Dataset Integrity Check for the

CHEIRO (DCCT/EDIC) Data Files

**Prepared by Michael Spriggs**

**IMS Inc.**

3901 Calverton Blvd, Suite 200 Calverton MD 20705

**April 9, 2015**

**Table of 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

Attachment A: SAS Code..............................................................................................................................6

**Table A**: Variables used to replicate Characteristics of subjects with and without cheiroarthropathy .……4

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

**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 DDCT 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 files that were provided by the Data Coordinating Center (DCC) for this replication are located in the “EDIC\EDIC\_Analysis\_Datasets\SAS\_DATA\EDIC\_CHEIRO” folder in the data package.

**4 Statistical Methods**

Analyses were performed to duplicate results for the data published by The DCCT/EDIC Research Group in Diabetes Care, July 2014. To verify the integrity of the datasets, a table from the paper was checked (Tables B).

**5 Results**

Table 1 in the publication [1], Characteristics of subjects with and without cheiroarthropathy. 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 1. The results of the replication are very similar to published results.

**6 Conclusions**

The NIDDK repository are confident that the CHEIRO data files to be distributed are a copy of the manuscript data.

**7 References**

[1] Musculoskeletal Complications in Type 1 Diabetes Larkin ME, Barnie A, Braffett BH, Cleary PA, Diminick L, Harth J, Gatcomb P, Golden E, Lipps J, Lorenzi G, Mahony C, Nathan DM; Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. Diabetes Care. 2014 Jul;37(7):1863-9. doi: 10.2337/dc13-2361. Epub 2014 Apr 10.

**Table A:** Variables used to replicate Characteristics of subjects with and without cheiroarthropathy

|  |  |
| --- | --- |
| Characteristic | Variable(s) |
| Age (years) | CHIROAGE |
| Female sex | SEX |
| Menopause\* | MENO |
| Married or remarried | MARRY\_18 |
| Duration of diabetes (years) | CHRDURAT |
| BMI (kg/m2) | C\_BMI |
| Obese (BMI ≥30 kg/m2) | C\_BMI |
| DCCT INT therapy | GROUP |
| Primary cohort | RETBASE |
| Current smoker | SMOKYR18 |
|  HbA1c [% (mmol/mol)] Time-weighted DCCT/EDIC | DTED\_HBA |
|  HbA1c [% (mmol/mol)] During DCCT | DCCT\_HBA\_MEAN |
|  HbA1c [% (mmol/mol)] During EDIC | EDIC\_HBA |
| Skin intrinsic fluorescence (AU) | SIF1 |
| Neuropathy | CCNEDIC |
| Nephropathy | SAER30 |
| Retinopathy | PDR |

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

|  |  |  |  |
| --- | --- | --- | --- |
| Characteristic | Total (Manuscript N = 1217) | Total (DSIC N=1217) | Total (DIFF N=0) |
| Age (years) | 52.2 ± 6.9 | 52.2 ± 6.9 | 0,0 |
| Female sex | 584 (48) | 584 (48) | 0,0 |
| Menopause\* | 300 (55) | 300 (55) | 0,0 |
| Married or remarried | 880 (73) | 880 (73) | 0,0 |
| Duration of diabetes (years) | 31.1 ± 4.9 | 31.1 ± 4.9 | 0,0 |
| BMI (kg/m2) | 28.8 ± 5.5 | 28.8 ± 5.5 | 0,0 |
| Obese (BMI ≥30 kg/m2) | 411 (35) | 414 (35) | -3,0 |
| DCCT INT therapy | 616 (51) | 616 (51) | 0,0 |
| Primary cohort | 607 (50) | 607 (50) | 0,0 |
| Current smoker | 136 (11) | 136 (11) | 0,0 |
|  HbA1c [% (mmol/mol)] Time-weighted DCCT/EDIC | 8.0 ± 1.0 (63.8 ± 10.5) | 8.0 ± 1.0 (63.8 ± 10.5) | 0,0,0,0 |
|  HbA1c [% (mmol/mol)] During DCCT | 8.1 ± 1.4 (64.8 ± 15.3) | 8.1 ± 1.4 (64.8 ± 15.3) | 0,0,0,0 |
|  HbA1c [% (mmol/mol)] During EDIC | 8.0 ± 1.0 (63.4 ± 11.1) | 8.0 ± 1.0 (63.4 ± 11.1) | 0,0,0,0 |
| Skin intrinsic fluorescence (AU) | 22.6 ± 4.7 | 22.6 ± 4.7 | 0,0 |
| Neuropathy | 327 (29) | 327 (29) | 0,0 |
| Nephropathy | 168 (14) | 168 (14) | 0,0 |
| Retinopathy | 255 (21) | 255 (21) | 0,0 |

|  |  |  |  |
| --- | --- | --- | --- |
| Characteristic | Cheiroarthropathy present (Manuscript N = 807) | Cheiroarthropathy present (DSIC N=807) | Cheiroarthropathy present (DIFF N=0) |
| Age (years) | 52.7 ± 6.6 | 52.7 ± 6.6 | 0,0 |
| Female sex | 430 (53) | 430 (53) | 0,0 |
| Menopause\* | 232 (57) | 232 (57) | 0,0 |
| Married or remarried | 584 (73) | 584 (73) | 0,0 |
| Duration of diabetes (years) | 31.9 ± 5.0 | 31.9 ± 5.0 | 0,0 |
| BMI (kg/m2) | 28.7 ± 5.5 | 28.7 ± 5.5 | 0,0 |
| Obese (BMI ≥30 kg/m2) | 281 (36) | 276 (35) | 5,1 |
| DCCT INT therapy | 397 (49) | 397 (49) | 0,0 |
| Primary cohort | 351 (43) | 351 (43) | 0,0 |
| Current smoker | 95 (12) | 95 (12) | 0,0 |
|  HbA1c [% (mmol/mol)] Time-weighted DCCT/EDIC | 8.1 ± 1.0 (64.5 ± 10.5) | 8.1 ± 1.0 (64.5 ± 10.5) | 0,0,0,0 |
|  HbA1c [% (mmol/mol)] During DCCT | 8.1 ± 1.4 (65.3 ± 15.4) | 8.1 ± 1.4 (65.3 ± 15.4) | 0,0,0,0 |
|  HbA1c [% (mmol/mol)] During EDIC | 8.0 ± 1.0 (64.2 ± 11.1) | 8.0 ± 1.0 (64.2 ± 11.1) | 0,0,0,0 |
| Skin intrinsic fluorescence (AU)‡ | 22.9 ± 4.7 | 22.9 ± 4.7 | 0,0 |
| Neuropathy | 250 (34) | 250 (34) | 0,0 |
| Nephropathy | 112 (14) | 112 (14) | 0,0 |
| Retinopathy | 201 (25) | 201 (25) | 0,0 |

|  |  |  |  |
| --- | --- | --- | --- |
| Characteristic | Cheiroarthropathy absent (Manuscript N = 410) | Cheiroarthropathy absent (DSIC N=410) | Cheiroarthropathy absent (DIFF N=0) |
| Age (years) | 51.3 ± 7.3 | 51.3 ± 7.3 | 0,0 |
| Female sex | 154 (38) | 154 (38) | 0,0 |
| Menopause\* | 68 (48) | 68 (48) | 0,0 |
| Married or remarried | 296 (73) | 296 (73) | 0,0 |
| Duration of diabetes (years) | 29.5 ± 4.3 | 29.5 ± 4.3 | 0,0 |
| BMI (kg/m2) | 28.8 ± 5.4 | 28.8 ± 5.4 | 0,0 |
| Obese (BMI ≥30 kg/m2) | 130 (33) | 138 (35) | -8,-2 |
| DCCT INT therapy | 219 (53) | 219 (53) | 0,0 |
| Primary cohort | 256 (62) | 256 (62) | 0,0 |
| Current smoker | 41 (10) | 41 (10) | 0,0 |
|  HbA1c [% (mmol/mol)] Time-weighted DCCT/EDIC | 7.9 ± 0.9 (62.3 ± 10.3) | 7.9 ± 0.9 (62.3 ± 10.3) | 0,0,0,0 |
|  HbA1c [% (mmol/mol)] During DCCT | 8.0 ± 1.4 (63.9 ± 15.0) | 8.0 ± 1.4 (63.9 ± 15.0) | 0,0,0,0 |
|  HbA1c [% (mmol/mol)] During EDIC | 7.8 ± 1.0 (61.8 ± 11.0) | 7.8 ± 1.0 (61.8 ± 11.0) | 0,0,0,0 |
| Skin intrinsic fluorescence (AU) | 22.1 ± 4.7 | 22.1 ± 4.7 | 0,0 |
| Neuropathy | 77 (21) | 77 (21) | 0,0 |
| Nephropathy | 56 (14) | 56 (14) | 0,0 |
| Retinopathy | 54 (13) | 54 (13) | 0,0 |

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

title2 " ";

%global caser;

%let caser=CHEIRO;

\*\*\* Macro \*\*\*;

%macro freqdata1(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.))||")";

 ORDERER=&order;

 if level in &level then output data1;

data accumfreq1;

 set accumfreq1 data1;

%mend freqdata1;

%macro meandata1(order=, invar=, roundvar=, digit=);

proc means data=table1 mean stddev noprint;

 var &invar;

 class &caser;

 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 &caser;

 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);

 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 mediandata1;

%macro rangedata1(order=, invar=, roundvar=, digit=);

proc means data=table1 median p25 p75 min max noprint;

 var &invar;

 class &caser;

 output out=data1 min=min max=max;

 run;

data data1(drop=\_TYPE\_ \_FREQ\_ min max);

 set data1;

 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 rangedata1;

%macro inertdata1(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 inertdata1;

%macro datachunk();

%meandata1(order=1, invar=CHIROAGE, roundvar=.1, digit=1);

%freqdata1(order=2, invar=SEX, level=("F"));

%freqdata1(order=3, invar=MENO, level=("1"));

%freqdata1(order=4, invar=MARRY\_18, level=("1"));

%meandata1(order=5, invar=CHRDURAT, roundvar=.1, digit=1);

%meandata1(order=6, invar=C\_BMI, roundvar=.1, digit=1);

%freqdata1(order=7, invar=OBESE, level=("Y"));

%freqdata1(order=8, invar=GROUP, level=("EXPERIMENTAL"));

%freqdata1(order=9, invar=RETBASE, level=("PRIM"));

%freqdata1(order=10, invar=SMOKYR18, level=("1"));

%meandata1(order=11, invar=DTED\_HBA, roundvar=.1, digit=1);

%meandata1(order=11.5, invar=DTED\_HBA\_CONV, roundvar=.1, digit=1);

%meandata1(order=12, invar=DCCT\_HBA\_MEAN, roundvar=.1, digit=1);

%meandata1(order=12.5, invar=DCCT\_HBA\_MEAN\_CONV, roundvar=.1, digit=1);

%meandata1(order=13, invar=EDIC\_HBA, roundvar=.1, digit=1);

%meandata1(order=13.5, invar=EDIC\_HBA\_CONV, roundvar=.1, digit=1);

%meandata1(order=14, invar=SIF1, roundvar=.1, digit=1);

%freqdata1(order=15, invar=CCNEDIC, level=("1"));

%freqdata1(order=16, invar=SAER30, level=("1"));

%freqdata1(order=17, invar=PDR, level=("1"));

%mend datachunk;

\*\*\* Input \*\*\*\*;

libname cheiro "/prj/niddk/ims\_analysis/DCCT\_EDIC/private\_created\_data/edic\_new\_data/CHEIRO/";

data shoulder; set cheiro.shoulder;

proc cimport data=cheiro2 infile='/prj/niddk/ims\_analysis/DCCT\_EDIC/private\_orig\_data/CHEIRO2.XPT';

data cheiro; set cheiro2;

\*\*\* Processing \*\*\*\*;

proc sort data=cheiro;

 by PATIENT;

proc sort data=shoulder;

 by PATIENT;

data table1;

 merge cheiro shoulder;

 by PATIENT;

 length OBESE $1;

 if C\_BMI>30 then OBESE="Y";

 else if C\_BMI>=0 then OBESE="N";

 else OBESE=" ";

 DTED\_HBA\_CONV=(DTED\_HBA-2.152)\*10.931;

 DCCT\_HBA\_MEAN\_CONV=(DCCT\_HBA\_MEAN-2.152)\*10.931;

 EDIC\_HBA\_CONV=(EDIC\_HBA-2.152)\*10.931;

\*\*\* Column processing;

data accumfreq1 accummean1 accummedian1 accuminert1;

 set \_null\_;

%datachunk();

data accumtab1;

 set accumfreq1 accummean1 accummedian1 accuminert1;

 if &caser=. then delete;

\*\*\* Total processing \*\*\*;

proc freq data=table1;

 tables &caser/missing list;

 title3 'Case Counts';

data table1;

 set table1;

 &caser=-1;

data accumfreq1 accummean1 accummedian1 accuminert1;

 set \_null\_;

%datachunk();

data accumtab2;

 set accumfreq1 accummean1 accummedian1 accuminert1;

 if &caser=. then delete;

\*\*\* Display processing \*\*\*;

proc sort data=accumtab2;

 by descending &caser orderer;

proc print data=accumtab2 noobs;

 by descending &caser;

 pageby &caser;

 title3 'Table 1 Total stats (list)';

 where &caser=-1;

proc sort data=accumtab1;

 by descending &caser orderer;

proc print data=accumtab1 noobs;

 by descending &caser;

 pageby &caser;

 title3 'Table 1 stats (list)';

 where &caser in (0 1);