Dataset Integrity Check for the Association of Proteinuria with Race, Cause of Chronic Kidney Disease, and Glomerular Filtration Rate in the Chronic Kidney Disease in Children Study(CKiD)

Version 1 Prepared by
IMS
3901 Calverton Blvd
Calverton MD 20705
January 27, 2014

Revision History

Version	Author/Title	Date	Comments
1.0	M Spriggs	January 27, 2014	Original

Contents

1 Standard Disclaimer

The intent of this DSIC is to provide confidence that the data distributed by the National Institute of Diabetes and Digestive and Kidney Diseases (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, and other factors. Experience suggests that most discrepancies can ordinarily be resolved by consulting 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 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

The CKiD Study is a multi-center, prospective cohort study of children aged 1 to 16 years with mild to moderate impaired kidney function. Two clinical coordinating centers at Children's Mercy Hospital in Kansas and at Children's Hospital of Philadelphia in Philadelphia, PA (previously at the Johns Hopkins Medical Institutions in Baltimore, MD), a central laboratory at the University of Rochester, and a data coordinating center at Johns Hopkins School of Public Health have formed a cooperative agreement to conduct a prospective study of chronic kidney disease in children. The scientific aims of CKiD have been to determine the risk factors for decline in kidney function and to define how progressive decline in kidney function impacts biomarkers of risk factors for cardiovascular disease; growth failure and its associated morbidity; and neurocognitive function and behavior.

Proteinuria is associated with chronic kidney disease (CKD), and heavy proteinuria predicts a rapid decline in kidney function. However, the epidemiologic distribution of this important biomarker study is not well described in the pediatric CKD population.

2.1 Study Methods

Briefly, children enrolled in CKiD were 1 to 16 yr of age and had a Schwartz-estimated GFR (12,13) between 30 and 90ml/min per 1.73m2. Exclusion criteria included: renal, other solid-organ, bone marrow, or stem cell transplantation; dialysis treatment within the past 3 mo; cancer/leukemia diagnosis or HIV diagnosis/treatment within the past 12 mo; current pregnancy or pregnancy within the past 12 mo; history of structural heart disease; genetic syndromes involving the central nervous system; and history of severe to profound mental retardation.

The diagnoses of CKD were reviewed by the members of the CKiD Steering Committee and categorized as either glomerular or nonglomerular. Glomerular diagnoses include chronic glomerulonephritis, congenital nephrotic syndrome, diffuse mesangial sclerosis (Denys-Drash syndrome), diabetic nephropathy, familial nephritis, focal segmental glomerulosclerosis, hemolytic uremic syndrome, Henoch Schonlein nephritis, idiopathic crescentic glomerulonephritis, IgA nephropathy, membranoproliferative glomerulonephritis types I and II, membranous nephropathy, sickle cell nephropathy, and systemic immunological disease including systemic lupus erythematosis. Nonglomerular diagnoses included aplastic-, hypoplastic, and dysplastic kidneys; cystinosis; medullary cystic disease/juvenile nephronophthisis, obstructive uropathy; oxalosis; autosomal dominant and recessive polycystic kidney disease; pyelonephritis/interstitial nephritis;

reflux nephropathy; renal infarct; syndrome of agenesis of abdominal musculature; and Wilm′s Tumor. CKD diagnoses not subscribing to one of the groups listed above were reviewed by the Steering Committee and, if necessary, discussed with the clinical site, so as to be properly categorized. The duration of a participants′ CKD was determined as the time between their baseline study visit and when they first became aware of their CKD diagnosis.

3 Archived Datasets

The DCC submitted 1 dataset that was used for the analysis for this paper that we used for this DSIC: WONGCJASN2009R. Contents of the archived dataset match descriptions provided in the document, Codebook for SAS dataset wongcjasn2009r.pdf

4 Statistical Methods

We compared our DSIC results to the published results in:

- Table 1. Patient characteristics

Our DSIC analyses were conducted in SAS v9 (Appendix 1). The SAS code and output used to support the findings of the DSIC appear as Appendix 1.

Patient characteristics between protein uria groups are provided in Table 1, which presents study Ns and percentages as well as means ± standard deviations where appropriate.

5 Results

Variables used to replicate Table 1. **Patient characteristics** are shown in Table A.

Table A: Variables Used to Replicate Table 1.

Measure	Variable
Protein Uria	PCRG
Age	AGE
Sex	FEMALE
Race	RACECAT
Hispanic Ethnicity	HISP
BMI	BMI
BMI standard deviation score	BMIZ
Obese, BMI 95th percentile (SDS)	OBESE
Serum albumin, g/dl	ALBUMIN
Hypoalbuminemia (albumin 4 g/dl)	HYPOALB
Glomerular CKD diagnosis, % (n)	GLOMDX
Iohexol-based GFR, ml/min per 1.73 m2	IGFR
K-DOQI CKD Stage, % (n)	STAGE
CKD duration, years	DURATION
Uncontrolled Systolic Hypertension, %	SHYP
Uncontrolled Diastolic Hypertension, %	DHYP
ACE-Inhibitor/ARB use, % (n)	ACEARB

DSIC Results: Table 1. The published manuscript results and the DSIC results for Table 1 are shown below (Table B). The base Ns and medians and interquartile rangers for the patient characteristics and histology results calculated by the DSIC correspond to published values, with only inconsequential discrepancies.

Table B: Table 1. Baseline Characteristics of Study Participants.

	Wong et al (2009)		**DSIC**	
	Overall (n 419)	Normal (n 101)	Overall (n 419)	Normal (n 101)
Age, years	11 [7,14]	9 [7,13]	11 [7,14]	9 [7,13]
Male, % (n)	62 (258)	61 (62)	62% (258)	61% (62)
Race, % (n)				
Caucasian	68 (285)	75 (76)	68% (285)	75% (76)
African-American	16 (67)	13 (13)	16% (67)	13% (13)
Multi-racial or other	16 (66)	12 (12)	16% (66)	12% (12)
Hispanic ethnicity	14 (59)	10 (10)	14% (59)	10% (10)
BMI	18 [16, 21]	17 [16, 20]	18 [16,21]	17 [16,20]
BMI standard deviation score	0.36 [-0.38,1.19]	0.24 [-0.41,0.91]	0.36 [-0.38,1.19]	0.24 [-0.41,0.91]
Obese, BMI 95th percentile (SDS)	16% (66)	14% (14)	16% (66)	14% (14)
Serum albumin, g/dl	4.3 [4.1,4.5]	4.4 [4.3,4.6]	4.3 [4.1,4.5]	4.4 [4.3,4.6]
Hypoalbuminemia (albumin 4 g/dl)	14% (60)	4% (4)	14% (60)	4% (4)
Glomerular CKD diagnosis, % (n)	22% (91)	12% (12)	22% (91)	12% (12)
Iohexol-based GFR, ml/min per 1.73 m2	42 [32,52]	50 [38,59]	42 [32,52]	50 [38,59]
K-DOQI CKD Stage, % (n)				
I	<1% (3)	1% (1)	1% (3)	1% (1)
II	15% (61)	24% (24)	15% (61)	24% (24)
III	65% (274)	65% (66)	65% (274)	65% (66)
IV	19% (81)	10% (10)	19% (81)	10% (10)
CKD duration, years	6 [3,10]	6 [3,10]	6 [3,10]	6 [3,10]
Uncontrolled Systolic Hypertension, % (n)	14% (59)	13% (13)	14% (59)	13% (13)
Uncontrolled Diastolic Hypertension, % (n)	15% (60)	13% (13)	15% (60)	13% (13)
ACE-Inhibitor/ARB use, % (n)	54% (228)	51% (51)	54% (228)	50% (51)

Table B: Table 1. Baseline Characteristics of Study Participants.

	Wong et al (2009)		**DSIC**	
	Significant (n 258)	Nephrotic (n 60)	Significant (n 258)	Nephrotic (n 60)
Age, years	11 [7,14]	12 [9,15]	11 [7,14]	12 [9,15]
Male, % (n)	62 (160)	60 (36)	62% (160)	60% (36)
Race, % (n)				
Caucasian	68 (174)	58 (35)	68% (174)	58% (35)
African-American	16 (41)	22 (13)	16% (41)	22% (13)
Multi-racial or other	16 (42)	20 (12)	16% (42)	20% (12)
Hispanic ethnicity	15 (39)	17 (10)	15% (39)	17% (10)
BMI	18 [16, 21]	20 [17, 23]	18 [16,21]	20 [17,23]
BMI standard deviation score	0.32 [-0.43,1.25]	0.66 [0.05,1.25]	0.32 [-0.43,1.25]	0.66 [0.05,1.25]
Obese, BMI 95th percentile (SDS)	17% (42)	18% (10)	17% (42)	18% (10)
Serum albumin, g/dlf	4.3 [4.1,4.5]	3.9 [3.3,4.2]	4.3 [4.1,4.5]	3.9 [3.3,4.2]
Hypoalbuminemia (albumin 4 g/dl)	9% (24)	54% (32)	9% (24)	54% (32)
Glomerular CKD diagnosis, % (n)	19% (50)	48% (29)	19% (50)	48% (29)
Iohexol-based GFR, ml/min per 1.73 m2	43 [32,51]	33 [28,40]	43 [32,51]	33 [28,40]
K-DOQI CKD Stage, % (n)				
I	<1% (2)	0% (0)	1% (2)	0% (0)
II	13% (33)	7% (4)	13% (33)	7% (4)
III	66% (171)	62% (37)	66% (171)	62% (37)
IV	20% (52)	32% (19)	20% (52)	32% (19)
CKD duration, years	6 [3,10]	7 [3,9]	6 [3,10]	7 [3,9]
Uncontrolled Systolic Hypertension, % (n)	14% (35)	19% (11)	14% (35)	19% (11)
Uncontrolled Diastolic Hypertension, % (n)	13% (33)	24% (14)	13% (33)	24% (14)
ACE-Inhibitor/ARB use, % (n)	56% (144)	55% (33)	56% (144)	55% (33)

6 Conclusions

The results of these DSIC analyses provide confidence that the CKiD data distributed by the NIDDK repository are a true copy of the study data.

7 References

Wong CS, Pierce CB, Cole SR, Warady BA, Mak RH, Benador NM, et al. Association of proteinuria with race, cause of chronic kidney disease, and glomerular filtration rate in the chronic kidney disease in children study. Clin J Am Soc Nephrol. 2009;4(4):812–819.

Appendix 1. SAS Output used to Replicate Manuscript Results.

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

proc format;
  value pcrg
    .="Overall"
    0="Normal"
    1="Significant"
    2="Nephrotic"
    ;

%macro meandata2(invar=, roundvar=, digit=);
proc means data=wongcjasn2009r nmiss median q1 q3 noprint;
      var &invar;
      class pcrg;
      output out=data1 nmiss=nmiss median=median q1=q1 q3=q3;
      run;

data data1(drop=_TYPE_ median q1 q3 rename=(_FREQ_=COUNT));
  set data1;
  length name CHARALL $100;
  name=upcase("&invar");
  median=round(median,&roundvar);
  q1=round(q1,&roundvar);
  q3=round(q3,&roundvar);
  CHARALL=compress(put(median,8.&digit))||" ["||compress(put(q1,8.&digit))||","||compress(put(q3,8.&digit))||"]";


data accummeans2;
  set accummeans2 data1;

%mend meandata2;

%macro freqdata(invar=);
proc freq data=wongcjasn2009r compress noprint;
      tables &invar/out=data1;
      run;

data data1(keep=pcrg LEVEL name CHARALL);
  set data1(rename=(&invar=LEVEL));
  length name $100 CHARALL $100;
  name=upcase("&invar");
  PCT_DISP=round(PERCENT);
  pcrg=.;
  CHARALL=compress(put(PCT_DISP,8.))||'% ('||compress(put(COUNT,8.))||')';

data accumfreq1;
  set accumfreq1 data1;
%mend freqdata;

%macro freqdata2(invar=);
proc freq data=wongcjasn2009r compress noprint;
      tables pcrg*&invar/out=data1 outpct;
      run;
```

```
data data1(keep=pcrg LEVEL name CHARALL);
  set data1(rename=(&invar=LEVEL));
  length name $100;
  name=upcase("&invar");
  PCT_DISP=round(PCT_ROW);
  CHARALL=compress(put(PCT_DISP,8.))||'% ('||compress(put(COUNT,8.))||')';

data accumfreq2;
  set accumfreq2 data1;
%mend freqdata2;

libname jandataa "/prj/niddk/ims_analysis/CKiD/private_orig_data/CKiD Upload 01-16-14/analytical files 01/wong.cjasn_2009/";

data wongcjasn2009r        ; set jandataa.wongcjasn2009r;

data accumfreq1;
  set _null_;

data accumfreq2;
  set _null_;

data accummeans1;
  set _null_;

data accummeans2;
  set _null_;

%freqdata(invar=pcrg);
%freqdata(invar=female);
%freqdata(invar=racecat);
%freqdata(invar=hisp);
%freqdata(invar=obese);
%freqdata(invar=hypoalb);
%freqdata(invar=glomdx);
%freqdata(invar=stage);
%freqdata(invar=shyp);
%freqdata(invar=dhyp);
%freqdata(invar=ACEARB);

%freqdata2(invar=female);
%freqdata2(invar=racecat);
%freqdata2(invar=hisp);
%freqdata2(invar=obese);
%freqdata2(invar=hypoalb);
%freqdata2(invar=glomdx);
%freqdata2(invar=stage);
%freqdata2(invar=shyp);
%freqdata2(invar=dhyp);
%freqdata2(invar=ACEARB);

%meandata2(invar=age, roundvar=1, digit=0);
%meandata2(invar=bmi, roundvar=1, digit=0);
%meandata2(invar=bmiz, roundvar=0.01, digit=2);
%meandata2(invar=albumin, roundvar=.1, digit=1);
%meandata2(invar=iGFR, roundvar=1, digit=0);
%meandata2(invar=duration, roundvar=1, digit=0);

data accumfreq;
  set accumfreq1 accumfreq2;
/*
proc print data=accumfreq noobs;
```

```
  format pcrg pcrg.;
  title 'Accum freq';
*/
data accummeans;
  set accummeans1 accummeans2;
/*
proc print data=accummeans noobs;
  format pcrg pcrg.;
  title 'Accum means';
*/
data accumfreq;
  set accumfreq;
  if name="PCRG" then delete;
  if name="FEMALE" and level ne 0 then delete;
  if name="RACECAT" and level =. then delete;
  if name="HISP" and level ne 1 then delete;
  if name="OBESE" and level ne 1 then delete;
  if name="HYPOALB" and level ne 1 then delete;
  if name="GLOMDX" and level ne 1 then delete;
  if name="SHYP" and level ne 1 then delete;
  if name="DHYP" and level ne 1 then delete;
  if name="ACEARB" and level ne 1 then delete;

data accummeans;
  set accummeans(drop=COUNT nmiss);

data accuminert;
    orderer=3;
    pcrg=.;
    output;

    orderer=3;
    pcrg=0;
    output;

    orderer=3;
    pcrg=1;
    output;

    orderer=3;
    pcrg=2;
    output;

    orderer=15;
    pcrg=.;
    output;

    orderer=15;
    pcrg=0;
    output;

    orderer=15;
    pcrg=1;
    output;

    orderer=15;
    pcrg=2;
    output;

data accumfreqmeans;
  set accumfreq accummeans;
```

```
  if name="AGE" then orderer=1;
  if name="FEMALE" then orderer=2;
  if name="RACECAT" and level=1 then orderer=4;
  if name="RACECAT" and level=2 then orderer=5;
  if name="RACECAT" and level=3 then orderer=6;
  if name="HISP" then orderer=7;
  if name="BMI" then orderer=8;
  if name="BMIZ" then orderer=9;
  if name="OBESE" then orderer=10;
  if name="ALBUMIN" then orderer=11;
  if name="HYPOALB" then orderer=12;
  if name="GLOMDX" then orderer=13;
  if name="IGFR" then orderer=14;
  if name="STAGE" and level=1 then orderer=16;
  if name="STAGE" and level=2 then orderer=17;
  if name="STAGE" and level=3 then orderer=18;
  if name="STAGE" and level=4 then orderer=19;
  if name="DURATION" then orderer=20;
  if name="SHYP" then orderer=21;
  if name="DHYP" then orderer=22;
  if name="ACEARB" then orderer=23;
  output;

data accumall;
  set accumfreqmeans accuminert;

proc sort data=accumall;
  by orderer;

proc print data=accumall noobs;
  var LEVEL name CHARALL;
  where pcrg=.;
  title 'accumall Overall';

proc print data=accumall noobs;
  var LEVEL name CHARALL;
  where pcrg=0;
  title 'accumall Normal';

proc print data=accumall noobs;
  var LEVEL name CHARALL;
  where pcrg=1;
  title 'accumall Significant';

proc print data=accumall noobs;
  var LEVEL name CHARALL;
  where pcrg=2;
  title 'accumall Nephrotic';
```

LEVEL	name	CHARALL
.	AGE	11 [7,14]
0	FEMALE	62% (258)
.		
1	RACECAT	68% (285)
2	RACECAT	16% (67)
3	RACECAT	16% (66)
1	HISP	14% (59)
.	BMI	18 [16,21]
.	BMIZ	0.36 [-0.38,1.19]
1	OBESE	16% (66)
.	ALBUMIN	4.3 [4.1,4.5]
1	HYPOALB	14% (60)
1	GLOMDX	22% (91)
.	IGFR	42 [32,52]
.		
1	STAGE	1% (3)
2	STAGE	15% (61)
3	STAGE	65% (274)
4	STAGE	19% (81)
.	DURATION	6 [3,10]
1	SHYP	14% (59)
1	DHYP	15% (60)
1	ACEARB	54% (228)

LEVEL	name	CHARALL
.	AGE	9 [7,13]
0	FEMALE	61% (62)
.		
1	RACECAT	75% (76)
2	RACECAT	13% (13)
3	RACECAT	12% (12)
1	HISP	10% (10)
.	BMI	17 [16,20]
.	BMIZ	0.24 [-0.41,0.91]
1	OBESE	14% (14)
.	ALBUMIN	4.4 [4.3,4.6]
1	HYPOALB	4% (4)
1	GLOMDX	12% (12)
.	IGFR	50 [38,59]
.		
1	STAGE	1% (1)
2	STAGE	24% (24)
3	STAGE	65% (66)
4	STAGE	10% (10)
.	DURATION	6 [3,10]
1	SHYP	13% (13)
1	DHYP	13% (13)
1	ACEARB	50% (51)

LEVEL	name	CHARALL
.	AGE	11 [7,14]
0	FEMALE	62% (160)
.		
1	RACECAT	68% (174)
2	RACECAT	16% (41)
3	RACECAT	16% (42)
1	HISP	15% (39)
.	BMI	18 [16,21]
.	BMIZ	0.32 [-0.43,1.25]
1	OBESE	17% (42)
.	ALBUMIN	4.3 [4.1,4.5]
1	HYPOALB	9% (24)
1	GLOMDX	19% (50)
.	IGFR	43 [32,51]
.		
1	STAGE	1% (2)
2	STAGE	13% (33)
3	STAGE	66% (171)
4	STAGE	20% (52)
.	DURATION	6 [3,10]
1	SHYP	14% (35)
1	DHYP	13% (33)
1	ACEARB	56% (144)

LEVEL	name	CHARALL
.	AGE	12 [9,15]
0	FEMALE	60% (36)
.		
1	RACECAT	58% (35)
2	RACECAT	22% (13)
3	RACECAT	20% (12)
1	HISP	17% (10)
.	BMI	20 [17,23]
.	BMIZ	0.66 [0.05,1.25]
1	OBESE	18% (10)
.	ALBUMIN	3.9 [3.3,4.2]
1	HYPOALB	54% (32)
1	GLOMDX	48% (29)
.	IGFR	33 [28,40]
.		
2	STAGE	7% (4)
3	STAGE	62% (37)
4	STAGE	32% (19)
.	DURATION	7 [3,9]
1	SHYP	19% (11)
1	DHYP	24% (14)
1	ACEARB	55% (33)