location of treatment can be determined by linking the Treatment History SAF or the Transplant SAF with the Facility SAF. The Facility SAF contains the ESRD Network, State, County, and ZIP Code of the facility. For 95 percent of the patients in the database who were alive at the end of 1990, the state of residence was the same as the state in which treatment was received, although the county of residence frequently differs from the county in which treatment is received. In the vast majority of cases in which the states differ, the states are adjacent. This topic is discussed further in Chapter XV.

The ZIP Codes of place of residence and place of treatment can be used to estimate the distance from the residence of the patient to the place of treatment. Such an analysis requires the latitude

and longitude of the ZIP Codes, and files giving the latitude and longitude of ZIP Code centroids are available commercially. The fact that ZIP Code boundaries change, sometimes extensively, over time must be taken into account when performing such analyses.

The Treatment History

The Treatment History, or Modality Sequence, Standard Analysis File is one of the most useful tools for making the wealth of data in the USRDS database accessible for research. The process of determining the sequence of treatment modalities for an individual patient in the USRDS database requires examining almost all of the data for that patient.

In this section, we will describe the Treatment History file and then summarize the primary data about events occurring to patients upon which the Treatment History is based. These events are divided into several categories in the database, including dates for first ESRD service, death, dialysis, hospital inpatient stays, and transplants.

Modality Classifications Used in USRDS Analyses

Code	Modality	Description
1	Death	Death does not appear in the year-end modality tables, which report only living patients. This category is used when looking at the subsequent modality history of patients alive at some point, as in the time series charts in Chapter IV of the 1991 ADR.
2	Transplant	Transplant Patients with a functioning transplant.
4	Center Hemodialysis	Center Hemodialysis is the most common category. Center Self Hemodialysis is usually combined with this category.
5	Home Hemodialysis	Home Hemodialysis. This category should be treated with some caution since home hemodialysis patients cannot always be reliably identified.
6	Center Self Hemodialysis	Center Self Hemodialysis is a very small category of patients who administer their own hemodialysis at a dialysis center. This category is usually combined with Center Hemodialysis.
7	CAPD	CAPD Continuous Ambulatory Peritoneal Dialysis. CCPD is usually combined with CAPD.
8	CCPD	CCPD, Continuous Cycling Peritoneal Dialysis, is usually combined with CAPD.
9	Other Peritoneal Dialysis	Other Peritoneal Dialysis includes primarily intermittent peritoneal dialysis (IPD). It is a small category except among very young children. It is sometimes combined with unknown dialysis and unknown modality to form an other/unknown dialysis
3 A	Unknown Dialysis	Unknown Dialysis occurs when a dialysis or Medical Evidence Form indicates dialysis but does not indicate the type of dialysis. It also occurs when we have a first service date or a transplant failure date followed by a period in which there are no data.
B	Unstable Dialysis Modality	Unstable Dialysis Modality is used when aggregating a series of short spells (less than 60 days) of differing dialysis modalities.
X	Lost to Follow-Up	Lost to Follow-Up is used to classify periods of at least one year during which we have no dialysis data about a patient, unless the patient is classified as having a functioning transplant.

Table XIV-6

The Treatment History SAF records the chronological history of ESRD treatments for each patient as derived from the primary event data. This SAF has one record for each spell on a given treatment from a given provider for each patient. The file also defines continuous periods of treatment with a modality, based on a definition which ignores changes in provider and ignores short periods on a different modality. A "60-day rule" is used; that is, spells of less than 60 days on a modality are not considered a true change in modality.

The Treatment History SAF is used for analyses which involve determining treatment modality at a point in time, such as six months or two years

after first service, which analyze switching between modalities, or which summarize treatment experience over periods other than a calendar year.

Table XIV-5 shows the variables in the Treatment History SAF.

Most USRDS analyses (see Reference Table Section C) use the ten modality classifications shown in Table XIV-6 or the aggregations indicated in that table. These ten categories are derived from a set of more detailed categories, which are based on the dialysis type and setting fields on the dialysis bills. Both the Modality Sequence and Patient Annual Summary tables

Principal Variables in the Hospital SAF

Variable Group	Variable Name	Variable Description
Identifier	USRDS_ID	USRDS patient ID number for use when merging with other files.
Stay Dates	ADMITTED DISCHARG	Beginning and ending date of this stay.
Stay Description	HSDIAG1 to HSDIAG5 SURG1 to SURG3	Up to 5 diagnosis codes. Up to 3 surgery codes
EPO Usage	HSDIALTP	Type of inpatient dialysis
	HSDIAL	Number of dial
	DRG_CODE	DRG code.
	DRG_DEST	Destination at discharge.
	EPOADMIN	EPO administrations
	EPOCHARG	EPO charges
	EPODOSE	EPO dose

Table XIV-7

include variables classifying modality at both levels of detail.

The methods for determining treatment modality are described in Chapters XI and XIII of the 1992 ADR and in the *Researcher's Guide to the USRDS Database*.

Lost To Follow-up Status

There are major gaps in the dialysis histories of many patients for a number of reasons. As reported in the 1992 ADR, a lost to follow-up modality category has been added to cover periods of more than one year during which we have no data. Lost to follow-up is not a permanent state, and patients frequently return from this category to another category.

A period of at least 365 days without any dialysis bills is classified as a lost to follow-up period, with two exceptions. First, a functioning transplant period is allowed to continue indefinitely until a transplant failure is encountered. We do not expect dialysis bills while a transplant is functioning, and Medicare eligibility may be lost after three years with a functioning transplant. The second exception occurs when we have a first service date and then a delay before dialysis bills start. We assume that the first service date is accurate and that it is

followed by at least a year of dialysis not covered by Medicare unless the data indicate otherwise.

Therefore, the earliest that a lost to follow-up period can start is 366 days after the first service date. If dialysis bills start during the first or second year, then a lost to follow-up period is not generated, and the initial dialysis modality (unknown dialysis, or as indicated on the Medical Evidence Form) is carried forward until the start of dialysis bills. If dialysis bills do not start until after the second year, then the lost to follow-up period begins at day 366 and continues until the start of dialysis bills.

Dialysis

Data about dialysis are summarized in quarterly increments and are stored in the Quarterly Dialysis table in the database. Additionally, the date and type of first chronic maintenance dialysis is recorded for each patient. The Quarterly Dialysis data are not available as a SAF, but they represent the largest source of data used in creating the Treatment History SAF.

The quarterly dialysis portion of the USRDS database contains about 5 million records, each representing the services performed by a given provider to a given patient in a given quarter. A patient will typically have one record per quarter, but a patient who changes dialysis centers or who

uses another center (e.g., for two weeks while on vacation) will have more than one record for a quarter. The records indicate the type and setting (center or home) of the last dialysis treatment performed in that quarter, but do not record changes in the type of dialysis within the period covered. Thus, it is possible to miss a switch, for example from hemodialysis to CAPD, of short duration.

Hospital SAF

The Hospital SAF has one record for each hospital inpatient stay covered by Medicare for each patient. Data recorded for inpatient hospital stays include dates, discharge status, DRG code (after 1983), up to five diagnostic codes, and up to three surgical codes. Before 1992, one-day hospital stays were not captured by the database. For about 15 percent of the patients in the database, there is a delay of up to three years or more after the first service date before the system begins capturing hospitalization data, resulting in incomplete data for these patients. This is particularly important for elderly patients who survive only a short time after the start of ESRD. The variable HIMPOST table indicates the date before which hospitalization data are likely to be incomplete. Below is a list of selected variables from the INPATIENT STAY table.

Transplant

Transplants are a subset of hospital stays, but much additional information is collected about each transplant a patient receives. The transplant data form was changed in 1982, and more detail is available starting with that year. Starting in July 1994, the transplant data will be collected by the United Network for Organ Sharing as part of the National Organ Procurement and Transplantation Network.

Follow-up data on transplants are collected at the time of hospital discharge, at six and twelve months after discharge, and annually thereafter for at least three years. Transplant patients under age 65 may lose Medicare eligibility after three years of a successful graft, but annual follow-up forms continue to be filed for many, if not most, patients after this point. The transplant follow-up

data are incomplete for many transplants, but reporting has improved in recent years.

Transplant Follow-Up

The date of transplant failure is not reported definitively by any single data source. It is reported on the Transplant Follow-up Form, but Follow-up Forms are missing for some transplants, and the failure dates reported on the follow-ups are sometimes inconsistent with the patient's dialysis records.

Because of these problems, the USRDS derives a transplant failure date from the following:

- Date of death
- Date of subsequent transplant
- Graft failure date reported on a Follow-up Form
- Date of graft nephrectomy as reported on the Transplant Form or for a hospital stay
- Date returned to dialysis as reported on the Chronic Renal Disease Medical Evidence Form
- Return to regular dialysis as indicated by a continuous spell of more than 22 dialysis treatments reported in the billing data (quarterly dialysis records)

The general procedure is to assume that the date reported on the Transplant Follow-up Form (if available) is correct unless death or a new transplant occurs before this date. If the Transplant Follow-up Form is not available, then the earliest of the above dates (if any is reported) is the failure date. Chapter IX in the 1992 ADR includes a validation of the transplant failure dates.

The transplant records contain data about the HLA antigens of donor and recipient. Since matching of the antigens between donor and recipient is of considerable interest, the number of HLA matches and mismatches is computed for each transplant, using the method described below.

Principal Variables in the Transplant SAF

Variable Group	Variable Name	Variable Description
Identifiers	USRDS_ID	USRDS patient ID number for use when merging with other files.
	TDATE	Date of transplant
Provider	INCCOUNT	Sequence number for transplants for this patient
	TRANSPLA	Transplant hospital
Tx Failure	FAILDATE	Date of transplant failure, derived from a variety of data
	GFAIL1, GFAIL2	Primary and secondary reasons for graft failure, from the Transplant Follow-ups. These variables are present only if they were recorded on a follow-up form.
Patient Characteristics	BORN, DIED, CDEATH, RXSTOP, FIRST_SE, DISGRP, SEX, RACE	Most variables from the Patients SAF are included
	RETHNIC	Hispanic ethnicity
Donor Characteristics	RBLOODTY	Recipient blood type
	RHBSAG	Recipient HBS positive
	DTYPEDET	Detailed donor type (cadaveric, living related, etc)
	LRTYPE	Living related donor type
	DAGE	Age
	DBLOODTY	Blood type
	DRACE	Race
	DETHNIC	Hispanic ethnicity
	DSEX	Sex
	DCANCER	Cancer at harvest
TX Procedure	DCMV	CMV antibody
	DHBS	HBS AG positive
	PULPERF	Pulsatile perfusion time
	COLDTIME	Cold time
	INPATDAY	Number of inpatient hospital days
	WARMTIME	Warm ischemia time
	XFUSNUMB	Number of transfusions
	NEPHREC	Nephrectomy
	SPLENECT	Splenectomy
	Lab Values	PRA
RCMV		Recipient CMV antibody
RES1WAY, RES2WAY		Relative response 1 way and 2 way
CREATDEC		Creatinine decline
MLC1WAY, MLC2WAY, NO_MLC		MLC 1 way, 2 way, or not done
STIMIND1, STIMIND2		STIM index 1 way and 2 way
HLA Matching		Matches and mismatches for A, B, adr DR loci (12 variables)

Table XIV-8

To determine the HLA A, B, and DR mismatch counts in this report, the following algorithm was used. In the file used to create the HLA mismatch tables, there are six variables containing values for recipient loci and six comparable variables containing values for donor loci. Each set of six variables is divided into three pairs of two variables each for the A, B, and DR loci. Both recipient and donor variables are first checked to insure they contain information (i.e., values greater than zero). Counts are kept of how many variables in each of the three categories on a given observation meet this test (the maximum count is two, the minimum count is zero). Missing values and zeroes are not considered in the determination of mismatches for the tables.

The number of matches in each category (A, B, and DR) is then computed by comparing each of the two recipient loci values with each of the donor loci values of the same type. Using this method the maximum number of matches within a category is two, while the minimum number is, of course, zero.

The number of mismatches can then be determined by subtracting the number of matches from the number of valid (zero or greater) entries in the recipient's category. The combination mismatch totals are simply a matter of adding together the individual categories (i.e., $AB \text{ mismatches} = A \text{ mismatches} + B \text{ mismatches}$). If all values for a given locus are invalid (missing or zero), the locus is counted in the row labeled INVALID in Reference Tables, Section F.

Chapter IX of the 1992 ADR contains a validation of the HLA A, B, and DR mismatch codes.

Data About Institutional Providers

The Facility SAF contains data on the characteristics of institutional providers of dialysis and transplantation services since 1981. These data are obtained via the Annual Facility Survey of dialysis and transplant providers conducted by the 18 HCFA Networks. Data about characteristics of the providers (such as size, profit/nonprofit) and about the volume and type of services provided are recorded. Data also include the number of transplants performed by each transplant center and a count of the patients on that facility's transplant waiting list. This waiting list count may be overstated due to multiple listings, although comparisons with UNOS waiting list data suggest that this bias was minimal in 1991. The Annual Facility Survey is the primary source of information about counts of non-Medicare dialysis patients and transplants, and a second source for counts of services and Medicare patients. A master provider data file includes additional provider characteristics, such as unit type, ownership, certification status, and address.

As discussed earlier, the USRDS has assigned a randomly generated ID number to the ESRD facilities, and this ID number is used to link data from the Facility SAF with facilities recorded in the Treatment History or Transplant files. The Facility SAF is provided with the Treatment History and Transplant SAFs.

Principal Variables in the Transplant Follow-Up SAF

Variable Group	Variable Name	Variable Description
Identifiers	USRDS_ID	USRDS patient ID number for use when merging with other files.
	TDATE	Date of transplant
	FUNUM	Follow-up number: 1 = discharge, 2 = 6 months, 3 = 12 months, 4 = 2 years, 5 = 3 years, etc.
Patient status	DIALYSIS	Dialysis performed during this period
	REJECTN	Rejection during this follow-up period
	CURCRET	Current serum creatinine level
	MAXCRET	Maximum serum creatinin level during this follow-up period
Immunosuppressive therapies	REHAB	Rehabilitation status
	ALG	Antithymocyte globulin
	SOLUMED	Solumedrol
	PREDNSNE	Prednisone
	IMURAN	Imuran
	CYSPORIN	Cyclosporin
	CYTOXAN	Cytoxan
	IRRAD	Irradiation
	OTH_IMMU	Other immumosupression. This is a free text field with many different entries which require interpretation

Table XIV-9

Patient Annual Summary SAF

The Patient Annual Summary table of the database, like the Treatment History described earlier, is another way in which the data are summarized in order to make the large volume of event data more usable for analysis

The Patient Annual Summary table provides calendar year summaries of the treatments received by each patient. These summaries are the basis for many of the Reference Tables in this ADR. The Patient Annual Summary table has one row for each calendar year in which a patient was alive at any time on renal replacement therapy. Patients who died during the year are included with the end-of-year codes indicating death. The following information is included:

End of Year Status

End of year status variables indicate treatment modality, provider, state and ESRD Network of treatment, detailed and grouped modality, and

grouped modality using the “60-day rule.” These summaries are used for the prevalence tables in Section B and many of the tables in Section C of the ADR.

Activity During the Year

These variables include Counts of days and providers during the year in the following categories. All will be zero if the whole year was a lost to follow-up period.

- Days of ESRD
- Dialysis days total and after first transplant (used in computing transplant rates for first and subsequent after first transplant)
- Billed dialysis days
- Number of different providers and treatment modalities
- Primary treatment modality and provider for the year, and days with this provider and modality.

Mortality and Hospital Rate Variables

These are the variables needed for doing annual mortality analyses such as appear in Section D of the Reference Tables for this ADR. These variables include status on January 1, days at risk during the year, and whether the patient died during this period. A similar set of variables for doing prevalent hospitalization rate analyses such as appear in Section H of the Reference Tables.

References

Webb R, Bergsman A, Port FK. Diagnostic Codes In The USRDS Database. May 31, 1989. USRDS Technical Memo.