Dataset Integrity Check for The Chronic Kidney Disease in Children Cohort Study (CKiD) Pierce et al. Analysis 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 – Demographic and Clinical Characteristics of the Study Population According to Chronic Kidney Disease Diagnosis, Chronic Kidney Disease in Children Study, 2005-2010 4

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

Attachment A: SAS Code 6

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.

The Pierce et al. 2011 analysis sought to characterize the effect of glomerular CKD versus nonglomerular CKD diagnoses on annualized GFR ratio.

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 "pierceaje2011_gfrchange_pub.sas7bdat" dataset.

4 Statistical Methods

Analyses were performed to replicate results for the data published by Pierce et al. [1] for Methods for Characterizing Differences in Longitudinal Glomerular Filtration Rate Changes Between Children With Glomerular Chronic Kidney Disease and Those With Nonglomerular Chronic Kidney Disease. 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 the Study Population According to Chronic Kidney Disease Diagnosis, Chronic Kidney Disease in Children Study, 2005–2010, 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] Pierce CB, Cox C, Saland JM, Furth SL, Muñoz A. Methods for Characterizing Differences in Longitudinal Glomerular Filtration Rate Changes Between Children with Glomerular Chronic Kidney Disease and Those With Nonglomerular Chronic Kidney Disease. American Journal of Epidemiology, 174(5), 604-612, August 2011. doi: https://doi.org/10.1093/aje/kwr121
PMCID: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3202151/

Table A: Variables used to replicate Table 1 – Demographic and Clinical Characteristics of the Study Population According to Chronic Kidney Disease Diagnosis, Chronic Kidney Disease in Children Study, 2005-2010

Table Variable	**dataset.variable**
Age	pierceaje2011_gfrchange_pub.age
Female Sex	pierceaje2011_gfrchange_pub.male1fe0
Race	pierceaje2011_gfrchange_pub.race_1w2b3o
Hispanic ethnicity	pierceaje2011_gfrchange_pub.hisp
Duration of CKD	pierceaje2011_gfrchange_pub.ckddur
Number of GFR measurements per subject	pierceaje2011_gfrchange_pub.iecount
Years of follow-up	pierceaje2011_gfrchange_pub.yrsfu
GFR level	pierceaje2011_gfrchange_pub.iegfr
Annualized percentage change in GFR	pierceaje2011_gfrchange_pub.iegfryrratio

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

Characteristic	Manuscript Total (n=529)	DSIC Total (n=529)	Diff. (n=0)	Manuscript Glomerular CKD (n=109)	DSIC Glomerular CKD (n=109)	Diff. (n=0)	Manuscript Non-glomerular CKD (n=420)	DSIC Non-glomerular CKD (n=420)	Diff. (n=0)	Manuscript *p*-value	DSIC *p*-value	Diff.
Age, years	11 (7 to 14)	11 (7 to 14)	0 (0 to 0)	14 (11 to 16)	14 (11 to 16)	0 (0 to 0)	10 (6 to 14)	10 (6 to 14)	0 (0 to 0)	<0.01	<0.01	0
Female sex	203 (38)	203 (38)	0 (0)	52 (48)	52 (48)	0 (0)	151 (36)	151 (36)	0 (0)	0.03	0.03	0
Race										<0.01	<0.01	0
White	354 (67)	354 (67)	0 (0)	58 (53)	58 (53)	0 (0)	296 (70)	296 (70)	0 (0)			
African American	116 (22)	116 (22)	0 (0)	37 (34)	37 (34)	0 (0)	79 (19)	79 (19)	0 (0)			
Other	59 (11)	59 (11)	0 (0)	14 (13)	14 (13)	0 (0)	45 (11)	45 (11)	0 (0)			
Hispanic ethnicity	77 (15)	77 (15)	0 (0)	19 (17)	19 (17)	0 (0)	58 (14)	58 (14)	0 (0)	0.36	0.36	0
Duration of CKD, years	6.4 (3.1 to 10.3)	6.0 (3.0 to 10.0)	0.4 (0.1 to 0.3)	3.8 (1.7 to 7.4)	4.0 (2.0 to 7.0)	0.2 (0.3 to 0.4)	7.1 (3.5 to 10.8)	7.0 (4.0 to 11.0)	0.1 (0.5 to 0.2)	<0.01	<0.01	0
Number of GFR measurements per subject										0.35	0.35	0
2	59 (11)	58 (11)	1 (0)	17 (16)	17 (16)	0 (0)	41 (10)	41 (10)	0 (0)			
3	156 (29)	156 (29)	0 (0)	35 (32)	35 (32)	0 (0)	121 (29)	121 (29)	0 (0)			
4	173 (33)	173 (33)	0 (0)	32 (29)	32 (29)	0 (0)	141 (34)	141 (34)	0 (0)			
5	118 (22)	118 (22)	0 (0)	22 (20)	22 (20)	0 (0)	96 (23)	96 (23)	0 (0)			
6	24 (5)	24 (5)	0 (0)	3 (3)	3 (3)	0 (0)	21 (5)	21 (5)	0 (0)			
Years of follow-up	3.0 (2.1 to 3.9)	3.0 (2.1 to 3.9)	0 (0 to 0)	2.8 (1.8 to 3.7)	2.8 (1.8 to 3.7)	0 (0 to 0)	3.0 (2.1 to 3.9)	3.0 (2.1 to 3.9)	0 (0 to 0)	<0.01	<0.01	0
GFR level, mL/minute/ 1.73 m^2	44 (33 to 57)	44 (33 to 57)	0 (0 to 0)	48 (34 to 63)	48 (34 to 63)	0 (0 to 0)	44 (33 to 55)	44 (33 to 55)	0 (0 to 0)	0.04	0.04	0
Annualized percentage change in GFR, % per year	-4.7 (-13.1 to 1.7)	-4.7 (-13.1 to 1.7)	0 (0 to 0)	-10.5 (-23.4 to -1.2)	-10.5 (-23.4 to -1.2)	0 (0 to 0)	-3.9 (-11.4 to 2.2)	-3.9 (-11.4 to 2.2)	0 (0 to 0)	<0.01	<0.01	0

[1] Expressed as number of cases (%) or median (IQR), as appropriate

Attachment A: SAS Code

```
libname ckid "Z:\NIDDK\niddk-dr_studies1\CKiD\private_orig_data\CKiD Upload 04-22-16\analytical
files 02\pierce.aje_2011";

data public_data;
	set ckid.pierceaje2011_gfrchnge_pub;
	FSGS1_HUS2_OTHR3=1*(PRIMDX=10)+2*(PRIMDX=11)+3*(11<PRIMDX<50);
	if glomdx=0 then FSGS1_HUS2_OTHR3=.;
run;

*** THIS DATA SET CONTAINS 2010 ID-VISIT RECORDS FOR 529 SUBJECTS ***;

** visit distribution - N=529 baseline visits (v10);
proc freq data=public_data;
	table visit;
	run;

*** RESTRICT TO BASELINE DATA (VISIT=10) ***;
data allr_bs;
	set public_data;
	if visit=10;
	run;

****************************************
** Table 1:  N=529, Overall statistics **
****************************************;

title "Table 1:  N=529, Overall statistics";
proc means data=allr_bs n nmiss median q1 q3;
	var age ckddur yrsFU ieGFR ieGFRyrratio;
	run;

proc freq;
	table  male1fe0 race_1w2b3o  hisp iecount;
	run;
title;

****************************************;

****************************************
** Table 1:   N=529: by CKD diagnosis **
****************************************;

title "Table 1:  N=529, statistics by CKD diagnosis";
proc sort; by glomdx;run;
proc means data=allr_bs n nmiss median q1 q3;
```

```
        var age ckddur yrsFU ieGFR ieGFRyrratio;
        by glomdx;
        run;

* Wilcoxon/ Kruskal-Wallis tests;
proc npar1way wilcoxon;
        class glomdx;
        var age ckddur yrsFU ieGFR ieGFRyrratio;
        run;

proc freq;
        table  (male1fe0 race_1w2b3o hisp iecount)*glomdx/fisher;
        run;
```