Norma Pugh
March 5, 2008

Integrity Check for the Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) Volume Progression Analysis Files

As a partial check of the integrity of the CRISP volume progression 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 Grantham et al [1] in the *New England Journal of Medicine* in May 2006. The results of this integrity check are described below. The full text of the *New England Journal of Medicine* 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 primary analysis paper reports on magnetic resonance imaging to determine total kidney volume and total cyst volume. Renal enlargement, the hallmark of autosomal dominant polycystic kidney disease (ADPKD), was quantified to determine the rate of disease progression. ADPKD is the most common renal disorder involving a single gene and the fourth leading cause of end-stage renal disease in adults. The goal was to make prospective, longitudinal measurements of cyst and kidney growth in a large cohort of patients with ADPKD [1].

Demographic and Baseline Characteristics. The results section of the primary paper [1] reports on demographic and baseline characteristics of the study participants. All variables summarized are taken from the GRANTHAM2006 and EXPORT052005 analysis datasets created for this study. Table A lists the variables we used in our replication of these variables.

Table A: Variables Used to Replicate Results Section

Table Variable	Variables Used in Replication
Age	GRANTHAM2006: age, where 'vis' variable=0
Gender	GRANTHAM2006: sex, where 'vis' variable=0
Hypertension	GRANTHAM2006: hdyn, where 'vis' variable=0
ACE inhibitor	EXPORT052005: aceyn, where 'vis' variable=0
Angiotensin-receptor blocker	EXPORT052005: arayn, where 'vis' variable=0
Iothalamate clearance	GRANTHAM2006: uic, where 'vis' variable=0
Iothalamate clearance, adjusted	GRANTHAM2006: cic_c, where 'vis' variable=0
Total kidney volume	GRANTHAM2006: mrskvs, where 'vis' variable=0
Total cyst volume	GRANTHAM2006: mrrcvs, 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 3 below regarding the discrepancies).

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

Table Variable	Group: Overall		
	Grantham et al (2006)	Integrity Check	Difference
Age (years)	32.4 ± 8.9 (241)	32.4 ± 8.9 (241)	0 (0)
Gender (female, %)	60.0 (241)	60.0 (241)	0 (0)
Hypertension (%)	61.4 (241)	61.8 (241)	0.4 (0)
ACE inhibitor (%)	43.5 (241)	35.8 (240)	7.7 (1)
Angiotensin-receptor blocker (%)	21.1 (241)	13.8 (240)	7.3 (1)
Iothalamate clearance (ml/minute)	107 ± 28 (236)	107 ± 28 (236)	0 (0)
Iothalamate clearance, adjusted (ml/min/1.73 m^2)	98 ± 25 (236)	98 ± 25 (236)	0 (0)
Total kidney volume (ml)	1076 ± 670 (241)	1076 ± 670 (241)	0 (0)
Total cyst volume (ml)	534 ± 529 (241)	534 ± 529 (240)	0 (1)
Note: Values are means ± standard deviations or percentages, and sample sizes.			

Total Kidney Volume and Glomerular Filtration Rate Characteristics. Table 1 of the 2006 *New England Journal of Medicine* article reports on total kidney volume and glomerular filtration rate characteristics. Variables summarized in this table (Table 1. Relationship between Total Kidney Volume and Glomerular Filtration Rate) are taken from the GRANTHAM2006 analysis dataset created for this study. Table C lists the variables used for replication of the Table 1 variables.

Table C: Variables Used to Replicate Table 1

Table Variable	Variables Used in Replication
Age	age
Total kidney volume	mrskvs
Glomerular filtration rate	uic

In Table D, we compare the results for characteristics calculated from the archived dataset to the results published in the results section. As Table D shows, most results obtained from the archived data are similar to those in the published tabulations (see Note 3 below, regarding the discrepancies).

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

Total Kidney Volume and Age	Grantham et al (2006)	Integrity Check	Difference
	Total Kidney Volume, Baseline Intercept (ml)		
<750 ml and <30 yr	506 ± 109 (45)	507 ± 110 (48)	1 ± 1 (3)
<750 ml and ≥30 yr	572 ± 130 (48)	574 ± 130 (49)	2 ± 0 (1)
750-1500 ml and <30 yr	978 ± 193 (28)	984 ± 197 (28)	6 ± 4 (0)
750-1500 ml and ≥30 yr	1052 ± 191 (61)	1052 ± 190 (62)	0 ± 1 (1)
>1500 ml and <30 yr	1859 ± 333 (12)	1878 ± 393 (13)	19 ± 60 (1)
>1500 ml and ≥30 yr	2155 ± 543 (38)	2187 ± 592 (41)	32 ± 49 (3)
Note: Values are means ± standard deviations, and sample sizes.			

Total Kidney Volume and Age	Grantham et al (2006)	Integrity Check	Difference
	Glomerular Filtration Rate, Baseline (ml/min)		
<750 ml and <30 yr	114 ± 24.7 (47)	115 ± 24.7 (47)	1 ± 0 (0)
<750 ml and ≥30 yr	108 ± 24.2 (49)	108 ± 24.2 (49)	0
750-1500 ml and <30 yr	122 ± 30.8 (28)	122 ± 30.8 (28)	0
750-1500 ml and ≥30 yr	101 ± 26.8 (61)	101 ± 26.8 (61)	0
>1500 ml and <30 yr	99.6 ± 23.8 (13)	99.6 ± 23.8 (13)	0
>1500 ml and ≥30 yr	94.0 ± 29.2 (38)	94.0 ± 29.2 (38)	0
Note: Values are means ± standard deviations, and sample sizes.			

Notes

1. Dataset GRANTHAM2006 does not contain all the analysis variables needed to replicate the Grantham publication. EXPORT052005 needs to be used for the following variables:
   aceyn: ACE inhibitor
   arayn: Angiotensin-Receptor Blocker (Antagonist)

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

3. 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.

4. In addition to the analysis datasets examined in this replication analysis (GRANTHAM2006 and EXPORT052005), the repository houses raw datasets and two additional analysis datasets from the CRISP cohort.

5. 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. Jared J. Grantham, M.D., Vicente E. Torres, M.D., Arlene B. Chapman, M.D., Lisa M. Guay-Woodford, M.D., Kyongtae T. Bae, M.D., Ph.D., Bernard F. King, Jr., M.D., Louis H. Wetzel, M.D., Deborah A. Baumgarten, M.D., Philip J. Kenney, M.D, Peter C. Harris, Ph.D., Saulo Klahr, M.D., William M. Bennett, M.D., Gladys H. Hirschman, M.D., Catherine M. Meyers, M.D., Xiaoling Zhang, M.S., Fang Zhu, M.D., and John P. Miller, A.B., for the CRISP Investigators, **Volume Progression in Polycystic Kidney Disease**, The New England Journal of Medicine, 2006, 354(20):2122-30.

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

Norma Pugh
March 5, 2008

Appendix A:
Unresolved Data Issues in GRANTHAM2006 and EXPORT052005
(Analysis datasets for primary paper)

Missing Labels

Dataset GRANTHAM2006 (variable name, total number of observations)

```
curstat, 469
userl, 469
risk, 962
s7, 906
s8, 906
s9, 904
racecorrect, 962
MDRD_gfr, 914
eureae_ca_mmol, 894
```

Dataset EXPORT052005 (variable name, total number of observations)

```
xxbvdate, 909
uswdes, 469
userl, 469
setrm, 906
rexmeas, 903
eureae_ca_mmol, 887
usimage, 959
uercdate, 0
mercdate, 0
rmrscvl, 0
TKVIntercept, 959
pTKVslope, 959
TCVIntercept, 959
pTCVslope, 959
```

Repeated Labels

Dataset GRANTHAM2006 (label, variables with label)

```
Creatinine/Clearance: creatclr, creatinine_clearance
Participant/ID Number: fhfcnt, pkdid
Visit/Date: basedate, visdate
```

Dataset EXPORT052005 (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

Dataset GRANTHAM2006, variables missing across ALL observations

RAC	Renal Artery Clearance
mrscvr	MR C Vol/Right Ster
mrscvl	MR C Vol/Left Ster
mrscvs	MR C Vol/Sum Ster
mrscvm	MR C Vol/Mean Ster
nls3	L Advs Ev3/Series #

Dataset GRANTHAM2006, variables missing at all baseline observations (vis=0)

rsurgpyn	Renal Surgery yes/no
rsim	Renal Surgery/Month
rsid	Renal Surgery/Day
rsiy	Renal Surgery/Year
rsidate	Renal Surgery/Date
ccafunit	Coffee\Tea units
scafunit	Other Caff units
creatser	Serum Creat
liveyn	Live birth?/ # 29
ndvdate	Visit/Date/#7
RAC	Renal Artery Clearance
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
rsidesc	Renal Surgery/ Descrip
hnme	Hospital
acv_hdiag	Hospital diagnosis
nls3	L Advs Ev3/Series #

Dataset EXPORT052005, 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	

Dataset EXPORT052005, 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 in both GRANTHAM2006 and EXPORT052005 have very few observations, including several variables with only 1 observation.

Note: In general, some of these discrepancies may be valid. For example, it makes sense that a variable related to cyst reduction (e.g., `creducyn`) 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.

Jared J. Grantham, M.D., Vicente E. Torres, M.D., Arlene B. Chapman, M.D., Lisa M. Guay-Woodford, M.D., Kyongtae T. Bae, M.D., Ph.D., Bernard F. King, Jr., M.D., Louis H. Wetzel, M.D., Deborah A. Baumgarten, M.D., Philip J. Kenney, M.D, Peter C. Harris, Ph.D., Saulo Klahr, M.D., William M. Bennett, M.D., Gladys H. Hirschman, M.D., Catherine M. Meyers, M.D., Xiaoling Zhang, M.S., Fang Zhu, M.D., and John P. Miller, A.B., for the CRISP Investigators, Volume Progression in Polycystic Kidney Disease, The New England Journal of Medicine, 2006, 354(20):2122-30.

ATTACHMENT 2

SAS Code for Tabulations from the CRISP Volume Progression Analysis Datasets in the NIDDK Repository

```
options nofmterr;
/*****************************************************************************/
/* Program: R:\05_Users\Norma\CRISP\PrimaryPaper\primary.sas
/* Author:  Norma Pugh
/* Date:    05 March 2008
/* Purpose: Replicate 'Results' section & Table 1 from NEJM article: Volume Progression
/*           in Polycystic Kidney Disease. (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.sas';


/*******************/
/* Results Section */
/*******************/
data baseline; set data.grantham2006(where=(vis=0)); run;
data baseline_v2; set data.export052005(where=(vis=0)); run;

title'Baseline Descriptives';

/* Age */
proc means data=baseline n mean std;
 var age; title2'Age';
run;

/* Gender, Hypertension */
proc freq data=baseline;
 tables sex hdyn; title2'Gender, Hypertension';
run;

/* ACE inhibitor, Angiotensin-Receptor Blocker (Antagonist) */
proc freq data=baseline_v2;
 tables aceyn arayn; title2'ACE inhibitor, Angiotensin-Receptor Blocker (Antagonist)';
run;

/*Iothalamate Clearance, Adjusted Iothalamate Clearance, Total kidney volume, Total cyst volume*/
proc means data=baseline n mean std;
 var uic cic_c mrskvs mrrcvs;
 title2'Iothalamate Clearance, Adjusted Iothalamate Clearance, Total kidney volume, Total cyst
volume';
run;


/***********/
/* Table 1 */
/***********/
/* Create sub-groups, based on age and total kidney volume at baseline */
data subgrps; set baseline;
 if mrskvs<750 & age<30 then subgrp=1;
 if mrskvs<750 & age>=30 then subgrp=2;
 if 750<=mrskvs<=1500 & age<30 then subgrp=3;
 if 750<=mrskvs<=1500 & age>=30 then subgrp=4;
 if mrskvs>1500 & age<30 then subgrp=5;
```

```
 if mrskvs>1500 & age>=30 then subgrp=6;
run;

/* Merge subgroup dataset with full Grantham dataset */
proc sort data=subgrps; by pkdid; run;
proc sort data=data.grantham2006 out=all_grantham; by pkdid; run;

data table1; merge subgrps(keep=pkdid subgrp) all_grantham; by pkdid; run;
proc sort data=table1; by subgrp; run;

/* Calculate: */
/* baseline intercept (total kidney volume), baseline intercept (glomerular filtration rate) */
title'Table 1';
title2'Baseline intercepts: total kidney volume & glomerular filtration rate';
proc means data=table1(where=(vis=0)) n mean std; by subgrp; var mrskvs uic; run;
```