S. Tan
May 3, 2010

<h1 style="text-align:center">Dataset Integrity Check (DSIC) for the
Comprehensive Dialysis Study (CDS) Data Files</h1>

<p style="text-align:center">Reference paper: Kutner NG, et al. The Comprehensive Dialysis Study (CDS):
A USRDS Special Study
Clin J Am Soc Nephrol 4 [2009]: 645-650.</p>

The Comprehensive Dialysis Study (CDS) is a study in which data on patient demographics, contact information, treatment, laboratory values, quality of life (QOL) survey interviews, and nutrition survey interviews were collected for U.S. dialysis patients starting treatment 2005-2007 at 335 randomly selected dialysis facilities. Data from the CDS were submitted to the NIDDK Data Repository in October, 2009.

As a partial check of the integrity of the CDS data files archived in the Repository, a dataset integrity check (DSIC) was performed to verify that selected published results from the CDS study could be reproduced using the archived datasets. This DSIC (results of which are described below) consists of a small number of analyses performed to duplicate published results reported by Kutner NG, et al. [1] in Clinical Journal of the American Society of Nephrology in 2009.

The intent of this DSIC is to provide confidence that the data distributed by the NIDDK repository is a true copy of the study data. Our intent is not to assess the integrity of the statistical analyses reported by study investigators. As with all statistical analyses of complex datasets, complete replication of a set of statistical results should not be expected on a first (or second) exercise in secondary analysis. This occurs for a number of reasons including differences in the handling of missing data, restrictions on cases included in samples for a particular analysis, software coding used to define complex variables, etc. Experience suggests that most discrepancies can ordinarily be resolved by consultation with the study data coordinating center (DCC), however this process is labor-intensive for both DCC and Repository staff. It is thus not our policy to resolve every discrepancy that is observed in an integrity check. Specifically, we do not attempt to resolve minor or inconsequential discrepancies with published results or discrepancies that involve complex analyses, unless staff of the NIDDK Repository suspect that the observed discrepancy suggests that the dataset may have been corrupted in storage, transmission, or processing by repository staff. We do, however, document in footnotes to the integrity check those instances in which our secondary analyses produced results that were not fully consistent with those reported in the target publication.

Archived Dataset Contents. Participants of the CDS were selected via a monthly Renal Management Information System (REMIS)/Standard Information Management System (SIMS) database custom extract process using the following criteria:

- Incident dialysis patients (age >= 18) who had survived at least two months,
- No prior transplantation,
- Dialyze at one of the pre-determined 335 facilities

11,292 patients were selected over a period of two years. Of these, 1677 consented and participated in surveys or had lab assays performed. The CDS data files contain the data for these 1677 patients. Initially, the following four *SAS version 7-8-9* data files were submitted, comprising the "raw" data files associated with the CDS:

(1) CDS Patient file
This file contains one record for each of 1677 participants. Each record includes a unique identifier "USRDS_ID" (which can also be used to link CDS files with other USRDS data files), demographic data, and a patient category indicator ("PAT_CAT_BL") which indicates patient's baseline survey and lab participation status. DCC documentation states that 1279 participated in the QOL-only survey; 364 participated in the QOL and nutrition survey (231 with lab data, 133 without lab data); and 34 patients did not participate in either survey, but have lab data[a].

(2) CDS QOL Baseline file
This file contains the Baseline QOL survey data from QOL only and QOL/Nutrition participants (1643 patient records).

(3) 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. Here, we note there are 360 patient records in the archived dataset, though the DCC states 364 participated in the QOL/nutrition survey[b].

(4) CDS Lab file
This file contains up to 5 sets of lab data (976 total patient-visit records) for 265 participants. Of the 265, 231 had a merged participation status of "QOL/Nutrition Participant – Survey With Lab", consistent with documentation. We note that 227 participants with Lab data had matching records in the Food Baseline file, contrary to the documented 231[c].

In addition to these four files, a file containing the race indicator from linked USRDS files (CMS form 2728) was subsequently submitted so published racial distributions may be replicated using archived data.

The DCC also submitted electronic files of data collection forms associated with the QOL Baseline and the Food Baseline datasets, as well as a data dictionary describing the contents of the dataset files. *SAS* variables were labeled with a detailed description including form name and item number. *SAS* formats (value labels) were provided as a separate *SAS v8.02* formats file (<formats.sas7bcat>). Missing data values were qualified with labels such as "Don't Know", "Refused", "Missing", or "N/A".

---

[a] In archived files, it was found that Four IDs with QOL and Lab data did NOT have records in the Food Baseline file, despite being classified as PAT_CAT_BL=2 ("QOL/Nutrition Participant - Survey With Lab"). These IDs are: 2341351, 2354530, 2464373, 2514115. The DCC confirmed these IDs do not have food records; thus, the revised counts should be: among 1677 patients, **1283** participated in QOL-only survey; **360** participated in QOL and the nutrition survey (**227** with lab data, 133 without lab data); 34 patients did not participate in either survey but have lab data.
[b] See footnote (a). The same four participants account for this discrepancy.
[c] See footnote (a).

DSIC Analysis Methods. The DCC informed the Repository that results of analyses reported in the Kutner *et al.* [2009] publication included 3 participants subsequently found to be ineligible. Datasets submitted to the Repository do NOT include these 3 ineligible participants. Thus, results of analyses of archived data are expected to differ slightly from published results. The publication was primarily descriptive, outlining the aims and methods of the CDS, along with limited baseline descriptive statistics. For purposes of this DSIC, all available descriptive statistics (means and frequencies) were attempted for replication using archived datasets.

The lab file contained multiple measures per participant, often drawn over the course of a year or more. For replication of published baseline analysis, the earliest available lab measure was retained for each participant.

DSIC Results: Demographics. Distributions of age (18-44, 45-64, 65-74, 75+), gender, African-American race, ethnicity (Hispanic), and baseline modality (hemodialysis) closely matched published breakdowns. Differences may be attributed to the exclusion of three ineligible participants from archived datasets. Racial distributions using the CDS race variable did not closely replicate published results. In archived data, 55% of all participants self-identify as "White" and 14% self-identify as "Other", while the publication reports that 69% of all participants self-identify as "White" and 0% as "Other". The DCC provided the explanation that racial frequencies from linked USRDS files (CMS Form 2728) were reported, instead of from the CDS. When archived data were analyzed using the linked USRDS race variable, racial distributions matched published breakdowns [Table 1].[d]

---

[d] The DCC provided the linked USRDS race variable, along with the following explanation: In the CDS, the question on race ("How do you describe yourself") *preceded* the question on ethnicity ("Do you describe yourself as Hispanic or Latino"). Thus, it is likely that many Hispanics identified themselves as "Other (race)", as they hadn't yet had the opportunity to self-identify their ethnicity. Due to this issue, the publication reports race frequencies from linked USRDS (CMS Form 2728) instead of from the CDS.

**Table 1. Sociodemographic characteristics of CDS participants
(based on Kutner NG, *et al.* [2009], p.649, Table 2).**

	All participants				Nutrition Study Subset		
	Archived	Published	Difference		Archived	Published	Difference
	n = 1643	N = 1646	n=3[e]		n = 360	n = 361	n=1[f]
Age at ESRD enrollment (yr; mean [SD])	60.4 (14.2)	60.2 (14.2)	-0.2 (0.0)		61.5 (13.7)	61.2 (13.7)	-0.3 (0.0)
Age (%)							
18 to 44	13.6	13.9	0.3		11.4	11.6	0.2
45 to 64	47.0	47.8	0.8		46.1	47.1	1
65 to 74	22.8	22.1	-0.7		24.4	23.6	-0.8
75+	16.6	16.2	-0.4		18.1	17.7	-0.4
Male (%)	55.0	55.1	0.1		51.1	51.3	0.2
Race[g]							
Native American	0.8	0.9	0.1		0.6	0.6	0.0
Asian	1.4	1.4	0.0		3.3	3.3	0.0
Black	28.0	28.1	0.1		26.1	26.1	0.0
White	68.9	68.8	-0.1		68.6	68.6	0.0
Pacific Islander / Other	0.8	0.8	0.0		1.4	1.4	0.0
Hispanic	16.5	16.5	0		15.0	15.0	0
HD (%)	89.6	89.7	0.1		85.8	85.9	0.1

---

[e] The CDS paper was published before it was discovered that 3 participants were not eligible and should not have been included. The analysis dataset submitted for archival does NOT include these 3 participants.
[f] See above footnote (e). It is probable that one of the 3 ineligible participants also participated in the food survey.
[g] These are racial breakdowns using the linked USRDS race variable (from CMS Form 2728), subsequently provided so published racial distributions may be replicated. See footnote (d).

DSIC Results: Lab Values. The publication reported 269 nutrition subset patients provided serum samples. Analysis of archived lab data indicated 227 participants with nutrition data provided serum samples.[h] Distributions of baseline lab measures for these 227 participants closely matched published distributions. [Table 2]

Table 2. Distributions[i] of lab measures for participants of the CDS Nutrition sub-study (based on Kutner NG, *et al.* [2009], p.649, text).

	Nutrition Study Subset, with Lab Measures		
	Archived	Published	Difference
	n = 227	n = 269	**n=42**
Albumin (g/dl[j], mean [SD])	3.47 (0.47)	3.47 (0.47)	**0.0 (0.0)**
Prealbumin (mg/dl, mean [SD])	31.5 (9.3)	31.5 (9.6)	**0.0 (0.3)**
A-1-AG (mg/dl, mean [SD])	116 (35)	115 (36)	**-1 (1)**
C-Reactive Protein (mg/L, median [IQR])	7.4 (3.95, 12.5)	7.4 (3.96, 12.7)	**0.0 (0.2)**

DSIC Results: QOL, Food Intake. No statistics on the various QOL and food intake measures (from the QOL Baseline and Food Baseline datasets, respectively) were reported in the publication, and therefore are not included in the current DSIC.

References

[1] Nancy KG, Johansen KL, Kaysen GA, Pederson S, Chen S-C, Agodoa LY, Eggers PW, Chertow GM. The Comprehensive Dialysis Study (CDS): A USRDS Special Study. Clin J Am Soc Nephrol 4 [2009]: 645-50.

---

[h] The published number "269" is not consistent with DCC data file documentation, which states 231 QOL/Nutrition participants have lab data available. The DCC confirmed via E-mail that the wording needed to be changed: 265 of CDS patients provided lab data, of which 227 completed the nutrition survey.

[i] Distributions are from the first draw for each participant (in the case of multiple draws). The DCC confirmed the data were analyzed this way.

[j] Archived variable for serum albumin is labeled "g/L", though the range of the variable indicates units are g/dl, as published. The DCC confirmed the units are g/dl.

Attachment 1

**"The full text of the article referenced will be provided to approved data requestors
along with the data archived."**

Kutner NG, Johansen KL, Kaysen GA, Pederson S, Chen SC, Agodoa LY, Eggers PW, Chertow GM.
The Comprehensive Dialysis Study (CDS): A USRDS Special Study.
Clinical Journal of the American Society of Nephrology 4: 645-650 [2009]

Attachment 2

SAS 9.1 Log
for programming code submitted
for the replication of results
in Table 2 and Text in
Kutner NG, et al. [2009]

NOTE: Copyright (c) 2002-2003 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) 9.1 (TS1M3)
      Licensed to RTI INTL MAIN, Site 0047670001.
NOTE: This session is executing on the XP_PRO platform.

NOTE: SAS 9.1.3 Service Pack 4

NOTE: SAS initialization used:
      real time   1.50 seconds
      cpu time   0.42 seconds

```
1           ************************************************************
2           * CDS DSIC.sas                                            *
3           * Purpose:  to perform Data Set Integrity Analyses        *
4           * on CDS study legacy dataset                             *
5           * comparison study paper: Kutner NG, et.al, Clin J Am Soc Nephrol, Vol 4
[2009] *
6           * Programmed by: S. Tan, NIDDK Repository                 *
7           ***********************************************************;
8           options ps=55 ls=78 nonumber formchar='|----|+\---+=|-^<>*' mprint;
9
10          libname cds
10        ! '\\rcdubuntu01.rtp.rti.org\niddk\03_Data_And_Tools\Studies\CDS\offi
10        ! cial archive\Data';
NOTE: Libref CDS was successfully assigned as follows:
      Engine:        V9
      Physical Name:
      \\rcdubuntu01.rtp.rti.org\niddk\03_Data_And_Tools\Studies\CDS\official
      archive\Data
11          libname library
11        ! '\\rcdubuntu01.rtp.rti.org\niddk\03_Data_And_Tools\Studies\CDS\offi
11        ! cial archive\Data';
NOTE: Libname LIBRARY refers to the same physical library as CDS.
NOTE: Libref LIBRARY was successfully assigned as follows:
      Engine:        V9
      Physical Name:
      \\rcdubuntu01.rtp.rti.org\niddk\03_Data_And_Tools\Studies\CDS\official
      archive\Data
12          libname addendum
12        ! '\\rcdubuntu01.rtp.rti.org\niddk\03_Data_And_Tools\Studies\CDS\DCC_
12        ! Delivery\20100106-USRDS-race';
NOTE: Libref ADDENDUM was successfully assigned as follows:
      Engine:        V9
      Physical Name:
      \\rcdubuntu01.rtp.rti.org\niddk\03_Data_And_Tools\Studies\CDS\DCC_Delive
```

```
      ry\20100106-USRDS-race
13
14         proc format;
14      !                value agecatf 1='18-44' 2='45-64' 3='65-74' 4='75+';
NOTE: Format AGECATF has been output.
14      !                                                                      r
14      ! un;

NOTE: PROCEDURE FORMAT used (Total process time):
      real time           0.12 seconds
      cpu time            0.00 seconds


15
16         data patient; set cds.cds_saf_patient;
17            age=(ENROLLMENT_DATE-dob)/365.25;
18            ageyr=floor(age);
19            agecat=1+(ageyr>44)+(ageyr>64)+(ageyr>74); format agecat agecatf.
19      ! ;
20            where PAT_CAT_BL^=4; run;

NOTE: There were 1643 observations read from the data set CDS.CDS_SAF_PATIENT.
      WHERE PAT_CAT_BL not = 4;
NOTE: The data set WORK.PATIENT has 1643 observations and 20 variables.
NOTE: DATA statement used (Total process time):
      real time           1.28 seconds
      cpu time            0.09 seconds


21            proc sort; by usrds_id;
22
23            * additional race indicator from USRDS CMS form 2728 *;
24            * submitted by DCC 02-08-10 *;
25            * see DSIC documentation *;

NOTE: There were 1643 observations read from the data set WORK.PATIENT.
NOTE: The data set WORK.PATIENT has 1643 observations and 20 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.07 seconds
      cpu time            0.01 seconds


26            data raceadd; set addendum.cds_race2728;

NOTE: There were 1677 observations read from the data set
      ADDENDUM.CDS_RACE2728.
NOTE: The data set WORK.RACEADD has 1677 observations and 2 variables.
NOTE: DATA statement used (Total process time):
      real time           0.26 seconds
      cpu time            0.00 seconds


27            proc sort; by usrds_id;
```

28


```
NOTE: There were 1677 observations read from the data set WORK.RACEADD.
NOTE: The data set WORK.RACEADD has 1677 observations and 2 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.03 seconds
      cpu time            0.01 seconds


29         data patient; merge patient(in=in1 rename=(race=race_cds)) raceadd;
30            by usrds_id; if in1;
31            title Replicate Table 2, column 1;

NOTE: There were 1643 observations read from the data set WORK.PATIENT.
NOTE: There were 1677 observations read from the data set WORK.RACEADD.
NOTE: The data set WORK.PATIENT has 1643 observations and 21 variables.
NOTE: DATA statement used (Total process time):
      real time           0.75 seconds
      cpu time            0.01 seconds


32            proc means data=patient; var age; run;

NOTE: There were 1643 observations read from the data set WORK.PATIENT.
NOTE: The PROCEDURE MEANS printed page 1.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.10 seconds
      cpu time            0.01 seconds


33            proc freq data=patient; tables agecat gender RACE_cds race2728
33       ! ethn  TREATMO_BL; run;

NOTE: There were 1643 observations read from the data set WORK.PATIENT.
NOTE: The PROCEDURE FREQ printed pages 2-3.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.51 seconds
      cpu time            0.06 seconds


34            * frequencies match published #s closely except for distns of
34       ! CDS race:  white, other *;
35            * per DCC, need to use race2728 (from USRDS CMS form 2728)
35       ! instead *;
36
37         data cds_saf_lab; set cds.cds_saf_lab;

NOTE: There were 976 observations read from the data set CDS.CDS_SAF_LAB.
NOTE: The data set WORK.CDS_SAF_LAB has 976 observations and 12 variables.
NOTE: DATA statement used (Total process time):
      real time           0.29 seconds
      cpu time            0.00 seconds
```

```
38            proc sort; by usrds_id specimen_date;
```

NOTE: There were 976 observations read from the data set WORK.CDS_SAF_LAB.
NOTE: The data set WORK.CDS_SAF_LAB has 976 observations and 12 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time   0.04 seconds
      cpu time   0.00 seconds


```
39          data firstlab; set cds_saf_lab; by usrds_id specimen_date;
40             if first.usrds_id; run;
```

NOTE: There were 976 observations read from the data set WORK.CDS_SAF_LAB.
NOTE: The data set WORK.FIRSTLAB has 265 observations and 12 variables.
NOTE: DATA statement used (Total process time):
      real time   0.00 seconds
      cpu time   0.00 seconds


```
41            * to just get baseline measures, assuming the earliest measure is
41        !  the baseline measure *;
42            * 01-08-10:  DCC confirmed the analysis was performed on the
42        ! first draw only *;
43            * n=265 with lab data *;
44
45          data food_base; set cds.cds_saf_food_baseline;
```

NOTE: There were 360 observations read from the data set
      CDS.CDS_SAF_FOOD_BASELINE.
NOTE: The data set WORK.FOOD_BASE has 360 observations and 274 variables.
NOTE: DATA statement used (Total process time):
      real time   4.39 seconds
      cpu time   0.21 seconds


```
46            proc sort; by usrds_id; run;
```

NOTE: There were 360 observations read from the data set WORK.FOOD_BASE.
NOTE: The data set WORK.FOOD_BASE has 360 observations and 274 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time   0.07 seconds
      cpu time   0.00 seconds


```
47          data pat_food_base; merge patient(in=in1) firstlab(in=in2)
47        ! food_base(in=in3) ;
48             by usrds_id; if in1;
49             if in2 then lab=1; else lab=0;
50             if in3 then food=1; else food=0;
51             medcrp=median(crp1,crp2);
52             meanalb=mean(alb1,alb2);
53             meanpab=mean(pab1,pab2);
```

```
54            meanaag=mean(aag1,aag2);
55             label meanalb='Lab: Mean Albumin Result1 (g/dL)' /* note DCC has
55       ! confirmed Albumin units are
56                                g/dL despite contrary indications in
56       ! dataset and pub */
57                 meanpab='Lab: Mean Prealbumin Result1 (mg/dL)'
58                 meanaag='Lab: Mean Alpha-1-acid glycoprotein (mg/dL)'
59                 medcrp='Lab: Median C-Reactive Protein (mg/dL)'; run;
```

```
NOTE: Character values have been converted to numeric
      values at the places given by: (Line):(Column).
      53:16   53:21   54:16   54:21
NOTE: Missing values were generated as a result of performing an operation on
      missing values.
      Each place is given by: (Number of times) at (Line):(Column).
      1412 at 52:11   1412 at 53:11   1412 at 54:11
NOTE: There were 1643 observations read from the data set WORK.PATIENT.
NOTE: There were 265 observations read from the data set WORK.FIRSTLAB.
NOTE: There were 360 observations read from the data set WORK.FOOD_BASE.
NOTE: The data set WORK.PAT_FOOD_BASE has 1643 observations and 309 variables.
NOTE: DATA statement used (Total process time):
      real time          3.78 seconds
      cpu time           0.04 seconds
```

```
60            title Replicate Table 2, column 2;
61          proc means data=pat_food_base; var age; where Food=1; run;
```

```
NOTE: There were 360 observations read from the data set WORK.PAT_FOOD_BASE.
      WHERE Food=1;
NOTE: The PROCEDURE MEANS printed page 4.
NOTE: PROCEDURE MEANS used (Total process time):
      real time          0.01 seconds
      cpu time           0.01 seconds
```

```
62          proc freq data=pat_food_base; tables agecat gender RACE_cds
62       ! race2728 ethn TREATMO_BL; where Food=1; run;
```

```
NOTE: There were 360 observations read from the data set WORK.PAT_FOOD_BASE.
      WHERE Food=1;
NOTE: The PROCEDURE FREQ printed pages 5-6.
NOTE: PROCEDURE FREQ used (Total process time):
      real time          0.51 seconds
      cpu time           0.03 seconds
```

```
63              * frequencies match published #s closely except for distns of
63       ! CDS race:  white, other *;
64              * per DCC, need to use race2728 (from USRDS CMS form 2728)
64       ! instead *;
65
66            title "Replicate descriptive stats reported in text on p.649 of
```

```
66          ! publication";
67            proc means data=pat_food_base; var meanalb meanpab meanaag; where
67          ! Food=1; run;
```

```
NOTE: There were 360 observations read from the data set WORK.PAT_FOOD_BASE.
      WHERE Food=1;
NOTE: The PROCEDURE MEANS printed page 7.
NOTE: PROCEDURE MEANS used (Total process time):
      real time            0.01 seconds
      cpu time             0.01 seconds
```

```
68            proc univariate data=pat_food_base; var medcrp; where Food=1; run;
```

```
NOTE: The PROCEDURE UNIVARIATE printed pages 8-9.
NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time            0.06 seconds
      cpu time             0.01 seconds
```

```
69
70
71
72
73
```

```
NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414
NOTE: The SAS System used:
      real time           14.85 seconds
      cpu time             1.01 seconds
```

Attachment 3

**SAS 9.1 Output
for programming code submitted
for the replication of results
in Table 2 and Text in
Kutner NG, et al. [2009]**

19:10 Thursday, May 6, 2010

The MEANS Procedure

Analysis Variable : age

N	Mean	Std Dev	Minimum	Maximum
1643	60.3731576	14.1672359	19.3675565	94.2806297

The FREQ Procedure

agecat	Frequency	Percent	Cumulative Frequency	Cumulative Percent
18-44	223	13.57	223	13.57
45-64	773	47.05	996	60.62
65-74	375	22.82	1371	83.44
75+	272	16.56	1643	100.00

Patient: Gender

GENDER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Male	904	55.02	904	55.02
Female	739	44.98	1643	100.00

Patient: Race

race_cds	Frequency	Percent
White	909	55.33
Black or African American	450	27.39
American Indian or Alaska	19	1.16
Asian	29	1.77
Native Hawaiian or other P	5	0.30
Other	230	14.00
Don't know	1	0.06

Patient: Race

race_cds	Cumulative Frequency	Cumulative Percent
White	909	55.33
Black or African American	1359	82.71
American Indian or Alaska	1378	83.87
Asian	1407	85.64
Native Hawaiian or other P	1412	85.94
Other	1642	99.94
Don't know	1643	100.00

The FREQ Procedure

race indicator from CMS 2728, 1=AMERICAN INDIAN/ALASKAN NATIVE,
2=ASIAN, 3=BLACK, 4=WHITE, 5=UNKNOWN,   6=PACIFIC
ISLANDER, 7=MIDEAST, 8=INDIAN SUBCO, 9=OTHER/MULTI-RACIAL

race2728	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	14	0.85	14	0.85
2	23	1.40	37	2.25
3	460	28.00	497	30.25
4	1132	68.90	1629	99.15
6	10	0.61	1639	99.76
9	4	0.24	1643	100.00

Patient: Participant Hispanic

ETHN	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Yes	271	16.49	271	16.49
No	1371	83.44	1642	99.94
Don't know	1	0.06	1643	100.00

Patient: Modality at the time of baseline survey

TREATMO_BL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Hemodialysis	1473	89.65	1473	89.65
Peritoneal dialysis	169	10.29	1642	99.94
Missing	1	0.06	1643	100.00

The MEANS Procedure

Analysis Variable : age

N	Mean	Std Dev	Minimum	Maximum
360	61.4815651	13.7076708	20.3148528	90.1273101

The FREQ Procedure

agecat	Frequency	Percent	Cumulative Frequency	Cumulative Percent
18-44	41	11.39	41	11.39
45-64	166	46.11	207	57.50
65-74	88	24.44	295	81.94
75+	65	18.06	360	100.00

Patient: Gender

GENDER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Male	184	51.11	184	51.11
Female	176	48.89	360	100.00

Patient: Race

race_cds	Frequency	Percent
White	204	56.67
Black or African American	94	26.11
American Indian or Alaska	2	0.56
Asian	13	3.61
Native Hawaiian or other P	3	0.83
Other	44	12.22

Patient: Race

race_cds	Cumulative Frequency	Cumulative Percent
White	204	56.67
Black or African American	298	82.78
American Indian or Alaska	300	83.33
Asian	313	86.94
Native Hawaiian or other P	316	87.78
Other	360	100.00

The FREQ Procedure

race indicator from CMS 2728, 1=AMERICAN INDIAN/ALASKAN NATIVE,
 2=ASIAN, 3=BLACK, 4=WHITE, 5=UNKNOWN,   6=PACIFIC
   ISLANDER, 7=MIDEAST, 8=INDIAN SUBCO, 9=OTHER/MULTI-RACIAL

race2728	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	2	0.56	2	0.56
2	12	3.33	14	3.89
3	94	26.11	108	30.00
4	247	68.61	355	98.61
6	4	1.11	359	99.72
9	1	0.28	360	100.00

Patient: Participant Hispanic

ETHN	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Yes	54	15.00	54	15.00
No	306	85.00	360	100.00

Patient: Modality at the time of baseline survey

TREATMO_BL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Hemodialysis	309	85.83	309	85.83
Peritoneal dialysis	51	14.17	360	100.00

The MEANS Procedure

Variable	Label	N	Mean
meanalb	Lab: Mean Albumin Result1 (g/dL)	227	3.4746696
meanpab	Lab: Mean Prealbumin Result1 (mg/dL)	227	31.5469163
meanaag	Lab: Mean Alpha-1-acid glycoprotein (mg/dL)	227	115.7422907

Variable	Label	Std Dev
meanalb	Lab: Mean Albumin Result1 (g/dL)	0.4679169
meanpab	Lab: Mean Prealbumin Result1 (mg/dL)	9.3295854
meanaag	Lab: Mean Alpha-1-acid glycoprotein (mg/dL)	35.1794369

Variable	Label	Minimum
meanalb	Lab: Mean Albumin Result1 (g/dL)	1.7500000
meanpab	Lab: Mean Prealbumin Result1 (mg/dL)	7.3500000
meanaag	Lab: Mean Alpha-1-acid glycoprotein (mg/dL)	37.0500000

Variable	Label	Maximum
meanalb	Lab: Mean Albumin Result1 (g/dL)	4.9500000
meanpab	Lab: Mean Prealbumin Result1 (mg/dL)	61.9000000
meanaag	Lab: Mean Alpha-1-acid glycoprotein (mg/dL)	273.5000000

The UNIVARIATE Procedure
Variable: medcrp (Lab: Median C-Reactive Protein (mg/dL))

Moments

N	227	Sum Weights	227
Mean	1.33361233	Sum Observations	302.73
Std Deviation	1.9990284	Variance	3.99611455
Skewness	3.79484991	Kurtosis	17.0175774
Uncorrected SS	1306.84635	Corrected SS	903.121888
Coeff Variation	149.895764	Std Error Mean	0.13268018

Basic Statistical Measures

Location		Variability	
Mean	1.333612	Std Deviation	1.99903
Median	0.745000	Variance	3.99611
Mode	0.100000	Range	14.95000
		Interquartile Range	0.85500

Tests for Location: Mu0=0

Test	-Statistic-		-----p Value------	
Student's t	t	10.05133	Pr > |t|	<.0001
Sign	M	113.5	Pr >= |M|	<.0001
Signed Rank	S	12939	Pr >= |S|	<.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	15.050
99%	11.100
95%	5.860
90%	2.820
75% Q3	1.250
50% Median	0.745
25% Q1	0.395
10%	0.230
5%	0.100
1%	0.100
0% Min	0.100

The UNIVARIATE Procedure
Variable: medcrp (Lab: Median C-Reactive Protein (mg/dL))

Extreme Observations

----Lowest----		-----Highest----	
Value	Obs	Value	Obs
0.1	1412	8.455	62
0.1	533	10.220	23
0.1	477	11.100	66
0.1	434	11.450	1169
0.1	433	15.050	1243

Missing Values

Missing Value	Count	-----Percent Of-----	
		All Obs	Missing Obs
.	133	36.94	100.00