Dataset Integrity Check for the HALT Progression of Polycystic Kidney Disease (HALT PKD) Allele Data

Prepared by NIDDK-CR
June 6, 2022

Contents

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 HALT Progression of Polycystic Kidney Disease (HALT PKD) studies were two simultaneous multicenter clinical trials designed to test the efficacy of interruption of the renin-angiotensin-aldosterone system (RAAS) on the progression of cystic disease and the decline in renal function in autosomal dominant polycystic kidney disease (ADPKD). Specifically, the studies assessed the effects of ACE-I/ARB combination therapy as compared to ACE-I monotherapy in hypertensive ADPKD subjects.

3 Archived Datasets

All data files, as provided by the Data Coordinating Center (DCC) and ancillary researchers, are located in the HALT PKD folder in the data package. For this replication, variables were taken from the “halt_pkd_ids_genetic_data.sas7bdat” dataset.

4 Statistical Methods

Analyses were performed to replicate results for the data published by Heyer et al. [1] for Predicted Mutation Strength of Nontruncating PKD1 Mutations Aids Genotype-Phenotype Correlations in Autosomal Dominant Polycystic Kidney Disease. To verify the integrity of the dataset, only descriptive statistics were computed for the provided variables pertaining to the HALT PKD allele data (i.e., demographic, medical conditions/results, and information on families from both the HALT PKD and CRISP studies were excluded). Participant IDs of the participants included in the publication for the HALT PKD allele data were provided by the DCC for the purposes of this replication.

5 Results

For Table 1 in the publication [1], Clinical and genetic characteristics of the studied populations, 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 within expected variation to the published results for only the HALT PKD allele data (as CRISP allele data was not included in this replication).

6 Conclusions

The NIDDK Central Repository is confident that the HALT PKD allele data files to be distributed are a true copy of the study data as the results of the replication are within expected variation to the published results (and excluded CRISP allele data).

7 References

[1] Heyer CM, Sundsbak JL, Abebe KZ, Chapman AB, Torres VE, Grantham JJ, Bae KT, Schrier RW, Perrone RD, Braun WE, Steinman TI, Mrug M, Yu ASL, Brosnahan G, Hopp K, Irazabal MV, Bennett WM, Flessner MF, Moore CG, Landsittel D, Harris PC. Predicted Mutation Strength of Nontruncating PKD1 Mutations Aids Genotype-Phenotype Correlations in Autosomal Dominant Polycystic Kidney Disease. Journal of the American Society of Nephrology, 27(9), 2872-2884, September 2016. doi: https://doi.org/10.1681/ASN.2015050583

Table A: Variables used to replicate Table 1 – Clinical and genetic characteristics of the studied populations

Table Variable	**dataset.variable**
Patients with no mutation detected (NMD)	halt_pkd_ids_genetic_data.Final_Gene
Patients with PKD2	halt_pkd_ids_genetic_data.Final_Gene
Patients with PKD1	halt_pkd_ids_genetic_data.Final_Gene
Total patients with PKD1, mutation type	halt_pkd_ids_genetic_data.Final_Mutation_Type
Patients with PKD1, mutation strength group (MSG)	halt_pkd_ids_genetic_data.Final_Gene.Final_Mutation_Fuctional_Effect halt_pkd_ids_genetic_data.Final_Gene.Final_Mutation_Mutation_Strength

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

Variable	**Total (n=1141)**	**Total DSIC (n=1121)**	**Diff. (n=20)**
Patients with NMD (%)	85 (7.6)	43 (3.85)	42 (3.75)
Patients with PKD2 (%)	165 (14.7)	166 (14.8)	1 (0.1)
Total patients with PKD1 (%)	869 (77.7)	897 (80.2)	28 (2.5)
Total patients with PKD1, mutation type (%)			
Frameshift, D/I	282 (32.5)	253 (28.2)	29 (4.3)
Splice	96 (11.0)	102 (11.4)	6 (0.4)
Nonsense	214 (24.6)	222 (24.7)	8 (0.1)
Missense	223 (25.7)	227 (25.3)	4 (0.4)
In-frame, D/I	54 (6.2)	56 (6.2)	2 (0)
Patients with PKD1, MSG (%)			
PKD1 truncating, MSG1	575 (66.3)	595 (66.3)	20 (0)
PKD1 nontruncating, MSG2	172 (19.8)	181 (20.2)	9 (0.4)
PKD nontruncating, MSG3	120 (13.8)	115 (12.8)	5 (1.0)

Attachment A: SAS Code

```
libname halt "X:\NIDDK\niddk-dr_studies2\HALT_PKD\private_created_data\All data for DEID";
libname crisp "X:\NIDDK\niddk-dr_studies1\CRISP\private_created_data\CRISP Allele";
libname allele "X:\NIDDK\niddk-dr_studies2\HALT_PKD\private_created_data\HALT_PKD Allele Data";
libname cr "X:\NIDDK\niddk-dr_studies1\CRISP\private_created_data\Redacted Data";

proc contents data=halt.halt_pkd_ids_genetic_data;
run;

proc contents data=crisp.crisp_ids_genetic_data;
run;

proc freq data=halt.halt_pkd_ids_genetic_data;
tables haltid;
run;

proc contents data=allele.subset_ids;
run;

*HALT Ids and CRISP Ids;
data ids; set allele.subset_ids;
keep haltid crispid;
run;

data ids_1; set ids;
pkdid = crispid;
if haltid ^= "" AND pkdid ^= . then haltid = "";
run;

data halt_ids; set ids_1;
keep haltid;
if haltid = "" then delete;
run;

data pkdids; set ids_1;
keep pkdid;
if pkdid = . then delete;
run;

*combining crisp and halt allele data;
data halt; set halt.halt_pkd_ids_genetic_data;
run;

data crisp; set crisp.crisp_ids_genetic_data;
run;
```

```
proc sort data= halt;
by haltid;
run;

proc sort data= crisp;
by pkdid;
run;

proc sort data=halt_ids;
by haltid;
run;

data one;
merge
halt_ids (in=a)
halt (in=b);
by haltid;
if a=1;
run;

proc sort data=pkdids;
by pkdid;
run;

data two;
merge
pkdids (in=a)
crisp (in=b);
by pkdid;
if a=1;
run;

data one; set one;
pkdid = .;
run;

data two; set two;
haltid = "";
run;

data paper; set one two;
run;

*Table 1 info from the allele datasets;
proc contents data=paper;
run;

proc freq data=paper;
```

```
tables Final_Gene Final_Gene2;
run;

proc freq data=allele.subset_ids;
tables Gene;
run;

proc freq data=paper;
tables Final_Mutation_Type;
where Final_Gene = "PKD1";
run;

proc freq data=paper;
run;

proc freq data=paper;
tables Final_Mutation_Fuctional_Effect*Final_Mutation_Mutation_Strength ;
where Final_Gene = "PKD1";
run;
```