Dataset Integrity Check for The Chronic Kidney Disease in Children Cohort Study (CKiD) Data Update through July 13, 2018 Data

Prepared by NIDDK-CR
October 29, 2021

Contents

1 Standard Disclaimer 2

2 Study Background 2

3 Archived Datasets 2

4 Statistical Methods 3

5 Results 3

6 Conclusions 3

7 References 3

Table A: Variables used to replicate Table 1 – Baseline characteristics by CKD diagnosis 4

Table B: Comparison of values computed in integrity check to reference article Table 1 values 5

Attachment A: SAS Code 7

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/1.73 m^2). Cohort 2 includes 280 children with mildly impaired kidney function (defined as an estimated GFR between 45-90 mL/min/1.73 m^2). 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.

This update contains follow-up data through July 31, 2018.

3 Archived Datasets

All 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 “saland.sas7bdat”, “kidhist.sas7bdat”, “socdem.sas7bdat”, “gfrcalibratedsummary.sas7bdat”, “labmarkers.sas7bdat”, “growth.sas7bdat”, and “medsum_short.sas7bdat” datasets.

4 Statistical Methods

Analyses were performed to replicate results for the data published by Saland et al. [1] for Change in Dyslipidemia with Declining Glomerular Filtration Rate and Increasing Proteinuria in Children with CKD. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For Table 1 in the publication [1], Baseline characteristics by CKD diagnosis, 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 within expected variation to the published results.

6 Conclusions

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

7 References

[1] Saland JM, Kupferman JC, Pierce CB, Flynn JT, Mitsnefes MM, Warady BA, Furth SL. Change in Dyslipidemia with Declining Glomerular Filtration Rate and Increasing Proteinuria in Children with CKD. Clinical Journal of the American Society of Nephrology, 14(12), 1711-1718, November 2019.
doi: https://doi.org/10.2215/cjn.03110319
PMCID: http://www.ncbi.nlm.nih.gov/pmc/articles/pmc6895497/

Table A: Variables used to replicate Table 1 – Baseline characteristics by CKD diagnosis

Table Variable	**dataset.variable**
Age	saland.age
Male	kidhist.gender
Black race	socdem.race
Hispanic ethnicity	socdem.hisp
Years of life with CKD	saland.CKDduration
Percent of life with CKD	saland.pctoflifewithCKD
GFR, mL/minute/1.73 m^2	gfrcalibratedsummary.bedgfr
Urine protein-to-creatinine, mg/mg	labmarkers.upcratio
Nephrotic range proteinuria	labmarkers.upcratio
BMI percentile	growth.bmipctag
BMI >85th percentile	growth.bmipctag
Total cholesterol, mg/dl	labmarkers.totalchol
Triglycerides, mg/dl	labmarkers.trig
HDL cholesterol, mg/dl	labmarkers.hdl
Non-HDL cholesterol, mg/dl	labmarkers.nonhdl
Non-HDL cholesterol >190 mg/dl	labmarkers.nonhdl
Dyslipidemia	labmarkers.trig, labmarkers.hdl, labmarkers.nonhdl
Lipid-lowering medication use	medsum_short.liplowrx
Years of follow-up	saland.FUyears
Number of lipid measures	saland.caseid

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

Baseline Characteristic	Manuscript Nonglomerular CKD (n=385)	DSIC Nonglomerular CKD (n=385)	Diff. (n=0)	Manuscript Glomerular CKD (n=123)	DSIC Glomerular CKD (n=123)	Diff. (n=0)
Age, years	11 [8-14]	11 [8-14]	0 [0-0]	15 [12-17]	15 [12-17]	0 [0-0]
Male	64% (248)	64% (248)	0 (0)	55% (68)	55% (68)	0 (0)
Black race	18% (70)	18% (70)	0 (0)	28% (34)	28% (34)	0 (0)
Hispanic ethnicity	14% (52)	14% (52)	0 (0)	13% (16)	13% (16)	0 (0)
Years of life with CKD	10 [6-14]	10 [7-14]	0 [1-0]	4 [2-8]	4 [2-8]	0 [0-0]
Percent of life with CKD	100% [100%-100%]	100% [100%-100%]	0 [0-0]	34% [18%-64%]	38% [21% -80%]	4% [3%-16%]
GFR, mL/minute/1.73 m^2	51 [38-65]	51 [38-65]	0 [0-0]	63 [51-79]	63 [51-79]	0 [0-0]
Urine protein-to-creatinine, mg/mg	0.25 [0.11-0.65]	0.23 [0.10-0.64]	0.02 [0.01-0.01]	0.41 [0.15-1.14]	0.40 [0.12-1.08]	0.01 [0.03-0.06]
Nephrotic range proteinuria	5% (20)	5% (20)	0 (0)	9% (11)	9% (11)	0 (0)
BMI percentile	61 [31-83]	61 [31-83]	0 [0-0]	81 [52-97]	81 [52-97]	0 [0-0]
BMI >85th percentile	23% (87)	23% (87)	0 (0)	46% (57)	46% (57)	0 (0)
Total cholesterol, mg/dl	169 [147-190]	169 [147-190]	0 [0-0]	176 [148-201]	176 [148-201]	0 [0-0]
Triglycerides, mg/dl	95 [71-135]	95 [71-135]	0 [0-0]	96 [68-146]	96 [68-146]	0 [0-0]
HDL cholesterol, mg/dl	49 [42-59]	49 [42-59]	0 [0-0]	49 [41-60]	49 [41-60]	0 [0-0]
Non-HDL cholesterol, mg/dl	118 [95-137]	118 [95-137]	0 [0-0]	122 [97-153]	122 [97-153]	0 [0-0]
Non-HDL cholesterol >190 mg/dl	3% (10)	3% (10)	0 (0)	5% (6)	5% (6)	0 (0)
Dyslipidemia	35% (134)	35% (134)	0 (0)	43% (53)	43% (53)	0 (0)
Lipid-lowering medication use	<1% (3)	<1% (3)	0 (0)	5% (6)	5% (6)	0 (0)
Years of follow-up	4.0 [2.2-6.1]	4.0 [2.2-6.1]	0 [0-0]	2.6 [2.0-4.4]	2.6 [2.0-4.4]	0 [0-0]

Baseline Characteristic	Manuscript Nonglomerular CKD (n=385)	DSIC Nonglomerular CKD (n=385)	Diff. (n=0)	Manuscript Glomerular CKD (n=123)	DSIC Glomerular CKD (n=123)	Diff. (n=0)
Number of lipid measures						
2	37% (143)	37% (143)	0 (0)	55% (68)	55% (68)	0 (0)
3	34% (129)	34% (129)	0 (0)	26% (32)	26% (32)	0 (0)
4	18% (68)	18% (68)	0 (0)	12% (15)	12% (15)	0 (0)
5	9% (35)	9% (35)	0 (0)	5% (6)	5% (6)	0 (0)
6	3% (10)	3% (10)	0 (0)	2% (2)	2% (2)	0 (0)

[1] Expressed as median (IQR) of % (*n*), as appropriate

Attachment A: SAS Code

```
* CKiD DSIC - Data through July 31, 2018;
* Saland et al. 2019;

*****************************************************
* Set up directories and load in data              *
*****************************************************;

* directory for updated data through July 31, 2018;
libname ckid "Z:\NIDDK\niddk-dr_studies1\CKiD\private_orig_data\CKiD Upload 08-18-2020\P06\data";

libname ids "Z:\NIDDK\niddk-dr_studies1\CKiD\private_created_data\CKiD Update through 073118";

*****************************************************
* Create list of baseline IDs and Visit*IDs        *
*****************************************************;
data saland;
	set ids.saland;
run;

proc contents data=saland; run; *1514 obs, 24 vars;

* Scrape IDs and visits from analysis data;
data ids;
	set saland (keep = caseid visit age CKDduration pctoflifewithCKD FUyrs);
run; *1514 person-visits;

proc freq data=ids nlevels; *508 unique IDs;
	tables caseid visit;
run;

proc sort data=ids; by caseid; run;

* Age --> age
* Years of life with CKD --> CKDduration
* Percent of life with CKD --> pctoflifewithCKD
* Years of follow-up --> FUyrs;

*****************************************************
* Study Data                                        *
*****************************************************;

* labmarkers is a summary file with info on laboratory markers, including lipid measures;
proc contents data = ckid.labmarkers; run; *5776 obs, 50 vars;

data lab;
```

```
        set ckid.labmarkers;
run;

proc contents data = lab; run; *5776 obs, 50 vars;
proc sort data=lab; by caseid; run;

* Cholesterol --> TOTALCHOL
* Triglycerides --> TRIG
* HDL --> HDL;
* Non-HDL cholesterol --> NonHDL;
* urine protein/creatinine ratio --> UPCRATIO;

* kidhist is a summary file with info on diagnosis;
proc contents data=ckid.kidhist; run; *891 obs, 15 vars;

data kidhist;
        set ckid.kidhist;
run;

proc contents data=kidhist; run; *891 obs, 15 vars;
proc sort data=kidhist; by caseid; run;

* Glomerular diagnosis --> GNGDIAG;
* Gender --> male1fe0;

* socdem is a summary file with info on sociodemographics;
proc contents data=ckid.socdem; run; *5775 obs, 10 vars;

data socdem;
        set ckid.socdem;
        where visit=20;
run;

proc contents data=socdem; run; *795 obs, 10 vars;
proc sort data=socdem; by caseid; run;

* Race --> race;
* Hispanic ethnicity --> hisp;

* growth is a summary file with info on height/weight/development;
proc contents data=ckid.growth; run; *6679 obs, 21 vars;

data growth;
        set ckid.growth;
        where visit=20;
run;

proc contents data=growth; run; *825 obs, 21 vars;
```

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

* BMI percentile --> BMIPCTAG;

* gfrcalibratedsummary is a summary file with info on gfr;
proc contents data=ckid.gfrcalibratedsummary; run; *6074 obs, 17 vars;

data gfr;
	set ckid.gfrcalibratedsummary;
	where visit=20;
run;

proc contents data=gfr; run; *822 obs, 17 vars;
proc sort data=gfr; by caseid; run;

* GFR --> BEDGFR;

* medsum_short is a summary file with info on medications;
proc contents data=ckid.medsum_short; run; *5172 obs, 28 vars;

data meds;
	set ckid.medsum_short;
	where visit=20;
run;

proc contents data=meds; run; *802 obs, 28 vars;
proc sort data=meds; by caseid; run;

* Lipid-lowering medication --> LIPLOWRX;

*****************************************************
* Merge Data                                        *
*****************************************************;

data dsic; *create dataset with all included ids and visits;
	merge	ids (in=a)
		lab (in=b keep=caseid visit TOTALCHOL TRIG HDL nonHDL UPCRATIO);
	by caseid visit;
	if a then output dsic;
run;

proc sort data=dsic; by caseid; run;
proc contents data=dsic; run;

*restrict to lipid baseline visit (one ob per participant --> N=508 obs);
proc sort data=dsic out=dsic_bs nodupkey;
	by caseid;
run;
```

```
*confirm lipid baseline visit (visit=20);
proc freq data=dsic_bs;
	tables visit;
run;

data dsic_bs_all; *one obs per participant, all baseline data;
	merge	dsic_bs (in=a)
				kidhist (keep=caseid gngdiag male1fe0)
				socdem (keep=caseid race hisp)
				growth (keep=caseid BMIPCTAG)
				gfr (keep=caseid bedgfr)
				meds (keep=caseid LIPLOWRX);
	by caseid;
	if a then output dsic_bs_all;
run;

proc freq data=dsic_bs_all;
	tables gngdiag;
run;

*****************************************************
* Recreate Table 1                                                    *
*****************************************************;

* create variables;
data dsic_all_bs2;
	set dsic_bs_all;

	if gngdiag = 1 OR gngdiag = 2 then g_diag = "G ";
	if gngdiag = 3 OR gngdiag = 4 then g_diag = "NG";

	if RACE = 2 OR RACE = 8 then race2=1;
		else race2=0;

	CKDduration2 = input(CKDduration, 8.);
	pctoflifewithCKD2 = input(pctoflifewithCKD, 8.);

	if UPCRATIO ge 2 then nephUPCR = 1;
		else nephUPCR = 0;

	if BMIPCTAG > 85 then BMI_cat = 1;
		else BMI_cat = 0;

	if nonHDL > 190 then nonHDL_cat = 1;
		else nonHDL_cat = 0;

	if TRIG > 130 OR HDL < 40 OR nonHDL > 160 then dyslipid = 1;
```

```
                    else dyslipid = 0;
run;

proc freq data=dsic_all_bs2;
        tables g_diag race2 nephUPCR BMI_cat nonHDL_cat dyslipid;
run;

proc freq data=dsic_all_bs2;
        tables (male1fe0 race2 hisp nephUPCR BMI_cat nonHDL_cat dyslipid LIPLOWRX)*g_diag;
run;

proc sort data=dsic_all_bs2; by g_diag; run;
proc means data=dsic_all_bs2 n mean median q1 q3;
        var age CKDduration2 pctoflifewithCKD2 BEDGFR uPCRatio BMIPCTAG TOTALCHOL TRIG HDL
nonHDL FUyrs;
        by g_diag;
run;

* create no. of lipid measures variable;
data count;
        set dsic;
        by caseid;
        retain N;
        if first.caseid then N = 1;
        else N = N+1;
        if last.caseid then output;
run;

proc sort data=count; by caseid; run;
proc sort data=dsic_all_bs2; by caseid; run;
data count2;
        merge   count (in=a)
                dsic_all_bs2 (in=b keep=caseid g_diag);
        by caseid;
        if a then output count2;
run;

proc freq data=count2;
        tables N*g_diag;
run;
```