

Dataset Integrity Check for the Diabetes Autoimmunity Study in the Young (DAISY)

Prepared by NIDDK-CR
January 4, 2023

Contents

1 Standard Disclaimer	2
2 Study Background	2
3 Archived Datasets	2
4 Statistical Methods	2
5 Results	2
6 Conclusions	3
7 References	3
Table A: Variables used to replicate Supplemental Table 2 – Characteristics of T1DI study cohort.....	4
Table B: Comparison of values computed in integrity check to reference article Supplemental Table 2....	5
Attachment A: SAS Code	6

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

Children determined to be at an increased risk for development of type 1 diabetes (T1D), either due to known genetic markers or due to family history of T1D, are followed from birth and monitored on an annual basis for development of T1D-related autoimmunity. At annual clinic visits, participants are tested for development of these autoantibodies and data are collected related to environmental exposures. In addition to blood collection for determination of autoantibodies, serum and plasma are frozen and stored for future analyses, along with other biological samples such as throat swabs, rectal swabs, saliva, and urine. If T1D-related autoantibodies are detected, the participant is asked to increase the frequency of clinic visits to 2-4 visits per year in order to monitor for the onset of disease more closely. The endpoint of the study is diagnosis of T1D per the American Diabetes Association (ADA) criteria.

3 Archived Datasets

All data files, as provided by the Data Coordinating Center (DCC), are located in the DAISY folder in the data package. For this replication, variables were taken from the “subjectleveldata.sas7bdat” dataset.

4 Statistical Methods

Analyses were performed to replicate results for the data in the publication by Anand et al. [1]. To verify the integrity of the data, only descriptive statistics were computed.

5 Results

For Supplemental Table 2 in the publication [1], Characteristics of T1DI study cohort, 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 Supplemental Table 2. The results of the replication are within expected variation to the published results.

6 Conclusions

The NIDDK Central Repository is confident that the DAISY data files to be distributed are a true copy of the study data.

7 References

[1] Anand V, Li Y, Liu B, Ghalwash M, Koski E, Ng K, Dunne JL, Jönsson J, Winkler C, Knip M, Toppari J, Ilonen J, Killian MB, Frohnert BI, Lundgren M, Ziegler AG, Hagopian W, Veijola R, Rewers M. Islet Autoimmunity and HLA Markers of Presymptomatic and Clinical Type 1 Diabetes: Joint Analyses of Prospective Cohort Studies in Finland, Germany, Sweden, and the U.S. *Diabetes Care*, 44(10), 2269-2276, June 2021. doi: <https://doi.org/10.2337/dc20-1836>

Table A: Variables used to replicate Supplemental Table 2 – Characteristics of T1DI study cohort

Table Variable	dataset.variable
Sex	subjectleveldata.sex
Race	subjectleveldata.race
Ethnicity	subjectleveldata.spanish_1
Breastfed ever	subjectleveldata.bfever
Family history of T1D	subjectleveldata.t1dfamhis
First-degree relative relationship	subjectleveldata.t1d_reln2

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

Characteristic	Publication: DAISY Cohort (n=2544)	DSIC: DAISY Cohort (n=2544)	Diff. (n=0)
Sex (Female)	1220 (48%)	1220 (48%)	0 (0)
Race			
White	2296 (90%)	2296 (90%)	0 (0)
Black	62 (2%)	62 (2%)	0 (0)
Other	27 (1%)	27 (1%)	0 (0)
Biracial	139 (6%)	139 (6%)	0 (0)
Hispanic	544 (21%)	545 (21%)	1 (0)
Breast fed ever (Yes)	2077 (82%)	2077 (82%)	0 (0)
Family history of T1D (Yes)	1126 (44%)	1126 (44%)	0 (0)
First-degree relative relationship			
Father	414 (16%)	414 (16%)	0 (0)
Mother	324 (13%)	324 (13%)	0 (0)
Sibling or Half-sibling	362 (14%)	362 (14%)	0 (0)
Multiple	25 (1%)	25 (1%)	0 (0)

Attachment A: SAS Code

```
libname dsic "X:\NIDDK\niddk-dr_studies1\DAISY\private_orig_data\Ready for submission\Dataset";
```

```
/******  
/* DSIC for DAISY */  
/******
```

```
proc contents data=dsic.subjectleveldata;  
run;
```

```
proc contents data=dsic.visitleveldata;  
run;
```

```
*Sex;  
proc freq data=dsic.subjectleveldata;  
tables sex;  
run;
```

```
*Race;  
proc freq data=dsic.subjectleveldata;  
tables race/missing;  
run;
```

```
*Hispanic;  
proc freq data=dsic.subjectleveldata;  
tables Spanish_1/missing;  
run;
```

```
*Breastfed;  
proc freq data=dsic.subjectleveldata;  
tables bfever/missing;  
run;
```

```
*Family history;  
proc freq data=dsic.subjectleveldata;  
tables T1DFamHis/missing;  
run;
```

```
*relative;  
proc freq data=dsic.subjectleveldata;  
tables T1D_reln2/missing;  
run;
```