

Dataset Integrity Check for EDIC CMRI Data File

Prepared by Michael Spriggs
3901 Calverton Blvd, Suite 200 Calverton MD 20705
November 5, 2015

Contents

1 Standard Disclaimer	2
2 Study Background	2
3 Archived Datasets	2
4 Statistical Methods	2
5 Results	3
6 Conclusions	3
7 References	3
Table A: Variables used to replicate Table 1. Clinical Characteristics of Epidemiology of Diabetes Interventions and Complications Participants at the Time of the Cardiac Magnetic Resonance Examination.	4
Table B: Comparison of values computed in integrity check to reference article Table 1 values.....	6
Attachment A: SAS Code	10

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

The Epidemiology of Diabetes Interventions and Complications (EDIC) study was initiated as follow-up to examine the long-term effects of the original DCCT interventions on diabetic complications such as cardiovascular events and advanced retinal and renal disease. Over 90 percent of participants from the DCCT study were followed by the EDIC study. Similar to the DCCT study, glycosylated hemoglobin values, fasting lipid levels, serum creatinine values, and other risk factors for cardiovascular disease were measured at different intervals for participants. Cardiovascular complications were assessed with standardized means and classified by an independent committee. The EDIC study has found that intensive diabetes therapy reduced risk of cardiovascular disease in patients with type 1 diabetes and that the differences in outcomes between the intensive and conventional therapy groups persist after long-term study.

3 Archived Datasets

The SAS data file, as provided by the Data Coordinating Center (DCC), are located in the data package. For this replication, variables were taken from the SAS file nih_cmri_risk1240.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Turkbey, et al [1] in *Circulation* in September 2011. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For Table 1 in the publication [1], Clinical Characteristics of Epidemiology of Diabetes Interventions and Complications Participants at the Time of the Cardiac Magnetic Resonance Examination, Table A lists the variables that can be used in the replication. Table B compares the results calculated from the archived data file to the results published in Table 1. The results of the replication are an exact match when available.

6 Conclusions

The NIDDK repository is confident that the EDIC data files to be distributed are an exact match to the manuscript data when available.

7 References

Turkbey EB, Backlund JYC, Genuth S, et al.; DCCT/EDIC Research Group. Myocardial structure, function, and scar in patients with type 1 diabetes mellitus. *Circulation* 2011;124:1737–1746.

Table A: Variables used to replicate Table 1. Clinical Characteristics of Epidemiology of Diabetes Interventions and Complications Participants at the Time of the Cardiac Magnetic Resonance Examination.

Table Variable	Variables Used in Replication from the Dataset(s)
Age, y	AGE
Duration of diabetes mellitus, y	IDDMYR
Smoking, %	SMOKE
BMI, kg/m ²	BMI
Mean BMI, kg/m ²	DTED_BMI
SBP, mm Hg	SBP
Mean SBP, mm Hg	DTED_SBP
DBP, mm Hg	DBP
Mean DBP, mm Hg	DTED_DBP
Hypertension, %	HT
Antihypertensive medication, %	HYPERTEN
Total cholesterol, mg/dL	TCHOL
Mean total cholesterol, mg/dL	DTED_CHL
Recent HDL, mg/dL	HDL
Mean HDL, mg/dL	DTED_HDL
Recent LDL, mg/dL	LDL
Mean LDL, mg/dL	DTED_LDL
Recent triglycerides, mg/dL	TRIG
Mean triglycerides, mg/dL	DTED_TRG
Hypercholesterolemia, %	HLIP
Lipid-lowering medication, %	LIPID
Recent HbA1c, %	HBA1C
Mean HbA1c, %	WTMHBA
Framingham Risk Score	FSCORE
Clinical or silent MI	-Not included in the dataset
Adjudicated clinical MI events	-Not included in the dataset
Silent MI	-Not included in the dataset
CAC score >0 (years 7–9)	CTGT0
CAC score >200 (years 7–9)	CTGT200
Common IMT (year 12)	COMMNW12
PDR or worse	ANYPDR
Macroalbuminuria/ESRD	ANYAE300
Sustained microalbuminuria/ESRD	ANYSAE30

Table Variable	Variables Used in Replication from the Dataset(s)
Autonomic neuropathy	AFTEDIC13
Peripheral neuropathy	CCNEDIC
Participants with 0 complications	COMP_SCORE
Participants with 1 complication	COMP_SCORE
Participants with 2 complications	COMP_SCORE
Participants with >=3 complications	COMP_SCORE

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

Characteristic	CMR-Screened Participants With Current EDIC Data (n=1240) [Manuscript]	CMR-Screened Participants With Current EDIC Data (n=1240) [DSIC]	CMR-Screened Participants With Current EDIC Data (n=0) [Diff]	Participants With CMR (n=1017) [Manuscript]	Participants With CMR (n=1017) [DSIC]	Participants With CMR (n=0) [Diff]
Age, y	49±7	49 ± 7	0	49±7	49 ± 7	0
Duration of diabetes mellitus, y	27.6±4.9	27.6 ± 4.9	0	27.6±4.9	27.6 ± 4.9	0
Smoking, %	12.0	12.0	0	11.5	11.5	0
BMI, kg/m ²	28.4±5.1	28.4 ± 5.1	0	28.1±4.7	28.1 ± 4.7	0
Mean BMI, kg/m ²	26.7±3.7	26.7 ± 3.7	0	26.5±3.4	26.5 ± 3.4	0
SBP, mm Hg	120±14	120 ± 14	0	120±14	120 ± 14	0
Mean SBP, mm Hg	118±8	118 ± 8	0	118±8	118 ± 8	0
DBP, mm Hg	72±9	72 ± 9	0	72±9	72 ± 9	0
Mean DBP, mm Hg	74±5	74 ± 5	0	74±5	74 ± 5	0
Hypertension, %	51.3	51.3	0	50.3	50.3	0
Antihypertensive medication, %	42.3	42.3	0	41.3	41.3	0
Total cholesterol, mg/dL	173±36	173 ± 36	0	173±36	173 ± 36	0
Mean total cholesterol, mg/dL	182±24	182 ± 24	0	182±24	182 ± 24	0
Recent HDL, mg/dL	60±18	60 ± 18	0	60±18	60 ± 18	0
Mean HDL, mg/dL	55±12	55 ± 12	0	55±13	55 ± 13	0
Recent LDL, mg/dL	97±30	97 ± 30	0	97±30	97 ± 30	0
Mean LDL, mg/dL	110±21	110 ± 21	0	110±20	110 ± 20	0
Recent triglycerides, mg/dL	81±56	81 ± 56	0	80±56	80 ± 56	0
Mean triglycerides, mg/dL	85±41	85 ± 41	0	83±40	83 ± 40	0
Hypercholesterolemia, %	63.5	63.5	0	63.7	63.7	0
Lipid-lowering medication, %	56.7	56.7	0	57.3	57.3	0
Recent HbA1c, %	7.9±1.2	7.9 ± 1.2	0	7.9±1.2	7.9 ± 1.2	0
Mean HbA1c, %	8.0±1.0	8.0 ± 1.0	0	8.0±1.0	8.0 ± 1.0	0
Framingham Risk Score	0.06±0.05	0.06 ± 0.05	0	0.06±0.04	0.06 ± 0.04	0
Cardiovascular disease : Clinical or silent MI	52 (4.2)	-	-	37 (3.6)	-	-
Cardiovascular disease : Adjudicated clinical MI events	24 (1.9)	-	-	14 (1.4)	-	-
Cardiovascular disease : Silent MI	30 (2.4)	-	-	23 (2.3)	-	-
Cardiovascular disease : CAC score >0 (years 7–9)	343 (30.6)	343 (30.6)	0	282 (30.3)	282 (30.3)	0
Cardiovascular disease: CAC score >200 (years 7–9)	88 (7.9)	88 (7.9)	0	67 (7.2)	67 (7.2)	0
Cardiovascular disease: Common IMT (year 12)	0.69±0.15	0.69 ± 0.15	0	0.68±0.14	0.68 ± 0.14	0
Retinopathy: PDR or worse	252 (20.3)	252 (20.3)	0	206 (20.3)	206 (20.3)	0

Characteristic	CMR-Screened Participants With Current EDIC Data (n=1240) [Manuscript]	CMR-Screened Participants With Current EDIC Data (n=1240) [DSIC]	CMR-Screened Participants With Current EDIC Data (n=0) [Diff]	Participants With CMR (n=1017) [Manuscript]	Participants With CMR (n=1017) [DSIC]	Participants With CMR (n=0) [Diff]
Nephropathy: Macroalbuminuria/ESRD	124 (10.0)	124 (10.0)	0	98 (9.6)	98 (9.6)	0
Nephropathy: Sustained microalbuminuria/ESRD#	338 (27.3)	338 (27.3)	0	269 (26.5)	269 (26.5)	0
Neuropathy: Autonomic neuropathy	377 (32.2)	377 (32.2)	0	310 (31.7)	310 (31.7)	0
Neuropathy: Peripheral neuropathy	342 (30.0)	342 (30.0)	0	282 (29.4)	282 (29.4)	0
All complications: Participants with 0 complications	694 (56.0)	694 (56.0)	0	578 (56.8)	578 (56.8)	0
All complications: Participants with 1 complication	351 (28.3)	351 (28.3)	0	277 (27.2)	277 (27.2)	0
All complications: Participants with 2 complications	138 (11.1)	138 (11.1)	0	116 (11.4)	116 (11.4)	0
All complications: Participants with >=3 complications	57 (4.6)	57 (4.6)	0	46 (4.5)	46 (4.5)	0

Characteristic	Participants With Gadolinium CMR (n=741) [Manuscript]	Participants With Gadolinium CMR (n=741) [DSIC]	Participants With Gadolinium CMR (n=741) [Diff]
Age, y	49±7	49 ± 7	0
Duration of diabetes mellitus, y	27.5±4.9	27.5 ± 4.9	0
Smoking, %	11.6	11.6	0
BMI, kg/m ²	28.2±4.7	28.2 ± 4.7	0
Mean BMI, kg/m ²	26.5±3.4	26.5 ± 3.4	0
SBP, mm Hg	120±14	120 ± 14	0
Mean SBP, mm Hg	118±8	118 ± 8	0
DBP, mm Hg	73±9	73 ± 9	0
Mean DBP, mm Hg	74±5	74 ± 5	0
Hypertension, %	48.0	48.0	0
Antihypertensive medication, %	39.1	39.1	0
Total cholesterol, mg/dL	171±36	171 ± 36	0
Mean total cholesterol, mg/dL	180±23	180 ± 23	0
Recent HDL, mg/dL	59±17	59 ± 17	0
Mean HDL, mg/dL	54±12	54 ± 12	0
Recent LDL, mg/dL	97±30	97 ± 30	0
Mean LDL, mg/dL	110±21	110 ± 21	0
Recent triglycerides, mg/dL	77±52	77 ± 52	0
Mean triglycerides, mg/dL	81±38	81 ± 38	0
Hypercholesterolemia, %	64.8	64.8	0
Lipid-lowering medication, %	58.2	58.2	0
Recent HbA1c, %	7.9±1.2	7.9 ± 1.2	0
Mean HbA1c, %	7.9±0.9	7.9 ± 0.9	0
Framingham Risk Score	0.06±0.04	0.06 ± 0.04	0
Cardiovascular disease : Clinical or silent MI	28 (3.8)	-	-
Cardiovascular disease : Adjudicated clinical MI events	10 (1.4)	-	-
Cardiovascular disease : Silent MI	18 (2.4)	-	-
Cardiovascular disease : CAC score >0 (years 7–9)	202 (29.5)	202 (29.5)	0
Cardiovascular disease: CAC score >200 (years 7–9)	43 (6.3)	43 (6.3)	0
Cardiovascular disease: Common IMT (year 12)	0.68±0.13	0.68 ± 0.13	0
Retinopathy: PDR or worse	123 (16.6)	123 (16.6)	0
Nephropathy: Macroalbuminuria/ESRD	37 (5.0)	37 (5.0)	0
Nephropathy: Sustained microalbuminuria/ESRD	157 (21.2)	157 (21.2)	0

Characteristic	Participants With Gadolinium CMR (n=741) [Manuscript]	Participants With Gadolinium CMR (n=741) [DSIC]	Participants With Gadolinium CMR (n=741) [Diff]
Neuropathy: Autonomic neuropathy	207 (28.9)	207 (28.9)	0
Neuropathy: Peripheral neuropathy	178 (25.3)	178 (25.3)	0
All complications: Participants with 0 complications	451 (60.9)	451 (60.9)	0
All complications: Participants with 1 complication	204 (27.5)	204 (27.5)	0
All complications: Participants with 2 complications	69 (9.3)	69 (9.3)	0
All complications: Participants with >=3 complications	17 (2.3)	17 (2.3)	0

Attachment A: SAS Code

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

%global caser;

*** Frequency Macro: % only ***;
%macro freqdatal(order=, invar=, level=);

data data0 data1;
  set _null_;

  proc freq data=table1 noprint;
    tables &invar*&caser/out=data0 outpct;
    format _all_;
  run;

data data1;
  set data0;
  length LEVEL $100;
  LEVEL=strip(&invar);

  data data1(keep=LEVEL &caser name CHARALL ORDERER);
    set data1;
    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 data1;

data accumfreq1;
  set accumfreq1 data1;

%mend freqdatal;

*** Frequency Macro, N and % ***;
%macro freqdata2(order=, invar=, level=);

data data0 data1;
  set _null_;

  proc freq data=table1 noprint;
    tables &invar*&caser/out=data0 outpct;
    format _all_;
  run;

data data1;
  set data0;
  length LEVEL $100;
  LEVEL=strip(&invar);
```

```

data datal(keep=LEVEL &caser name CHARALL ORDERER);
  set datal;
  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 datal;

data accumfreq1;
  set accumfreq1 datal;

%mend freqdata2;

%macro meandatal(order=, invar=, roundvar=, digit=);
proc means data=table1 mean stddev noprint;
  var &invar;
  class &caser;
  output out=dat1 mean=mean stddev=stddev;
run;

data datal(drop=_TYPE_ _FREQ_ mean stddev);
  set datal;
  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 datal;

%mend meandatal;

%macro mediandatal(order=, invar=, roundvar=, digit=);
proc means data=table1 median p25 p75 min max noprint;
  var &invar;
  class &caser;
  output out=dat1 median=median p25=p25 p75=p75 min=min max=max;
run;

data datal(drop=_TYPE_ _FREQ_ median p25 p75 min max);
  set datal;
  length name CHARALL $100;
  name=upcase("&invar");
  median=round(median,&roundvar);
  min=round(min,&roundvar);
  max=round(max,&roundvar);
  ORDERER=&order;
  CHARALL=compress(put(median,8.&digit));
  output;
  ORDERER=ORDERER+.01;
  CHARALL=compress(put(min,8.&digit)||"-"||put(max,8.&digit));
  output;

```

```

data accummedian1;
  set accummedian1 data1;

%mend mediandatal;

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

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

data accummedian1;
  set accummedian1 data1;

%mend rangedatal;
%macro inertdatal(order=);

data inert1;
  length orderer &caser 8.;
  orderer=&order.;
  &caser=-1;
  output;
  orderer=&order.;
  &caser=0;
  output;
  orderer=&order.;
  &caser=1;
  output;

data accuminert1;
  set accuminert1 inert1;

%mend inertdatal;

%macro datachunk();

%meandatal(order=1 , invar=AGE , roundvar=1, digit=0);
%meandatal(order=2 , invar=IDDMYR , roundvar=.1, digit=1);
%freqdatal(order=3 , invar=SMOKE , level=("1"));
%meandatal(order=4 , invar=BMI , roundvar=.1, digit=1);
%meandatal(order=5 , invar=DTED_BMI , roundvar=.1, digit=1);
%meandatal(order=6 , invar=SBP , roundvar=1, digit=0);
%meandatal(order=7 , invar=DTED_SBP , roundvar=1, digit=0);
%meandatal(order=8 , invar=DBP , roundvar=1, digit=0);

```

```

%meandatal(order=9 , invar=DTED_DBP , roundvar=1, digit=0);
%freqdata1(order=10, invar=HT , level=("1"));
%freqdata1(order=11, invar=HYPERTEN , level=("1"));
%meandatal(order=12, invar=TCHOL , roundvar=1, digit=0);
%meandatal(order=13, invar=DTED_CHL , roundvar=1, digit=0);
%meandatal(order=14, invar=HDL , roundvar=1, digit=0);
%meandatal(order=15, invar=DTED_HDL , roundvar=1, digit=0);
%meandatal(order=16, invar=LDL , roundvar=1, digit=0);
%meandatal(order=17, invar=DTED_LDL , roundvar=1, digit=0);
%meandatal(order=18, invar=TRIG , roundvar=1, digit=0);
%meandatal(order=19, invar=DTED_TRG , roundvar=1, digit=0);
%freqdata1(order=20, invar=HLIP , level=("1"));
%freqdata1(order=21, invar=LIPID , level=("1"));
%meandatal(order=22, invar=HBA1C , roundvar=.1, digit=1);
%meandatal(order=23, invar=WTHMBA , roundvar=.1, digit=1);
%meandatal(order=24, invar=FSCORE , roundvar=.01, digit=2);
%inertdata1(order=24.1);
%inertdata1(order=24.2);
%inertdata1(order=24.3);
%freqdata2(order=25, invar=CTGT0 , level=("1"));
%freqdata2(order=26, invar=CTGT200 , level=("1"));
%meandatal(order=27, invar=COMMNW12 , roundvar=.01, digit=2);
%freqdata2(order=28, invar=ANYPDR , level=("1"));
%freqdata2(order=28.1, invar=ANYAE300 , level=("1"));
%freqdata2(order=28.2, invar=ANYSAE30 , level=("1"));
%freqdata2(order=28.3, invar=AFTEDIC13 , level=("1"));
%freqdata2(order=28.4, invar=CCNEDIC , level=("1"));
%freqdata2(order=29, invar=COMP_SCORE_CAT , level=("0"));
%freqdata2(order=30, invar=COMP_SCORE_CAT , level=("1"));
%freqdata2(order=31, invar=COMP_SCORE_CAT , level=("2"));
%freqdata2(order=32, invar=COMP_SCORE_CAT , level=("3"));

%mend datachunk;

*** Input ***;

libname edic "/prj/niddk/ims_analysis/DCCT_EDIC/private_orig_data/EDIC_06_02_2015/";

data nih_cmri_risk1240; set edic.nih_cmri_risk1240;

proc contents data=nih_cmri_risk1240;

data nih_cmri_risk1240;
  set nih_cmri_risk1240;
  if COMP_SCORE in(0 1 2 3) then COMP_SCORE_CAT=COMP_SCORE;
  else if COMP_SCORE=4 then COMP_SCORE_CAT=3;
  else abort return;

proc freq data=nih_cmri_risk1240;
  tables COMP_SCORE*COMP_SCORE_CAT CCNEDIC/missing list;
  title3 'Construction checks';

*** Column processing: CMR;

%let caser=MRICOMP;

```

```

data accumfreq1 accummean1 accummedian1 accuminert1;
  set _null_;

data table1;
  set nih_cmri_risk1240;

%datachunk();

data accumtab1;
  set accumfreq1 accummean1 accummedian1 accuminert1;
  if &caser=. then delete;

*** Total processing ***;

proc freq data=nih_cmri_risk1240;
  tables &caser/missing list;
  title3 'Case Counts';

data accumfreq1 accummean1 accummedian1 accuminert1;
  set _null_;

data table1;
  set nih_cmri_risk1240;
  &caser=-1;

%datachunk();

data accumtab2;
  set accumfreq1 accummean1 accummedian1 accuminert1;
  if &caser=. then delete;
run;

*** Column processing: GAD CMR;

%let caser=MRI_SCAR;

data accumfreq1 accummean1 accummedian1 accuminert1;
  set _null_;

data table1;
  set nih_cmri_risk1240;

%datachunk();

data accumtab3;
  set accumfreq1 accummean1 accummedian1 accuminert1;
  if &caser=. then delete;

*** Display processing ***;

proc sort data=accumtab1;
  by orderer;

proc sort data=accumtab2;
  by orderer;

```

```
proc sort data=accumtab3;
  by orderer;

proc print data=accumtab2 noobs;
  var name charall orderer;
  title3 'Table 1 Total stats (list)';
  where MRICOMP=-1;

proc print data=accumtab1 noobs;
  var name charall orderer;
  by MRICOMP;
  pageby MRICOMP;
  title3 'Table 1 stats';
  where MRICOMP=(1);

proc print data=accumtab3 noobs;
  var name charall orderer;
  by MRI_SCAR;
  pageby MRI_SCAR;
  title3 'Table 1 stats';
  where MRI_SCAR=(1);
```