Norma Pugh
January 10, 2008

Integrity Check for the Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) Baseline Analysis File

As a partial check of the integrity of the CRISP baseline analysis dataset archived in the NIDDK data repository, a set of tabulations was performed to verify that published results can be reproduced using the archived dataset. Analyses were performed to duplicate published results for the data reported by Chapman et al [1] in *Kidney International* in September 2003. The results of this integrity check are described below. The full text of the *Kidney International* article can be found in Attachment 1, and the SAS code for our tabulations is included in Attachment 2.

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.

Background. This five-year prospective cohort study was designed to determine if changes in anatomic characteristics of the kidneys of patients with polycystic kidney disease (PKD) as measured by radiologic imaging techniques are useful in providing surrogate measures for disease progression [2].

Comprising four participating clinical centers and a data-coordinating and imaging-analysis center, the consortium has developed and is implementing studies nationwide to test whether imaging techniques can provide accurate and reproducible markers of progression of renal disease in patients with PKD. Participating clinical centers are Emory University, the Mayo Clinic, University of Kansas, and the University of Alabama at Birmingham. The data coordinating and imaging analysis center is at Washington University [2].

Over the five-year period of CRISP, several cohorts of patients, at different stages of disease and with varying rates of disease progression, were studied in interrelated investigations [2].

The baseline paper reports on longitudinal observations of autosomal-dominant polycystic kidney disease (ADPKD) individuals. ADPKD is characterized by gradual renal enlargement and cyst growth prior to loss of renal function. Standard radiographic imaging has not provided the resolution and accuracy necessary to detect small changes in renal volume or to reliably measure renal cyst volumes. CRISP longitudinally observed ADPKD individuals using high-resolution magnetic resonance imaging to determine if change in renal and cyst volumes can be detected over a short period of time, and if they correlate with decline in renal function early in disease [1].

Preliminary Review. Review of the baseline paper [1] determined Tables 1-3 were not suitable for replication. Table 1, True and measured whole (renal) and balloon (cystic) volumes in large and small phantoms, relies on phantom data which the Data Coordinating Center (DCC) confirms is not part of the

dataset suite housed at the repository. Table 2, Interclass coefficients (reliability measures) of clinical variables measured in the standardization protocol subjects, and Table 3, Mean renal and cyst volumes of four subjects measured at each participating clinical center, summarize data on four subjects used for standardization. Information about these four subjects is not included in the datasets currently housed at the repository.

Demographic and Clinical Characteristics. Tables 4 and 5 of the baseline paper [1] report on demographic and clinical characteristics overall, by gender and by hypertensive status. All variables summarized in these baseline tables are taken from the CHAPMAN2003 analysis dataset created for this study. Table A lists the variables we used in our replication of these variables.

Table A: Variables Used to Replicate Tables 4 and 5

Table Variable	Variables Used in Replication
Sample size	Overall: used all records; By gender: sex; By hypertensive status: hdyn
Age (years)	age
Weight (kg)	weight_c
BMI (m^2)	bmi_c
SBP (mm Hg)	systold
DBP (mm Hg)	diastold
MAP (mm Hg)	mapn
Serum creatinine (mg/dL)	serumcreat
GFR (mL/mm/1.73 m^2)	cic
Sodium excretion (mEq/day)	esode_cc
Potassium excretion (mEq/day)	epote_ca
Albumin excretion (mg/day)	albe_ca
Total cholesterol (mg/dL)	lptce_ca
HDL (mg/dL)	lphdle_ca
LDL (mg/dL)	lpldle_ca
Triglycerides (mg/dL)	lptri

In Table B, we compare the results for characteristics calculated from the archived dataset to the results published in Table 4, Clinical characteristics of autosomal-dominant polycystic kidney disease (ADPKD) men and women enrolled in the Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) Cohort. As Table B shows, the results obtained from the archived data are similar to those in the published tabulations (see Note 1 below, regarding the discrepancies).

Table B: Comparison of Values Computed in Integrity Check to Reference Article Table 4 Values

Table Variable	Group: Overall		
	Chapman et al (2003)	Integrity Check	Difference
Age (years)	33.8 ± 8.9	32.4 ± 8.9	1.4 ± 0
Weight (kg)	77.0 ± 18.4	77.0 ± 18.4	0
BMI (m^2)	25.9 ± 5.2	25.9 ± 5.3	0 ± 0.1
SBP (mm Hg)	125.1 ± 13.6	125.1 ± 13.8	0 ± 0.2
DBP (mm Hg)	83.4 ± 10.7	83.4 ± 10.7	0
MAP (mm Hg)	98.2 ± 10.9	98.3 ± 11.0	0.1 ± 0.1
Serum creatinine (mg/dL)	1.0 ± 0.2	1.0 ± 0.2	0
GFR (mL/mm/1.73 m^2)	98.2 ± 24.9	98.0 ± 24.7	0.2 ± 0.2
Sodium excretion (mEq/day)	185.5 ± 87.2	185.9 ± 86.5	0.4 ± 0.7
Potassium excretion (mEq/day)	56.5 ± 24.7	56.4 ± 24.7	0.1 ± 0
Albumin excretion (mg/day)	42.4 ± 60.9	41.8 ± 59.8	0.6 ± 1.1
Total cholesterol (mg/dL)	172.9 ± 37.0	174.4 ± 35.4	1.5 ± 1.6
HDL (mg/dL)	49.2 ± 22.6	47.7 ± 13.3	1.5 ± 9.3
LDL (mg/dL)	102.4 ± 34.3	101.7 ± 34.1	0.7 ± 0.2
Triglycerides (mg/dL)	120.2 ± 95.0	121.1 ± 92.8	0.9 ± 2.2

Table Variable	Group: Women		
	Chapman et al (2003)	Integrity Check	Difference
Sample size	145	145	0
Age (years)	34.0 ± 9.0	32.6 ± 9.0	1.4 ± 0
Weight (kg)	69.7 ± 15.8	69.7 ± 15.8	0
BMI (m^2)	25.5 ± 5.6	25.5 ± 5.6	0
SBP (mm Hg)	122.7 ± 14.0	122.8 ± 14.2	0.1 ± 0.2
DBP (mm Hg)	83.2 ± 11.0	83.3 ± 11.0	0.1 ± 0
MAP (mm Hg)	97.4 ± 11.5	97.6 ± 11.7	0.2 ± 0.2
Serum creatinine (mg/dL)	0.9 ± 0.2	0.9 ± 0.2	0
GFR (mL/mm/1.73 m^2)	99.8 ± 26.6	99.4 ± 26.2	0.4 ± 0.4
Sodium excretion (mEq/day)	162.0 ± 74.8	162.7 ± 73.7	0.7 ± 1.1
Potassium excretion (mEq/day)	51.1 ± 21.3	50.9 ± 21.3	0.2 ± 0
Albumin excretion (mg/day)	39.4 ± 48.2	39.1 ± 47.2	0.3 ± 1.0
Total cholesterol (mg/dL)	169.2 ± 34.1	169.8 ± 32.5	0.6 ± 1.6
HDL (mg/dL)	51.8 ± 12.5	51.6 ± 12.4	0.2 ± 0.1
LDL (mg/dL)	97.6 ± 28.9	97.0 ± 28.5	0.6 ± 0.4
Triglycerides (mg/dL)*	120.4 ± 63.0	102.4 ± 61.3	18.0 ± 1.7
* The DCC believes published results may have been transposed for this variable.			

Table B: Comparison of Table 4 Values Computed in Integrity Check to Reference Article Values (cont.)

Table Variable	Group: Men		
	Chapman et al (2003)	Integrity Check	Difference
Sample size	96	96	0
Age (years)	33.5 ± 8.8	32.0 ± 8.8	1.5 ± 0
Weight (kg)	88.1 ± 16.6	88.1 ± 16.6	0
BMI (m^2)	26.5 ± 4.7	26.5 ± 4.7	0
SBP (mm Hg)	128.5 ± 12.3	128.5 ± 12.3	0
DBP (mm Hg)	83.6 ± 10.4	83.6 ± 10.4	0
MAP (mm Hg)	99.4 ± 9.7	99.4 ± 9.7	0
Serum creatinine (mg/dL)	1.1 ± 0.2	1.1 ± 0.2	0
GFR (mL/mm/1.73 m^2)	95.8 ± 22.1	95.8 ± 22.1	0
Sodium excretion (mEq/day)	220.9 ± 92.8	220.9 ± 92.8	0
Potassium excretion (mEq/day)	64.7 ± 27.3	64.7 ± 27.3	0
Albumin excretion (mg/day)	47.0 ± 76.2	46.0 ± 75.3	1.0 ± 0.9
Total cholesterol (mg/dL)	178.6 ± 40.6	181.3 ± 38.4	2.7 ± 2.2
HDL (mg/dL)	45.3 ± 31.96	41.9 ± 12.4	3.4 ± 19.56
LDL (mg/dL)	109.8 ± 40.2	108.9 ± 40.3	0.9 ± 0.1
Triglycerides (mg/dL)	147.5 ± 125.0	148.8 ± 121.0	1.3 ± 4.0

Table Variable	P-values (Women vs Men)		
	Chapman et al (2003)	Integrity Check	Difference
Age (years)	0.6500	0.6186	0.0314
Weight (kg)	<0.0001	<0.0001	0
BMI (m^2)	0.1230	0.1230	0
SBP (mm Hg)	0.0011	0.0016	0.0005
DBP (mm Hg)	0.7836	0.8032	0.0196
MAP (mm Hg)	0.1773	0.2064	0.0291
Serum creatinine (mg/dL)	<0.0001	<0.0001	0
GFR (mL/mm/1.73 m^2)	0.2199	0.2642	0.0443
Sodium excretion (mEq/day)	<0.0001	<0.0001	0
Potassium excretion (mEq/day)	<0.0001	<0.0001	0
Albumin excretion (mg/day)	0.3961	0.3905	0.0056
Total cholesterol (mg/dL)	0.0559	0.0146	0.0413
HDL (mg/dL)	0.0674	<0.0001	0.0673
LDL (mg/dL)	0.0130	0.0112	0.0018
Triglycerides (mg/dL)	0.0026	0.0001	0.0025

In Table C, we compare the results for characteristics calculated from the archived dataset to the results published in Table 5, Clinical characteristics of hypertensive and normotensive autosomal-dominant polycystic kidney disease (ADPKD) individuals enrolled in the Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) Cohort. As Table C shows, the results obtained from the archived data are similar to those in the published tabulations (see Note 1 below, regarding the discrepancies).

Table C: Comparison of Table 5 Values Computed in Integrity Check to Reference Article Values

Table Variable	Group: Normotensives		
	Chapman et al (2003)	Integrity Check	Difference
Sample size	93	93	0
Age (years)	31.0 ± 9.3	29.7 ± 9.3	1.3± 0
Weight (kg)	73.6 ± 15.7	73.3 ± 15.6	0.3 ± 0.1
BMI (m^2)	24.8 ± 4.5	24.7 ± 4.5	0.1 ± 0
SBP (mm Hg)	121.3 ± 13.5	120.8 ± 13.1	0.5 ± 0.4
DBP (mm Hg)	79.9 ± 10.4	79.6 ± 10.0	0.3 ± 0.4
MAP (mm Hg)	94.5 ± 10.8	94.2 ± 10.6	0.3 ± 0.2
Serum creatinine (mg/dL)	0.9 ± 0.2	0.9 ± 0.2	0
GFR (mL/mm/1.73 m^2)	106.8 ± 20.2	106.2 ± 19.7	0.6 ± 0.5
Sodium excretion (mEq/day)	183.1 ± 86.8	182.9 ± 85.1	0.2 ± 1.7
Potassium excretion (mEq/day)	56.5 ± 28.2	56.2 ± 28.2	0.3 ± 0
Albumin excretion (mg/day)	33.0 ± 45.2	32.0 ± 43.6	1.0 ± 1.6
Total cholesterol (mg/dL)	168.2 ± 36.7	169.6 ± 34.6	1.4 ± 2.1
HDL (mg/dL)	48.6 ± 13.7	48.5 ± 13.5	0.1 ± 0.2
LDL (mg/dL)	101.3 ± 33.0	98.4 ± 31.2	2.9 ± 1.8
Triglycerides (mg/dL)	106.5 ± 99.9	108.5 ± 95.0	2.0 ± 4.9

Table Variable	Group: Hypertensives		
	Chapman et al (2003)	Integrity Check	Difference
Sample size	148	148	0
Age (years)	35.5 ± 8.3	34.0 ± 8.2	1.5 ± 0.1
Weight (kg)	79.2 ± 19.6	79.4 ± 19.7	0.2 ± 0.1
BMI (m^2)	26.6 ± 5.6	26.6 ± 5.6	0
SBP (mm Hg)	127.4 ± 13.2	127.8 ± 13.4	0.4 ± 0.2
DBP (mm Hg)	85.5 ± 10.4	85.8 ± 10.5	0.3 ± 0.1
MAP (mm Hg)	100.4 ± 10.3	100.8 ± 10.5	0.4 ± 0.2
Serum creatinine (mg/dL)	1.0 ± 0.2	1.0 ± 0.2	0
GFR (mL/mm/1.73 m^2)	92.8 ± 26.1	92.8 ± 26.1	0
Sodium excretion (mEq/day)	187.0 ± 87.7	187.7± 87.6	0.7 ± 0.1
Potassium excretion (mEq/day)	56.5 ± 22.5	56.5 ± 22.5	0
Albumin excretion (mg/day)	48.4 ± 68.4	48.1 ± 67.7	0.3 ± 0.7
Total cholesterol (mg/dL)	175.9 ± 37.1	177.4 ± 35.6	1.5 ± 1.5
HDL (mg/dL)	49.6 ± 26.7	47.2 ± 13.1	2.4 ± 13.6
LDL (mg/dL)	103.2 ± 35.1	103.7 ± 35.6	0.5 ± 0.5
Triglycerides (mg/dL)	128.2 ± 91.4	128.8 ± 90.8	0.6 ± 0.6

Table C: Comparison of Table 5 Values Computed in Integrity Check to Reference Article Values (cont.)

Table Variable	P-values (Normotensives vs Hypertensives)		
	Chapman et al (2003)	Integrity Check	Difference
Age (years)	<0.0001	0.0002	0.0001
Weight (kg)	0.0157	0.0122	0.0035
BMI (m^2)	0.0049	0.0060	0.0029
SBP (mm Hg)	0.0008	<0.0001	0.0007
DBP (mm Hg)	<0.0001	<0.0001	0
MAP (mm Hg)	<0.0001	<0.0001	0
Serum creatinine (mg/dL)	<0.0001	<0.0001	0
GFR (mL/mm/1.73 m^2)	<0.0001	<0.0001	0
Sodium excretion (mEq/day)	0.7431	0.6820	0.0611
Potassium excretion (mEq/day)	0.9990	0.9297	0.0693
Albumin excretion (mg/day)	0.0430	0.0482	0.0052
Total cholesterol (mg/dL)	0.1260	0.1049	0.0211
HDL (mg/dL)	0.7111	0.4583	0.2528
LDL (mg/dL)	0.6830	0.2713	0.4117
Triglycerides (mg/dL)	0.1073	0.1058	0.0015

Notes

1. The discrepancies documented in this report are likely due to data corrections and updates made between the paper data freeze and the final data freeze. The DCC has confirmed that the appropriate variables were used for this replication analysis.

2. The baseline analysis dataset, CHAPMAN2003, was examined in this replication analysis. In addition to this baseline dataset, the repository houses raw datasets and two additional analysis datasets from the CRISP cohort. An additional replication analysis will be completed to verify the integrity of at least one other analysis dataset.

3. The SAS datasets provided to the NIDDK Data Repository are in an archival format. In order to use SAS Viewer, limit CPU resources and increase performance when using these datasets, they must be converted back to an un-archived state. One method to do this is via PROC MIGRATE, as follows:

```
/* Location of Archived CRISP SAS Data Files */
LIBNAME OLD 'R:\CRISP\CRISP_20070706';

/* Location for Un-archived CRISP SAS Data Files */
LIBNAME NEW 'R:\CRISP\CRISP_20070706\MigratedData';

/* Migrate the datasets */
PROC MIGRATE IN=OLD OUT=NEW; RUN;
```

   Un-archived versions of all the archived datasets in the 'OLD' location will then be created in the 'NEW' location.

References

1. Arlene B. Chapman, Lisa M. Guay-Woodford, Jared J. Grantham, Vicente E. Torres, Kyongtae T. Bae, Deborah A. Baumgarten, Philip J. Kenney, Bernard F. King, Jr., James F. Glockner, Louis H. Wetzel, Marijn E. Brummer, W. Charles O-Neil, Michelle L. Robbin, William M. Bennett, Saulo Klahr, Gladys H. Hirschman, Paul L. Kimmel, Paul A. Thompson, and J. Philip Miller, **Renal structure in early autosomal-dominant polycystic kidney disease (ADPKD): The Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) cohort**, Kidney International, Volume 64 (2003), pp. 1035-1045.

2. NIDDK Website: CRISP page. Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) : NIDDK

Attachment 1

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

Arlene B. Chapman, Lisa M. Guay-Woodford, Jared J. Grantham, Vicente E. Torres, Kyongtae T. Bae, Deborah A. Baumgarten, Philip J. Kenney, Bernard F. King, Jr., James F. Glockner, Louis H. Wetzel, Marijn E. Brummer, W. Charles O-Neil, Michelle L. Robbin, William M. Bennett, Saulo Klahr, Gladys H. Hirschman, Paul L. Kimmel, Paul A. Thompson, and J. Philip Miller, Renal structure in early autosomal-dominant polycystic kidney disease (ADPKD): The Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) cohort, Kidney International, Volume 64 (2003), pp. 1035-1045.

ATTACHMENT 2

SAS Code for Tabulations from the CRISP Baseline Analysis Dataset in the NIDDK Repository

```
/*****************************************************************************/
/*
/* Program: R:\05_Users\Norma\CRISP\BaselinePaper\Updated\descriptives.sas
/* Author:  Norma Pugh
/* Date:    09 January 2008
/* Purpose: Table replication results for CRISP baseline paper -
/*           Chapman et al: Renal structure in early ADPKD: Kindney Int. 2003
/*           Tables 4 and 5
/*           NOTE: Tables 1-3 cannot be replicated (see 7/10 e-mail from P. Thompson)
/*
/*****************************************************************************/
/****************************************/
/* Location of NIDDK Repository SAS files */
/****************************************/
libname data 'R:\03_Data_And_Tools\Studies\CRISP\DCC_Delivery\CRISP_20080104';

/***********/
/* Formats */
/***********/
%include
'R:\03_Data_And_Tools\Studies\CRISP\DCC_Delivery\CRISP_20070505\rti\documents\pkdformat.sas';
options nofmterr;

/************/
/* Get data */
/************/
data test; set data.chapman2003; run;

/***********/
/* TABLE 4 */
/***********/
/* Overall counts */
title'CRISP Baseline Paper: Table 4'; run;
title2'Overall counts'; run;

proc means mean std;
 var age weight_c bmi_c systold diastold mapn serumcreat cic esode_cc epote_ca albe_ca lptce_ca
lphdle_ca lpldle_ca lptri;
run;

/* Counts & p-values, by gender */
title2'Counts by Gender'; run;

proc sort; by sex; run;

proc freq; tables sex; run;

proc means mean std; by sex;
 var age weight_c bmi_c systold diastold mapn serumcreat cic esode_cc epote_ca albe_ca lptce_ca
lphdle_ca lpldle_ca lptri;
run;

title2'P-values by Gender'; run;

proc anova;
 class sex;
```

```
 model age weight_c bmi_c systold diastold mapn serumcreat cic esode_cc epote_ca albe_ca lptce_ca
lphdle_ca lpldle_ca lptri = sex;
run;

/*
NOTE: T-test confirms ANOVA results
proc ttest; class sex;
var age weight_c bmi_c systold diastold mapn serumcreat cic esode_cc epote_ca albe_ca lptce_ca
lphdle_ca lpldle_ca lptri;
run;


/*****************************************/
/* Counts & p-values, by hypertensive status */
/*****************************************/
title'CRISP Baseline Paper: Table 5'; run;
title2'Counts by Hypertensive Status'; run;

proc sort; by hdyn; run;

proc freq; tables hdyn; run;

proc means mean std; by hdyn;
 var age weight_c bmi_c systold diastold mapn serumcreat cic esode_cc epote_ca albe_ca lptce_ca
lphdle_ca lpldle_ca lptri;
run;

title2'P-values by Hypertensive Status'; run;

proc anova;
 class hdyn;
 model age weight_c bmi_c systold diastold mapn serumcreat cic esode_cc epote_ca albe_ca lptce_ca
lphdle_ca lpldle_ca lptri = hdyn;
run;
```