Dataset Integrity Check for the CRISP2 Data Files

Prepared by David Ruggieri
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton MD 20705
November 3, 2014

Table of Contents

1 Standard Disclaimer

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 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 NIDDK Repository staff 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.

2 Study Background

The Consortium for Radiological Imaging Studies of Polycystic Kidney Disease (CRISP) studied the progression of autosomal dominant polycystic kidney disease (ADPKD), and compared radiological techniques for measuring increases in renal volume during the progression of ADPKD. The CRISP study tested whether magnetic resonance (MR) can detect changes in renal volume, cyst volume, or changes in % cystic involvement in ADPKD individuals over a short period of time (1 to 2 years). CRISP participants had ADPKD with relatively normal renal function and creatinine clearances. The recruitment goal was for two thirds of the participants to have a high risk of progression to ESRD and one third to have an absence of risk factors for ESRD.

3 Archived Datasets

All SAS data files, as provided by the Data Coordinating Center (DCC), are located in the CRISP2 data package. For this replication, variables were taken from the “CJASN_Revised_April2014” dataset.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Chapman et al [1] in the Clinical Journal of the American Society of Nephrology in March 2012.

To verify the integrity of the dataset, descriptive statistics were computed (table 1, supplemental tables 1 and 2, and some numbers directly from the paper).

5 Results

Table 1 in the publication [1], Frequency of renal insufficiency at baseline and Year 8. Our Table A lists the variables we used in our replication and Table B compare the results calculated from the archived data file to the results published in Table 1. The results of the replication are identical to those listed in the table.

Supplemental Table 1 in the publication [1] Supplemental Table 1. Effect of referencing TKV and GFR to baseline height, weight, BSA, and BMI. Our Table C lists the variables we used in our replication and Tables D and E compare the results calculated from the archived data file to the results published in supplemental table 1. The results of the replication are within expected results.

The data for supplemental table 2 in the publication [1] Supplemental Table 2. Baseline through year 8 visit htTKV and GFR (mL/min/1.73m2) values and percentage change in htTKV and GFR. Our Table F lists the variables we used in our replication and Tables G, H, I, and J compare the results calculated from the archived data file to the results published in supplemental Table 2. The results of the replication are within expected results.

6 Conclusions

The NIDDK repository is confident that the CRISP2 data files to be distributed are within expected results.

7 References

Arlene B. Chapman, James E. Bost, Vicente E. Torres, Lisa Guay-Woodford, Kyongtae Ty Bae, Douglas Landsittel, Jie Li, Bernard F. King, Diego Martin, Louis H. Wetzel, Mark E. Lockhart, Peter C. Harris, Marva Moxey-Mims, Mike Flessner, William M. Bennett, Jared J. Grantham. Kidney Volume and Functional Outcomes in Autosomal Dominant Polycystic Kidney Disease. Clin J Am Soc Nephrol. 2012 March; 7(3): 479–486

Table A: Variables used to replicate Table 1: Table 1 Frequency of renal insufficiency at baseline and Year 8

Table Variable	Variables Used in Replication from the "Table 1" Dataset
GFR [baseline]	gfrstg
eGFR(MDRD) [baseline]	egfrstg
GFR [year 8]	last_gfrstg
eGFR(MDRD) [year 8]	last_egfrstg

Table B: Comparison of values computed in integrity check to reference article Table 1 values

GFR	CKD Stage	Baseline% [Manuscript]	Baseline% [DSIC]	Year 8(%) [Manuscript]	Year 8 (%) [DSIC]
<30	Stage 4	0 (0.0)	0 (0.0)	26 (10.8)	26 (10.8)
30-60	Stage 3	10 (4.2)	10 (4.2)	54 (22.4)	54 (22.4)
60-<90	Stage 2	94 (39.8)	94 (39.8)	66 (27.4)	66 (27.4)
90+	Stage 1	132 (55.9)	132 (55.9)	95 (39.4)	95 (39.4)

eGFR	Baseline% [Manuscript]	Baseline% [DSIC]	Year 8(%) [Manuscript]	Year 8 (%) [DSIC]
<30	0 (0.0)	0 (0.0)	36 (14.9)	36 (14.9)
30-60	24 (10.0)	24 (10.0)	64 (26.6)	64 (26.6)
60-<90	121 (50.2)	121 (50.2)	76 (31.5)	76 (31.5)
90+	96 (39.8)	96 (39.8)	65 (27.0)	65 (27.0)

Table C: Variables used to replicate supplemental table 1: Supplemental Table 1. Effect of referencing TKV and GFR to baseline height, weight, BSA, and BMI

Table Variable	Variables Used in Replication from the "Table 2" Dataset
TKV	tkv
GFR	Uic0
Height	Height_c0
Weight	Weight_c0
BSA	Bsa_c0*1.73
BMI	Bmi_c0
M, F, M/F	gender

Table D: Comparison of values computed in integrity check to reference article supplemental table 1 values; TKV

TKV							
Ref	Units	M [Manuscript]	M [DSIC]	F [Manuscript]	F [DSIC]	M/F [Manuscript]	M/F [DSIC]
None	cc	1161.1	1161.1	1014.1	1014.1	1.145	1.145
Height	cc/m	633.9	633.9	611.1	611.1	1.037	1.037
Weight	cc/kg	13.2	13.2	14.9	14.9	0.886	0.882
BSA	cc/1.73m2	949.8	949.8	994.2	994.2	0.955	0.955
BMI	cc/m2	44.1	44.1	41	41	1.076	1.074

Table E: Comparison of values computed in integrity check to reference article supplemental table 1 values; GFR

GFR							
Ref	Units	M [Manuscript]	M [DSIC]	F [Manuscript]	F [DSIC]	M/F [Manuscript]	M/F [DSIC]
None	ml/min	114.9	114.9	100.7	100.7	1.141	1.141
Height	ml/min/m	63.1	63.1	60.9	60.9	1.036	1.035
Weight	ml/min/kg	1.3	1.3	1.5	1.5	0.867	0.905
BSA	ml/min/1.73m2	95.7	95.7	99.2	99.2	0.965	0.964
BMI	ml/min/m2	4.5	4.5	4.1	4.1	1.098	1.099

Table F: Variables used to replicate supplemental table 2: Supplemental Table 2. Baseline through year 8 visit htTKV and GFR (mL/min/1.73m^2) values and percentage change in htTKV and GFR

Table Variable	Variables Used in Replication from the "Table 2" Dataset
Visit	Vis
htTKV (cc/m)	Tkvhb
GFR (mL/min/1.73m2)	Cic_new

Table G: Comparison of values computed in integrity check to reference article supplemental Table 2 values ; All Crisp Patients, htTKV

All CRISP Patients						
			htTKV (cc/m)			
Visit	N [Manuscript]	N [DSIC]	Mean +/- SD [Manuscript]	Mean +/- SD [DSIC]	% from BL [Manuscript]	% from BL [DSIC]
Baseline	236	236	615.6 +/- 371.0	615.6 +/- 371.0	--	--
Year 1	227	227	650.8 +/- 391.0	650.8 +/- 391.0	5.7	5.7
Year 2	214	214	712.9 +/- 443.3	712.9 +/- 443.3	15.8	15.8
Year 3	219	219	737.6 +/- 466.9	737.6 +/- 466.9	19.8	19.8
Year 6	176	176	903.2 +/- 650.3	903.2 +/- 650.3	46.7	46.7
Year 8	128	128	992 +/- 711.3	992 +/- 711.3	61.2	61.2

Table H: Comparison of values computed in integrity check to reference article supplemental Table 2 values ; All Crisp Patients, GFR

All CRISP Patients						
			GFR			
Visit	N [Manuscript]	N [DSIC]	Mean +/- SD [Manuscript]	Mean +/- SD [DSIC]	% from BL [Manuscript]	% from BL [DSIC]
Baseline	236	236	97.8 +/- 24.7	97.8 +/- 24.7	--	--
Year 1	227	227	99.1 +/- 25.6	99.1 +/- 25.6	1.4	1.4
Year 2	214	214	97.8 +/- 30.2	97.8 +/- 30.2	0	0
Year 3	219	219	95.8 +/- 29.5	95.8 +/- 29.5	-2.1	-2.1
Year 6	176	176	85.6 +/- 40.9	85.6 +/- 40.9	-12.5	-12.5
Year 8	128	128	74.9 +/- 35.2	74.9 +/- 35.2	-23.4	-23.4

Table I: Comparison of values computed in integrity check to reference article supplemental Table 2 values; CRISP participants with measurements at every visit, htTKV

CRISP participants with measurements from every visit						
			htTKV (cc/m)			
Visit	N [Manuscript]	N [DSIC]	Mean +/- SD [Manuscript]	Mean +/- SD [DSIC]	% from BL [Manuscript]	% from BL [DSIC]
Baseline	108	108	636.3 +/- 380.3	636.3 +/- 380.3	--	--
Year 1	108	108	66.4 +/- 405.9	66.4 +/- 405.9	4.7	4.7
Year 2	108	108	729.3 +/- 460.1	729.3 +/- 460.1	14.6	14.6
Year 3	108	108	759.0 +/- 475.1	759.0 +/- 475.1	19.3	19.3
Year 6	108	108	895.7 +/- 621.2	895.7 +/- 621.2	40.8	40.8
Year 8	108	108	1029.3 +/- 738.5	1029.3 +/- 738.5	61.8	61.8

Table J: Comparison of values computed in integrity check to reference article supplemental table 2 values; CRISP participants with measurements at every visit GFR

CRISP participants with measurements from every visit						
			GFR			
Visit	N [Manuscript]	N [DSIC]	Mean +/- SD [Manuscript]	Mean +/- SD [DSIC]	% from BL [Manuscript]	% from BL [DSIC]
Baseline	108	108	94.4 +/- 22.5	94.4 +/- 22.5	--	--
Year 1	108	108	97.0 +/- 25.2	97.0 +/- 25.2	2.7	2.7
Year 2	108	108	95.3 +/- 27.7	95.3 +/- 27.7	0.9	0.9
Year 3	108	108	92.9 +/- 27.2	92.9 +/- 27.2	-1.6	-1.6
Year 6	108	108	93.0 +/- 38.6	93.0 +/- 38.6	-12.1	-12.1
Year 8	108	108	71.8 +/- 33.1	71.8 +/- 33.1	24	24

```
title1 "Program Saved As: /prj/niddk/ims_analysis/CRISP2/prog_initial_analysis/recreate.Revised.CJASN.final.paper.m05d05y2014.sas";
title2 "NIDDK Effort: CRISP II study - Check Revised.CJASN.final.paper.m05d05y2014 paper";

options nofmterr;

/*************************************************************************************************************************************
 Programmer: Dave Ruggieri
 Date: 30 May 2014
 Billing Code: DDKDR1.T01

 Function/Notes: This program recreates the numbers found on Chapman CJASN 032012.pdf paper. It is titled Kidney volume and
 Functional Outcomes in Autosomal Dominat Polycystic Kidney disease.
*************************************************************************************************************************************
 Programmer: Dave Ruggieri
 Date: 27 June 2014
 Billing Code: DDKDR1.T01
 *************************************************************************************************************************************/
*  Input Files  *;
*****************;
*** Primary analysis dataset ***;
filename analy pipe "gunzip -c /prj/niddk/ims_analysis/CRISP2/private_created_data/sasfile/CJASN_Revised_April2014.v9x.gz";
proc cimport data=analy infile=analy;
run;

proc contents data = analy;
  title4 'Contents of analysis dataset';

*** Long 2 dataset ***;
libname sas_data '/prj/niddk/ims_analysis/CRISP2/private_orig_data/5_-_Other_data_files/';

data long1;
  set sas_data.crisp2_long_04172012_1     ;

data long3;
  set sas_data.crisp2_long_04172012_3     ;

/*proc contents data = long1;
  title4 'Contents of long1 dataset';*/
*************;
*  Formats  *;
*************;
proc format;
  value genfmt /* gender */
    0 = '0=Female'
    1 = '1=Male'
    ;

  value racefmt /* racewhite */
    1 = '1=White'
    0 = '0=Non-white'
    ;

  value stagefmt /* gfrstg egfrstg */
    1 = '1=Stage 1'
    2 = '2=Stage 2'
    3 = '3=Stage 3'
```

```
    4 = '4=Stage 4'
    ;

  value agegrp /* gfr20_age_grp gfr40_age_grp gfrstg3_age_grp gfrstg4_age_grp */
    3 = "3=15-20"
    4 = "4=20-25"
    5 = "5=25-30"
    6 = "6=30-35"
    7 = "7=35-40"
    8 = "8=40+"
    ;

  value visfmt /* vis */
    0  = 'BV01:Baseline Visit'
    10 = 'FV01:Clinic Visit Year 1'
    20 = 'FV02:Clinic Visit Year 2'
    30 = 'FV03:Clinic Visit Year 3'
    60 = 'FV06:Biannual Clinic Visit Year 6'
    65 = 'FV06.6:Semi-Annual Phone Call post  Year 6 visit'
    70 = 'FV07:Lab Visit Year 7 '
    75 = 'FV07.6:Semi-Annual Phone Call post  Year 7 visit'
    80 = 'FV08:Biannual Clinic Visit Year 8'
    85 = 'FV08.6:Semi-Annual Phone Call post Year 8 visit'
    90 = 'FV09:Lab Visit Year 9 '
    95 = 'FV09.6:Semi-Annual Phone Call post  Year 9 visit'
    96 = 'N/A:6 months or more post vis 95'
    99 = 'As needed'
    40 = 'Unknown: Clinic visit year 4?'
    45 = 'Unknown: Phone call post year 4?'
    50 = 'Unknown: Clinic visit year 5?'
    ;

  value stgcheck
    . = 'Missing'
    low-29.999999 = '<30'
    30-59.999999 = '30-<60'
    60-89.999999 = '60-<90'
    90-high = '90+'
    ;

  value missfmt
    . = 'Missing'
    other = 'Other'
    ;

************;
*  Page 1  *;
************;
*** First check if all of these people signed consent ***;
proc freq data = long1;
  tables sigcon /missing list;
  title4 'Signed consent flag on long1 dataset';

proc sort data = long1 nodupkey;
  where sigcon = 1;
  by pkdid;
```

```
proc freq data = long1;
  tables sigcon /missing list;
  title4 'Signed consent flag on long1 dataset after subset';

proc sort data = analy;
  by pkdid;

data analy_mod;
  merge analy (in = ina)
        long1 (in = inl keep = pkdid);
  by pkdid;

  inanaly = ina;
  signed_consent = inl;

proc freq data = analy_mod;
  tables inanaly*signed_consent /missing list;
  title4 'Did all people in analysis file sign consent?';

*****************************************************;
*  Establish population of patients - should be 241  *;
*****************************************************;
title4 'Patient level counts';

proc sort data = analy;
  by pkdid vis dvdate;

data _null_;
  set analy end=eof;
  by pkdid vis;

  file print;

  retain pidcount womcount mancount firstenrl last3yr vis60 vis60men vis60wom vis60agetot lastyr;

  if _N_ = 1 then do;
    pidcount = 0;
    womcount = 0;
    mancount = 0;
    firstenrl = 999999999;
    last3yr = 0;
    vis60 = 0;
    vis60men = 0;
    vis60wom = 0;
    vis60agetot = 0;
    lastyr = 0;
  end;

  if first.pkdid then do;
    pidcount + 1;
    if gender = 0 then womcount + 1;
    else if gender = 1 then mancount + 1;
    else put 'WAR' 'NING: Unknown gender: ' pkdid= gender= ;
  end;
```

```
  if vis = 0 and dvdate < firstenrl then firstenrl = dvdate;
  else if vis = 30 and dvdate > last3yr then last3yr = dvdate;

  if dvdate > lastyr then lastyr = dvdate;

  if vis = 60 then do;
    vis60 + 1;
    vis60agetot = vis60agetot + age;
    if gender = 0 then vis60wom + 1;
    else if gender = 1 then vis60men + 1;
  end;

  if NOT(first.vis AND last.vis) then do;
    put 'ER' 'ROR: Analysis dataset is not 1 rec/visit. Aborting: ' pkdid= vis= ;
    abort return;
  end;

  if eof then do;
    vis60_agemean = vis60agetot/vis60;
    put 'Total number of patients in the study(looking for 241): ' pidcount;
    put 'Men: ' mancount ;
    put 'Women: ' womcount;
    put / 'Number enrolled in CRISP2: ' vis60;
    put 'Number of men in CRISP2: ' vis60men;
    put 'Number of women in CRISP2: ' vis60wom;
    put 'Mean age at CRISP2 enrollment: ' vis60_agemean;
    put / 'First enrollment date: ' firstenrl;
    put 'Last 3 year visit date: ' last3yr;
    put 'Last visit date: ' lastyr;
  end;

  format firstenrl last3yr date9.;

proc means data = analy;
  where vis = 0;
  var age;
  title4 'Means of age variable at CRISP I enrollment';

proc means data = analy;
  where vis = 60;
  var age;
  title4 'Means of age variable at CRISP II enrollment';

proc freq data = analy;
  tables dvdate
         dvdate*vis
         /missing list;
  format dvdate year.;
  title4 'Visit date formatted to year only';

proc freq data = analy;
  tables vis /missing list;
  format vis visfmt.;
  title4 'Visit on analysis file';

proc freq data = analy;
```

```
  where (vis <= 30) and (gfrstg = 3);
  tables vis*gfrstg*gfrstg3 /missing list;
  format gfrstg stagefmt.;
  title4 'Stage 3 CKD at or before 3 years of f-up';
  title5 'Looking for 29 patients';

data analyst3;
  set analy;
  where vis <=30 and gfrstg = 3;
  by pkdid;

  if last.pkdid;

proc freq data = analyst3;
  where (vis <= 30) and (gfrstg = 3);
  tables vis*gfrstg*gfrstg3 /missing list;
  format gfrstg stagefmt.;
  title4 'Stage 3 CKD at or before 3 years of f-up after subset to 1 rec/paitent';
  title5 'Looking for 29 patients';

proc print data = analy;
  where (vis <= 30) and (gfrstg = 3);
  by pkdid;
  id pkdid;
  var vis gfrstg gfrstg3;
  format gfrstg stagefmt.;
  title4 'Listing of the 38 patients who reached stage 3 CKD (GFR)';
  title5 '38 participants were found, 29 were listed in the paper.';

**********************************;
*  Determine mean follow-up time  *;
**********************************;
proc sort data = analy;
  by pkdid visc;

data lastvis;
  set analy;
  by pkdid;

  if last.pkdid;

proc means data = lastvis;
  var visc;
  title4 'Follow-up time based on last visit attended';

proc freq data = lastvis;
  tables vis /missing list;
  title4 'Last visit attended by entire cohort';

data _null_;
  set analy end=eof;
  by pkdid;

  file print;

  retain hasstg3 numstg3 numstg3mean numstg3wrong hasstg3mean hasstg3wrong hadckd numckd;
```

```
  if _N_ = 1 then do;
    numstg3 = 0;
    numstg3mean = 0;
    numstg3wrong = 0;
    numckd = 0;
  end;

  if first.pkdid then do;
    hasstg3 = 0;
    hasstg3mean = 0;
    hasstg3wrong = 0;
    hadckd = 0;
  end;


  if (gfrstg ^= .) and (hadckd = 0) then do;
    numckd + 1;
    hadckd = 1;
  end;

  if gfrstg = 3 then do;
    if hasstg3 = 0 then numstg3 + 1;
    hasstg3 = 1;
    if visc > 8.1956695 then do;
      if hasstg3mean = 0 then numstg3mean + 1;
      hasstg3mean = 1;
    end;
    if visc > 7.8 then do;
      if hasstg3wrong = 0 then numstg3wrong + 1;
      hasstg3wrong = 1;
    end;
  end;

  if eof then do;
    pctstg3 = (numstg3/numckd)*100;
    pctmean = (numstg3mean/numckd)*100;
    pctwrong = (numstg3wrong/numckd)*100;
    put / 'Total Number of patients who ever had stage 3 CKD? ' numstg3 ' Percent (over ' numckd ') ' pctstg3;
    put 'Number of patients who ever had stage 3 CKD after observed mean followup? ' numstg3mean ' Percent (over ' numckd ') ' pctmean;
    put 'Number of patients who ever had stage 3 CKD after 7.8 years mean followup? ' numstg3wrong ' Percent (over ' numckd ') ' pctwrong;
  end;

proc freq data = analy;
  where vis = 0 and gfrstg ^= .;
  tables gfrstg
         /missing list;
  format gfrstg egfrstg stagefmt.;
  title4 'Table 1: Stage at enrollment visit';

************;
*  Page 2  *;
************;
proc freq data = analy;
  where vis = 0;
  tables genetype /missing list;
```

```
  title4 'Genetype at visit 0';

proc print data = analy;
  where vis = 85;
  var pkdid;
  title4 'Listing of PKDIDs that have post-8 year phone call';

proc print data = analy;
  where pkdid in(201877,209281,244831,246620,294511,301372,313195,313893,327325,386758,392316) and vis = 80;
  var pkdid vis;
  title4 'Checking if all of these patients had a visit 80';
  title5 'PKDIDs: 201877,209281,244831,246620,294511,301372,313195,313893,327325,386758,392316';

*** Determine 7.9 +/- .6 years f-up after enrollment ***;
data fupset;
  set analy;
  where vis = 0;

  earlydate = dvdate + (365.25 * 7.3);
  latedate = dvdate + (365.25 * 8.5);

  format earlydate latedate date9.;

proc print data = fupset (obs = 25);
  var pkdid dvdate earlydate latedate;
  format dvdate earlydate latedate date9.;
  title4 'Listing checking the creation of earlydate and latedate';

data analy_merged;
  merge analy (in = ina)
        fupset (in = inf keep = pkdid earlydate latedate);
  by pkdid;

  if NOT(ina AND inf) then abort return;

title4 'Checking number of patients who have a visit 7.9 +/- .6 years after enrollment based on DV Dates';

data analy_merged;
  set analy_merged end=eof;
  by pkdid;

  file print;

  retain numvis viscount pidviscount visccount pidvisccount numvisc;

  if _N_ = 1 then do;
    viscount = 0;
    pidviscount = 0;
    visccount = 0;
    pidvisccount = 0;
  end;

  if first.pkdid then do;
    numvis = 0;
    numvisc = 0;
  end;
```

```
  if (earlydate <= dvdate <= latedate) then do;
    numvis + 1;
    viscount + 1;
    if numvis = 1 then pidviscount + 1;
  end;

  if (7.3 <= visc <= 8.5) then do;
    numvisc + 1;
    visccount + 1;
    if numvisc = 1 then pidvisccount + 1;
  end;

  if eof then do;
    put 'Total number of visits attended in the range specified: ' viscount;
    put 'Number of patients who attended at least 1 visit in the range specified: ' pidviscount;
    put 'Total number of visits attended in the range specified based on visc: ' visccount;
    put 'Number of patients who attended at least 1 visit in the range specified based on visc: ' pidvisccount;
  end;

  if last.pkdid then output;

proc freq data = analy_merged;
  tables numvis numvisc /missing list;
  title4 'Distribution of the number of visits attended 7.9 +/- .6 years after enrollment for each patient';

proc means data = analy;
  where vis = 0 and gender = 0;
  var tkv;
  title4 'TKV for women at baseline. Looking for 1014';

proc means data = analy;
  where vis = 0 and gender = 1;
  var tkv;
  title4 'TKV for men at baseline. Looking for 1161';

*************;
*  Table 1  *;
*************;
proc freq data = analy;
  where vis = 0 and gfrstg ^= .;
  tables gfrstg
         /missing list;
  format gfrstg egfrstg stagefmt.;
  title4 'Table 1: Stage at enrollment visit';

proc freq data = analy;
  where vis = 0;
  tables egfrstg*egfr
         /missing list;
  format gfrstg egfrstg stagefmt. egfr stgcheck.;
  title4 'Table 1: Stage at enrollment visit';

proc freq data = analy;
  tables vis*visname*gfrstg*egfrstg
         /missing list;
```

```
  format egfrstg gfrstg missfmt.;
  title4 'Vis by visname';

proc sort data = analy;
  by pkdid vis;

proc freq data = analy;
  where vis = 80 and gfrstg ^= .;
  tables gfrstg
         /missing list;
  format gfrstg egfrstg stagefmt.;
  title4 'Table 1: Stage at 8 year visit';

proc freq data = analy;
  where vis = 80;
  tables egfrstg
         /missing list;
  format gfrstg egfrstg stagefmt.;
  title4 'Table 1: Stage at 8 year visit';

proc freq data = analy;
  tables vis*gfr20*gfr40
         vis*egfr20*egfr40
         /missing list;
  title4 'Checking visit by GFR and eGFR decrease rates';

*****************************************************;
*  Determine the last visit with GFR or eGFR results  *;
*****************************************************;
proc freq data = analy;
  tables vis*gfrstg*egfrstg
         /missing list;
  format gfrstg egfrstg missfmt.;
  title4 'Visit type by has gfrstg or egfrstg results flags';

data analy_coll checkset;
  set analy;
  by pkdid;

  array visnum (12) (0,10,20,30,40,45,50,60,65,70,75,80);
  array hasvis (12) has0 has10 has20 has30 has40 has45 has50 has60 has65 has70 has75 has80;
  array haspkd (12) pkd0 pkd10 pkd20 pkd30 pkd40 pkd45 pkd50 pkd60 pkd65 pkd70 pkd75 pkd80;
  array hasepkd (12) epkd0 epkd10 epkd20 epkd30 epkd40 epkd45 epkd50 epkd60 epkd65 epkd70 epkd75 epkd80;
  array dates (12) date0 date10 date20 date30 date40 date45 date50 date60 date65 date70 date75 date80;

  retain hasvis haspkd hasepkd dates worstgfr worstegfr;

  if first.pkdid then do;
    do i = 1 to 12;
      hasvis[i] = 0;
      haspkd[i] = .;
      hasepkd[i] = .;
      dates[i] = .;
    end;
    worstgfr = 0;
    worstegfr = 0;
```

```
end;

do i = 1 to 12;
  if visnum[i] = vis then do;
    dates[i] = dvdate;
    hasvis[i] = 1;
    if (gfrstg ^= .) then haspkd[i] = gfrstg;
    if (egfrstg ^= .) then hasepkd[i] = egfrstg;
  end;
end;

if gfrstg > worstgfr then worstgfr = gfrstg;
if egfrstg > worstegfr then worstegfr = egfrstg;

if last.pkdid then do;
  if worstgfr = 4 then do;
    gfr4 = 1;
    gfr3 = 1;
    gfr2 = 1;
    gfr1 = 1;
  end;
  else if worstgfr = 3 then do;
    gfr4 = 0;
    gfr3 = 1;
    gfr2 = 1;
    gfr1 = 1;
  end;
  else if worstgfr = 2 then do;
    gfr4 = 0;
    gfr3 = 0;
    gfr2 = 1;
    gfr1 = 1;
  end;
  else if worstgfr = 1 then do;
    gfr4 = 0;
    gfr3 = 0;
    gfr2 = 0;
    gfr1 = 1;
  end;
  else do;
    gfr4 = 0;
    gfr3 = 0;
    gfr2 = 0;
    gfr1 = 0;
  end;

  if worstegfr = 4 then do;
    egfr4 = 1;
    egfr3 = 1;
    egfr2 = 1;
    egfr1 = 1;
  end;
  else if worstegfr = 3 then do;
    egfr4 = 0;
    egfr3 = 1;
    egfr2 = 1;
```

```
      egfr1 = 1;
    end;
    else if worstegfr = 2 then do;
      egfr4 = 0;
      egfr3 = 0;
      egfr2 = 1;
      egfr1 = 1;
    end;
    else if worstegfr = 1 then do;
      egfr4 = 0;
      egfr3 = 0;
      egfr2 = 0;
      egfr1 = 1;
    end;
    else do;
      egfr4 = 0;
      egfr3 = 0;
      egfr2 = 0;
      egfr1 = 0;
    end;
    do i = 1 to 12;
      if (haspkd[i] ^= .) then do;
        last_gfrstg = haspkd[i];
        last_gfrstg_vis = visnum[i];
      end;
      if (hasepkd[i] ^= .) then do;
        last_egfrstg = hasepkd[i];
        last_egfrstg_vis = visnum[i];
      end;
    end;
    output analy_coll;
  end;
  output checkset;

  format date0 date10 date20 date30 date40 date45 date50 date60 date65 date70 date75 date80 date9.;

proc print data = checkset (obs = 50);
  by pkdid;
  id pkdid;
  var vis last_gfrstg last_gfrstg_vis last_egfrstg last_egfrstg_vis dvdate has0 has10 has20 has30 has40 has45 has50 has60 has65 has70 has75 has80
                pkd0 pkd10 pkd20 pkd30 pkd40 pkd45 pkd50 pkd60 pkd65 pkd70 pkd75 pkd80
                epkd0 epkd10 epkd20 epkd30 epkd40 epkd45 epkd50 epkd60 epkd65 epkd70 epkd75 epkd80
                date0 date10 date20 date30 date40 date45 date50 date60 date65 date70 date75 date80;
  title4 'Listing checking creation of collapsed variables';

/*proc freq data = analy_coll;
  tables */

proc means data = analy;
  where gfrstg = 3;
  var httkv18 httkv60;
  title4 'HTTKV where patients have stage 3 CKD';

proc freq data = analy_coll;
  tables worstgfr worstegfr /missing list;
  title4 'Worst GF and worst EGFR by patient';
```

```
proc freq data = analy_coll;
  tables last_gfrstg
         last_egfrstg
         /missing list;
  format last_gfrstg last_egfrstg stagefmt.;
  title4 'Table 1: Stage at 8 year visit using last gfr/egfr numbers';

************************;
*  Supplemental Table 1  *;
************************;
proc means data = analy;
  where vis = 0 and gender = 1;
  var tkv;
  title4 'TKV for males at baseline';

proc means data = analy;
  where vis = 0 and gender = 0;
  var tkv;
  title4 'TKV for women at baseline';

proc means data = analy;
  where vis = 0;
  var tkv;
  title4 'TKV for both genders at baseline';
run;

proc print data = analy (obs = 10);
  var height_c0 weight_c0 bsa_c0 gender;
  title4 'Check listing of height and BSA';
run;

proc contents data = long3;
  title4 'Contents of long3 file';
run;

proc sort data = long3;
  by pkdid vis;

proc sort data = analy;
  by pkdid vis;

title4 'Check supplemental table 1';

data baseline;
  merge analy (in = ina)
        long3 (in = inl keep = pkdid vis bsa_c);
  by pkdid vis;
  where vis = 0;

  if ina;
  inlong3 = inl;

proc freq data = baseline;
  tables inlong3 /missing list;
  title4 'Distribugtion of inlong3 in baseline dataset';
```

```
data baseline;
  set baseline end=eof;
  where vis = 0;

  file print;

  retain tkv_men tkv_wom height_men height_wom weight_men weight_wom bsa_men bsa_wom bmi_men bmi_wom men wom bsa_denom_men bsa_denom_wom
         bsac_men bsac_wom
         gfr_men gfr_wom height_gfr_men height_gfr_wom weight_gfr_men weight_gfr_wom bsa_gfr_men bsa_gfr_wom bmi_gfr_men bmi_gfr_wom
         bsa_gfrdenom_men bsa_gfrdenom_wom bsac_gfr_men bsac_gfr_wom gfrmen gfrwom;

  if _N_ = 1 then do;
    tkv_men    = 0;
    tkv_wom    = 0;
    height_men = 0;
    height_wom = 0;
    weight_men = 0;
    weight_wom = 0;
    bsa_men    = 0;
    bsa_wom    = 0;
    bmi_men    = 0;
    bmi_wom    = 0;
    men        = 0;
    wom        = 0;
    bsa_denom_men = 0;
    bsa_denom_wom = 0;
    bsac_men = 0;
    bsac_wom = 0;
    gfr_men        = 0;
    gfr_wom        = 0;
    height_gfr_men = 0;
    height_gfr_wom = 0;
    weight_gfr_men = 0;
    weight_gfr_wom = 0;
    bsa_gfr_men    = 0;
    bsa_gfr_wom    = 0;
    bmi_gfr_men    = 0;
    bmi_gfr_wom    = 0;
    bsa_gfrdenom_men = 0;
    bsa_gfrdenom_wom = 0;
    bsac_gfr_men     = 0;
    bsac_gfr_wom     = 0;
    gfrmen = 0;
    gfrwom = 0;
  end;

  bsa_denom = 1.73*(((1/100)*(height_c0))**2);

  tkv_height = tkv/(height_c0/100);
  tkv_weight = tkv/weight_c0;
  tkv_bsa    = tkv/bsa_c0;
  tkv_bmi    = tkv/bmi_c0;
  tkv_bsaden = tkv/bsa_denom;
  tkv_bsac   = tkv/bsa_c;
  tkv0_bsa   = tkv/bsa_c0*1.73;
```

```
if uic0 ^= . then do;
  gfr_height = uic0/(height_c0/100);
  gfr_weight = uic0/weight_c0;
  gfr_bsa    = uic0/bsa_c0;
  gfr_bmi    = uic0/bmi_c0;
  gfr_bsaden = uic0/bsa_denom;
  gfr_bsac   = uic0/bsa_c;
  gfr0_bsa   = uic0/bsa_c0*1.73;
end;

if gender = 1 then do;
  men + 1;
  tkv_men = tkv_men + tkv;
  height_men = height_men + tkv_height;
  weight_men = weight_men + tkv_weight;
  bsa_men    = bsa_men    + tkv0_bsa   ;
  bsac_men    = bsac_men    + tkv_bsac  ;
  bmi_men    = bmi_men    + tkv_bmi   ;
  bsa_denom_men = bsa_denom_men + tkv_bsaden;

  if uic0 ^= . then do;
    gfrmen + 1;
    gfr_men = gfr_men + uic0;
    height_gfr_men = height_gfr_men + gfr_height;
    weight_gfr_men = weight_gfr_men + gfr_weight;
    bsa_gfr_men    = bsa_gfr_men    + gfr0_bsa   ;
    bsac_gfr_men    = bsac_gfr_men    + gfr_bsac  ;
    bmi_gfr_men    = bmi_gfr_men    + gfr_bmi   ;
    bsa_gfrdenom_men = bsa_gfrdenom_men + gfr_bsaden;
  end;
end;
else if gender = 0 then do;
  wom + 1;
  tkv_wom = tkv_wom + tkv;
  height_wom = height_wom + tkv_height;
  weight_wom = weight_wom + tkv_weight;
  bsa_wom    = bsa_wom    + tkv0_bsa   ;
  bsac_wom    = bsac_wom    + tkv_bsac   ;
  bmi_wom    = bmi_wom    + tkv_bmi   ;
  bsa_denom_wom = bsa_denom_wom + tkv_bsaden;

  if uic0 ^= . then do;
    gfrwom + 1;
    gfr_wom = gfr_wom + uic0;
    height_gfr_wom = height_gfr_wom + gfr_height;
    weight_gfr_wom = weight_gfr_wom + gfr_weight;
    bsa_gfr_wom    = bsa_gfr_wom    + gfr0_bsa   ;
    bsac_gfr_wom    = bsac_gfr_wom    + gfr_bsac  ;
    bmi_gfr_wom    = bmi_gfr_wom    + gfr_bmi   ;
    bsa_gfrdenom_wom = bsa_gfrdenom_wom + gfr_bsaden;
  end;
end;

if eof then do;
  mean_tkv_men    = tkv_men/men;
```

```
mean_height_men = height_men /men;
mean_weight_men = weight_men /men;
mean_bsa_men    = bsa_men    /men;
mean_bmi_men    = bmi_men    /men;
mean_bsaden_men    = bsa_denom_men    /men;
mean_bsac_men   = bsac_men/men;

mean_tkv_wom = tkv_wom/wom;
mean_height_wom = height_wom /wom;
mean_weight_wom = weight_wom /wom;
mean_bsa_wom    = bsa_wom    /wom;
mean_bmi_wom    = bmi_wom    /wom;
mean_bsaden_wom    = bsa_denom_wom    /wom;
mean_bsac_wom   = bsac_wom /wom;

rat_tkv    = mean_tkv_men   /mean_tkv_wom   ;
rat_height = mean_height_men/mean_height_wom;
rat_weight = mean_weight_men/mean_weight_wom;
rat_bsa    = mean_bsa_men   /mean_bsa_wom   ;
rat_bmi    = mean_bmi_men   /mean_bmi_wom   ;
rat_bsaden = mean_bsaden_men    /mean_bsaden_wom;
rat_bsac   = mean_bsac_men / mean_bsac_wom;

mean_gfr_men        = gfr_men/gfrmen;
mean_gfr_height_men = height_gfr_men /gfrmen;
mean_gfr_weight_men = weight_gfr_men /gfrmen;
mean_gfr_bsa_men    = bsa_gfr_men    /gfrmen;
mean_gfr_bmi_men    = bmi_gfr_men    /gfrmen;
mean_gfr_bsaden_men = bsa_gfrdenom_men    /gfrmen;
mean_gfr_bsac_men   = bmi_gfr_men/gfrmen;

mean_gfr_wom    = gfr_wom/gfrwom;
mean_gfr_height_wom = height_gfr_wom /gfrwom;
mean_gfr_weight_wom = weight_gfr_wom /gfrwom;
mean_gfr_bsa_wom    = bsa_gfr_wom    /gfrwom;
mean_gfr_bmi_wom    = bmi_gfr_wom    /gfrwom;
mean_gfr_bsaden_wom    = bsa_gfrdenom_wom    /gfrwom;
mean_gfr_bsac_wom   = bmi_gfr_wom/gfrwom;

rat_gfr        = mean_gfr_men       /mean_gfr_wom   ;
rat_gfr_height = mean_gfr_height_men/mean_gfr_height_wom;
rat_gfr_weight = mean_gfr_weight_men/mean_gfr_weight_wom;
rat_gfr_bsa    = mean_gfr_bsa_men   /mean_gfr_bsa_wom   ;
rat_gfr_bmi    = mean_gfr_bmi_men   /mean_gfr_bmi_wom   ;
rat_gfr_bsaden = mean_gfr_bsaden_men/mean_gfr_bsaden_wom;
rat_gfr_bsac   = mean_gfr_bsac_men  /mean_gfr_bsac_wom;

put 'Supplemental Table 1. Effect of referencing TKV and GFR to baseline height, weight, BSA, and BMI' /;
put 'TKV';
put @1 'Ref' @20 'Units' @70 'M' @90 'F' @110 'M/F';
put @1 'None' @20 'cc' @70 mean_tkv_men @90 mean_tkv_wom @110 rat_tkv;
put @1 'Height' @20 'cc/m' @70 mean_height_men @90 mean_height_wom @110 rat_height;
put @1 'Weight' @20 'cc/kg' @70 mean_weight_men @90 mean_weight_wom @110 rat_weight;
put @1 'BSA' @20 'cc/1.73m2(corrected)' @70 mean_bsa_men @90 mean_bsa_wom @110 rat_bsa;
put @1 'BSA' @20 'cc/1.73m2' @70 mean_bsaden_men @90 mean_bsaden_wom @110 rat_bsaden;
put @1 'BSAC' @20 'cc/?' @70 mean_bsac_men @90 mean_bsac_wom @110 rat_bsac;
```

```
    put @1 'BMI' @20 'cc/m2' @70 mean_bmi_men @90 mean_bmi_wom @110 rat_bmi;
    put / 'GFR';
    put @1 'Ref' @20 'Units' @70 'M' @90 'F' @110 'M/F';
    put @1 'None' @20 'ml/min' @70 mean_gfr_men @90 mean_gfr_wom @110 rat_gfr;
    put @1 'Height' @20 'ml/min/m' @70 mean_gfr_height_men @90 mean_gfr_height_wom @110 rat_gfr_height;
    put @1 'Weight' @20 'ml/min/kg' @70 mean_gfr_weight_men @90 mean_gfr_weight_wom @110 rat_gfr_weight;
    put @1 'BSA' @20 'ml/min/1.73m2(corrected)' @70 mean_gfr_bsa_men @90 mean_gfr_bsa_wom @110 rat_gfr_bsa;
    put @1 'BSA' @20 'ml/min/1.73m2' @70 mean_gfr_bsaden_men @90 mean_gfr_bsaden_wom @110 rat_gfr_bsaden;
    put @1 'BSAC' @20 'ml/min/?' @70 mean_gfr_bsac_men @90 mean_gfr_bsac_wom @110 rat_gfr_bsac;
    put @1 'BMI' @20 'ml/min/m2' @70 mean_gfr_bmi_men @90 mean_gfr_bmi_wom @110 rat_gfr_bmi;
  end;

  output baseline;

  format mean_tkv_men mean_height_men mean_weight_men mean_bsa_men mean_bmi_men mean_tkv_wom mean_height_wom mean_weight_wom
         mean_bsa_wom mean_bmi_wom mean_gfr_men mean_gfr_height_men mean_gfr_weight_men mean_gfr_bsa_men mean_gfr_bmi_men
         mean_gfr_bsaden_men mean_gfr_wom mean_gfr_height_wom mean_gfr_weight_wom mean_gfr_bsa_wom mean_gfr_bmi_wom
         mean_gfr_bsaden_wom mean_gfr_bsac_men mean_bsac_men mean_bsac_wom mean_bsaden_men mean_bsaden_wom 8.1 rat_tkv rat_height
         rat_weight rat_bsa rat_bmi rat_gfr rat_gfr_height rat_gfr_weight rat_gfr_bsa rat_gfr_bmi rat_gfr_bsaden rat_gfr_bsac
         rat_bsac rat_bsaden 8.3;


proc means data = baseline;
  where vis = 0 and gender = 1;
  var tkv uic0;
  title4 'TKV for males at baseline not adjusted';
  title5 'Supplemental table 1';

proc means data = baseline;
  where vis = 0 and gender = 0;
  var tkv uic0 ;
  title4 'TKV for women at baseline not adjusted';
  title5 'Supplemental table 1';

proc means data = baseline;
  where vis = 0 and gender = 1;
  var tkv_height cic_new;
  title4 'TKV for males at baseline adjusted for height';
  title5 'Supplemental table 1';

proc means data = baseline;
  where vis = 0 and gender = 0;
  var tkv_height cic_new ;
  title4 'TKV for women at baseline adjusted for height';
  title5 'Supplemental table 1';

proc means data = baseline;
  where vis = 0;
  var tkv_height cic_new;
  title4 'TKV for both genders at baseline adjusted for height';
  title5 'Supplemental table 1';

proc means data = baseline;
  where vis = 0 and gender = 1;
  var tkv_weight cic_new;
  title4 'TKV for males at baseline adjusted for weight';
```

```
  title5 'Supplemental table 1';

proc means data = baseline;
  where vis = 0 and gender = 0;
  var tkv_weight cic_new;
  title4 'TKV for women at baseline adjusted for weight';
  title5 'Supplemental table 1';

proc means data = baseline;
  where vis = 0;
  var tkv_weight cic_new;
  title4 'TKV for both genders at baseline adjusted for weight';
  title5 'Supplemental table 1';

proc means data = baseline;
  where vis = 0 and gender = 1;
  var tkv_weight cic_new;
  title4 'TKV for males at baseline adjusted for BSA';
  title5 'Supplemental table 1';

proc means data = baseline;
  where vis = 0 and gender = 0;
  var tkv_weight cic_new;
  title4 'TKV for women at baseline adjusted for BSA';
  title5 'Supplemental table 1';

proc means data = baseline;
  where vis = 0;
  var tkv_weight cic_new;
  title4 'TKV for both genders at baseline adjusted for BSA';
  title5 'Supplemental table 1';

proc means data = baseline;
  where vis = 0 and gender = 1;
  var tkv_weight cic_new;
  title4 'TKV for males at baseline adjusted for BMI';
  title5 'Supplemental table 1';

proc means data = baseline;
  where vis = 0 and gender = 0;
  var tkv_weight cic_new;
  title4 'TKV for women at baseline adjusted for BMI';
  title5 'Supplemental table 1';

proc means data = baseline;
  where vis = 0;
  var tkv_weight cic_new;
  title4 'TKV for both genders at baseline adjusted for BMI';
  title5 'Supplemental table 1';

************************;
*  Supplemental table 2  *;
************************;
proc sort data = analy;
  by pkdid vis;
```

```
data analy_coll;
  set analy end=eof;
  by pkdid;
  where cic_new ^= . and tkvhb ^= .;

  file print;

  array viss (6) (0,10,20,30,60,80);
  array has (6) has0 has10 has20 has30 has60 has80;
  array hbs (6) tkvhb0 tkvhb10 tkvhb20 tkvhb30 tkvhb60 tkvhb80;
  array cics (6) cic_new0 cic_new10 cic_new20 cic_new30 cic_new60 cic_new80;

  array sumhas  (6) sum_has0    sum_has10    sum_has20    sum_has30    sum_has60    sum_has80;
  array sumhbs  (6) sum_tkvhb0  sum_tkvhb10  sum_tkvhb20  sum_tkvhb30  sum_tkvhb60  sum_tkvhb80;
  array sumcics  (6) sum_cic_new0  sum_cic_new10  sum_cic_new20  sum_cic_new30  sum_cic_new60  sum_cic_new80;

  array meanhbs  (6) mean_tkvhb0  mean_tkvhb10  mean_tkvhb20  mean_tkvhb30  mean_tkvhb60  mean_tkvhb80;
  array meancics  (6) mean_cic_new0  mean_cic_new10  mean_cic_new20  mean_cic_new30  mean_cic_new60  mean_cic_new80;

  array allhbs  (6) sum_tkvhb0_all  sum_tkvhb10_all  sum_tkvhb20_all  sum_tkvhb30_all  sum_tkvhb60_all  sum_tkvhb80_all;
  array allcics  (6) sum_cic_new0_all  sum_cic_new10_all  sum_cic_new20_all  sum_cic_new30_all  sum_cic_new60_all  sum_cic_new80_all;

  array meanhbsa  (6) mean_tkvhb0_all  mean_tkvhb10_all  mean_tkvhb20_all  mean_tkvhb30_all  mean_tkvhb60_all  mean_tkvhb80_all;
  array meancisa  (6) mean_cic_new0_all  mean_cic_new10_all  mean_cic_new20_all  mean_cic_new30_all  mean_cic_new60_all  mean_cic_new80_all;

  retain has hbs cics sumhas sumhbs sumcics num_hasall allcics allhbs;

  if _N_ = 1 then do i = 1 to 6;
    sumhas[i]  = 0;
    sumhbs[i]  = 0;
    sumcics[i] = 0;
    allhbs[i] = 0;
    allcics[i] = 0;
    num_hasall = 0;
  end;

  if first.pkdid then do i = 1 to 6;
    has[i] = 0;
    hbs[i] = .;
    cics[i] = .;
  end;

  do i = 1 to 6;
    if vis = viss[i] then do;
      has[i] = 1;
      hbs[i] = tkvhb;
      cics[i] = cic_new;
    end;
  end;

  if last.pkdid then do;
    tothas = 0;
    hasall = 0;
    do i = 1 to 6;
      if has[i] = 1 then do;
        sumhas[i] = sumhas[i] + has[i];
```

```
        sumhbs[i] = sumhbs[i] + hbs[i];
        sumcics[i] = sumcics[i] + cics[i];
        tothas = tothas + 1;
      end;
    end;
    if tothas = 6 then do;
      hasall = 1;
      num_hasall = num_hasall + 1;
      do i = 1 to 6;
        allhbs[i] = allhbs[i] + hbs[i];
        allcics[i] = allcics[i] + cics[i];
      end;
    end;
    output analy_coll;
  end;

  if eof then do;
    put 'Supplemental Table 2. Baseline through year 8 visit htTKV and GFR values and percentage change in htTKV and GFR' /;
    put 'htTKV' / ;
    put @1 'Visit' @30 'N' @40 'Mean htTKVb' @60 'Mean GFR' @80 'Mean htTKVb all visits' @115 'Mean htTKVb all visits';
    do i = 1 to 6;
      meanhbs[i]  = sumhbs[i]  / sumhas[i];
      meancics[i]  = sumcics[i]  / sumhas[i];
      meanhbsa[i]  = allhbs[i]  / num_hasall;
      meancisa[i]  = allcics[i]  / num_hasall;
      put @1 viss[i] @30 sumhas[i] @40 meanhbs[i] @60 meancics[i] @80 meanhbsa[i] @115 meancisa[i];
    end;
    put / 'Number of patients who has all visits: ' num_hasall;
  end;

  format mean_tkvhb0  mean_tkvhb10  mean_tkvhb20  mean_tkvhb30  mean_tkvhb60  mean_tkvhb80
         mean_cic_new0  mean_cic_new10  mean_cic_new20  mean_cic_new30  mean_cic_new60  mean_cic_new80
         mean_tkvhb0_all  mean_tkvhb10_all  mean_tkvhb20_all  mean_tkvhb30_all  mean_tkvhb60_all  mean_tkvhb80_all
         mean_cic_new0_all  mean_cic_new10_all  mean_cic_new20_all  mean_cic_new30_all  mean_cic_new60_all  mean_cic_new80_all 8.1;

data analy_mod;
  merge analy analy_coll (keep = pkdid hasall);
  by pkdid;

proc sort data = analy;
  by vis;

proc means data = analy;
  where (vis in(0,10,20,30,60,80)) and (cic_new ^= .) and (tkvhb ^= .);
  by vis;
  var tkvhb
      cic_new;
  title4 'Height adjusted TKV by visit';
  title5 'Supplemental table 2';

proc sort data = analy_mod;
  by vis;

proc means data = analy_mod;
  where (vis in(0,10,20,30,60,80)) and (cic_new ^= .) and (tkvhb ^= .) and (hasall = 1);
  by vis;
```

```
var tkvhb
    cic_new;
title4 'Height adjusted TKV by visit, subset to patients who had all visits';
title5 'Supplemental table 2';
```

```
title1 "Program Saved As: /prj/niddk/ims_analysis/CRISP2/prog_initial_analysis/recreate.Revised.CJASN.final.paper.m08d06y2014.sas";
title2 "NIDDK Effort: CRISP II study - Check Revised.CJASN.final.paper.m05d05y2014 paper";

options nofmterr;

/*************************************************************************************************************************************
 Programmer: Dave Ruggieri
 Date: 30 May 2014
 Billing Code: DDKDR1.T01

 Function/Notes: This program recreates the numbers found on Chapman CJASN 032012.pdf paper. It is titled Kidney volume and
 Functional Outcomes in Autosomal Dominat Polycystic Kidney disease.
*************************************************************************************************************************************
 Programmer: Dave Ruggieri
 Date: 27 June 2014
 Billing Code: DDKDR1.T01

 Function/Notes: Confirming Doug's responses to 1 and 2 below.

*************************************************************************************************************************************
 Programmer: Dave Ruggieri
 Date: 06 August 2014
 Billing Code: DDKDR1.T01

Function/notes: This a re-run trying to calculate eGFR numbers from table 1.
*************************************************************************************************************************************/
*  Input Files  *;
*****************;
*** Primary analysis dataset ***;
filename analy pipe "gunzip -c /prj/niddk/ims_analysis/CRISP2/private_created_data/sasfile/CJASN_Revised_April2014.v9x.gz";
proc cimport data=analy infile=analy;
run;

proc contents data = analy;
  title4 'Contents of analysis dataset';

*** Long 2 dataset ***;
libname sas_data '/prj/niddk/ims_analysis/CRISP2/private_orig_data/5_-_Other_data_files/';

data long1;
  set sas_data.crisp2_long_04172012_1     ;

data long3;
  set sas_data.crisp2_long_04172012_3     ;

/*proc contents data = long1;
  title4 'Contents of long1 dataset';*/
*************;
*  Formats  *;
*************;
proc format;
  value genfmt /* gender */
    0 = '0=Female'
    1 = '1=Male'
    ;
```

```
  value racefmt /* racewhite */
    1 = '1=White'
    0 = '0=Non-white'
    ;

  value stagefmt /* gfrstg egfrstg */
    1 = '1=Stage 1'
    2 = '2=Stage 2'
    3 = '3=Stage 3'
    4 = '4=Stage 4'
    ;

  value agegrp /* gfr20_age_grp gfr40_age_grp gfrstg3_age_grp gfrstg4_age_grp */
    3 = "3=15-20"
    4 = "4=20-25"
    5 = "5=25-30"
    6 = "6=30-35"
    7 = "7=35-40"
    8 = "8=40+"
    ;

  value visfmt /* vis */
    0  = 'BV01:Baseline Visit'
    10 = 'FV01:Clinic Visit Year 1'
    20 = 'FV02:Clinic Visit Year 2'
    30 = 'FV03:Clinic Visit Year 3'
    60 = 'FV06:Biannual Clinic Visit Year 6'
    65 = 'FV06.6:Semi-Annual Phone Call post  Year 6 visit'
    70 = 'FV07:Lab Visit Year 7 '
    75 = 'FV07.6:Semi-Annual Phone Call post  Year 7 visit'
    80 = 'FV08:Biannual Clinic Visit Year 8'
    85 = 'FV08.6:Semi-Annual Phone Call post Year 8 visit'
    90 = 'FV09:Lab Visit Year 9 '
    95 = 'FV09.6:Semi-Annual Phone Call post  Year 9 visit'
    96 = 'N/A:6 months or more post vis 95'
    99 = 'As needed'
    40 = 'Unknown: Clinic visit year 4?'
    45 = 'Unknown: Phone call post year 4?'
    50 = 'Unknown: Clinic visit year 5?'
    ;

  value stgcheck
    . = 'Missing'
    low-29.999999 = '<30'
    30-59.999999 = '30-<60'
    60-89.999999 = '60-<90'
    90-high = '90+'
    ;

  value missfmt
    . = 'Missing'
    other = 'Other'
    ;

**********************;
*  Table 1: New code  *;
```

```
**********************;

/*
Based on the following STATA code found in /prj/niddk/ims_analysis/CRISP2/private_created_data/CJASN_Table 1 Stata Code/CJASN without HALT follow-up_for
NIDDK.do

557   *[TABLE 1] ~~~~~~~
558   *FREQUENCIES FOR GFR AND EGFR AT BASELINE AND YEAR 8
559   *-----------------------------------------------------
560   use tmp_cjasn, clear
561
562   **GFR
563   tab gfrstg if vis == 0
564   tab gfr20 if vis == 0
565   tab gfr40 if vis == 0
566
567   *"year 8" values are based on last visit gfr stage
568   keep if gfrstg != .
569   bysort pkdid: egen maxvis = max(vis)
570   bysort pkdid: gen laststg = gfrstg if vis == maxvis
571   tab laststg
572
573    *determine the average follow-up for the gfrstg3 participants...(page 3)
574    sum visc if laststg >2 & laststg <.
575    /*
576           Variable |       Obs        Mean    Std. Dev.       Min        Max
577    -------------+--------------------------------------------------------
578                  visc |        80    7.790761     2.17608          0   10.06982
579    *|
580
581
582   **eGFR
583   use tmp_cjasn, clear
584
585   tab egfrstg if vis == 0
586   tab egfr20 if vis == 0
587   tab egfr40 if vis == 0
588
589   *"year 8" values are based on last visit egfr stage
590   keep if egfrstg != .
591   bysort pkdid: egen maxvis = max(vis)
592   bysort pkdid: gen laststg = egfrstg if vis == maxvis
593   tab laststg
*/
proc freq data = analy;
  where vis = 0;
  tables gfrstg
         gfr20
         gfr40
         /missing list;
  title4 'Enrollment gfr numbers';

proc sort data = analy out=analygfr;
  where gfrstg ^= .;
  by pkdid vis;
```

```
proc freq data = analygfr;
  tables vis /missing list;
  title4 'Visits that have gfr';

data analy_lastvis checklast;
  set analygfr;
  by pkdid vis;

  if last.pkdid then do;
    lastvis = 1;
    output;
  end;
  else output checklast;

proc print data = checklast (obs = 50);
  var pkdid vis lastvis gfrstg;
  title4 'Checking last visit?';

proc freq data = analy_lastvis;
  tables vis
         gfrstg
         gfr20
         gfr40
         /missing list;
  title4 '8 year gfr numbers';

proc freq data = analy;
  where vis = 0;
  tables egfrstg
         egfr20
         egfr40
         /missing list;
  title4 'Enrollment egfr numbers';

data analy_lastvis;
  set analy;
  where egfrstg ^= .;
  by pkdid;
  if last.pkdid;

proc freq data = analy_lastvis;
  tables vis
         egfrstg
         egfr20
         egfr40
         /missing list;
  title4 'Last visit egfr numbers';
```