

Dataset Integrity Check for A Multicenter Clinical Trial of Allopurinol to Prevent GFR Loss in Type 1 Diabetes (PERL) Study

Prepared by NIDDK-CR
October 5, 2021

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 Table 1 – Characteristics of the Trial Participants at Baseline	4
Table B1: Comparison of values computed in integrity check to reference article Table 1 values (Placebo and Allopurinol groups)	5
Table B2: Comparison of values computed in integrity check to reference article Table 1 values (Total) ..	6
Attachment A: SAS Code	7

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 PERL study was an international clinical trial conducted at eight diabetes centers. The study enrolled patients with type 1 diabetes that were at increased risk of developing kidney disease. Participants were randomly assigned to take allopurinol or placebo for three years, during which they were followed through periodical visits. Kidney function was measured at the beginning and at the end of the treatment period to assess the rate of loss of kidney function over time.

3 Archived Datasets

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

4 Statistical Methods

Analyses were performed to replicate results for the data published by Doria et al. [1] for Serum Urate Lowering with Allopurinol and Kidney Function in Type 1 Diabetes. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For Table 1 in the publication [1], Characteristics of the Trial Participants at Baseline, Table A lists the variables that were used in the replication, and Tables B1-B2 compares the results calculated from the archived data files to the results published in Table 1. The results of the replication are within expected variation to the published results.

6 Conclusions

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

7 References

[1] Doria A, Galecki AT, Spino C, Pop-Busui R, Cherney DZ, Lingvay I, Parsa A, Rossing P, Sigal RJ, Afkarian M, Aronson R, Caramori ML, Crandall JP, de Boer IH, Elliott TG, Goldfine AB, Haw JS, Hirsch IB, Karger AB, Maahs DM, McGill JB, Molitch ME, Perkins BA, Polsky S, Pragnell M, Robiner WN, Rosas SE, Senior P, Tuttle KR, Umpierrez GE, Wallia A, Weinstock RS, Wu C, Mauer M. Serum Urate Lowering with Allopurinol and Kidney Function in Type 1 Diabetes. *The New England Journal of Medicine*, 382(26), 2493-2503, June 2020. doi: <https://doi.org/10.1056/NEJMoa1916624>
PMCID: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7375708/>

Table A: Variables used to replicate Table 1 – Characteristics of the Trial Participants at Baseline

Table Variable	dataset.variable
Age (years)	perl_baseline.dmga3age
Male sex – no. (%)	perl_baseline.gender
Race – no. (%)	perl_baseline.race
Diabetes duration (years)	perl_baseline.diabetes_dur
Body mass index	perl_baseline.bphra3bmi_v04
Glycated hemoglobin (%)	perl_baseline.glyhb_v01
Serum urate (mg/dL)	perl_baseline.uric_acid_v04
Blood pressure (mmHg)	perl_baseline.sbp_mean_v04 perl_baseline.dbp_mean_v04
Iohexol-based GFR (mL/min/1.73 m ²)	perl_baseline.igfr_v04
Serum creatinine-based GFR (mL/min/1.73 m ²)	perl_baseline.egfr_web_v4
Median urinary albumin excretion rate (µg/min)	perl_baseline.aer_base34_gm
Use of renin-angiotensin system inhibitor – no. (%)	perl_baseline.onrasb_base

Table B1: Comparison of values computed in integrity check to reference article Table 1 values (Placebo and Allopurinol groups)

Variable	Publication: Placebo (n=263)	DSIC: Placebo (n=263)	Diff. (n=0)	Publication: Allopurinol (n=267)	DSIC: Allopurinol (n=267)	Diff. (n=0)
Age – yr	51.8 ± 10.6	51.8 ± 10.6	0 ± 0	50.4 ± 11.2	50.4 ± 11.2	0 ± 0
Male sex – no. (%)	168 (63.9)	168 (63.9)	0 (0)	183 (68.5)	183 (68.5)	0 (0)
Race – no. (%)						
White	216 (82.1)	216 (82.1)	0 (0)	230 (86.1)	230 (86.1)	0 (0)
Black	30 (11.4)	30 (11.4)	0 (0)	28 (10.5)	28 (10.5)	0 (0)
Other	17 (6.5)	17 (6.5)	0 (0)	9 (3.4)	9 (3.4)	0 (0)
Diabetes duration – yr	35.5 ± 12.5	35.3 ± 12.5	0.2 ± 0	33.8 ± 12.2	33.8 ± 12.2	0 ± 0
Body mass index	29.5 ± 5.9	29.5 ± 5.9	0 ± 0	29.5 ± 6.1	29.5 ± 6.1	0 ± 0
Glycated hemoglobin – %	8.2 ± 1.3	8.2 ± 1.3	0 ± 0	8.2 ± 1.3	8.2 ± 1.3	0 ± 0
Serum urate – mg/dL	6.1 ± 1.5	6.1 ± 1.5	0 ± 0	6.1 ± 1.5	6.1 ± 1.5	0 ± 0
Blood pressure – mmHg						
Systolic	126 ± 13.6	126.3 ± 13.6	0 ± 0	125.6 ± 14.7	125.6 ± 14.7	0 ± 0
Diastolic	71.3 ± 10.0	71.3 ± 10.0	0 ± 0	71.2 ± 10.4	71.2 ± 10.4	0 ± 0
Iohexol-based GFR – ml/min/1.73 m ²	67.3 ± 16.7	67.3 ± 16.7	0 ± 0	68.7 ± 17.1	68.7 ± 17.1	0 ± 0
Serum creatinine-based estimated GFR – ml/min/1.73 m ²	74.0 ± 19.4	74.0 ± 19.4	0 ± 0	75.4 ± 18.7	75.4 ± 18.7	0 ± 0
Median urinary albumin excretion rate (IQR) – µg/min	43.0 (9.0-198.0)	43.0 (9.0-198.0)	0 (0-0)	41.1 (7.7-216.0)	41.1 (7.7-216.0)	0 (0-0)
Use of renin-angiotensin system inhibitor – no. (%)	230 (87.5)	230 (87.5)	0 (0)	247 (92.5)	247 (92.5)	0 (0)

Table B2: Comparison of values computed in integrity check to reference article Table 1 values (Total)

Variable	Publication: Total (n=530)	DSIC: Total (n=530)	Diff. (n=0)
Age – yr	51.1 ± 10.9	51.1 ± 10.9	0 ± 0
Male sex – no. (%)	351 (66.2)	351 (66.2)	0 (0)
Race – no. (%)			
White	446 (84.2)	446 (84.2)	0 (0)
Black	58 (10.9)	58 (10.9)	0 (0)
Other	26 (4.9)	26 (4.9)	0 (0)
Diabetes duration – yr	34.6 ± 12.3	34.6 ± 12.3	0 ± 0
Body mass index	29.5 ± 6.0	29.5 ± 6.0	0 ± 0
Glycated hemoglobin – %	8.2 ± 1.3	8.2 ± 1.3	0 ± 0
Serum urate – mg/dL	6.1 ± 1.5	6.1 ± 1.5	0 ± 0
Blood pressure – mmHg			
Systolic	126.0 ± 14.2	126.0 ± 14.2	0 ± 0
Diastolic	71.2 ± 10.2	71.2 ± 10.2	0 ± 0
Iohexol-based GFR – ml/min/1.73 m ²	68.0 ± 16.9	68.0 ± 16.9	0 ± 0
Serum creatinine-based estimated GFR – ml/min/1.73 m ²	74.7 ± 19.1	74.7 ± 19.1	0 ± 0
Median urinary albumin excretion rate (IQR) – µg/min	41.6 (8.5-207.5)	41.6 (8.5-207.5)	0 (0-0)
Use of renin-angiotensin system inhibitor – no. (%)	477 (90.0)	477 (90.0)	0 (0)

Attachment A: SAS Code

```
libname dsic "X:\NIDDK\niddk-dr_studies6\PERL\private_orig_data\PERL Data\PERL
Data\Analytical_Data";
```

```
/* ***** /
/* Calling Datasets and checking variables */
/* ***** /
```

```
data secondary; set dsic.secondary_outcome;
run;
```

```
data safe_sae; set dsic.safety_sae;
run;
```

```
data safe_ae; set dsic.safety_ae;
run;
```

```
data visit; set dsic.perl_visit;
run;
```

```
data dispo; set dsic.perl_disposition;
run;
```

```
data base; set dsic.perl_baseline;
run;
```

```
/* ***** /
/* Replicating Table 1 */
/* ***** /
```

```
*Albuminuria Status at baseline for stratification;
proc freq data=base;
tables trt_grp;
run;
```

```
/* ***** /
/* Placebo and Treatment Groups Table 1 columns */
/* ***** /
```

```
*Age;
Proc means data=base n mean std;
var dmga3age;
class trt_grp;
run;
```

```
*Sex;
```



```
proc freq data=base;  
tables gender*trt_grp;  
run;
```

```
*Race;  
data base2; set base;  
if race = "White" then race2 = "White";  
if race = "Black" then race2 = "Black";  
if race ^= "White" and race ^= "Black"  
then race2 = "Other";  
run;
```

```
proc freq data=base2;  
tables race2*trt_grp;  
run;
```

```
*Diabetes Duration;  
proc means data=base n mean std;  
var Diabetes_Dur;  
class trt_grp;  
run;
```

```
*BMI;  
proc means data=base mean std;  
var BPHRA3BMI_v04;  
class trt_grp;  
run;
```

```
*Glycated hemoglobin - %;  
proc means data=base mean std;  
var GLYHB_v01;  
class trt_grp;  
run;
```

```
*SUA;  
proc means data=base mean std;  
var uric_acid_v04;  
class trt_grp;  
run;
```

```
*Blood Pressure;  
proc means data=base mean std;  
var sbp_mean_v04 dbp_mean_v04;  
class trt_grp;  
run;
```

```
*iGFR;  
proc means data=base mean std;
```

```

var igfrm_v04;
class trt_grp;
run;

*Baseline eGFR;
proc means data=base mean std;
var egfr_web_V4;
class trt_grp;
run;

*UAER;
proc means data=base median p25 p75;
var aer_base34_gm;
class trt_grp;
run;

*RASI use;
data base3; set base;
if onrasb_base = "None (Contraindications)" or onrasb_base = "None (Normotensive)" or onrasb_base =
"None" then rasi = "none";
else rasi = "yes";
run;

proc freq data=base3;
tables rasi*trt_grp;
run;

/*****/
/* Replication of Total Cohort column of Table 1 */
/*****/

*Age;
Proc means data=base n mean std;
var dmga3age;
run;

*Sex;
proc freq data=base;
tables gender;
run;

*Race;
data base2; set base;
if race = "White" then race2 = "White";
if race = "Black" then race2 = "Black";
if race ^= "White" and race ^= "Black"
then race2 = "Other";
run;

```

```
proc freq data=base2;  
tables race2;  
run;
```

```
*Diabetes Duration;  
proc means data=base n mean std;  
var Diabetes_Dur;  
run;
```

```
*BMI;  
proc means data=base mean std;  
var BPHRA3BMI_v04;  
run;
```

```
*Glycated hemoglobin - %;  
proc means data=base mean std;  
var GLYHB_v01;  
run;
```

```
*SUA;  
proc means data=base mean std;  
var uric_acid_v04;  
run;
```

```
*Blood Pressure;  
proc means data=base mean std;  
var sbp_mean_v04 dbp_mean_v04;  
run;
```

```
*iGFR;  
proc means data=base mean std;  
var igfrm_v04;  
run;
```

```
*Baseline eGFR;  
proc means data=base mean std;  
var egfr_web_V4;  
run;
```

```
*UAER;  
proc means data=base median p25 p75;  
var aer_base34_gm;  
run;
```

```
*RASI use;  
data base3; set base;
```

```
if onrasb_base = "None (Contraindications)" or onrasb_base = "None (Normotensive)" or onrasb_base =  
"None" then rasi = "none";  
else rasi = "yes";  
run;
```

```
proc freq data=base3;  
tables rasi;  
run;
```