

Dataset Integrity Check for EDIC Optimret analysis dataset

Prepared by Michael Spriggs

IMS Inc.

3901 Calverton Blvd, Suite 200 Calverton, MD 20705

April 9, 2019

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 S1.Observed number of transitions of the retinopathy status from one visit to the next visit	4
Table B: Comparison of values computed in integrity check to reference article Table S1 values	4
Attachment A: SAS Code	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 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

All the SAS data files, as provided by the Data Coordinating Center (DCC), are located in the OPTIMRET folder in the data package. For this replication, variables were taken from the optimret datasets.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by the DCCT/EDIC research group in *New England Journal of Medicine* in 2017.

To verify the integrity of the dataset, Supplementary Table 1 descriptive statistics were computed.

5 Results

For Supplementary Table 1 in the publication [1], Table S1.Observed number of transitions of the retinopathy status from one visit to the next visit, 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 identical to the published results.

6 Conclusions

The NIDDK repository is confident that the Optimret data files to be distributed are a true copy of the study data.

7 References

[1] Frequency of Evidence-Based Screening for Retinopathy in Type 1 Diabetes, DCCT/EDIC Research Group et al., N Engl J Med 2017; 376:1507-1516 DOI: 10.1056/NEJMoa1612836

Table A: Variables used to replicate Table S1. Observed number of transitions of the retinopathy status from one visit to the next visit

Table Variable	dataset.variable
Current retinopathy state(1-5)	Optimret.RET_STATE
Prior retinopathy state(1-4)	Constructed using Optimret.RET_STATE Optimret.time and Optimret.MASK_PAT, see appendix A.

Table B: Comparison of values computed in integrity check to reference article Table S1 values

Retinopathy Status between visits	Manuscript (n= 22520)	DSIC (n= 22520)	Diff. (n=0)
1 to 1	4898	4898	0
1 to 2	1823	1823	0
1 to 3	9	9	0
1 to 4	1	1	0
1 to 5	3	3	0
2 to 1	1216	1216	0
2 to 2	11510	11510	0
2 to 3	854	854	0
2 to 4	17	17	0
2 to 5	227	227	0
3 to 1	3	3	0
3 to 2	505	505	0
3 to 3	1076	1076	0
3 to 4	56	56	0
3 to 5	238	238	0
4 to 1	0	0	0
4 to 2	4	4	0
4 to 3	33	33	0
4 to 4	11	11	0
4 to 5	36	36	0

Attachment A: SAS Code

```
options nocenter validvarname=upcase;

title '/prj/niddk/ims_analysis/DCCT_EDIC/prog_initial_analysis/table1_optimret_22mar2019.sas';
run;

*****;
* INPUT ;
*****;
libname pclib '/prj/niddk/ims_analysis/DCCT_EDIC/private_orig_data/OPTIMRET/';

data analysis;
    set pclib.optimret;

proc sort data=analysis;
    by mask_pat time;

data analysis(keep=mask_pat RET_STATE_PRIOR RET_STATE);
    set analysis;
    by mask_pat;
    retain RET_STATE_PRIOR;
    if first.mask_pat then RET_STATE_PRIOR=.F;
    output;
    RET_STATE_PRIOR=RET_STATE;

proc freq data=analysis;
    tables RET_STATE_PRIOR*RET_STATE/nocum norow nocol nopercnt;
    title3 'RET_STATE checks (First visits excluded)';
```