Dataset Integrity Check for the Immune Regulation and Costimulation in Natural History of Chronic Hepatitis B (HBRN Immunology Cohort)

Prepared by NIDDK-CR
January 31, 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

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 Immune Regulation and Costimulation in Natural History of Chronic Hepatitis B (HBRN Immunology Cohort) study aimed to assess whether the balance between immune regulatory and effector responses in HBV-infected individuals defines the level of viremia, liver inflammation, and treatment outcomes.

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 Immunology Cohort folder in the data package. For this replication, variables were taken from the "immuno_cohort_received.sas7bdat", "immune_cohort_ds.sas7bdat", "controls.sas7bdat", and "ic.sas7bdat" datasets from the HBRN Immunology Cohort data package, and merged with the "baselinechar.sas7bdat" dataset from the HBRN Adult Cohort data package.

4 Statistical Methods

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

5 Results

For Supplemental Table 2 in the publication [1], Patient characteristics relative to physician-assigned CHB phenotype, 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 Supplemental Table 2. Only overall

statistics were replicated. The results of the replication are within expected variation to the published results.

6 Conclusions

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

7 References

[1] Park JJ, Wong DK, Wahed AS, Lee WM, Feld JJ, Terrault N, Khalili M, Sterling RK, Kowdley KV, Bzowej N, Lau DT, Kim WR, Smith C, Carithers RL, Torrey KW, Keith JW, Levine DL, Traum D, Ho S, Valiga ME, Johnson GS, Doo E, Lok ASF, Chang KM. Hepatitis B Virus—Specific and Global T-Cell Dysfunction in Chronic Hepatitis B. Gastroenterology, 150(3), 684-695, March 2016. doi: https://doi.org/10.1053/j.gastro.2015.11.050

Table A: Variables used to replicate Supplementary Table 2 – Patient characteristics relative to physician-assigned CHB phenotype

Table Variable	**dataset.variable**
Age	baselinechar.age_erl controls.con_age
Sex	baselinechar.sex controls.con_sex
Race	baselinechar.race_new controls.con_race
ALT	baselinechar.alt
Total bilirubin	baselinechar.tbili
FIB-4	baselinechar.fib4
APRI	baselinechar.apri
$\log_{10}$ HBV DNA	baselinechar.hbvdna_log_i
HBV genotype	baselinechar.gen_cat
Prior HBV therapy	baselinechar.txstat_erl

Table B: Comparison of values computed in integrity check to reference article Supplementary Table 2

Characteristic	Publication: All (n=200)	DSIC: All (n=193)	Diff. (n=7)
Median Age, years (25^{th}, 75^{th} IQR)	42 (29, 51)	42 (29.8, 52.5)	0 (0.8, 1.5)
% Males : % Females	55% : 45%	50% : 50%	5% : 5%
Caucasians (%)	22 (11%)	22 (11%)	0 (0)
African Americans (%)	9 (5%)	6 (3%)	3 (2%)
Asians (%)	166 (83%)	153 (79%)	13 (4%)
Hawaiians (%)	1 (1%)	0 (0)	1 (1%)
Mixed (%)	2 (1%)	4 (2%)	2 (1%)
ALT, IU/mL (IQR)	38 (24, 78)	35 (24, 74)	3 (0, 4)
Min : Max	11 : > threshold	11 : 808	0 : --
Total Bilirubin, mg/dL (IQR)	0.6 (0.5, 0.9)	0.6 (0.5, 0.9)	0 (0, 0)
Min : Max	0.2 : 7.6	0.2 : 5.0	0 : 2.6
FIB-4 (IQR)	1 (0.6, 1.4)	0.9 (0.6, 1.3)	0.1 (0, 0.1)
Min : Max	0.1 : 17.6	0.2 : 9.0	0.1 : 8.6
APRI (IQR)	0.4 (0.3, 0.7)	0.4 (0.3, 0.6)	0 (0, 0.1)
Min : Max	0.1 : 30.1	0.1 : 4.5	0 : 25.6
log_{10} HBV DNA, IU/mL (IQR)	4.9 (3.5, 7.8)	5.0 (3.2, 7.9)	0.1 (0.3, 0.1)
Min : Max	0 : 12	0.3 : 9.1	0.3 : 2.9
HBV Genotype			
A	18 (9%)	15 (7%)	3 (2%)
B	93 (47%)	82 (43%)	11 (4%)
C	59 (30%)	58 (30%)	1 (0%)
D	12 (6%)	11 (6%)	1 (0%)
E	5 (3%)	7 (4%)	2 (1%)
Not Available	13 (7%)	20 (11%)	7 (4%)
Prior HBV Therapy (%)	28 (14%)	25 (14%)	3 (0%)

Attachment A: SAS Code

```
libname immu "X:\NIDDK\niddk-dr_studies2\HBRN\private_created_data\HBRN Immunology
Cohort\Redacted Data";
libname adult "X:\NIDDK\niddk-dr_studies2\HBRN\private_created_data\Adult
Cohort\HBRN_Adult_V2\Data\Analytic Datasets";

/*****************************/
/* HBRN Immunology Cohort */
/* DSIC Park et al. */
/*****************************/

*identifying participants included in the pub;

proc sql;
select count(distinct id) as distinct_var1
from immu.immuno_cohort_received;
quit;

proc freq data=immu.ic;
tables incons*immelig inphen inphens*incons studyid;
run;

proc freq data=immu.immuno_cohort_received;
tables immstudy ;
run;

proc freq data=immu.cd3hi;
tables enroll imm_group;
run;

proc freq data=immu.controls;
tables studygrp_char ssi ;
run;

proc freq data=immu.immuno_cohort_ds;
tables clinicalgrp immstudy;
run;

*identifying the 180 participants (not controls);
data ic_one; set immu.ic;
if immelig = 1 AND incons = 1;
run;

proc freq data=ic_one;
tables inphens*incons;
run;
```

```
data received; set immu.immuno_cohort_received;
if immstudy = 1;
run;

data ds; set immu.immuno_cohort_ds;
if clinicalgrp = "CHB" and immstudy = 1;
run;

proc sort data=ic_one;
by id ;
run;

proc sort data=received;
by id visit;
run;

proc sort data=ds;
by id Visit;
run;

data participants; merge
ic_one (in=a)
received (in=b)
ds (in=c);
by id;
if a=b=c;
run;

proc freq data=participants;
tables inphens*incons visit;
run;

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

proc freq data=participants;
tables inphens;
run;

proc freq data=immu.controls;
tables studygrp_char ;
run;

data controls; set immu.controls;
run;
```

```
*using the participant dataset for replication n=183 and 20 controls and merging the adult cohort data
with the immunology cohort data;
data one; set participants;
run;

data base; set adult.baselinechar;
run;

*sorting adult data by id;
proc sort data=base;
by orig_id;
run;

proc sort data=one;
by orig_id;
run;

*merging the adult cohort data with the participants in the immunology cohort;
data two; merge
one (in=a)
base (in=b);
by orig_id;
if a=b;
run;

*lost 10 participants not in the Adult analytic dataset;
*need to add in the controls;

data three; set two controls;
run;
*n=193;

data four; set three;
keep orig_id id age_cat age_erl con_age con_race con_sex alt tbili tbili_i alt_i fib4 apri hbvdna_i
hbvdna_log_i hbvdna gen_cat gen_comb
bdna_log_cat asian race_new race_oms txstat_erl sex;
run;

*age;
data five; set four;
age = .;
if con_age = . then age = age_erl;
if age_erl = . then age = con_age;
run;

proc means data=five n median q1 q3;
var age;
run;
```

```
*sex;
proc freq data=five;
tables con_sex sex;
run;

data six; set five;
sex_new = .;
if con_sex = "F" then sex_new = 2;
if con_sex = "M" then sex_new = 1;
if sex = 1 then sex_new = 1;
if sex = 2 then sex_new = 2;
run;

proc freq data=six;
tables sex_new;
run;

*race;
proc freq data=six;
tables con_race race_new;
run;

data seven; set six;
if con_race = "W" then race_dsic = "White";
if con_race = "B" then race_dsic = "Black";
if con_race = "C" then race_dsic = "C";
if con_race = "A" then race_dsic = "Asian";
if race_new = 1 then race_dsic = "White";
if race_new = 2 then race_dsic = "Black";
if race_new = 3 then race_dsic = "Asian";
if race_new = 6 then race_dsic = "Mixed";
run;

proc freq data=seven;
tables race_dsic;
run;

*ALT, total bilirubin, fib4, apri, log10 hbv dna;
proc means data=seven n median q1 q3 min max;
var alt alt_i tbili tbili_i fib4 apri hbvdna_log_i;
run;

*HBV genotype;
proc freq data=seven;
tables gen_cat gen_comb/missing;
run;
```

```
*prior treatment;
proc freq data=seven;
tables txstat_erl;
run;
```