

Dataset Integrity Check for the Hepatitis B Research Network Pediatric Cohort Study (HBRN Cohort P)

Prepared by NIDDK-CR
October 20, 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 Table 1 – Baseline Characteristics.....	4
Table B: Comparison of values computed in integrity check to reference article Table 1	5
Attachment A: SAS Code	8

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 Hepatitis B Research Network (HBRN) was a multicenter network to investigate the etiology and progression of the disease and to test the safety and efficacy of treatment approaches. The HBRN Pediatric Cohort Study (HBRN Cohort P) was designed to describe participants 6 months to < 18 years of age with hepatitis B virus (HBV) infection in a prospective cohort in the United States and Canada, and to identify predictors of disease activation and progression.

3 Archived Datasets

A full listing of archived datasets included in the package can be found in the Roadmap document. All data files, as provided by the Data Coordinating Center (DCC), are located in the HBRN Pediatric Cohort folder in the data package. For this replication, variables were taken from the “analytic.sas7bdat” dataset.

4 Statistical Methods

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

5 Results

For Table 1 in the publication [1], [Baseline Characteristics](#), 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 in Table 1. The results of the replication are an exact match to the published results.

6 Conclusions

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

7 References

[1] Ling SC, Lin HS, Murray KF, Rosenthal P, Mogul D, Rodriguez-Baez N, Schwarzenberg SJ, Teckman J, Schwarz KB. Chronic Hepatitis Is Common and Often Untreated Among Children with Hepatitis B Infection in the United States and Canada. *The Journal of Pediatrics*, 237, 24-33, October 2021. doi: <https://doi.org/10.1016/j.jpeds.2021.05.035>

Table A: Variables used to replicate Table 1 – Baseline Characteristics

Table Variable	dataset.variable
Sex	analytic.sex
Age at baseline	analytic.age_visit
Race	analytic.race_new
Mode of transmission of HBV	analytic.tranmode_com
Place of birth	analytic.pob
Patient adopted	analytic.adopt
HBV genotype	analytic.gen_cat
Platelets x 10 ³ /mm ³	analytic.plat_i
ALT x ULN	analytic.alt_norm_new
AST x ULN	analytic.ast_norm
HBeAg	analytic.eag
HBV DNA, log ₁₀ IU/mL	analytic.hbv dna_log_imp
qHBeAg, log ₁₀ IU/mL	analytic.equant_log_i
qHBsAg, log ₁₀ IU/mL	analytic.squant_log_i
Anti-HBe	analytic.hbe
Anti-HBs	analytic.antihbs
BMI z-score	analytic.bmi_z
Height z-score	analytic.height_z
Weight z-score	analytic.weight_z
Tanner Stage: Physical growth	analytic.tanphy
Tanner stage: Pubic hair growth	analytic.tanpub

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

Characteristic	Pub: Baseline Characteristics (n=362)	DSIC: Baseline Characteristics (n=362)	Diff. (n=0)
Sex			
Male	148 (41%)	148 (41%)	0 (0)
Female	214 (59%)	214 (59%)	0 (0)
Age at baseline			
Mean (SD)	9.8 (4.4)	9.8 (4.4)	0 (0)
Median (IQR)	10.0 (6.1:13.5)	10.0 (6.1:13.5)	0 (0)
Min:Max	1.0:17.6	1.0:17.6	0
Age at baseline			
6 months to < 5 years	67 (19%)	67 (19%)	0 (0)
5 to < 13 years	194 (54%)	194 (54%)	0 (0)
13 to < 18 years	101 (28%)	101 (28%)	0 (0)
Race			
White	32 (9%)	32 (9%)	0 (0)
Black	36 (10%)	36 (10%)	0 (0)
Asian	286 (79%)	286 (79%)	0 (0)
Other/Mixed	7 (2%)	7 (2%)	0 (0)
Mode of Transmission of HBV			
Vertical transmission	286 (98%)	286 (98%)	0 (0)
Horizontal/Household contact	6 (2%)	6 (2%)	0 (0)
Place of Birth			
U.S./Canada	93 (26%)	93 (26%)	0 (0)
Other North America/South America	6 (2%)	6 (2%)	0 (0)
Europe	21 (6%)	21 (6%)	0 (0)
Asia/Australia	215 (60%)	215 (60%)	0 (0)
Africa	26 (7%)	26 (7%)	0 (0)
Patient Adopted			
No	165 (46%)	165 (46%)	0 (0)
Yes	197 (54%)	197 (54%)	0 (0)
HBV Genotype			
A	13 (5%)	13 (5%)	0 (0)
B	118 (42%)	118 (42%)	0 (0)
C	88 (31%)	88 (31%)	0 (0)
D	52 (18%)	52 (18%)	0 (0)
E	11 (4%)	11 (4%)	0 (0)
Multiple	1 (0%)	1 (0%)	0 (0)
Platelets (x10 ³ /mm ³)			
Median (IQR)	267 (229:311)	267 (229:311)	0 (0)
Min:Max	122:582	122:582	0
ALT x ULN			
Median (IQR)	1.6 (1.1:2.3)	1.6 (1.1:2.3)	0 (0)
Min:Max	0.3:82.7	0.3:82.7	0

Characteristic	Pub: Baseline Characteristics (n=362)	DSIC: Baseline Characteristics (n=362)	Diff. (n=0)
ALT x ULN			
≤ 1 ULN	69 (19%)	69 (19%)	0 (0)
> 1-3 ULN	235 (66%)	235 (66%)	0 (0)
> 3-5 ULN	25 (7%)	25 (7%)	0 (0)
> 5 ULN	26 (7%)	26 (7%)	0 (0)
AST x ULN			
Median (IQR)	0.9 (0.8:1.2)	0.9 (0.8:1.2)	0 (0)
Min:Max	0.3:27.5	0.3:27.5	0
AST x ULN			
≤ 1 ULN	209 (59%)	209 (59%)	0 (0)
> 1-3 ULN	134 (38%)	134 (38%)	0 (0)
> 3-5 ULN	7 (2%)	7 (2%)	0 (0)
> 5 ULN	5 (1%)	5 (1%)	0 (0)
HBeAg			
Negative	89 (25%)	89 (25%)	0 (0)
Positive	261 (75%)	261 (75%)	0 (0)
HBV DNA (log ₁₀ IU/mL)			
Median (IQR)	8.1 (5.2:8.4)	8.1 (5.2:8.4)	0 (0)
Min:Max	BLD:9.3	0.6:9.3	0
qHBeAg (log ₁₀ IU/mL)			
Median (IQR)	3.1 (-0.6:3.3)	3.1 (-0.6:3.3)	0 (0)
Min:Max	BLD:3.8	-2.3:3.8	0
qHBsAg (log ₁₀ IU/mL)			
Median (IQR)	4.4 (4.0:4.8)	4.4 (4.0:4.8)	0 (0)
Min:Max	-0.2:5.6	-0.2:5.6	0
Anti-HBe			
Negative	227 (72%)	227 (72%)	0 (0)
Positive	90 (28%)	90 (28%)	0 (0)
Anti-HBs			
Negative	289 (95%)	289 (95%)	0 (0)
Positive	16 (5%)	16 (5%)	0 (0)
BMI z-score			
Median (IQR)	0.1 (-0.7:0.8)	0.1 (-0.7:0.8)	0 (0)
Min:Max	-2.9:2.2	-2.9:2.2	0
Height z-score			
Median (IQR)	-0.2 (-0.9:0.5)	-0.2 (0.9:0.5)	0 (0)
Min:Max	-3.6:2.6	-3.6:2.6	0
Weight z-score			
Median (IQR)	-0.1 (-0.9:0.7)	-0.1 (-0.9:0.7)	0 (0)
Min:Max	-3.4:2.6	-3.4:2.6	0

Characteristic	Pub: Baseline Characteristics (n=362)	DSIC: Baseline Characteristics (n=362)	Diff. (n=0)
Tanner Stage: Physical growth			
I	186 (55%)	186 (55%)	0 (0)
II	28 (8%)	28 (8%)	0 (0)
III	49 (14%)	49 (14%)	0 (0)
IV	58 (17%)	58 (17%)	0 (0)
V	18 (5%)	18 (5%)	0 (0)
Tanner Stage: Pubic hair growth			
I	204 (60%)	204 (60%)	0 (0)
II	23 (7%)	23 (7%)	0 (0)
III	37 (11%)	37 (11%)	0 (0)
IV	53 (16%)	53 (16%)	0 (0)
V	21 (6%)	21 (6%)	0 (0)

Attachment A: SAS Code

```
libname dsic "X:\NIDDK\niddk-dr_studies2\HBRN\private_created_data\Pediatric Cohort\Redacted Datasets";
```

```
/******  
/* HBRN Peds Cohort */  
/* DSIC Ling et al. */  
/******
```

```
data one; set dsic.analytic;  
where tmpt_num_p = 0;  
run;
```

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

```
*sex;  
proc freq data=one;  
tables sex;  
run;
```

```
*age;  
proc means data=one mean std median q1 q3 min max;  
var age_visit;  
run;
```

```
*age at baseline (cat);  
data two; set one;  
age_cat = 0;  
if age_visit < 5 then age_cat = 1;  
if age_visit >=5 AND age_visit < 13 then age_cat = 2;  
if age_visit >=13 then age_cat = 3;  
run;
```

```
proc freq data=two;  
tables age_cat;  
run;
```

```
*Race;  
proc freq data=two;  
tables race_new;  
run;
```

```
*Mode of transmission;  
proc freq data=two;
```

```
tables tranmode_com;  
run;
```

```
*Place of birth;  
proc freq data=two;  
tables pob;  
run;
```

```
*adopted;  
proc freq data=two;  
tables adopt;  
run;
```

```
*HBV genotype;  
proc freq data=two;  
tables gen_cat;  
run;
```

```
*platelets;  
proc means data=two n median q1 q3 min max;  
var plat_i;  
run;
```

```
*ALTxULN;  
proc means data=two n median q1 q3 min max;  
var alt_norm_new;  
run;
```

```
*ALTxULN cat;  
data three; set two;  
alt_cat = .;  
if alt_norm_new <=1 AND alt_norm_new > 0 then alt_cat = 1;  
if alt_norm_new > 1 AND alt_norm_new <= 3 then alt_cat = 2;  
if alt_norm_new > 3 AND alt_norm_new <= 5 then alt_cat = 3;  
if alt_norm_new > 5 then alt_cat = 4;  
run;
```

```
proc freq data=three;  
tables alt_cat;  
run;
```

```
*AST;  
proc means data=three n median q1 q3 min max;  
var ast_norm;  
run;
```

```
*ASTxULN;  
data four; set three;
```

```
ast_cat = .;
if ast_norm <= 1 AND ast_norm > 0 then ast_cat = 1;
if ast_norm > 1 AND ast_norm <=3 then ast_cat = 2;
if ast_norm > 3 AND ast_norm <=5 then ast_cat = 3;
if ast_norm > 5 then ast_cat = 4;
run;
```

```
proc freq data=four;
tables ast_cat;
run;
```

```
*HbeAG;
proc freq data=four;
tables eag;
run;
```

```
*HBV DNA;
proc means data=four n median q1 q3 min max;
var HBVDNA_log_imp;
run;
```

```
*qHBeAg;
proc means data=four n median q1 q3 min max;
var eQuant_log_i;
run;
```

```
*qHBsAg;
proc means data=four n median q1 q3 min max;
var sQuant_log_i;
run;
```

```
*Anti-Hbe;
proc freq data=four;
tables HBe;
run;
```

```
*Anti-Hbs;
proc freq data=four;
tables antiHBs;
run;
```

```
*BMI, Height, Weight zscore;
proc means data=four n median q1 q3 min max;
var bmi_z height_z weight_z;
run;
```

```
*Tanner stage physical and pubic hair growth;
proc freq data=four;
```

```
tables tanphy tanpub;  
run;
```