Dataset Integrity Check for Ambulatory Blood Pressure Patterns in Children With Chronic Kidney Disease Data Files

Prepared by Michael Spriggs
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton MD 20705
May 6, 2015

Table of 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 SAS data files, as provided by the Data Coordinating Center (DCC), are located in the data package. For this replication, variables were taken from nine datasets and a paper population dataset created by the DCC.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Samuels, et al [1] in Hypertension in June 2012. To verify the integrity of the dataset, descriptive statistics were computed (Table B and Table D).

5 Results

Table 1 in the publication [1], Baseline Demographics Comparing Attainable (n=332), Unattainable (n=70), and Not Attempted (n=115) ABPMs. Table A lists the variables that were used in the replication and Table B compares the results calculated from the archived data file to the results published in Table I. The results of the replication are similar to the published results.

Table 2 in the publication [1] Ambulatory Blood Pressure Parameters for Attainable ABPM Studies (n=332)). Table C lists the variables that were used in the replication and Table D compares the results calculated from the archived data file to the results published in Table II. 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 copy of the manuscript data.

7 References

Samuels, Joshua, et al. "Ambulatory blood pressure patterns in children with chronic kidney disease." Hypertension 60.1 (2012): 43-50.

Table A: Variables used to replicate Table 1: <u>Baseline Demographics Comparing Attainable (n=332), Unattainable (n=70), and Not Attempted (n=115) ABPMs</u>

Table Variable	Variables Used in Replication from the Datasets
Age, y	KIDHIST.DOB KIDHIST.BSDATE
Male	KIDHIST.MALE1FE0
Race	SOCDEM.RACE
Hispanic ethnicity	SOCDEM.HISP
Height percentile	GROWTH.HTPCTAG
Weight percentile	GROWTH.WTPCTAG
Weight, kg	GROWTH.AVWEIGHT
BMI percentile	GROWTH.BMIPCTAG
Percent obese, BMI >95th percentile	GROWTH.BMIPCTAG
Iohexol GFR,† mL/min per 1.73 m2	GFRSUMMARY.EGFR GFRSUMMARY.IGFR
Urine protein:creatinine†	L05.rlurprot L05.rlurcrea
Low birth weight	GROWTH.LBW
Premature birth	GROWTH.PREMATURE
Glomerular CKD	KIDHIST.GNGDIAG
Duration of CKD, y	KIDHIST.CKDONST
History of HTN	mh.MH_HBP
Current HTN by casual readings	Cardio.SHYPagh Cardio.DHYPagh
Use of antihypertensive	medsum_short.ANTIHYP
ACE inhibitor	medsum_short.ACEI
Calcium channel blocker	medsum_full.med_code
ARB	medsum_short.ARB

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

Characteristic	Attainable (n=332) [Manuscript]	Attainable (n=343) [DSIC]	Attainable (n=-11) [Diff]	Unattainable (n=70) [Manuscript]	Unattainable (n=70) [DSIC]	Unattainable (n=0) [Diff]
Age, y	12 [9–15]	12 [8 - 15]	0[0-0]	9 [6–15]	9 [5 - 15]	0[1-0]
Male	59% (195)	59% (201)	0(-6)	70% (49)	71% (50)	-1(-1)
Race						
White	73% (242)	72% (246)	1(-4)	57% (40)	57% (40)	0(0)
Black	18% (59)	19% (64)	-1(-5)	30% (21)	30% (21)	0(0)
Other	9% (31)	10% (33)	-1(-2)	13% (9)	13% (9)	0(0)
Hispanic ethnicity	15% (50)	15% (52)	0(-2)	11% (8)	11% (8)	0(0)
Height percentile	24 [8–51]	24 [9 - 52]	0[-1-1]	26 [7–48]	25 [6 - 48]	1[1-0]
Weight percentile	44 [19–78]	44 [19 - 78]	0[0-0]	45 [18–76]	46 [18 - 76]	-1[0-0]
Weight, kg	42 [29–58]	42 [28 - 57]	0[1-1]	35 [21–57]	35 [21 - 57]	0[0-0]
BMI percentile	61 [29–89]	61 [30 - 86]	0[-1-3]	56 [37–82]	60 [37 - 82]	4[0-0]
Percent obese, BMI >95th percentile	17% (55)	16% (55)	1{0}	14% (10)	13% (9)	1(1)
Iohexol GFR,† mL/min per 1.73 m2	44 [32–55]	44 [32 - 57]	0[0-2]	48 [34–64]	49 [34 - 64]	-1[0-0]
Urine protein:creatinine†	0.44 [0.18–1.07]	0.43 [0.18 - 1.07]	0.01[0-0]	0.44 [0.18–1.24]	0.44 [0.16 - 1.24]	0[0.02-0]
Low birth weight	17% (54)	17% (57)	0(-3)	19% (13)	18% (12)	1(1)
Premature birth	10% (31)	10% (34)	0(-3)	20% (13)	18% (12)	2(1)
Glomerular CKD	19% (63)	19% (65)	0(-2)	30% (21)	29% (20)	1(1)
Duration of CKD, y	9 [4–12]	9 [5 - 13]	0[1-1]	6 [5–10]	6 [4 - 9]	0[1-1]
History of HTN	46% (152)	46% (157)	0(6)	50% (34)	49% (34)	1(0)
Current HTN by casual readings	19% (61)	19% (63)	0(-2)	18% (12)	16% (11)	2(1)
Use of antihypertensive	68% (226)	69% (234)	-1(-8)	64% (45)	66% (46)	-2(-1)
ACE inhibitor	52% (173)	52% (178)	0(-5)	46% (32)	44% (31)	2(1)
Calcium channel blocker	14% (46)	15% (47)	-1(-1)	16% (11)	17% (11)	1(0)
ARB	9% (31)	10% (34)	-1(-3)	11% (8)	10% (7)	1(1)

Characteristic	No ABPM Study Attempted (n=115) [Manuscript]	No ABPM Study Attempted (n=104) [DSIC]	No ABPM Study Attempted (n=11) [Diff]
Age, y	10 [5–15]	9 [4 - 14]	1[1-1]
Male	61% (70)	61% (63)	0(7)
Race			
White	55% (63)	56% (58)	-1(5)
Black	30% (35)	28% (29)	2(6)
Other	15% (17)	16% (17)	-1(0)
Hispanic ethnicity	14% (16)	13% (13)	1(3)
Height percentile	32 [11–60]	32 [10 - 60]	0[1-0]
Weight percentile	53 [23–81]	51 [20 - 81]	2[3-0]
Weight, kg	36 [18–57]	35 [18 - 57]	1[0-0]
BMI percentile	66 [38–89]	62 [38 - 87]	4[0-2]
Percent obese, BMI >95th percentile	15% (17)	15% (16)	0(1)
Iohexol GFR,† mL/min per 1.73 m2	46 [36–60]	43 [34 - 61]	2[2—1]
Urine protein:creatinine†	0.41 [0.14–1.26]	0.44 [0.14 - 1.29]	0.03[0-0.03]
Low birth weight	22% (24)	22% (22)	0(2)
Premature birth	20% (21)	18% (18)	2(3)
Glomerular CKD	19% (22)	19% (20)	0(2)
Duration of CKD, y	7 [3–11]	8 [3 - 13]	-1[0—2]
History of HTN	38% (44)	38% (39)	0(5)
Current HTN by casual readings	14% (14)	14% (13)	0(1)
Use of antihypertensive	52% (60)	50% (52)	2(8)
ACE inhibitor	37% (43)	38% (39)	-1(-4)
Calcium channel blocker	11% (13)	15% (13)	4(0)
ARB	10% (11)	9% (9)	1(2)

Table C: Variables used to replicate Table 2: <u>Ambulatory Blood Pressure Parameters for</u>

<u>Attainable ABPM Studies (n=332)</u>

Table Variable	Variables Used in Replication
Wake Systolic	CARDIO.WKSYSINDX, CARDIO.WKSYSLOAD
Wake Diastolic	CARDIO.WKDIAINDX, CARDIO.WKDIALOAD
Sleep Systolic	CARDIO.SLSYSINDX, CARDIO.SLSYSLOAD
Sleep Diastolic	CARDIO.SLDIAINDX, CARDIO.SLDIALOAD

Table D: Comparison of values computed in integrity check to reference article Table 2 values

Characteristic	Index Median [IQR] [Manuscript]	Index Median [IQR] [DSIC]	Index Median [IQR] [Diff]	Load Median [IQR] [Manuscript]	Load Median [IQR] [DSIC]	Load Median [IQR] [Diff]
Wake Systolic	0.91 [0.85–0.96]	0.90 [0.85 - 0.96]	0.01[0-0]	10% [3%–28%]	9 [2 - 28]	1[1-0]
Wake Diastolic	0.83 [0.79–0.92]	0.83 [0.78 - 0.92]	0[0.01-0]	7% [2%–21%]	7 [2 - 20]	0[0-1]
Sleep Systolic	0.91 [0.86–0.98]	0.91 [0.85 - 0.98]	0[0.01-0]	12% [0%–38%]	11 [0 - 36]	1[0-2]
Sleep Diastolic	0.89 [0.81–0.98]	0.89 [0.80 - 0.97]	0[0.01-0.01]	17% [4%–40%]	16 [4 - 38]	1[0-2]

Characteristic	Index >1, % (n) [Manuscript]	Index >1, % (n) [DSIC]	Index >1, % (n) [Diff]	Load >25%, % (n) [Manuscript]	Load >25%, % (n) [DSIC]	Load >25%, % (n) [Diff]
Wake Systolic	13% (43)	10% (36)	3(7)	28% (93)	27% (93)	1(0)
Wake Diastolic	7% (24)	7% (23)	0(1)	22% (72)	21% (71)	1(1)
Sleep Systolic	18% (61)	17% (58)	1(3)	33% (110)	32% (109)	1(1)
Sleep Diastolic	20% (66)	18% (62)	2(4)	39% (129)	37% (128)	2(1)

```
*******************************************************************************************************
***Program:
***Programmer: Michael Spriggs
***Date Created: 4/21/2015
***Purpose: DSIC of CKiD data upload of 03-30-15
***
*******************************************************************************************************;

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

options nofmterr mprint source2;

libname sas_data "/prj/niddk/ims_analysis/CKiD/private_orig_data/CKiD Upload 03-30-15/P03/data";
libname add_data "/prj/niddk/ims_analysis/CKiD/private_orig_data/data_for_dsic";

*** File containing macro for examining each dataset ***;
%include '/prj/niddk/ims_analysis/sas_macros/redaction_data_summary.sas';

*** formats ***;

proc format;
  value ABPM
  0= "0= Not successful ABPM"
  1= "1= Successful ABPM"
  -1= "-1= NA"
  9= "9= missing"
  -9= "-9= missing"
  ;


*** Macros ***;

%macro freqdata1(order=, invar=, level=, popvar=, totalvl=);

%if &totalvl.=null %then %do;
  proc freq data=table1 noprint;
        tables &invar*casevar/out=data1 outpct;
        format _all_;
        run;

  data data1(keep=LEVEL casevar name CHARALL ORDERER);
    set data1;
    length name LEVEL CHARALL $100;
    name=upcase("&invar");
    level=strip(&invar);
    PCT_DISP=round(PCT_COL,1);
    CHARALL=compress(put(PCT_DISP,8.))||"% ("||compress(put(COUNT,8.))||")";
    ORDERER=&order;
    if level in &level then output;
%end;
%else %do;
  proc freq data=table1 noprint;
        tables &invar*casevar/out=data1 outpct;
        format _all_;
        where &popvar. in &totalvl.;
```

```
        run;

    data data1(keep=LEVEL casevar name CHARALL ORDERER);
        set data1(rename=(&invar=LEVEL));
        length name $100 CHARALL $100;
        name=upcase("&invar");
        PCT_DISP=round(PCT_COL,.1);
        CHARALL=compress(put(COUNT,8.))||" ("||compress(put(PCT_DISP,8.1))||")";
        ORDERER=&order;
        if level in &level then output;
%end;
data accumfreq1;
    set accumfreq1 data1;

%mend freqdata1;


%macro meandata1(order=, invar=, roundvar=, digit=);
proc means data=table1 mean stddev noprint;
        var &invar;
        class casevar;
        output out=data1 mean=mean stddev=stddev;
        run;

data data1(drop=_TYPE_ _FREQ_ mean stddev);
    set data1;
    length name CHARALL $100;
    name=upcase("&invar");
    mean=round(mean,&roundvar);
    stddev=round(stddev,&roundvar);
    CHARALL=compress(put(mean,8.&digit))||" ± "||compress(put(stddev,8.&digit));
    ORDERER=&order;

data accummean1;
    set accummean1 data1;

%mend meandata1;


%macro mediandata1(order=, invar=, roundvar=, digit=);
proc means data=table1 median p25 p75 min max noprint;
        var &invar;
        class casevar;
        output out=data1 median=median p25=p25 p75=p75 min=min max=max;
        run;

data data1(drop=_TYPE_ _FREQ_ median p25 p75 min max);
    set data1;
    length name CHARALL $100;
    name=upcase("&invar");
    median=round(median,&roundvar);
    min=round(min,&roundvar);
    max=round(max,&roundvar);
    p25=round(p25,&roundvar);
    p75=round(p75,&roundvar);
    ORDERER=&order;
    CHARALL=compress(put(median,8.&digit))||" ["||compress(put(p25,8.&digit))||" - "||compress(put(p75,8.&digit))||"]";
    output;
```

9

```
data accummedian1;
    set accummedian1 data1;

%mend mediandata1;

%macro inertdata1(order=);

    proc freq data=table1 noprint;
        tables casevar/out=data1;
        format _all_;
        run;

    data data1(keep=casevar name CHARALL ORDERER);
        set data1;
        length name $100 CHARALL $100;
        name=" ";
        CHARALL=" ";
        ORDERER=&order;

data accuminert1;
    set accuminert1 data1;

%mend inertdata1;

%macro table1core();

data accumfreq1 accummean1 accummedian1 accuminert1;
    set _null_;

%mediandata1(order=1,  invar=AGE, roundvar=1, digit=0);
%freqdata1(order=2,  invar=MALE1FE0,  level=("1"),popvar=, totalvl=null);
%inertdata1(order=3);
%freqdata1(order=4,  invar=RACE_CHAR, level=("White"),popvar=, totalvl=null);
%freqdata1(order=5,  invar=RACE_CHAR, level=("Black"),popvar=, totalvl=null);
%freqdata1(order=6,  invar=RACE_CHAR, level=("Other"),popvar=, totalvl=null);
%freqdata1(order=7,  invar=HISP, level=("1"),popvar=, totalvl=null);
%mediandata1(order=8,  invar=HTPCTAG, roundvar=1, digit=0);
%mediandata1(order=9,  invar=WTPCTAG, roundvar=1, digit=0);
%mediandata1(order=10,  invar=AVWEIGHT, roundvar=1, digit=0);
%mediandata1(order=11,  invar=BMIPCTAG, roundvar=1, digit=0);
%freqdata1(order=12,  invar=OBESE, level=("1"),popvar=, totalvl=null);
%mediandata1(order=13,  invar=IGFR_DENSE, roundvar=1, digit=0);
%mediandata1(order=14,  invar=uprcr, roundvar=.01, digit=2);
%freqdata1(order=15,  invar=LBW, level=("1"),popvar=, totalvl=null);
%freqdata1(order=16,  invar=PREMATURE, level=("1"),popvar=, totalvl=null);
%freqdata1(order=17,  invar=GLOM1NG0, level=("1"),popvar=, totalvl=null);
%mediandata1(order=18, invar=ckdduration, roundvar=1, digit=0);
%freqdata1(order=19,  invar=MH_HBP, level=("1"),popvar=, totalvl=null);
%freqdata1(order=20,  invar=CURRENTHTN, level=("1"),popvar=, totalvl=null);
%freqdata1(order=21,  invar=ANTIHYP, level=("1"),popvar=, totalvl=null);
%freqdata1(order=22,  invar=ACEI, level=("1"),popvar=, totalvl=null);
%freqdata1(order=23,  invar=CCB, level=("1"),popvar=, totalvl=null);
%freqdata1(order=24,  invar=ARB, level=("1"),popvar=, totalvl=null);

data accumtab1;
```

```
   set accumfreq1 accummean1 accummedian1 accuminert1;
   if casevar=" " then delete;

proc sort data=accumtab1;
   by casevar orderer;

%mend;


%macro table2core();

data accumfreq1 accummean1 accummedian1 accuminert1;
   set _null_;

%mediandata1(order=1,  invar=WKSYSINDX, roundvar=.01, digit=2);
%mediandata1(order=2,  invar=WKDIAINDX, roundvar=.01, digit=2);
%mediandata1(order=3,  invar=SLSYSINDX, roundvar=.01, digit=2);
%mediandata1(order=4,  invar=SLDIAINDX, roundvar=.01, digit=2);
%mediandata1(order=5,  invar=WKSYSLOAD, roundvar=1, digit=0);
%mediandata1(order=6,  invar=WKDIALOAD, roundvar=1, digit=0);
%mediandata1(order=7,  invar=SLSYSLOAD, roundvar=1, digit=0);
%mediandata1(order=8,  invar=SLDIALOAD, roundvar=1, digit=0);
%freqdata1(order=9,    invar=WKSYSINDX_F, level=("1"),popvar=, totalvl=null);
%freqdata1(order=10,   invar=WKDIAINDX_F, level=("1"),popvar=, totalvl=null);
%freqdata1(order=11,   invar=SLSYSINDX_F, level=("1"),popvar=, totalvl=null);
%freqdata1(order=12,   invar=SLDIAINDX_F, level=("1"),popvar=, totalvl=null);
%freqdata1(order=13,   invar=WKSYSLOAD_F, level=("1"),popvar=, totalvl=null);
%freqdata1(order=14,   invar=WKDIALOAD_F, level=("1"),popvar=, totalvl=null);
%freqdata1(order=15,   invar=SLSYSLOAD_F, level=("1"),popvar=, totalvl=null);
%freqdata1(order=16,   invar=SLDIALOAD_F, level=("1"),popvar=, totalvl=null);


data accumtab1;
   set accumfreq1 accummean1 accummedian1 accuminert1;
   if casevar=" " then delete;

proc sort data=accumtab1;
   by casevar orderer;

%mend;

*** Input files ***;

data cardio        ; set sas_data.cardio       ;
data mh            ; set sas_data.mh           ;
data gfrsummary    ; set sas_data.gfrsummary   ;
data growth        ; set sas_data.growth       ;
data kidhist       ; set sas_data.kidhist      ;
data l05           ; set sas_data.l05          ;
data medsum_full   ; set sas_data.medsum_full  ;
data medsum_short  ; set sas_data.medsum_short ;
data socdem        ; set sas_data.socdem       ;
data ids_for_samuels2012; set add_data.ids_for_samuels2012;

data socdem;
   set socdem;
   where visit=20;
```

```
data growth;
  set growth;
  where visit=20;

data l05;
  set l05;
  where visit=20;

*** pull medical history from earliest visit ***;
data mh;
  set mh;
  where visit=10;

data MEDSUM_SHORT;
  set MEDSUM_SHORT;
  where visit=20;

data CARDIO;
  set CARDIO;
  where visit=20;

data gfrsummary;
  set gfrsummary;
  where visit=20;

data medsum_full;
  set medsum_full;
  if visit=20 then output;

proc sort data=medsum_full;
  by caseid;

data medsum_full(keep=caseid ccb);
  set medsum_full;
  by caseid;
  retain ccb;
  if first.caseid then ccb=0;
  if med_code = 10600 then ccb = 1;
  if last.caseid then output;

proc sort data=growth;
  by caseid;

proc sort data=socdem;
  by caseid;

proc sort data=KIDHIST;
  by caseid;

proc sort data=MEDSUM_SHORT;
  by caseid;

proc sort data=CARDIO;
  by caseid;
```

```sas
data table1;
  merge growth socdem KIDHIST l05 mh MEDSUM_SHORT CARDIO gfrsummary(keep=caseid igfr egfr) medsum_full(keep=caseid ccb);
  by CASEID;
  length casevar $20.;
  if ABPMSUCCESS=1 then casevar='1: Attainable';
  else if ABPMSUCCESS=0 then casevar='2: Unattainable';
  else if ABPMSUCCESS=-9 then casevar='3: No ABPM attempted';

  if rlurprot >= 0 and rlurcrea > 0 then uprcr = rlurprot/rlurcrea;

  if RACE=1 then RACE_CHAR="White";
  else if RACE=2 or RACE=8 then RACE_CHAR="Black";
  else RACE_CHAR="Other";
  if BMIPCTAG>95 then OBESE=1;
  else OBESE=0;
  AGE=BSDATE-DOB;

  if WKSYSINDX>1 then WKSYSINDX_F=1; else WKSYSINDX_F=0;
  if WKDIAINDX>1 then WKDIAINDX_F=1; else WKDIAINDX_F=0;
  if SLSYSINDX>1 then SLSYSINDX_F=1; else SLSYSINDX_F=0;
  if SLDIAINDX>1 then SLDIAINDX_F=1; else SLDIAINDX_F=0;
  if WKSYSLOAD>25 then WKSYSLOAD_F=1; else WKSYSLOAD_F=0;
  if WKDIALOAD>25 then WKDIALOAD_F=1; else WKDIALOAD_F=0;
  if SLSYSLOAD>25 then SLSYSLOAD_F=1; else SLSYSLOAD_F=0;
  if SLDIALOAD>25 then SLDIALOAD_F=1; else SLDIALOAD_F=0;

  if gngdiag <= 2 then glom1ng0 = 1; /*Glomerular diagnosis*/
  if gngdiag >= 3 then glom1ng0 = 0; /*Non-glomerular diagnosis*/
  if gngdiag < 0  then glom1ng0 = .; /*Missing CKD diagnosis*/
  ckdduration = abs(ckdonst);

  if igfr>=0 then igfr_dense=igfr;
  else if egfr>=0 then igfr_dense=egfr;
  else if igfr < 0 and egfr < 0 then igfr_dense=.;
  else abort;

  if SHYPagh = 1 or DHYPagh = 1 then currenthtn = 1;
  else if SHYPagh = 0 and DHYPagh = 0 then currenthtn = 0;
  else currenthtn=.;

data table1;
  merge table1 ids_for_samuels2012(keep=caseid in=in_sam);
  by caseid;
  if in_sam then output;

proc freq data=table1;
  tables casevar/missing list;

%table1core();

proc print data=accumtab1 noobs;
  by casevar;
  title3 'Table 1 stats (list)';

%table2core();
```

13

```
proc print data=accumtab1 noobs;
   by CASEVAR;
   where CASEVAR='1: Attainable';
   title3 'Table 2 stats (list)';
```