Norma Pugh
May 17, 2008

Integrity Check for the Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) Analysis File, EXPORT102005

As a partial check of the integrity of the CRISP analysis datasets archived in the NIDDK data repository, a set of tabulations was performed to verify that published results can be reproduced using the archived datasets. Analyses were performed to duplicate published results for the data reported by Rule et al [1] in the *Journal of the American Society of Nephrology* in December 2005. The results of this integrity check are described below. The full text of the *Journal of the American Society of Nephrology* 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 implemented 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, the 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 the CRISP study, several cohorts of patients, at different stages of disease and with varying rates of disease progression, were studied in interrelated investigations [2].

The Rule paper compares methods for a decline in renal function for a cohort with early autosomal dominant polycystic kidney disease (ADPKD) [1].

Demographic and Baseline Characteristics. Table 1 [1] reports on demographic and baseline characteristics. All variables summarized are taken from the EXPORT102005 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 Table 1

Table Variable	Variables Used in Replication
age	age, where 'vis' variable=0
female	sex, where 'vis' variable=0
white, black	race, where 'vis' variable=0
weight	weight_c, where 'vis' variable=0
height	height_c, where 'vis' variable=0
hypertension	hdyn, where 'vis' variable=0
bilateral kidney volume	mrskvs, where 'vis' variable=0
bilateral cyst volume	mrrcvs, where 'vis' variable=0 (see Note 1 below)
albumin to creatinine ratio	albe_ca / creatclr, where 'vis' variable=0
current smoker	csyn, where 'vis' variable=0
history of urinary tract infection	ludyn, where 'vis' variable=0
abdominal pain	freqrp, where 'vis' variable=0
gross hematuria	ghdyn, where 'vis' variable=0
unstandardized iothalamate clearance	uic, where 'vis' variable=0
standardized iothalamate clearance	cic_c, where 'vis' variable=0
SCr	serumcreat, where 'vis' variable=0
MDRD equation	mdrd_gfr_c, where 'vis' variable=0
Cockcroft-Gault equation	cc_cg, where 'vis' variable=0
creatinine clearance	cc_su, where 'vis' variable=0

In Table B, we compare the results for characteristics calculated from the archived dataset to the results published in the results section. As Table B shows, most results obtained from the archived data are similar to those in the published tabulations (see Note 2 below regarding the discrepancies). Additionally, all variables examined in the published paper [1] are summarized in this baseline table.

Table B: Comparison of Values Computed in Integrity Check to Reference Article Table 1 Values

Table Variable	Group: Overall		
	Rule et al (2005)	**Integrity Check**	**Difference**
age (yr)	34 (25 to 40)	34 (25 to 40)	0
female	60% (140)	60% (145)	0 (5)
white	88% (206)	85% (206)	3 (0)
black	10% (24)	12% (28)	2 (4)
weight (kg)	74 (61 to 91)	74 (61 to 90)	0 (0,1)
height (cm)	170 (163 to 181)	170 (164 to 181)	0 (1,0)
hypertension	61% (143)	62% (149)	1 (6)
bilateral kidney volume (ml)	865(585 to 1340)	868(580 to 1342)	3 (5,2)
bilateral cyst volume (ml)	320 (166 to 727)	329 (166 to 736)	9 (0,9)
albumin to creatinine ratio (mg/g)	25 (11 to 49)	27 (11 to 51)	2 (0,2)
current smoker	17% (40)	17% (40)	0
history of urinary tract infection	45% (104)	45% (108)	0 (4)
abdominal pain	61% (142)	60% (145)	1 (3)
gross hematuria	32% (76)	33% (79)	1 (3)
unstandardized iothalamate clearance (ml/min)	107 (86 to 123)	107 (86 to 123)	0
standardized iothalamate clearance (ml/min per 1.73 m^2)	95 (79 to 115)	95 (79 to 115)	0
SCr (mg/dl)	1.03 (0.82 to 1.21)	0.90 (0.80 to 1.10)	0.13 (0.02,0.11)
MDRD equation (ml/min per 1.73 m^2)	79 (63 to 96)	77 (63 to 96)	2 (0,0)
Cockcroft-Gault equation (ml/min)	101 (82 to 126)	108 (88 to 129)	7 (6,3)
creatinine clearance (ml/min)	109 (89 to 130)	110 (90 to 131)	1 (1,1)
Note: Results given as percentage (count) or median (25^{th} to 75^{th} percentile).			

Finally, Appendix A documents the data issues to be aware of when using the EXPORT102005 analysis dataset. The Appendix details the variables with missing labels, cases where a unique label was assigned to more than one variable, variables with a missing value across all observations, and variables with a missing value across all baseline observations. As noted in the Appendix, it may be reasonable that some of the variables (e.g., 'cyst reduction indicator') were not recorded at baseline.

Notes

1. The variable expected for 'total cyst volume' (mrscvs) is missing for all observations. For this replication, mrrcvs was used instead.

2. 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.

3. In addition to the analysis dataset examined in this replication analysis (EXPORT102005), the repository houses raw datasets and two additional analysis datasets from the CRISP cohort.

4. 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. Andrew D. Rule, Vicente E. Torres, Arlene B. Chapman, Jared J. Grantham, Lisa M. Guay-Woodford, Kyongtae T. Bae, Saulo Klahr, William M. Bennett, Catherine M. Meyers, Paul A. Thompson, J. Philip Miller, for the CRISP Consortium, **Comparison of Methods for Determining Renal Function Decline in Early Autosomal Dominant Polycystic Kidney Disease: The Consortium of Radiologic Imaging Studies of Polycystic Kidney Disease Cohort**, Journal of the American Society of Nephrology, 17: 854-862, 2006.

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

Appendix A:
Unresolved Data Issues in EXPORT102005
(Analysis dataset for renal function paper)

Missing Labels

(variable name, total number of observations)

```
  xxbvdate, 916
  uswdes, 469
  userl, 469
  setrm, 918
  rexmeas, 916
  eureae_ca_mmol, 897
  usimage, 964
  uercdate, 0
  mercdate, 0
  rmrscvl, 0
```

Repeated Labels

(label, variables with label)

```
  Completion/Date: cddate, ucddate
  Creatinine/Clearance: creatclr, creatinine_clearance
  Data entry/Date: dedate, udedate
  MR C VOL/MEAN STER: mrscvm, rmrscvm
  MR C VOL/RIGHT STER: mrscvr, rmrscvr
  MR C VOL/SUM STER: mrscvs, rmrscvs
  MR K Vol/Left Ster: mrskvl, rmrskvl
  MR K Vol/Mean Ster: mrskvm, rmrskvm
  MR K Vol/Right Ster: mrskvr, rmrskvr
  MR K Vol/Sum Ster: mrskvs, rmrskvs
  Participant/ID Number: fhfcnt, pkdid
  Participant/ID Number/#7: npkdid, pkdidx
  Physician Visit/Date: pv2date, pvdate
  Physician Visited: pv2nme, pvnme
  Physician Visited Address: pv2adds, pvadds
  Reason for Physician Visited: pv2reason, pvreason
  Visit/Date: basedate, visdate, xbvdate
```

Missing Values

Variables missing across ALL observations

thtime	Last Hyd/time
rkidw	R Kidney/Vein Width
lkidw	L Kidney/Vein Width
rdvdate	Visit date
npkdid	Participant/ID Number/#7
ndvdate	Visit/Date/#7
rdedate	Reg Date Entry/Date
ndedate	Data entry/Date/#7
rimgdmb	R Kidney/Meas 2 Image 4
scis	Seconds:/contrast/injection/-scan
uercdate	
mercdate	
rmrscvr	MR C Vol/right ster
rmrscvl	
rmrscvs	MR C Vol/sum ster
rmrscvm	MR C Vol/mean ster
mrscvr	MR C Vol/Right Ster
mrscvl	MR C Vol/Left Ster
mrscvs	MR C Vol/Sum Ster
mrscvm	MR C Vol/Mean Ster
pmd9	Prescribe med discont 9
pmd10	Prescribe med discont 10
oma8	OTC med add 8
oma9	OTC med add 9
oma10	OTC med add 10
omd8	OTC med discont 8
omd9	OTC med discont 9
omd10	OTC med discont 10
nrs2	R Advs Ev2/Series #
nrs3	R Advs Ev3/Series #
nls3	L Advs Ev3/Series #
rmraid	
nmraid	
nusaid	
rusaid	

Variables missing at all baseline observations (vis=0)

ilyn	Ill?
pvyn	Physician Visit yes/no
pvdate	Physician Visit/Date
mvc1	Mult. Visit ind. 1
pv2date	Physician Visit/Date
mvc2	Mult. Visit ind. 2
rsurgpyn	Renal Surgery yes/no
rsidate	Renal Surgery/Date
hvyn	Hospital yes/no
hadate	Hospital admitted/Date
pipeyn	Pipe?
chewyn	Chewing Tobacco?
payn	Prescribed added?
pdyn	Prescribed stopped?
oayn	OTC drugs added?
odyn	OTC stopped?
rmail	regular mail?
phone	telephone?
aeyn	AE reported?
creatser	Serum Creat
oopdate	Oophorectomy/Date/ # 29
msyn	Menopausal Status changed?/ # 29
cmenos	Menopausal Current State/ # 29
pregyn	Pregnant last year?/ # 29
liveyn	Live birth?/ # 29
bfeedyn	Breast Feeding?/ # 29
thtime	Last Hyd/time
rkidw	R Kidney/Vein Width
lkidw	L Kidney/Vein Width
rdvdate	Visit date
npkdid	Participant/ID Number/#7
ndvdate	Visit/Date/#7
rdedate	Reg Date Entry/Date
ndedate	Data entry/Date/#7
rimgdmb	R Kidney/Meas 2 Image 4
scis	Seconds:/contrast/injection/-scan
uercdate	
mercdate	

rmrskvr	MR K Vol/Right Ster
rmrskvl	MR K Vol/Left Ster
rmrscvr	MR C Vol/right ster
rmrscvl	
rmrskvs	MR K Vol/Sum Ster
rmrskvm	MR K Vol/Mean Ster
rmrscvs	MR C Vol/sum ster
rmrscvm	MR C Vol/mean ster
mrscvr	MR C Vol/Right Ster
mrscvl	MR C Vol/Left Ster
mrscvs	MR C Vol/Sum Ster
mrscvm	MR C Vol/Mean Ster
creducyn	Cyst reduction indicator
ill	Illnesses
pvnme	Physician Visited
pvadds	Physician Visited Address
pvreason	Reason for Physician Visited
pv2nme	Physician Visited
pv2adds	Physician Visited Address
pv2reason	Reason for Physician Visited
rsidesc	Renal Surgery/ Descrip
hnme	Hospital
hadds	Hospital Address
phnme	Hospital Physician
phadds	Hospital Physician Address
acv_hdiag	Hospital diagnosis
pma1	Prescribe med add 1
pma2	Prescribe med add 2
pma3	Prescribe med add 3
pma4	Prescribe med add 4
pma5	Prescribe med add 5
pmd1	Prescribe med discont 1
pmd2	Prescribe med discont 2
pmd3	Prescribe med discont 3
pmd4	Prescribe med discont 4
pmd5	Prescribe med discont 5
oma1	OTC med add 1
oma2	OTC med add 2
oma3	OTC med add 3
oma4	OTC med add 4
oma5	OTC med add 5
omd1	OTC med discont 1
omd2	OTC med discont 2
omd3	OTC med discont 3
omd4	OTC med discont 4
omd5	OTC med discont 5
time	contact time
pma6	Prescribe med add 6
pma7	Prescribe med add 7
pma8	Prescribe med add 8
pma9	Prescribe med add 9
pma10	Prescribe med add 10
pmd6	Prescribe med discont 6
pmd7	Prescribe med discont 7
pmd8	Prescribe med discont 8
pmd9	Prescribe med discont 9
pmd10	Prescribe med discont 10
oma6	OTC med add 6
oma7	OTC med add 7
oma8	OTC med add 8
oma9	OTC med add 9
oma10	OTC med add 10
omd6	OTC med discont 6
omd7	OTC med discont 7
omd8	OTC med discont 8
omd9	OTC med discont 9
omd10	OTC med discont 10
nrs2	R Advs Ev2/Series #
nrs3	R Advs Ev3/Series #
nls3	L Advs Ev3/Series #
rmraid	
nmraid	
nusaid	
rusaid	

Several other variables have very few observations – including several variables with only 1 observation.

Note: Some of these may be valid. For example, it makes some sense if the ‘cyst reduction indicator’ was

not recorded at baseline.

ATTACHMENT 1

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

Andrew D. Rule, Vicente E. Torres, Arlene B. Chapman, Jared J. Grantham, Lisa M. Guay-Woodford, Kyongtae T. Bae, Saulo Klahr, William M. Bennett, Catherine M. Meyers, Paul A. Thompson, J. Philip Miller, for the CRISP Consortium, Comparison of Methods for Determining Renal Function Decline in Early Autosomal Dominant Polycystic Kidney Disease: The Consortium of Radiologic Imaging Studies of Polycystic Kidney Disease Cohort, Journal of the American Society of Nephrology, 17: 854-862, 2006.

ATTACHMENT 2

SAS Code for Tabulations from the Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) Analysis File EXPORT102005 in the NIDDK Repository

```
options nofmterr;
/*****************************************************************************************/
/* Program: R:\05_Users\Norma\CRISP\RenalFuncPaper\table1.sas
/* Author:  Norma Pugh
/* Date:    11 March 2008
/* Revised: 17 May 2008 per Paul Thompson response regarding how to properly calculate
/*           the albumin to creatinine ratio.
/* Purpose: Replicate results from Rule article, Journal of the American Society of
/*           Nephrology: Renal Function (2006)
/*****************************************************************************************/
/*************************/
/* Libnames and formats */
/*************************/
libname data 'R:\05_Users\Norma\CRISP\MigratedData';
%include
'R:\03_Data_And_Tools\Studies\CRISP\DCC_Delivery\CRISP_20070505\rti\documents\pkdformat.s
as';


/************/
/* Table 1 */
/************/
data table1; set data.export102005(where=(vis=0));
 /* Define albumin to creatinine ratio */
 AtoCRatio = albe_ca / creatclr;
 label AtoCRatio = 'Albumin/Creatinine ratio';
run;

title'Table 1: Demographics';

/* Age */
proc means data=table1 n median q1 q3;
 var age; title2'Age';
run;

/* Gender, Race */
proc freq data=table1;
 tables sex race; title2'Gender, Race';
run;

/* Weight, Height */
proc means data=table1 n median q1 q3;
 var weight_c height_c; title2'Weight, Height';
run;

title'Table 1: Predictors for a decline in renal function';

/* Hypertension */
proc freq data=table1;
 tables hdyn; title2'Hypertension';
run;

/* Bilateral kidney volume, Bilateral cyst volume, Albumin to creatinine ratio */
proc means data=table1 n median q1 q3;
 var mrskvs mrrcvs AtoCRatio;
 title2'Bilateral kidney volume, Bilateral cyst volume, Albumin to creatinine ratio';
run;

/* Smoker, Hx of UTI, Abdominal pain, Gross hematuria */
proc freq data=table1;
 tables csyn ludyn freqrp ghdyn; format freqrp frpfmt.; title2'Current smoker, History of
Urinary Tract Infection, Abdominal pain, Gross hematuria';
run;

title'Table 1: Renal function measures';

/* Unstandardized & Standardized iothalamate clearance, SCr, MDRD, Cockgroft-Gault,
Creatinine Clearance */
proc means data=table1 n median q1 q3;
 var uic cic_c serumcreat mdrd_gfr_c cc_cg cc_su;
 title2'Unstandardized & Standardized iothalamate clearance, SCr, MDRD, Cockgroft-Gault,
Creatinine Clearance';
run;
```