Dataset Integrity Check for Chronic Kidney Disease in Children (CKiD)

Derek K. Ng

Prepared by Sabrina Chen
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton, MD 20705
Nov 30, 2018

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.

3 Archived Datasets

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

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Derek K. Ng et al [1] in Kidney International in 2018. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For Table 1 in the publication [1], Demographic and clinical characteristics of young adults with a history of pediatric chronic kidney disease at time of first GFR after age 18 years. Median [interquartile range] or % (frequency), 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 similar to the published results.

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]
Derek K Ng, George J Schwartz, Michael F Schneider, Susan L Furth, Bradley A Warady. Kidney International 2018.

Table A: Variables used to replicate Table 1**:** Demographic and clinical characteristics of young adults with a history of pediatric chronic kidney disease at time of first GFR after age 18 years.

Table Variable	**dataset.variable**
Age	kidhist.age
Height (cm)	gfrcalibratedsummary.avheight
Weight (kg)	Gfrcalibratedsummary.avweight
Body Surface Area (m^2)	Gfrcalibratedsummary.bsa
Serum Creatinine (mg/dL)	Gfrcalibratedsummary.scr
Cystatin-C (mg/L)	Gfrcalibratedsummary.cyc_db
BUN (mg/dL)	Gfrcalibratedsummary.bun

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

	Manuscript (n=219)	DSIC (n=219)	Diff (n=0)
Age	18.5 (18.2, 18.9)	18.266 (18.076, 18.932)	0.234 (0.124, -0.032)
Female sex	41% (89)	45% (99)	4 (10)
Height (cm)	169 (161 , 177)	168.5 (159.9 , 176.1)	0.5 (1.1 , 0.9)
Weight (kg)	68 (57 , 84)	67.3 (57.1 , 84.1)	.7 (-0.1 , -0.1)
Body Surface Area (m^2)	1.8 (1.6 , 2)	1.809 (1.612 , 2.011)	-0.009 (-0.012 , -0.011)
Serum Creatinine (mg/dL)	1.6 (1.2 , 2.2)	1.61 (1.2, 2.18)	-0.01 (0, 0.02)
Cystatin-C (mg/L)	1.6 (1.2 , 2.3)	1.41 (1.05, 1.93)	0.19 (0.15 , 0.37)

Attachment A: SAS Code

```
 options nofmterr validvarname=upcase;

***********;
* INPUT  ;
***********;
libname sas_data "/prj/niddk/ims_analysis/CKiD/private_orig_data/CKiD Upload 12-22-17/P05/data";


* pre process;
data kidhist    ;
 set sas_data.kidhist   ;
run;

proc contents data=kidhist;
title3 'kidhist';
run;


data gfrcalibratedsummary    ;
 set sas_data.gfrcalibratedsummary   ;
run;

proc contents data=gfrcalibratedsummary;
title3 'gfrcalibratedsummary';
run;


data socdem    ;
 set sas_data.socdem   ;

 if race in (2,8) then black = 1;
 if race in (1,3,4,5,6,7) then black = 0;

run;

proc contents data=socdem;
title3 'kidhis';
run;

proc freq data=socdem;
 tables black*race/list missing;
 title3 'check binary var';
run;



* combine;
```

```
proc sort data=kidhist;
 by caseid ;
run;

proc sort data=gfrcalibratedsummary;
 where visstatus in (0,1);
 by caseid visit;
run;

proc sort data=socdem nodupkey;
 by caseid race;
run;

data combine;
 merge kidhist            (keep= caseid dob bsdate)
    gfrcalibratedsummary (keep= caseid visit visstatus gfrvdate male1fe0 avheight bsa scr bun bedgfr cyc_db igfrc avweight)
       socdem            (keep= caseid black)
       ;
 by caseid;

 age = (bsdate - dob) + gfrvdate;

 ckidscrcyc = 39.8 * (((avheight/100)/scr)**0.456) * ((1.8/cyc_db)**0.419) * ((30/bun)**0.079) * (1.076**male1fe0) *
(((avheight/100)/1.4)**0.179);

 if male1fe0 = 1 then do;
   ckdepiscr = 141 * (min((scr/0.9),1)**-0.411) * (max((scr/0.9),1)**-1.209) * (0.993**age) * 1.159**black;
 end;

 if male1fe0 = 0 then do;
   ckdepiscr = 141 * (min((scr/0.7),1)**-0.329) * (max((scr/0.7),1)**-1.209) * (0.993**age) * 1.018 * 1.159**black;
 end;

 ckdepicyc =  133 * (min((cyc_db/0.8),1)**-0.499) * (max((cyc_db/0.8),1)**-1.328) * (0.996**age) * (0.932**(1-male1fe0));

 if male1fe0 = 1 then do;
   ckdepiscrcyc = 135 * (min((scr/0.9),1)**-0.207) * (max((scr/0.9),1)**-0.601) * (min((cyc_db/0.8),1)**-0.375) *
(max((cyc_db/0.8),1)**-0.711) * (0.995**age) * (1.08**black);
 end;

 if male1fe0 = 0 then do;
   ckdepiscrcyc = 135 * (min((scr/0.7),1)**-0.248) * (max((scr/0.7),1)**-0.601) * (min((cyc_db/0.8),1)**-0.375) *
(max((cyc_db/0.8),1)**-0.711) * (0.995**age) * 0.969 * (1.08**black);
 end;

 * create subset flags for Table 1 "Participants contributing estimated GFR" ;
 if age >= 18 then subset_ge18 = 1;
 if bedgfr > 0 and ckidscrcyc > 0 and ckdepiscr > 0 and ckdepicyc > 0 and ckdepiscrcyc > 0 then subset_gfr=1;

 * create subset flags for Table 2;
 if igfrc > . then subset_igfrc=1;
```

```
 diff = ckidscrcyc - igfrc;
 avg = ((igfrc + ckidscrcyc)/2) - 50;

 if 0.7 <= (ckidscrcyc/igfrc) <= 1.3 then within30 = 1;
 if (ckidscrcyc/igfrc) > 1.3 or (ckidscrcyc/igfrc) < 0.7 then within30 = 0;

run;

proc freq data=combine;
 tables age*bsdate*dob*gfrvdate/list missing;
 tables subset_ge18*age/list missing;
 tables subset_gfr*bedgfr*ckidscrcyc*ckdepiscr*ckdepicyc*ckdepiscrcyc/list missing;
 tables subset_ge18*subset_gfr/list missing;
 tables subset_igfrc*igfrc/list missing;
 tables within30*ckidscrcyc*igfrc/list missing;
 title3 "check new vars";
run;

proc print data=combine (obs=100);
 by caseid;
 id caseid;
 var visit age bsdate dob gfrvdate subset_ge18 subset_gfr;
 title3 "print per visit";
run;


** subset to Table 1;
proc sort data=combine out=subset_t1 nodupkey;
 by caseid;
 where subset_ge18=1 and subset_gfr=1;
run;

proc print data=subset_t1 (obs=25);
 var caseid visit age bsdate dob gfrvdate subset_ge18 subset_gfr;
 title3 "print per subject";
run;


** Table 1;
proc means data=subset_t1 n p50 p25 p75;
 var age avheight avweight bsa scr cyc_db bun;
 title3 "Table 1";
run;

proc freq data=subset_t1;
 tables male1fe0/missing;
 title3 'Table 1';
 run;
```