Dataset Integrity Check for The Chronic Kidney Disease in Children Cohort Study (CKiD) – Data through July 2014

Prepared by Allyson Mateja
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton, MD 20705

November 21, 2016

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

Chronic kidney disease (CKD) is a life-long condition that often results in substantial morbidity and premature death due to complications from a progressive decrease in kidney function. The early detection of, and initiation of therapy for, CKD is key to delaying or preventing progression to end-stage renal disease (ESRD). The CKiD (Chronic Kidney Disease in Children) study is a prospective cohort study of children with CKD that investigates risk factors and outcomes of the disease. The study population consists of two cohorts. Cohort 1 includes 586 racially and ethnically diverse children recruited between the ages of 1 and 16 years with mild to moderately impaired kidney function (defined by an estimated GFR between 30-90 ml/min/1.73m2). Cohort 2 includes 280 children with mildly impaired kidney function (defined as an estimated GFR between 45-90 ml/min/1.73m2). At baseline, participants underwent a physical examination, in addition to assessments of kidney, cardiovascular, and neurocognitive symptoms and function. Similar measures of kidney function, neurocognitive function, markers of risk factors for cardiovascular disease, growth and other co-morbid conditions are assessed at regularly scheduled study visits. Biospecimens, including serum, plasma, and urine are also collected. The primary outcome measure is the rate of decline of GFR, which is measured repeatedly over time in cohort participants. A secondary outcome measure is the time to ESRD, defined by transplantation, dialysis, or a 50% decrease in GFR.

3 Archived Datasets

All the SAS data files, as provided by the Data Coordinating Center (DCC), are located in the data folder in the data package. For this replication, variables were taken from the “kidhist.sas7bdat”, “l05.sas7bdat”, gfrcalibratedsummary.sas7bdat”, “cardio.sas7bdat”, “socdem.sas7bdat”, “vert_datebase.sas7bdat”, and “medsum_short.sas7bdat” datasets.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Fathallah-Shaykh [1] in the Clinical Journal of the American Society of Nephrology in 2015. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For Table 1 in the publication [1], Baseline clinical and demographic characteristics for 522 nonglomerular CKD Children, Table A lists the variables that were used in the replication and Table B compares the results calculated from the archived data files to the results published in Table 1. The results of the replication are almost an exact match to the published results, with only minor discrepancies.

For Table 3 in the publication [1], Estimated declines in GFR by combined baseline urinary protein-to-creatinine ratio and BP status, Table C lists the variables that were used in the replication and Table D compares the results calculated from the archived data files to the results published in Table 3. The results of the replication are almost an exact match to the published results, with only minor discrepancies.

Note that some discrepancies are expected due to a difference in the lock date for the public use dataset and the data used for the publication.

6 Conclusions

The NIDDK repository is confident that the CKiD data files to be distributed are a true copy of the study data.

7 References

[1] Fathallah-Shaykh, S.A., Flynn, J.T., Pierce, C.B., Abraham, A.G., Blydt-Hansen, T.D., Massengill, S.F., Moxey-Mims, M.M., Warady, B.A., Furth, S.L., and Wong, C.S. "Progression of Pediatric CKD of Nonglomerular Origin in the CKiD Cohort". Clin J Am Soc Nephrol 10 (2015).

Table A: Variables used to replicate Table 1: Baseline clinical and demographic characteristics for 522 nonglomerular CKD Children

Table Variable	**dataset.variable**
Age	kidhist.bsdate, kidhist.dob
Male	kidhist.male1fe0
African American race	socdem.race
Duration of CKD	kidhist.ckdonst
GFR	gfrcalibratedsummary.bedgfr
Urine protein-to-creatinine ratio	l05.rlurcrea, l05.rlurprot
Casual BP (percentile) - Systolic	cardio.sbppctagh
Casual BP (percentile) - Diastolic	cardio.dbppctagh
Renin-angiotensin-system antagonist use	medsum_short.acei, medsum_short.arb
Other antihypertensive therapy	medsum_short.acei, medsum_short.arb, medsum_short.antihyp
Duration of follow-up	vert_datebase.year

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

Baseline Characteristic	Value Manuscript	Value DSIC	Difference
Age (yr)	10 (7, 14)	10 (7, 14)	0 (0, 0)
Male, % (n)	65 (340)	65 (340)	0 (0)
African American race, % (n)	19 (101)	19 (101)	0 (0)
Duration of CKD (yr)	10 (6, 13)	10 (6, 13)	0 (0, 0)
GFR (ml/min per 1.73 m^2)	48 (36, 64)	48 (35, 61)	0 (1, 3)
Urine protein-to-creatinine ratio (mg/ml)	0.29 (0.12, 0.82)	0.29 (0.12, 0.82)	0 (0, 0)
Casual BP (percentile)			
Systolic	66 (39, 86)	66 (39, 86)	0 (0, 0)
Diastolic	70 (47, 88)	70 (47, 88)	0 (0, 0)
Renin-angiotensin-system antagonist use, % (n)	45 (237)	45 (237)	0 (0)
Other antihypertensive therapy	9 (47)	9 (47)	0 (0)
Duration of follow-up (yr)	4.4 (1.7, 6.0)	4.6 (1.8, 5.9)	0.2 (0.1, 0.1)

Values are presented as median (interquartile range) for continuous variables, and percentage (frequency) for categorical variables

Table C: Variables used to replicate Table 3: Estimated declines in GFR by combined baseline urinary protein-to-creatinine ratio and BP status

Table Variable	**dataset.variable**
Baseline Up/c	l05.rlurcrea, l05.rlurprot
BP status	cardio.sbppctagh, cardio.dbppctagh
GFR	gfrcalibratedsummary.bedgfr, gfrcalibratedsummary.igfrc

Table D: Comparison of values computed in integrity check to reference article Table 3 values

Normotensive (BP < 90th Percentile) (n=358 Manuscript, n=359 DSIC, Difference = 1)

Baseline Up/c	Patients (n) Manuscript	Patients (n) DSIC	Diff.	Median (IQR) Baseline GFR (ml/min per 1.73 m²) Manuscript	Median (IQR) Baseline GFR (ml/min per 1.73 m²) DSIC	Diff.	Mean Change in GFR (ml/min per 1.73 m² per yr) (95% CI) Manuscript	Mean Change in GFR (ml/min per 1.73 m² per yr) (95% CI) DSIC	Diff.
< 0.5	229	230	1	53 (42, 69)	55 (41, 65)	2 (1, 4)	-0.8 (-1.2 to -0.4)	-1.2 (-1.7 to -0.7)	0.4 (0.5 to 0.3)
≥ 0.5	129	129	0	39 (29, 51)	40 (29, 50)	1 (0, 1)	-1.8 (-2.4 to -1.2)	-1.9 (-2.6 to -1.2)	0.1 (0.2 to 0)

Elevated BP (BP ≥ 90th Percentile) (n=164 Manuscript, n=163 DSIC, Difference = 1)

Baseline Up/c	Patients (n) Manuscript	Patients (n) DSIC	Diff.	Median (IQR) Baseline GFR (ml/min per 1.73 m²) Manuscript	Median (IQR) Baseline GFR (ml/min per 1.73 m²) DSIC	Diff.	Mean Change in GFR (ml/min per 1.73 m² per yr) (95% CI) Manuscript	Mean Change in GFR (ml/min per 1.73 m² per yr) (95% CI) DSIC	Diff.
< 0.5	94	93	1	58 (44, 71)	57 (44, 66)	1 (0, 5)	-1.9 (-2.5 to -1.2)	-2.0 (-2.8 to -1.2)	0.1 (0.3 to 0)
≥ 0.5	70	70	0	40 (30, 49)	37 (30, 48)	3 (0, 1)	-1.7 (-2.4 to -1.0)	-1.8 (-2.7 to -1.0)	0.1 (0.3 to 0)

Attachment A: SAS Code

```
*** DSIC for CKiD Data through July 2014;
***
*** Programmer: Allyson Mateja;
*** Date: 08/29/2016;
*** Modified: 10/10/16;

title1 "%sysfunc(getoption(sysin))";
title2 " ";

options nofmterr;

proc format;
       value glomf 1,2 = 'G'
                   3,4 = 'NG';
       value sexf 0 = 'F'
                  1 = 'M';
       value racef 2,8         = 1
                   1,3,4,5,6,7 = 0;

libname sas_data "/prj/niddk/ims_analysis/CKiD/private_orig_data/CKiD Upload 07-28-16/P04/data";

proc import datafile = '/prj/niddk/ims_analysis/CKiD/private_orig_data/CKiD Upload 10-07-16/analytical files 
03/Fathallah_Shaykh_S.UPCR_BP.CJASN2015.CASEID_VIS.csv'
       dbms = csv
       out = subjects;
       getnames = yes;
run;

data cardio       ; set sas_data.cardio      ;
data gfrcalibratedsummary   ; set sas_data.gfrcalibratedsummary  ;
data kidhist      ; set sas_data.kidhist     ;
data l05          ; set sas_data.l05         ;
data medsum_short ; set sas_data.medsum_short;
data socdem       ; set sas_data.socdem      ;
data vert_datebase; set sas_data.vert_datebase;
data growth       ; set sas_data.growth;

proc freq data = cardio;
       tables visit;

proc contents data = l05;

proc freq data = kidhist;
       tables gngdiag bsdate /list missing;
       format gngdiag glomf.;

proc freq data = l05;
```

```
       tables visit;

data baseline_l05;
       set l05;
       if visit = 10;

data visit2_l05;
       set l05;
       if visit = 20;

data baseline_gfr;
       set gfrcalibratedsummary;
       if visit = 10;

data baseline_cardio;
       set cardio;
       if visit = 10;

data visit2_cardio;
       set cardio;
       if visit = 20;

data socdem;
       set socdem;
       if visit = 10;

proc sort data = kidhist;
       by caseid;

proc sort data = baseline_l05;
       by caseid;;

proc sort data = baseline_gfr;
       by caseid;

proc sort data = baseline_cardio;
       by caseid;

proc sort data = subjects;
       by caseid visit;

data subject_ids;
       set subjects;
       by caseid;
       if first.caseid then output;

proc freq data = subject_ids;
       tables visit;

data last_visit;
       set subjects;
```

```
        by caseid;
        if last.caseid then output;

proc sort data = gfrcalibratedsummary;
        by caseid visit;

proc sort data = cardio;
        by caseid visit;

proc sort data=vert_datebase;
        by caseid visit;

data last_visit;
        merge last_visit (in=val1)
              vert_datebase (keep=caseid visit year);
        by caseid visit;
        followup = year+0.75;
        if val1 then output;

data baseline_medsum_short;
        set medsum_short;
        if visit=10;

proc sort data = baseline_medsum_short;
        by caseid;

data subjects_in_paper;
        merge kidhist      (in=val1)
              baseline_l05 (keep=caseid rlurcrea rlurprot)
              baseline_gfr (keep=caseid bedgfr igfrc)
              baseline_cardio (keep=caseid SBPPCTAGH DBPPCTAGH)
              socdem          (keep=caseid race)
              baseline_medsum_short (keep=caseid antihyp acei arb)
              subject_ids (in=val2 keep=caseid)
              last_visit  (keep=caseid followup);
        by caseid;
        age = bsdate-dob;
        if acei=1 or arb=1 then ras = 1;
        else ras=0;
        if ras=0 and antihyp=1 then other_antihypertensive=1;
        else other_antihypertensive=0;
        duration = round(abs(ckdonst), 1);
        upc = rlurprot/rlurcrea;
        if SBPPCTAGH < 90 and DBPPCTAGH < 90 then elev_bp = 0;
        else elev_bp = 1;
        if upc < 0.5 then upc_less_05 = 1;
        else upc_less_05 = 0;
        updated_followup = ldatrtfree-0.83;
        if val1 and val2 then output;

proc means data = subjects_in_paper n median p25 p75;
```

```
       var age;
       title3 'Table 1 - Age';

proc freq data = subjects_in_paper;
       tables male1fe0;
       format male1fe0 sexf.;
       title3 'Table 1 - Male %';

proc freq data = subjects_in_paper;
       tables race;
       format race racef.;
       title3 'Table 1 - African American race';

proc means data = subjects_in_paper n median p25 p75;
       var duration;
       title3 'Table 1 - Duration of CKD';

proc means data = subjects_in_paper n median p25 p75;
       var bedgfr;
       title3 'Table 1 - GFR';

proc means data = subjects_in_paper n median p25 p75;
       var upc;
       title3 'Table 1 - Urine protein-to-creatinine ratio';

proc means data = subjects_in_paper n median p25 p75;
       var SBPPCTAGH;
       title3 'Table 1 - Systolic BP Percentile';

proc means data = subjects_in_paper n median p25 p75;
       var DBPPCTAGH;
       title3 'Table 1 - Diastolic BP Percentile';

proc freq data = subjects_in_paper;
       tables ras;
       title3 'Table 1 - Renin-angiotensin-system antagonist use';

proc freq data = subjects_in_paper;
       tables other_antihypertensive;
       title3 'Table 1 - Other antihypertensive therapy';

proc means data = subjects_in_paper n median p25 p75;
       var followup;
       title3 'Table 1 - Duration of follow-up';

proc freq data = subjects_in_paper;
       tables elev_bp elev_bp*upc_less_05 /list missing;

proc means data = subjects_in_paper n median p25 p75;
       var bedgfr;
       where elev_bp = 0 and upc_less_05 = 1;
```

```
proc means data = subjects_in_paper n median p25 p75;
        var bedgfr;
        where elev_bp = 0 and upc_less_05 = 0;

proc means data = subjects_in_paper n median p25 p75;
        var bedgfr;
        where elev_bp = 1 and upc_less_05 = 1;

proc means data = subjects_in_paper n median p25 p75;
        var bedgfr;
        where elev_bp = 1 and upc_less_05 = 0;

proc sort data = growth;
        by caseid visit;

proc sort data = medsum_short;
        by caseid visit;

data table3;
        merge subjects             (in=val1)
              gfrcalibratedsummary (in=val2)
              growth               (keep=caseid visit bmizag)
              medsum_short         (keep=caseid visit acei arb);
        by caseid visit;
        if acei = 1 or arb = 1 then acei_arb = 1;
        else acei_arb = 0;
        years_from_baseline = (visit/10)-1;
        if val1 and val2 then output;

data table3;
        merge table3            (in=val1)
              subjects_in_paper (in=val2 keep=caseid elev_bp upc_less_05 upc race age male1fe0 SBPPCTAGH DBPPCTAGH);
        by caseid;
        if race in (2,8) then AArace = 1;
        else AArace = 0;
        if igfrc = . then igfrc = bedgfr;
        if val1 and val2 then output;

proc sort data=table3; by caseid years_from_baseline; run;

%macro mixmod(depvar,indvars);
        proc mixed data=table3 method=ml;
        class caseid;
        model &depvar.=&indvars./s cl;
        random int years_from_baseline/ subject=caseid type=un;
        ods exclude ClassLevels;
%mend mixmod;

title "TABLE 3";
%mixmod(igfrc,upc_less_05 elev_bp age male1fe0 AArace BMIzag acei_arb
                years_from_baseline years_from_baseline*upc_less_05 years_from_baseline*elev_bp
                years_from_baseline*elev_bp*upc_less_05);

        estimate "bsupclt05: bs norm BP" years_from_baseline 1 years_from_baseline*upc_less_05 1/cl;
        estimate "bsupclt05: bs elv BP" years_from_baseline 1 years_from_baseline*upc_less_05 1 years_from_baseline*elev_bp 1
years_from_baseline*elev_bp*upc_less_05 1;
        estimate "bsupclt05: change, elv vs. norm BP" years_from_baseline*elev_bp 1 years_from_baseline*elev_bp*upc_less_05 1;

        estimate "bsupc ge05: bs norm BP" years_from_baseline 1;
        estimate "bsupc ge05: bs elv BP" years_from_baseline 1 years_from_baseline*elev_bp 1;
        estimate "bsupc ge05: change, elv vs. norm BP" years_from_baseline*elev_bp 1;

        estimate "Norm BP: change, elevated vs. norm uP/C"  upc_less_05*years_from_baseline -1;
        estimate "Elevated BP: change, elevated vs. norm uP/C"  upc_less_05*years_from_baseline -1
years_from_baseline*elev_bp*upc_less_05 -1;
        estimate "Up/c <0.5: Elevated vs Normal BP" years_from_baseline*elev_bp 1 years_from_baseline*elev_bp*upc_less_05 1 ;
        estimate "Up/c >=0.5: Elevated vs Normal BP" years_from_baseline*elev_bp 1;
        estimate "elevated/elevated vs norm/norm" years_from_baseline*elev_bp 1 years_from_baseline*upc_less_05 -1;
        estimate "elv Upc/norm BP vs norm Upc/ elev BP" years_from_baseline*upc_less_05 -1 years_from_baseline*elev_bp -1
years_from_baseline*elev_bp*upc_less_05 -1;
run;
```