Chapter III

Structure and Flow of the USRDS Database

The USRDS database is the foundation upon which all other USRDS work rests. The previous chapter summarized the sources of data for the USRDS database. This chapter discusses the structure of the database and the flow of data into the database.

Structure of the Database

This section describes the structure of the database and the flows of data into the database. Figure III-1 shows the structure of the USRDS database and the flows of data to and from the database. Table III-1 and Table III-2 the Oracle table names and SAF file names for the components of the database. Table III-3 lists some additional secondary database tables. Cross-reference numbers are provided in these Figures and Tables as a guide.

Figure III-1 shows the structure of the database and the relationships among the tables in the database. The two root entities in this data structure are patients and facilities. The patient is the root of a hierarchy of entities, all of which are linked by the patient ID number. The links between patient data and facilities are show in Figure III-1.

The USRDS database exists in two forms: as an Oracle database and as the set of USRDS Standard Analysis Files (SAFs), which are described in the *Researcher's Guide to the USRDS Database*. Most elements of the database exist in both forms. The biggest exception is the detailed Medicare claim data, which exists only as SAFs. Table III-1 and Table III-2 list the principal tables in the Oracle database and shows the correspondence between the Oracle tables and the SAFs. The SAS Access for Oracle product is used to move the data between the Oracle tables and the SAS files. The Medicare ID numbers are replaced with the USRDS-assigned ID numbers in the SAFs so that individual patients cannot be identified in the files which the USRDS provides to other researchers. Table III-1 and Table III-2 list the Oracle table names and corresponding SAF file names and provides a brief description of each table or file. Reference numbers appear in Figure III-1, Figure III-2, Table III-1, and Table III-2 to provide a cross reference.

Table III-3 lists additional database table that define or provide labels, cross reference, and other information for the values of various codes which are used elsewhere in the database. These tables are not shown in Figure III-1.
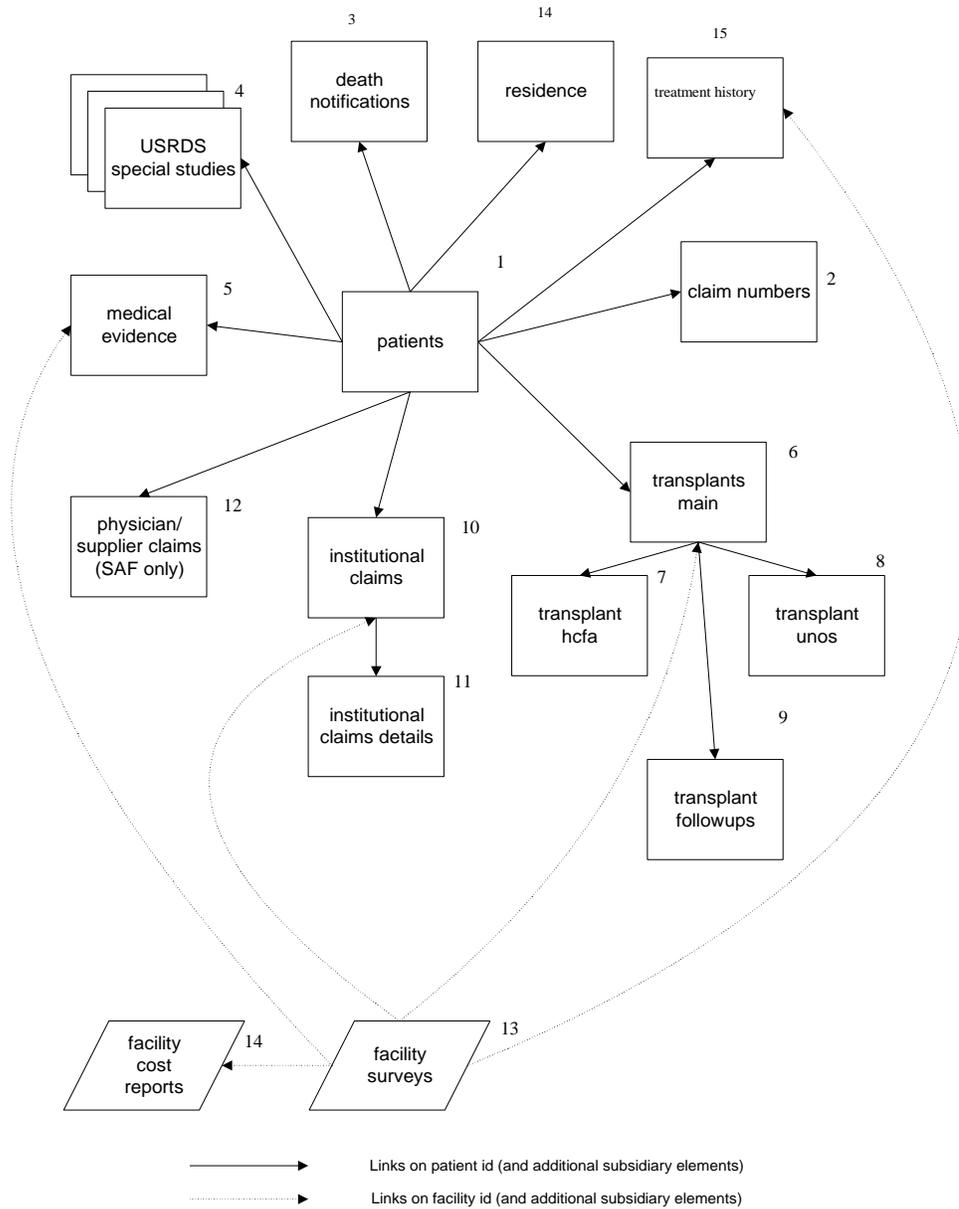
Structure of the USRDS Database

Links on patient id (and additional subsidiary elements)

Links on facility id (and additional subsidiary elements)

Figure III-1

Table III-1
Oracle Database Tables and Corresponding SAF Files
Where Data Source is HCFA

	Oracle Table Name	**SAF File Name**	**Contents**
2	TAB$CLAIM_NUMBERS		Cross reference between HCFA claim numbers and USRDS assigned patient ID numbers. A beneficiary may change Medicare ID over time. In this case, there will be multiple rows in this table for that patient. The SAF files that are made available to outside researchers contain only the USRDS ID as a patient identifier.
3	TAB$DEATH_NOTIFICATIONS	PATIENTS	Death Notifications from HCFA Form 2746. Patient death date come from this form and from Social Security records. Some deaths do not have a Death Notification form. The cause of death data from this form appear in the SAF.PATIENTS file.
13	TAB$FACILITY_SURVEYS	FACILITY	Facility Surveys Source is the HCFA ESRD Facility Survey, Form 2744
10	TAB$INSTITUTIONAL_CLAIMS	INCLAIM	Institutional Medicare Claims, Medicare Institutional Claims from HCFA Standard Analytical Files.
11		INDETAIL	Institutional Medicare Claims Details, Medicare Institutional Claims from HCFA Standard Analytical Files. See Researcher's Guide.
12		PSCLAIM	Physician/Supplier Medicare Claims. Because of the large volume of data in this file, it is maintained only as SAS files and is not brought into Oracle.
5	TAB$MEDICAL_EVIDENCE_REPORTS	MEDEVID, PATIENTS	Medical Evidence (HCFA 2728) The SAF includes only the new 2728 Form that went into use in April, 1995. The Oracle table includes all 2728 Forms. This is the source for the primary disease causing renal failure on the Patient table.
1	TAB$PATIENTS	PATIENTS	Patient Sources are the REBUS IDEN record, the Medicare Enrollment Database, and the 2728 and 2746 forms. Additional variables, such as date of first transplant, are computed from other data in the database.
15	TAB$RESIDENCE_PERIODS	RESIDENC	Residence This file attempts to reconstruct the patient's residence history, using multiple sources: Medicare Enrollment Database, REBUS, old PMMIS files, dialysis bills.
9	TAB$TRANSPLANT_FOLLOWUPS	TXFU	Transplant Followups. Source was the HCFA Transplant Followup Form. Starting with 1994, the source is UNOS.
7	TAB$TRANSPLANT_HCFA_DETAILS	TXPMMIS	Transplant HCFA details (pre 1994). Only a limited number of variables are kept in Oracle, with the remainder appearing only in the SAS file.
8	TAB$TRANSPLANT_UNOS	TXUNOS	Transplant UNOS details (starting with 1988). Only a limited number of variables are kept in Oracle, with the remainder appearing only in the SAS file.
6	TAB$TRANSPLANTS	TX	Transplants.. This is the master list of transplants. A limited set of variables are kept for each transplant. The transplant failure date computed by USRDS is kept on this file. Most transplant in this file will also appear in the HCFA or UNOS details files, or both.
17		TXWAIT	UNOS Transplant Waiting List
15	TAB$TREATMENT_HISTORY	RXHIST	Treatment History. This file is der
14		FCOSHOS FCOSIND	Facility cost reports filed with HCFA each year by all dialysis facilities
16		CPUS3R CPUS4R CPST4R	U.S. Census Population Counts for the U.S. and by State

Table III-2
USRDS Special Study Data Tables

4		ADQFACS	Case Mix Adequacy Special Study Facilities
4		ADQUACY	Case Mix Adequacy Special Study Patients
4		CAPD	CAPD Peritonitis Special Study
4		CASEMIXS	Case Mix Adequacy Special Study Patients
4		CMSFACS	Case Mix Adequacy Special Study Facilities
4		PEDGROW	Pediatric Growth and Development Special Study
4	TAB$DMMS_WAVE1_FACILITIES	DMMSFACS	DMMS Wave 1 Facility Form
4	TAB$DMMS_WAVE1_PATIENTS_CORE	DMMSWAV1	DMMS Wave 1 Patient Data
4		DMMSWAV2	DMMS Wave 2 Patient Data
4		DMMSFAC2	DMMS Wave 2 Facility Data
4		DMMSWAV3	DMMS Wave 3 & 4 Paitent Data
4		DMMSWAV4	DMMS Wave 2 & 4 Facility Data
4		DIALYZER	Characteristics of dialyzers and codes used in DMMS and Case Mix Adequacy

Table III-3
Oracle Code Definition Tables

Oracle Table	Contents
TAB$ESRD_FACILITIES	ESRD facility addresses
TAB$CENSUS_NETWORK	Census population estimates for ESRD Network areas
TAB$CENSUS_STATE	Census population estimates for states
TAB$CENSUS_US	Census population estimates for the US
TAB$COUNTIES	County codes: FIPS, SSA
TAB$DMMS_DIALYZER	Dialyzer codes and caracteristics
TAB$DMMS_WAVE1_DIALYZER	Dialyzer codes and caracteristics
TAB$ESRD_DEATH_CODES	Labels for cause of death codes
TAB$ESRD_DISEASE_CODES	Labels for codes for primary disease causing ESRD
TAB$ESRD_DISEASE_GROUPS	Groups primary disease codes into disease groups
TAB$ESRD_NETWORKS	Data about ESRD Network organizations
TAB$ICD9_DISEASE_CODES	Labels for ICD9 disease codes
TAB$MSA	Metropolitan Statistical Area codes
TAB$RACE_CODES	
TAB$SEX_CODES	
TAB$STATES	State codes: FIPS, SSA, Post Office
TAB$ZIPCODES	Zipcode geographic coordinates

Flow of Data

Figure VII-2 shows the flow of data to the USRDS from the various (primarily HCFA) sources. The Program Management and Medical Information System (PMMIS) of the HCFA End Stage Renal Disease Program always has been the primary source for the roster of patients included in the USRDS database and for much of the data about those patients. With the 1995 update, the USRDS began to get Medicare claims data for 1989 on directly from the HCFA Standard Analytical Files and Physician/Supplier files rather than from the PMMIS. With the 1997 update, transplant data for 1994 and later comes to the USRDS directly from tapes produced by UNOS. Figure 3 contains four copies of Figure 2, each with one of the principal data paths highlighted.
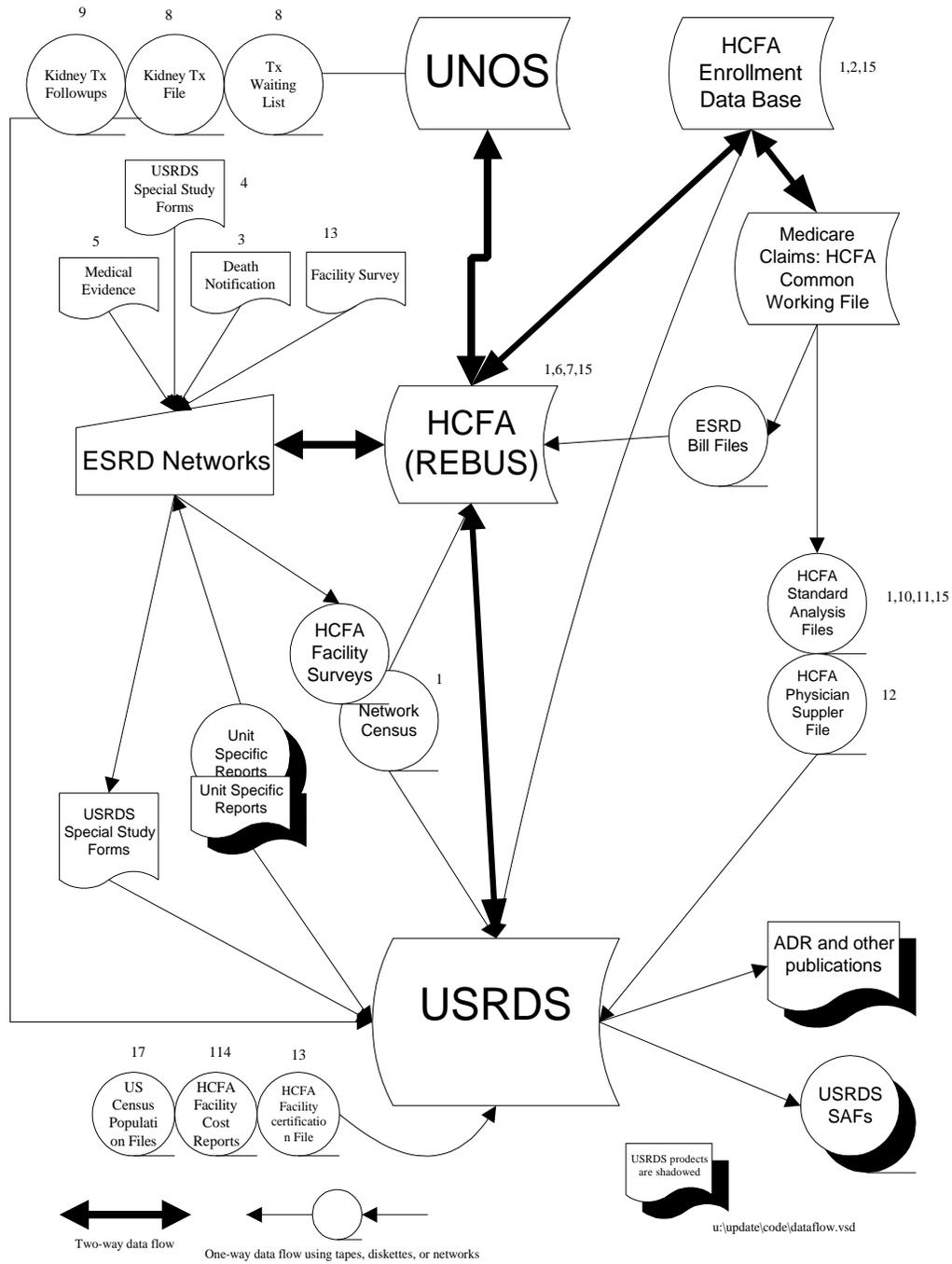
USRDS and Related Data Flows

Figure III-2

The Roster of Patients and Patient Characteristics

Most patients in the USRDS database come from the HCFA ESRD Program Medical and Management Information System (PMMIS). The current on-line front end system for PMMIS is the Renal Beneficiary and Utilization System (REBUS). Most of these patients are enrolled in Medicare and have data in the Medicare Enrollment Database (EDB). Some are not Medicare eligible and have no data in REBUS except the HCFA Form 2728, Medical Evidence Form. Some patients also enter the USRDS database because they have outpatient dialysis bills or a kidney transplant hospitalization in the Medicare claims but are not in REBUS or because they were non-Medicare patients in a USRDS Special Study. Reconciling these sets of patients is a continuing task.

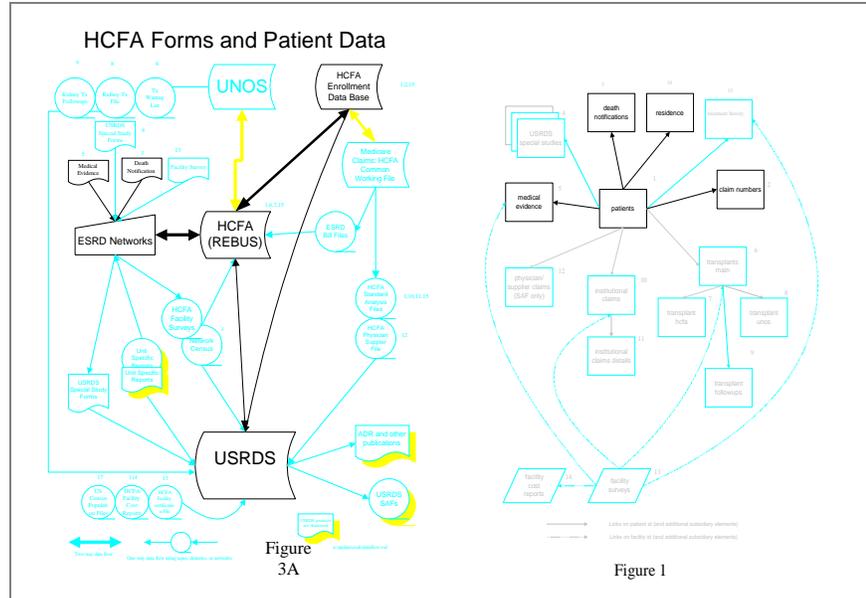Figure III-3 shows the paths by which data about patient characteristics enters the USRDS database.

Figure III-3

SAF.PATIENTS, SAF.RESIDENC, SAF.MEDEVID

These characteristics include dates of birth and death, sex, race, primary disease causing renal failure, and location residence. Non-Medicare patients who appear only in a USRDS Special Studies file may not have data for these items. Most of these items come to the USRDS database from the Medicare EDB or from the HCFA ESRD forms by way of REBUS. Some items, notably residence, the USRDS gets directly from REBUS.

USRDS Special Studies Data

Each year the USRDS conducts one or more Special Studies on a sample of approximately 5000 patients. The data for these studies are abstracted from medical records by staff of the 18 ESRD Networks or by staff of dialysis facilities. In either case the data collection forms pass through the ESRD Networks on their way to the USRDS database, as shown in Figure III-4.
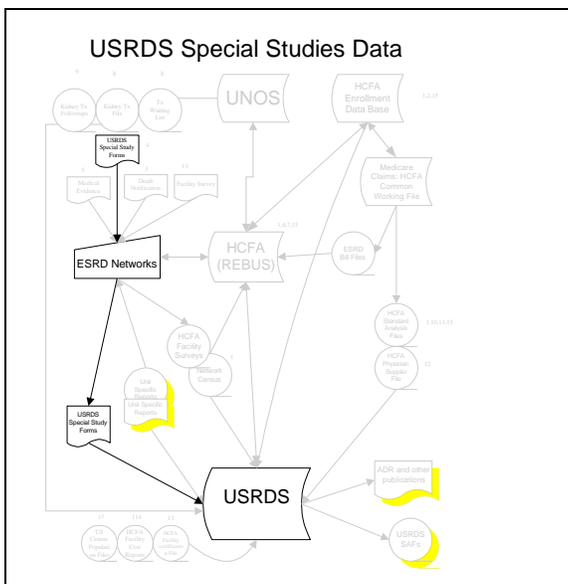

Figure III-4

USRDS Special Studies Data

Facility Data

Data from the HCFA Annual ESRD Facility Survey and from the HCFA Facility Cost Reports are obtained each year from HCFA and entered into the USRDS database.

Census Population Data

Census population data are needed for computing incidence and prevalence rates. Census population estimates by age, race, and gender are downloaded from the Census World Wide Web site each year. We use files of estimates by county in order to be able to split California into the two ESRD Network areas. These county estimates are then summed to the state, ESRD Network, and national levels. Estimates for the most recent year are generally not available by county. For this year we download estimates of total population by state and use those estimates to inflate the estimates by county from the previous year.

Transplant Data

From 1988 through 1993, data about renal transplantation and followup reports on renal transplants were collected by both HCFA and the United Network for Organ Sharing (UNOS). Starting in 1994, these two data collection efforts were consolidated under UNOS. The USRDS database contains data from both sources. In addition, the 2728 Medical Evidence Report and claims data can indicate the occurrence of a transplant. Since there are multiple sources of transplant data, the USRDS has decided to maintain these data in three separate files (see chapter 7). HCFA, UNOS and the Medical Evidence Reports are the source for the USRDS main transplant table,
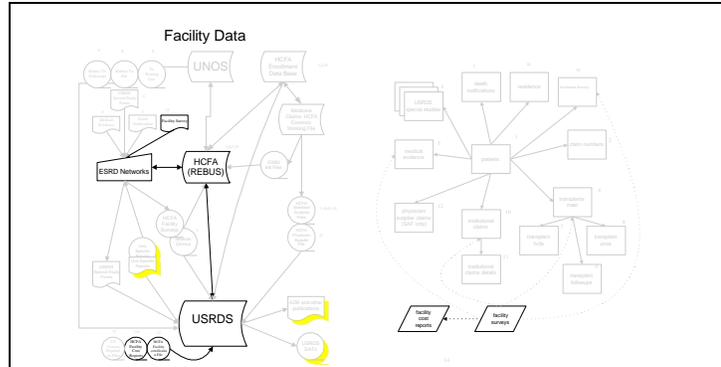

Figure III-5

SAF.FACILITY, SAF.FCOSHOS, SAF.FCOSIND

which is intended to contain at least minimal data for all transplants that we know about. The transplant details from the two sources of data collection, HCFA and UNOS are stored in two separate files. The USRDS supplies five SAFs for transplants. SAF.TX contains one record, with a small number of data items, for each transplant. This file also contains the transplant failure date computed by USRDS. SAF.TXWAIT includes the patient ID and the date of first entry onto the transplant waitlist. Both these files are included on the USRDS Core CD. SAF.TXPMMIS contains the transplant clinical
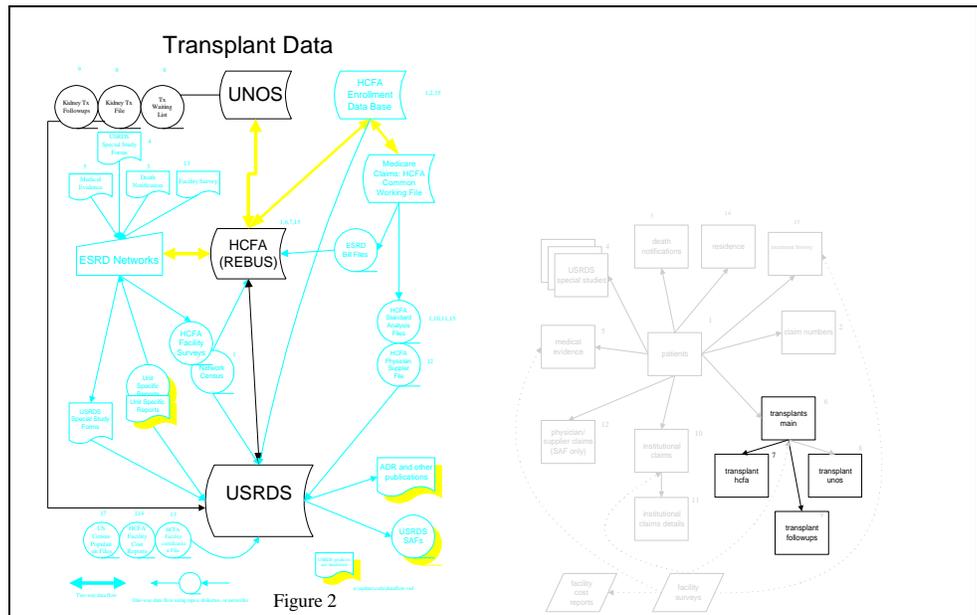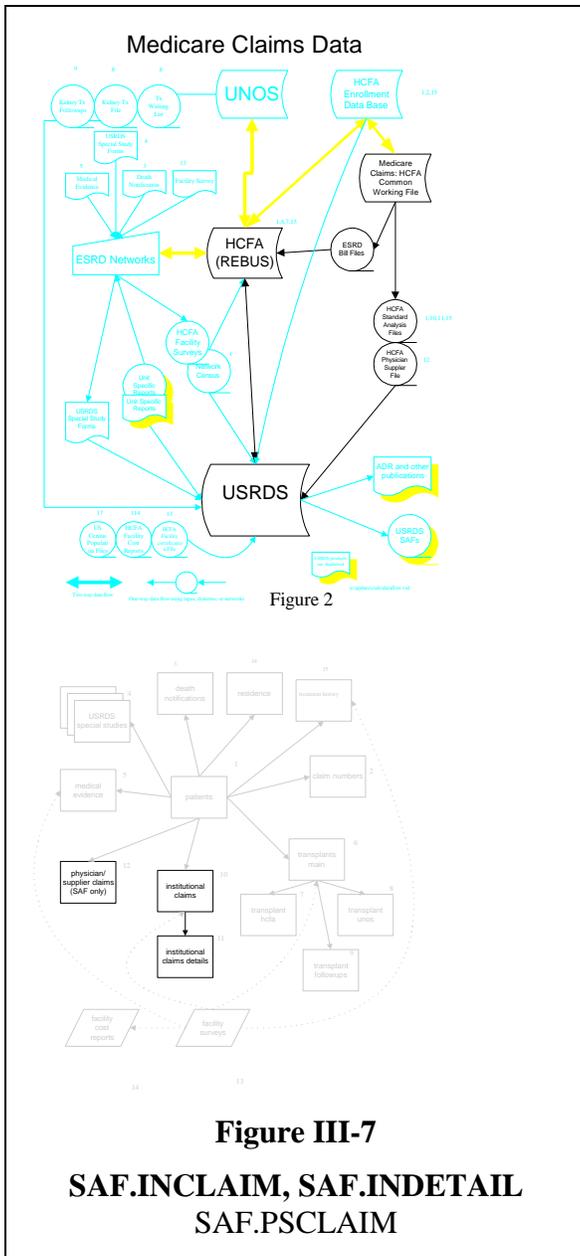

Figure III-6

SAF.TXMAIN, SAF.TXHCFA, SAF.TXUNOS

Medicare Claims Data

Figure 2

Figure III-7

SAF.INCLAIM, SAF.INDETAIL
SAF.PSCLAIM

details data from the HCFA forms. SAF.TXUNOS contains the UNOS clinical details data for collected on the UNOS forms. The researcher must decide which detailed source to use or how to reconcile the differences in the two sources.

SAF.TXUNOS contains the UNOS data for 1988 on. The researcher must decide which detailed source to use or how to reconcile the differences in the two sources.

Medicare Paid Claims Data

Data about Medicare paid claims are needed for economic analyses but also are the source of data

about the medical service used by ESRD patients. The flow of these data is shown in Figure III-7.

There are two types of Medicare claims: institutional and physician/supplier, as described in Chapter 3 of the *Researcher's Guide*. Physician/supplier claims make up about 20 percent of the claims but about 80 percent of the dollars paid.

HCFA freezes its claims analysis files for a given year with data processed through June of the following year. HCFA considers these files to be 97 to 98 percent complete. However, they are less complete for the last few months of the year. Some researchers have handled this incompleteness by analyzing only data through October or by including month of the year as a variable in a regression analysis.

Claims are identified by year by HCFA and by the USRDS based on the through date of service, not the from date of service. Generally this is only an issue with hospital inpatient and skilled nursing facility (SNF) claims, which frequently do span the end of a year.

The USRDS database has physician/supplier claims data from 1991 on.. The claims for patients in the USRDS database are extracted by the HCFA MANRLINE system in September or October of each year for the two preceeding years. For example, at the time of the 1997 database update in the Fall of 1997, the USRDS database has physician/supplier data through 1995. The MANRLINE runs in September or October of 1997 get claims for 1995 and 1996. The 1996 claims are added to the database, and the 1995 claims in the database are replaced with the newly extracted 1995 claims. The re-extraction of the 1995 claims is intended to catch claims for patients who had claims in 1995 but were not identified as ESRD patients and added to the USRDS database until late 1996 or 1997.

The USRDS database has data, including dollar amounts, about all types of institutional claims for 1989 and data excluding dollar amounts for inpatient stays and outpatient dialysis for years before 1989. The source for the data for 1989 on is the HCFA Standard Analytical Files (HCFASAFs). For years before 1989 the source is REBUS. The REBUS data about outpatient dialysis consists of summaries by calendar quarter and provider, so that some changes in treatment modality may be lost.

For 1989 on, the REBUS data are used to supplement the data from the HCFASAFs. If an inpatient stay from REBUS does not match with a

record from the HCFASAFs, then the REBUS record is added to the database. If there are no HCFASAF outpatient dialysis records in a quarter for which there is a REBUS Quarterly Dialysis record, then the REBUS record is added to the database. The primary HCFA claims data in the HCFA Common Working File are the source for both REBUS and the HCFASAFs. Because the HCFASAFs are frozen for a given year in June of the following year, some late claims may be missed by the HCFASAFs. Since REBUS looks at all claims as they are filed for patients in its roster, it will get these late claims. However, REBUS will miss claims which were processed before REBUS identified the patient as an ESRD patient.

The USRDS database update in 1995 extracted institutional claims data for 1989 through 1993. The 1996 update re-extracted data for 1989-1993 and added 1994 and 1995. The 1997 update re-extracted data for 1989 through 1996. For the 1998 update, we plan to retain the existing data for 1989 through 1992, re-extract data for 1993 through 1996, and add data for 1997. In the 1996 and 1997 updates, data were re-extracting to add additional variables from the HCFASAFs. We now believe that we have stable specifications for these files so that it will no longer be necessary to extract all years since 1989. We will continue to re-extract data for the most recent four years to get complete data for patients who are newly added to the database. Please recall that there sometimes is a delay in getting the patient into the USRDS database after the patient has reached ESRD. Re-extracting also allows us to get up to four years of pre-ESRD Medicare claims.

Computer System

The computer systems used by the USRDS CC has grown and evolved dramatically during the first nine years of the USRDS. During the first five-year contract, the database was maintained on Digital Equipment VAX computers, first a 3500 and later a 4300, operating as part of a VAX cluster. At first the project also used one 386 PC, in addition to the VAX terminals on each person's desk. By the end of the project all staff were using PCs. When the CC moved to Michigan, the database again was housed on a VAX 4300, with plans to evaluate other configurations as the project progressed. Staff members used 486 PCs running Windows for Workgroups on a Novelle network. In 1996 the database was moved from the VAX to Intel-based Servers, and the entire system was converted to

Windows NT. At the end of the first contract, the database occupied about 3.5 gigabytes (billions of characters) of disk storage. The KECC computer system servers now have about 140 Gigabytes of disk storage.

USRDS CC will continue to use the computer system of the Kidney Epidemiology and Cost Center (KECC), which is described in detail in section VIII. This system consists of twenty some Pentium workstations networked with five Pentium servers, all running Windows NT. The standard workstation is a Pentium Pro 200 with 64 megabytes of memory and 1 to 3 gigabytes of disk storage. About half of the workstations are connected to the servers by high speed "100 base T" networking. The combination of powerful workstations and fast networking allows large tasks to be run efficiently on the workstations.

Each of the five server computers is tuned for a specific type of work, including running the Oracle database, running SAS and providing file service for large database files, providing general office file service, and handling e-mail and communications. Various tape drives, printers, and modems are connected to the servers. One tape drive handles the IBM 3480 tape cartridges used by the HCFA computer, allowing efficient transfer of large amounts of data between the HCFA and USRDS computers.

The USRDS system was designed to be transportable to a new computer or to a different type of computer system. It utilizes widely used off-the-shelf software and minimizes the use of hardware- or software-dependent features.

The primary software used to store the database and to perform statistical analysis will be unchanged. The database is stored and maintained using the Oracle relational database management system. SAS is used for statistical analysis. The SAS ACCESS for Oracle product is used to move data in both directions between Oracle and SAS. FORTRAN is used for a few tasks, such as creating the Treatment History file based on the detailed data for each patient and processing the large volumes of claims data for Section K of the ADR Reference Tables. We expect to convert these programs from FORTRAN toC.

Documentation

Over the first nine years of the USRDS, three documents have evolved to meet the varying needs of different constituencies: Chapters in the ADR, the

Researcher's Guide to the USRDS Database and the *Operations and Forms Manual.*

Chapter I of the ADR, "The USRDS and Its Products", describes the Standard Analysis Files (SAFs) and how a researcher can obtain them. Thus it is the basic documentation of the USRDS database at it is available outside of the USRDS CC. Chapter I of the ADR also is used as Chapter I of the *Researcher's Guide* and is the basis for Chapter III, "Products of the USRDS" in this proposal. The last chapter of the ADR, "Analytical Methods" provides additional information about the sources of data for the USRDS database and about various technical issues. These chapters are available on the USRDS Web site and on the USRDS 1997 ADR CD.

The *Researcher's Guide* is the documentation supplied with the USRDS Standard Analysis Files (SAFs). It also is distributed to researchers who inquire about USRDS data files. Approximately 500 copies of the *Researcher's Guide* have been distributed over the four years of the current contract. It is available also on the USRDS Web site and on the USRDS 1997 ADR CD.

The *Researcher's Guide* has evolved to become the primary actively used documentation of the USRDS database. The *Operations and Forms Manual*, on the other hand, is the primary reference documentation for the USRDS database. Whereas hundreds of copies of the *Researcher's Guide* have been distributed, only about 5 copies of the *Manaual* are maintained.

The *Operations and Forms Manual* occupies eight loose- leaf volumes and contains approximately 1800 pages (all page numbers are approximate). The volumes of the *Operations and Forms Manua* are summarized below:

Book 1: Administrative Manual, 50 pp.

This book provides a copy of the USRDS organization chart; descriptions of how the various components of the organization are structured; copies of the policy statements on roles and responsibilities, and data access, authorship, publications and presentations; and the current directory of USRDS organizational participants.

Book 2: Operations Manual, 400 pp.

This was designed as a "stand-alone" document that describes the data sources; database structure; data dictionary; data import procedures; data security, integrity, and validation; ADR Reference Table procedures; and supporting technical notes for the entire data system. Many of the sections of this manual have been replaced by other documents.

Book 3: HCFA ESRD Forms, 250 pp

Provides copies of the forms and instructions for medical evidence information, death notification, transplant information, ESRD facility survey information, and followup information; and Appendix A -- miscellaneous pages from instruction booklets, Appendix B -- relationships between the forms and data sources to the USRDS, and Appendix C -- code lists.

Book 4: HCFA Billing Forms, 300 pp

Provides copies of the forms and instructions for UNI bills, pre-UNI bills billing information, and ESRD facility cost report information.

Book 5: File Layouts, 500 pp

Provides file layouts for PMMIS, transplant followup, facility survey and master provider file, facility survey (pre-1982), BMAD, MEDPAR, UNI bill inpatient, UNI bill outpatient, UNI bill data element glossary/codes, payment record, master identification record, and county census.

Book 6: USRDS Special Studies Forms, 200 pp

Provides copies of the forms and instructions for data collection for USRDS special studies.

Book 7: HCFA Standard Analytical Files and Physician/Supplier Files, 200 pp

Copies of the file layouts for the HCFA Standard Analytical Files and Physician/Supplier files.

Book 8: USRDS Oracle Data Dictionary

The Oracle Designer 2000 product is used to define and document the Oracle database structure. Designer 2000 maintains all of the database description elements in tables in the Oracle database. Thus if the database were to be moved to another computer system at another organization, the database description and documentation would be transferred along with the actual data. The reports listed below are produced using Designer 2000 as documentation of the Oracle database. These reports appear as a separate document entitled *USRDS Oracle Database Description Reports*. The table below lists the contents of these reports.

Report Name	Description
Table Definition	For each database table, this report lists the columns, the primary key, all foreign keys, and all indexes. This is the most comprehensive report.
Tables and Primary Key Derivations	This table presents some of the information from the Table Definition report in a more compact format. It lists the columns in each table and shows which columns compose the primary key and which columns are foreign keys pointing to other tables.
View Definition Report	Describes views which have been defined.