Dataset Integrity Check for the Prevalence and Clinical Features of Patients with Concurrent HBsAg and Anti-HBs: Evaluation of the Hepatitis B Research Network Cohort (HBRN Concurrent HBsAg/Anti-HBs)

Prepared by NIDDK-CR
January 2, 2024

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

An ancillary study from the Hepatitis B Research Network (HBRN) Adult and Pediatric Cohort studies, the Prevalence and Clinical Features of Patients with Concurrent HBsAg and Anti-HBs: Evaluation of the Hepatitis B Research Network Cohort, aimed to examine the prevalence and clinical and virological features of concurrent HBsAg and anti-HBs in children and adults with chronic hepatitis B virus (HBV) infection.

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 and HBRN Pediatric Cohort folders in their respective data packages. For this replication, variables were taken from datasets in the HBRN Adult Cohort study (“hbrn_screen_info , “hbrn_bp.sas7bdat”, “hbrn_bc.sas7bdat”, “hbrn_cdc_results.sas7bdat”, and “hbrn_fw.sas7bdat”), datasets in the HBRN Pediatric Cohort study (“hbrn_peds_ages.sas7bdat”, “hbrn_bpp.sas7bdat”, “hbrn_bcp.sas7bdat”, and “hbrn_cdc_results.sas7bdat”), and merged with the Concurrent HBsAg/Anti_HBs ancillary study dataset (“concur_hbsag.sas7bdat”).

4 Statistical Methods

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

5 Results

For Table 1 in the publication [1], Demographics, clinical and virologic characteristics among HBsAg positive children and adults with chronic HBV infection, and anti-HBs status by demographic, clinical and virologic 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 vary with the published results. The DCC informed the NIDDK Central Repository that not all samples tested were used in the analysis, and the NIDDK Central Repository does not have access to the program the publication used to analyze the datasets and variables.

6 Conclusions

The DCC informed the NIDDK Central Repository that the HBRN Concurrent HBsAg/Anti-HBs data files to be distributed are a true copy of the study data.

7 References

[1] Lee WM, King WC, Schwarz KB, Rule J, Lok ASF. Prevalence and Clinical Features of Patients with Concurrent HBsAg and Anti-HBs: Evaluation of the Hepatitis B Research Network Cohort. Journal of Viral Hepatitis, 27(9), 922-931, September 2020. doi: https://doi.org/10.1111/jvh.13312

Table A: Variables used to replicate Table 1 – Demographics, clinical and virologic characteristics among HBsAg positive children and adults with chronic HBV infection, and anti-HBs status by demographic, clinical and virologic characteristics

Table Variable	**dataset.variable**
Age	hbrn_screen_info.age hbrn_peds_ages.age
Gender	hbrn_bp.sex hbrn_bpp.sex
Race	hbrn_bp.racew hbrn_bpp.racew hbrn_bp.raceb hbrn_bpp.raceb hbrn_bp.racea hbrn_bpp.racea hbrn_bp.raceh hbrn_bpp.raceh hbrn_bp.racei hbrn_bpp.racei hbrn_bp.raceo hbrn_bpp.raceo
Place of birth	hbrn_bc.cborns hbrn_bcp.cborns
HDV+	hbrn_bc.hdv hbrn_bcp.hdv
HCV+	hbrn_bc.hcv hbrn_bcp.hcv
Known family history of chronic HBV	hbrn_bc.hxhbv hbrn_bcp.hxhbv
Prior HBV treatment	hbrn_bc.txhbv hbrn_bcp.txhbv
HBeAg	hbrn_bc.hbeag hbrn_bcp.hbeag
Genotype	hbrn_cdc_results.subgenotype
ALT x ULN	hbrn_fw.altu hbrn_fw.alt hbrn_bcp.altu hbrn_bcp.alt
AST x ULN	hbrn_fw.astu hbrn_fw.ast hbrn_bcp.astu hbrn_bcp.ast
Platelets	hbrn_bc.plat hbrn_bcp.plat

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

Characteristic	Publication: Total (n=1462)	DSIC: Total (n=1462)	Diff. (n=0)
Age, years			
Median (IQR)	41.3 (31.3, 52.0)	41.0 (32.0, 52.0)	0.3 (0.7, 0)
Min, Max	4.0, 80.2	3.2, 80	0.8, 0.2
Age, years			
< 18	124 (8.5)	126 (8.6)	2 (0.1)
18 - < 30	203 (13.9)	211 (14.4)	8 (0.5)
30 - < 40	345 (23.6)	346 (23.7)	1 (0.1)
40 - < 50	355 (24.3)	348 (23.8)	7 (0.5)
≥ 50	435 (29.8)	431 (29.4)	4 (0.4)
Gender, n (%)			
Male	763 (52.2)	746 (52.0)	17 (0.2)
Female	699 (47.8)	689 (48.0)	10 (0.2)
Race, n (%)			
White	164 (11.2)	162 (11.1)	2 (0.1)
Black	189 (13.0)	178 (12.2)	11 (0.8)
Asian	1066 (73.1)	1069 (73.1)	3 (0)
Other	39 (2.7)	53 (3.6)	14 (0.9)
Place of birth, n (%)			
United States/Canada	269 (18.4)	260 (18.1)	9 (0.3)
Other North America/South America	21 (1.4)	16 (1.1)	5 (0.3)
Europe	53 (3.6)	39 (2.7)	14 (0.9)
Asia/Australia	987 (67.6)	986 (68.8)	1 (1.2)
Africa	130 (8.9)	133 (9.3)	3 (0.4)
HDV+			
No	1419 (97.1)	598 (97.4)	821 (0.3)
Yes	43 (2.9)	16 (2.6)	27 (0.3)
HCV+			
No	1437 (98.3)	1012 (98.4)	425 (0.1)
Yes	25 (1.7)	17 (1.6)	8 (0.1)
Known family history of chronic HBV, n (%)			
No	410 (36.1)	409 (36.6)	1 (0.5)
Yes	727 (63.9)	710 (63.5)	17 (0.4)
Prior HBV treatment, n (%)			
No	1256 (85.9)	1235 (86.1)	21 (0.2)
Yes	206 (14.1)	199 (13.9)	7 (0.2)
HBeAg, n (%)			
Negative	1074 (73.6)	970 (72.9)	104 (0.7)
Positive	385 (26.4)	360 (27.1)	25 (0.7)

Characteristic	Publication: Total (n=1462)	DSIC: Total (n=1462)	Diff. (n=0)
Genotype, n (%)			
A-1	92 (6.8)	82 (6.3)	10 (0.5)
A-2	126 (9.3)	117 (9.0)	9 (0.3)
B	530 (39.0)	494 (37.8)	36 (1.2)
C	434 (31.9)	419 (32.0)	15 (0.1)
D	127 (9.3)	93 (7.1)	34 (2.2)
E	36 (2.6)	31 (2.4)	5 (0.2)
Other	14 (1.0)	5 (0.4)	9 (0.6)
ALT x ULN, Median (IQR)	1.3 (0.9:2.0)	1.3 (0.9:2.0)	0 (0:0)
AST x ULN, Median (IQR)	0.7 (0.6:1.0)	1.0 (0.8:1.4)	0.3 (0.2:0.4)
Platelets ($x10^3/mm^3$), Median (IQR)	221 (183.5:259)	220 (182:260)	1 (1.5:1)

Attachment A: SAS Code

```
libname dsic "X:\NIDDK\niddk-dr_studies2\HBRN\private_orig_data\HBRN Ancillary Studies\HBRN
Ancillary Studies\Concurrent HBsAg_antiHBs";
libname adult "X:\NIDDK\niddk-dr_studies2\HBRN\private_created_data\Adult
Cohort\HBRN_Adult_V1\Data" ;
libname peds "X:\NIDDK\niddk-dr_studies2\HBRN\private_created_data\Pediatric Cohort\Redacted
Datasets";
libname peds_one "X:\NIDDK\niddk-dr_studies2\HBRN\private_created_data\Pediatric
Cohort\HBRN_Pediatric_V1\Data";
libname id_list "X:\NIDDK\niddk-dr_studies2\HBRN\private_created_data\HBRN Ancillary Studies";

proc freq data=dsic.concur_hbsag;
run;

proc contents datadata=dsic.concur_hbsag;
run;

proc sql;
select count(distinct id) as distinct_var1
from dsic.concur_hbsag;
quit;

proc contents data=peds_one.hbrn_bpp;
run;

/*****************************/
/* DSIC for HBRN Concurrent */
/* Lee et al. */
/*****************************/

data id; set work.concurrent_id_list;
keep id adult ped;
run;

data concur; set dsic.concur_hbsag;
run;

data bp; set adult.hbrn_bp;
run;

data bpp; set peds_one.hbrn_bpp;
run;

proc sort data=id;
by id tmpt;
run;
```

```
proc sort data=concur;
by id;
run;

proc sort data=bp;
by id;
run;

proc sort data=bpp;
by id;
run;

data one; merge
id (in=e)
concur (in=a)
bp (in=b)
bpp (in=c);
by id;
if e=1;
run;

proc sort data=one nodupkey;
by id;
run;
*age;
*need analytic datsets for age;
data adco; set adult.hbrn_screen_info;
orig_id = id;
run;

data pedco; set peds_one.hbrn_peds_ages;
run;

proc sort data=adco;
by orig_id;
run;

proc sort data=pedco;
by id;
run;

data concur_one; set concur;
orig_id = id;
run;

data age; merge
one (in=e)
```

```
adco (in=a)
pedco (in=b)
concur_one (in=c);
by id;
if e=1;
run;

proc sort data=age nodupkey;
by id;
run;

proc means data=age n median q1 q3 min max;
var age;
run;

*age cat;
data age_cat; set age;
age_one = .;
if age < 18 then age_one = 1;
if age >=18 AND age < 30 then age_one = 2;
if age >=30 AND age < 40 then age_one = 3;
if age >=40 AND age < 50 then age_one = 4;
if age >=50 then age_one = 5;
run;

proc freq data=age_cat;
tables age_one;
run;

*sex;
data ped_sex; set peds_one.hbrn_slp;
run;

data ped_sex2; set peds_one.hbrn_bpp;
run;

data adult_sex; set adult.hbrn_sl;
run;

proc sort data=adult_sex;
by id;
run;

proc sort data=ped_sex ;
by id;
run;

proc sort data=ped_sex2;
```

```
by id;
run;

data sex; merge
one (in=e)
ped_sex (in=b)
ped_sex2 (in=d)
adult_sex (in=c);
by id;
if e=1;
run;

proc sort data=sex nodupkey;
by id scrsex sex ;
run;

data sex_1; set sex;
gender = 0;
if sex = 1 OR scrsex = 1 then gender = 1;
if sex = 2 OR scrsex = 2 then gender = 2;
run;

proc freq data=sex_1;
tables scrsex gender;
run;

*race;
data race; set sex_1;
race_dsic = .;
if racew = 1 then race_dsic = 1;
if raceb = 1 then race_dsic = 2;
if racea = 1 then race_dsic = 3;
if raceh = 1 OR racei = 1 OR raceo = 1 then race_dsic = 4;
if race_dsic = . then race_dsic = 4;
run;

proc freq data=race;
tables race_dsic/missing;
run;

*place of birth;
data pedcountry; set peds_one.hbrn_bcp;
keep id cborns;
run;

data adcountry; set adult.hbrn_bc;
keep id cborns;
run;
```

```
proc sort data=pedcountry;
by id;
run;

proc sort data=adcountry;
by id;
run;

data country; merge
one (in=d)
concur (in=a)
pedcountry (in=b)
adcountry (in=c);
by id;
if d = 1;
run;

proc sort data=country nodupkey ;
by id;
run;

proc freq data=country;
tables cborns;
run;

data country_one; set country;
cob = 5;
if cborns = "CANADA" OR cborns = "UNITED STATES OF AMERICA" then cob = 1;
if cborns = "BRAZIL" or cborns = "COLOMBIA" OR cborns = "CUBA" OR cborns = "GUYANA" OR cborns =
"HAITI" or cborns = "HONDURAS"
OR cborns = "JAMAICA" OR cborns = "PUERTO RICO" then cob = 2;
if cborns = "ALBANIA" OR cborns = "BELARUS" OR cborns = "BOSNIA AND HERZEGOVINA" OR cborns =
"BULGARIA" OR cborns = "CZECH REPUBLIC"
OR cborns = "ERITREA" OR cborns = "ESTONIA" OR cborns = "FRANCE" OR cborns = "GREECE" OR cborns
= "ITALY" OR cborns = "POLAND"
OR cborns = "ROMANIA" OR cborns = "SLOVAKIA" OR cborns = "UKRAINE" OR cborns = "UNITED
KINGDOM" then cob = 3;
if cborns = "AFGHANISTAN" OR cborns = "BANGLADESH" OR cborns = "CAMBODIA" OR cborns = "CHINA"
OR cborns = "HONG KONG"
OR cborns = "INDIA" OR cborns = "INDONESIA" OR cborns = "IRAN, ISLAMIC REPUBLIC OF" OR cborns =
"ISRAEL" OR cborns = "JAPAN"
OR cborns = "KAZAKHSTAN" OR cborns = "KOREA, (NORTH) DEMOCRATIC PEOPLES REPUBLIC OF" OR
cborns = "KOREA, (SOUTH) REPUBLIC OF"
OR cborns = "LAO PEOPLES DEMOCRATIC REPUBLIC" OR cborns = "MALAYSIA" OR cborns =
"MICRONESIA, FEDERATED STATES OF"
OR cborns = "MONGOLIA" OR cborns = "MYANMAR" OR cborns = "NEPAL" OR cborns = "PAKISTAN" or
cborns = "PHILIPPINES"
```

```
OR cborns = "RUSSIAN FEDERATION" OR cborns = "TAIWAN" OR cborns = "THAILAND" OR cborns =
"TURKEY" OR cborns = "UZBEKISTAN"
OR cborns = "VIETNAM" then cob = 4;
if cborns = "" then cob = .;
run;

proc freq data=country_one;
tables cob;
run;

*HDV and HCV;
data bc; set adult.hbrn_bc;
run;

data bcp; set peds_one.hbrn_bcp;
run;

proc sort data=bc;
by id;
run;

proc sort data=bcp;
by id;
run;

data hdv_hcv; merge
one (in=d)
bc (in=a)
bcp (in=b)
concur_one (in=c);
by id;
if d=1;
run;

proc sort data=hdv_hcv nodupkey;
by id;
run;

proc freq data=hdv_hcv;
tables hdv hcv;
run;

*family history;
proc freq data=hdv_hcv;
tables hxhbv;
run;

*history of treatment;
```

```
proc freq data=hdv_hcv;
tables txhbv;
run;

*HBeAG;
proc freq data=hdv_hcv;
tables hbeag;
run;

*genotype;
proc freq data=hdv_hcv;
tables bgen;
run;

data genoadult; set adult.hbrn_cdc_results;
run;

data genopeds; set peds_one.hbrn_cdc_results;
run;

proc sort data=genoadult;
by id;
run;

proc sort data=genopeds;
by id;
run;

data geno; merge
one (in=d)
genoadult (in=a)
genopeds (in=b)
concur_one (in=c);
by id;
if d = 1;
run;

proc freq data=geno;
tables genotype genotype2 subgenotype subgenotype2;
run;

*alt ast and platelets;
data fw; set adult.hbrn_fw;
run;

proc sort data=fw;
by id;
run;
```

```
data labs; merge
fw (in=a)
bcp (in=b)
concur_one (in=c);
by id;
if c=1;
run;

proc means data=labs n median q1 q3;
var alt altu altd altl altm alty;
run;

data labs_one; set labs;
alt_uln = (altu/alt);
ast_uln = (astu/ast);
run;

proc means data=labs_one n median q1 q3;
var alt_uln ast_uln plat;
run;
```