

Dataset Integrity Check for the Hepatitis B Research Network Adult Cohort Study (HBRN Cohort A)

Prepared by NIDDK-CR
August 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 Table 1 – Baseline Characteristics of Participants With and Without Cirrhosis	4
Table B: Comparison of values computed in integrity check to reference article Table 1	5
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 Hepatitis B Research Network (HBRN) is a multicenter network to investigate the etiology and progression of the disease and to test the safety and efficacy of current treatment approaches. The HBRN Adult Cohort Study (HBRN Cohort A) was designed to describe participants with hepatitis B virus (HBV) infection and identify predictors of disease activation and progression. The relationship of HBV genotype to clinical, biochemical, and histological characteristics and to pregnancy was also explored.

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 Adult Cohort folder in the data package. For this replication, variables were taken from the “baselinechar.sas7bdat” dataset.

4 Statistical Methods

Analyses were performed to replicate results for the data in the publication by Lok 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 of Participants With and Without Cirrhosis, 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 almost an exact match to the published results.

6 Conclusions

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

7 References

[1] Lok AS, Perrillo R, Lalama CM, Fried MW, Belle SH, Ghany MG, Khalili M, Fontana RJ, Sterling RK, Terrault N, Feld JJ, Di Bisceglie AM, Lau DTY, Hassan M, Janssen HLA. Low Incidence of Adverse Outcomes in Adults With Chronic Hepatitis B Virus Infection in the Era of Antiviral Therapy. *Hepatology*, 73(6), 2124-2140, June 2021. doi: <https://doi.org/10.1002/hep.31554>

Table A: Variables used to replicate Table 1 – Baseline Characteristics of Participants With and Without Cirrhosis

Table Variable	dataset.variable
Age	baselinechar.age_eri
Sex	baselinechar.sex
Race	baselinechar.race_new
Continent of birth	baselinechar.continp
HBeAg status	baselinechar.eag_st
HBV DNA	baselinechar.hbv dna_log_i
Quant HBsAg	baselinechar.sQuant_log_i
HBV genotype	baselinechar.gen_comb
APRI	baselinechar.apri_cat
FIB-4	baselinechar.fib4_cat
HBV phenotype	baselinechar.calc_pheno
Platelets	baselinechar.plat
AST × ULN	baselinechar.ast_cat
ALT × ULN	baselinechar.alt_cat
Albumin	baselinechar.alb
Total bilirubin	baselinechar.tbili
Ever received HBV treatment	baselinechar.txstat_eri
BMI	baselinechar.bmi_a
Diabetes	baselinechar.mxdiab
NIAAA alcohol risk level in past 12 months	baselinechar.alc_risk12

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

Characteristic	Pub: All (n=1418)	DSIC: All (n=1418)	Diff. (n=0)	Pub: With Baseline Cirrhosis (n=21)	DSIC: With Baseline Cirrhosis (n=21)	Diff (n=0)	Pub: Without Baseline Cirrhosis (n=1397)	DSIC: Without Baseline Cirrhosis (n=1397)	Diff. (n=0)
Age at enrollment (years) Median (IQR)	41.1 (32.9:51.5)	41.1 (32.9:51.5)	0 (0)	50.7 (42.4:57.3)	50.7 (42.4:57.3)	0 (0)	41.0 (32.9:51.4)	41.0 (32.9:51.4)	0 (0)
Sex (Female)	730 (51.5%)	730 (51.5%)	0 (0)	6 (28.6%)	6 (28.6%)	0 (0)	724 (51.8%)	724 (51.8%)	0 (0)
Race									
White	142 (10.0%)	142 (10.0%)	0 (0)	5 (23.8%)	5 (23.8%)	0 (0)	137 (9.8%)	137 (9.8%)	0 (0)
Black	179 (12.7%)	179 (12.7%)	0 (0)	3 (14.3%)	3 (14.3%)	0 (0)	176 (12.6%)	176 (12.6%)	0 (0)
Asian	1,055 (74.6%)	1,055 (74.6%)	0 (0)	11 (52.4%)	11 (52.4%)	0 (0)	1,044 (74.9%)	1,044 (74.9%)	0 (0)
Other/mixed	39 (2.8%)	39 (2.8%)	0 (0)	2 (9.5%)	2 (9.5%)	0 (0)	37 (2.7%)	37 (2.7%)	0 (0)
Continent of birth									
Africa	139 (9.8%)	139 (9.8%)	0 (0)	1 (4.8%)	1 (4.8%)	0 (0)	138 (9.9%)	138 (9.9%)	0 (0)
Asia	972 (68.7%)	972 (68.7%)	0 (0)	11 (52.4%)	11 (52.4%)	0 (0)	961 (69.0%)	961 (69.0%)	0 (0)
Europe	43 (3.0%)	43 (3.0%)	0 (0)	1 (4.8%)	1 (4.8%)	0 (0)	42 (3.0%)	42 (3.0%)	0 (0)
North America	249 (17.6%)	249 (17.6%)	0 (0)	8 (38.1%)	8 (38.1%)	0 (0)	241 (17.3%)	241 (17.3%)	0 (0)
South America	8 (0.6%)	8 (0.6%)	0 (0)	0 (0.0%)	0 (0.0%)	0 (0)	8 (0.6%)	8 (0.6%)	0 (0)
Australia	3 (0.2%)	3 (0.2%)	0 (0)	0 (0.0%)	0 (0.0%)	0 (0)	3 (0.2%)	3 (0.2%)	0 (0)
HBeAg status (Positive)	331 (24.0%)	331 (24.0%)	0 (0)	6 (30.0%)	6 (30.0%)	0 (0)	325 (23.9%)	325 (23.9%)	0 (0)
HBV DNA (log ₁₀ IU/mL) Median (IQR)	3.6 (2.6:5.6)	3.6 (2.6:5.6)	0 (0)	3.7 (1.8:5.8)	3.7 (1.8:5.8)	0 (0)	3.6 (2.6:5.5)	3.6 (2.6:5.5)	0 (0)
Quant HBsAg (log ₁₀ IU/mL) Median (IQR)	3.39 (2.67:4.11)	3.39 (2.69:4.11)	0 (0)	3.14 (2.45:3.92)	3.14 (2.41:3.90)	0 (0.04:0.02)	3.40 (2.67:4.11)	3.40 (2.67:4.11)	0 (0)
HBV genotype									
A	219 (16.8%)	219 (16.8%)	0 (0)	4 (23.5%)	4 (23.5%)	0 (0)	215 (16.7%)	215 (16.7%)	0 (0)
B	523 (40.2%)	523 (40.2%)	0 (0)	3 (17.6%)	3 (17.6%)	0 (0)	520 (40.5%)	520 (40.5%)	0 (0)
C	434 (33.4%)	434 (33.4%)	0 (0)	7 (41.2%)	7 (41.2%)	0 (0)	427 (33.3%)	427 (33.3%)	0 (0)
D	87 (6.7%)	87 (6.7%)	0 (0)	2 (11.8%)	2 (11.8%)	0 (0)	85 (6.6%)	85 (6.6%)	0 (0)
E	33 (2.5%)	33 (2.5%)	0 (0)	1 (5.9%)	1 (5.9%)	0 (0)	32 (2.5%)	32 (2.5%)	0 (0)
Mixed/other	5 (0.4%)	5 (0.4%)	0 (0)	0 (0.0%)	0 (0.0%)	0 (0)	5 (0.4%)	5 (0.4%)	0 (0)
APRI									
≤ 0.5	943 (78.0%)	943 (78.0%)	0 (0)	3 (14.3%)	3 (14.3%)	0 (0)	940 (79.1%)	940 (79.1%)	0 (0)
> 0.5-2.0	244 (20.2%)	244 (20.2%)	0 (0)	15 (71.4%)	15 (71.4%)	0 (0)	229 (19.3%)	229 (19.3%)	0 (0)
> 2	22 (1.8%)	22 (1.8%)	0 (0)	3 (14.3%)	3 (14.3%)	0 (0)	19 (1.6%)	19 (1.6%)	0 (0)

Characteristic	Pub: All (n=1418)	DSIC: All (n=1418)	Diff. (n=0)	Pub: With Baseline Cirrhosis (n=21)	DSIC: With Baseline Cirrhosis (n=21)	Diff (n=0)	Pub: Without Baseline Cirrhosis (n=1397)	DSIC: Without Baseline Cirrhosis (n=1397)	Diff. (n=0)
FIB-4									
< 1.45	966 (79.9%)	966 (79.9%)	0 (0)	6 (28.6%)	6 (28.6%)	0 (0)	960 (80.8%)	960 (80.8%)	0 (0)
1.45-3.25	222 (18.4%)	222 (18.4%)	0 (0)	11 (52.4%)	11 (52.4%)	0 (0)	211 (17.8%)	211 (17.8%)	0 (0)
> 3.25	21 (1.7%)	21 (1.7%)	0 (0)	4 (19.0%)	4 (19.0%)	0 (0)	17 (1.4%)	17 (1.4%)	0 (0)
HBV phenotype									
Immune tolerant	55 (4.1%)	55 (4.1%)	0 (0)	0 (0.0%)	0 (0.0%)	0 (0)	55 (4.2%)	55 (4.2%)	0 (0)
HBeAg+ CHB	230 (17.2%)	230 (17.2%)	0 (0)	5 (25.0%)	5 (25.0%)	0 (0)	225 (17.1%)	225 (17.1%)	0 (0)
HBeAg- CHB	233 (17.4%)	233 (17.4%)	0 (0)	3 (15.0%)	3 (15.0%)	0 (0)	230 (17.5%)	230 (17.5%)	0 (0)
Inactive carrier	324 (24.2%)	324 (24.2%)	0 (0)	4 (20.0%)	4 (20.0%)	0 (0)	320 (24.3%)	320 (24.3%)	0 (0)
Indeterminant	495 (37.0%)	495 (37.0%)	0 (0)	8 (40.0%)	8 (40.0%)	0 (0)	487 (37.0%)	487 (37.0%)	0 (0)
Platelets ($\times 10^3/\text{mm}^3$)									
Median (IQR)	219.0 (182.0:256.0)	219.0 (182.0:256.0)	0 (0)	142.0 (106.0:181.0)	142.0 (106.0:181.0)	0 (0)	220.0 (183.0:257.0)	220.0 (183.0:257.0)	0 (0)
AST \times ULN									
$\leq 1 \times$ ULN	1,103 (80.0%)	1,103 (80.0%)	0 (0)	8 (38.1%)	8 (38.1%)	0 (0)	1,095 (80.6%)	1,095 (80.6%)	0 (0)
> 1-2 \times ULN	211 (15.3%)	211 (15.3%)	0 (0)	9 (42.9%)	9 (42.9%)	0 (0)	202 (14.9%)	202 (14.9%)	0 (0)
> 2 \times ULN	65 (4.7%)	65 (4.7%)	0 (0)	4 (19.0%)	4 (19.0%)	0 (0)	61 (4.5%)	61 (4.5%)	0 (0)
ALT \times ULN									
$\leq 1 \times$ ULN	466 (33.3%)	466 (33.3%)	0 (0)	4 (19.0%)	4 (19.0%)	0 (0)	462 (33.5%)	462 (33.5%)	0 (0)
> 1-2 \times ULN	622 (44.4%)	622 (44.4%)	0 (0)	9 (42.9%)	9 (42.9%)	0 (0)	613 (44.4%)	613 (44.4%)	0 (0)
> 2 \times ULN	313 (22.3%)	313 (22.3%)	0 (0)	8 (38.1%)	8 (38.1%)	0 (0)	305 (22.1%)	305 (22.1%)	0 (0)
Albumin (g/dL)									
Median (IQR)	4.3 (4.1:4.6)	4.3 (4.1:4.6)	0 (0)	3.9 (3.8:4.3)	3.9 (3.8:4.3)	0 (0)	4.3 (4.1:4.6)	4.3 (4.1:4.6)	0 (0)
Total bilirubin (mg/dL)									
Median (IQR)	0.6 (0.4:0.8)	0.6 (0.4:0.8)	0 (0)	0.8 (0.6:1.0)	0.8 (0.6:1.0)	0 (0)	0.6 (0.4:0.8)	0.6 (0.4:0.8)	0 (0)
Ever received HBV treatment (Yes)	194 (13.7%)	194 (13.7%)	0 (0)	2 (9.5%)	2 (9.5%)	0 (0)	192 (13.7%)	192 (13.7%)	0 (0)
BMI (kg/m^2)									
Median (IQR)	24.1 (21.6:27.0)	24.1 (21.6:27.0)	0 (0)	26.2 (24.6:31.1)	26.2 (24.6:31.1)	0 (0)	24.0 (21.6:27.0)	24.0 (21.6:27.0)	0 (0)
Diabetes (Yes)	77 (5.4%)	77 (5.4%)	0 (0)	5 (23.8%)	5 (23.8%)	0 (0)	72 (5.2%)	72 (5.2%)	0 (0)
NIAAA alcohol risk level in past 12 months (At-risk)	100 (7.1%)	100 (7.1%)	0 (0)	2 (9.5%)	2 (9.5%)	0 (0)	98 (7.1%)	98 (7.1%)	0 (0)

Attachment A: SAS Code

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

```
/******  
/*DSIC for HBRN Adult*/  
/* Cohort Lok et al. */  
/******
```

```
*identifying the analysis group: Indicator variables tx_fl24 and gt24fup;
```

```
proc freq data=dsic.baselinechar;  
tables tx_fl24*gt24fup/missing;  
run;
```

```
*creating analysis dataset based on the indicator;
```

```
data base; set dsic.baselinechar;  
if tx_fl24 = . AND gt24fup = 1; /*these two together identify 1418 participants*/  
run;
```

```
proc freq data=base;  
tables cirr_erl; /*indicator for cirrhosis at baseline*/  
run;
```

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

```
*Age at enrollment;
```

```
proc means data=base median q1 q3;  
var age_erl;  
class cirr_erl;  
run;
```

```
*Sex;
```

```
proc freq data=base;  
tables sex*cirr_erl/norow;  
run;
```

```
*Race;
```

```
proc freq data=base;  
tables race_new*cirr_erl/norow;  
run;
```

```
*Continent of birth;
```

```
proc freq data=base;  
tables continp*cirr_erl/norow;  
run;
```

```
*HBeAG status;  
proc freq data=base;  
tables eag_st*cirr_erl/norow;  
run;
```

```
*HBV DNA;  
proc means data=base median q1 q3;  
var hbvdna_log_i;  
class cirr_erl;  
run;
```

```
*Quant HBsAg;  
proc means data=base median q1 q3;  
var sQuant_log_i;  
class cirr_erl;  
run;
```

```
*HBV genotype;  
proc freq data=base;  
tables gen_comb*cirr_erl/norow;  
run;
```

```
*APRI;  
proc freq data=base;  
tables apri_cat*cirr_erl/norow;  
run;
```

```
*FIB-4;  
proc freq data=base;  
tables fib4_cat*cirr_erl/norow;  
run;
```

```
*HBV phenotype;  
proc freq data=base;  
tables calc_pheno*cirr_erl/norow;  
run;
```

```
*Platelets;  
proc means data=base median q1 q3;  
var plat;  
class cirr_erl;  
run;
```

```
*ASTxULN;  
proc freq data=base;  
tables ast_cat*cirr_erl/norow;  
run;
```

```
*ALTxULN;  
proc freq data=base;  
tables alt_cat*cirr_erl/norow;  
run;
```

```
*Albumin;  
proc means data=base median q1 q3;  
var alb;  
class cirr_erl;  
run;
```

```
*Total Bilirubin;  
proc means data=base median q1 q3;  
var tbili;  
class cirr_erl;  
run;
```

```
*ever received HBV treatment;  
proc freq data=base;  
tables txstat_erl*cirr_erl/norow;  
run;
```

```
*BMI;  
proc means data=base median q1 q3;  
var bmi_a;  
class cirr_erl;  
run;
```

```
*diabetes;  
proc freq data=base;  
tables mxdiab*cirr_erl/norow;  
run;
```

```
*NIAA alcohol risk;  
proc freq data=base;  
tables alc_risk12*cirr_erl/norow;  
run;
```