Dataset Integrity Check for the Combination Entecavir and Peginterferon Therapy in HBeAg-Positive Immune Tolerant Adults with Chronic Hepatitis B (HBRN Immune Tolerant A)

Prepared by NIDDK-CR
October 20, 2023

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 Combination Entecavir and Peginterferon Therapy in HBeAg-Positive Immune Tolerant Adults With Chronic Hepatitis B (HBRN Immune Tolerant A) study was designed to evaluate the safety and efficacy of a short lead-in course (8 weeks) of entecavir followed by combination of entecavir plus peginterferon alfa-2a for 40 weeks.

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 Immune Tolerant A folder in the data package. For this replication, variables were taken from the “it_long.sas7bdat” dataset.

4 Statistical Methods

Analyses were performed to replicate results for the data in the publication by Feld 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 Immune Tolerant A data files to be distributed are a true copy of the study data.

7 References

[1] Feld JJ, Terrault NA, Lin HS, Belle SH, Chung RT, Tsai N, Khalili M, Perrillo R, Cooper SL, Ghany MG, Janssen HLA, Lok AS. Entecavir and Peginterferon Alfa-2a in Adults With Hepatitis B e Antigen-Positive Immune-Tolerant Chronic Hepatitis B Virus Infection. Hepatology, 69(6), 2338-2348, June 2019. doi: https://doi.org/10.1002/hep.30417

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

Table Variable	**dataset.variable**
Sex	it_long.sex
Race	it_long.race
Age	it_long.age_visit
BMI	it_long.bmi
HBV genotype	it_long.cdcgen_cat
HBeAg	it_long.hbeag
HBV DNA	it_long.hbvdna_c
Quantitative HBeAg	it_long.equant
Quantitative HBsAg	it_long.squant
ALT	it_long.alt
AST	it_long.ast
FIB-4	it_long.fib4
INR	it_long.inr
Total bilirubin	it_long.tbili
Albumin	it_long.alb
Platelets	it_long.plat
White blood cells	it_long.wbc

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

Characteristic	Pub: All (n=28)	DSIC: All (n=28)	Diff. (n=0)
Sex			
Male	15 (54)	15 (54)	0 (0)
Female	13 (46)	13 (46)	0 (0)
Race			
Black/African American	1 (4)	1 (4)	0 (0)
Asian	27 (96)	27 (96)	0 (0)
Age (years)	37.2 (22.2-61.2)	37.2 (22.2-61.2)	0 (0)
BMI (kg/m^2)	22.4 (18.0-28.3)	22.4 (18.0-28.3)	0 (0)
HBV Genotype			
B	9 (32)	9 (32)	0 (0)
C	18 (64)	18 (64)	0 (0)
E	1 (4)	1 (4)	0 (0)
HBeAg (+)	28 (100)	28 (100)	0 (0)
HBV DNA (log_{10} IU/mL) – Screening	8.3 (7.4-8.8)	8.3 (7.4-8.8)	0 (0)
HBV DNA (log_{10} IU/mL) – Baseline	8.2 (7.2-8.8)	8.2 (7.2-8.8)	0 (0)
Quantitative HBeAg (log_{10} IU/mL)	3.3 (2.3-3.6)	3.3 (2.3-3.6)	0 (0)
Quantitative HBsAg (log_{10} IU/mL)	4.7 (4.2-5.1)	4.7 (4.2-5.1)	0 (0)
ALT (U/L): Screening	21 (9-47)	21 (9-47)	0 (0)
Males	32 (16-47)	32 (16-47)	0 (0)
Females	18 (9-26)	18 (9-26)	0 (0)
ALT (U/L): Baseline	21 (9-47)	21 (9-47)	0 (0)
Males	27 (14-47)	27 (14-47)	0 (0)
Females	15 (9-30)	15 (9-30)	0 (0)
AST (U/L)	20 (15-30)	20 (15-30)	0 (0)
FIB-4	0.7 (0.4-2.3)	0.7 (0.4-2.3)	0 (0)
INR	1.0 (0.9-1.1)	1.0 (0.9-1.1)	0 (0)
Total Bilirubin (mg/dL)	0.6 (0.3-2.0)	0.6 (0.3-2.0)	0 (0)
Albumin (g/dL)	4.2 (3.5-4.8)	4.2 (3.5-4.8)	0 (0)
Platelets (x $10^3/mm^3$)	242.5 (137-325)	242.5 (137-325)	0 (0)
White Blood Cells (x $10^3/mm^3$)	5.0 (3.0-9.1)	5.0 (3.0-9.1)	0 (0)

Attachment A: SAS Code

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

/*************************/
/* HBRN Adult IT Study */
/* DSIC Felt et al. */
/*************************/

*enrollment dataset;
data enrol; set dsic.it_enrcrit;
if orig_id="110089UCW" then ITCONS=0;
run;

*longitudinal dataset;
data long; set dsic.it_long;
run;

*sex, using the screening timepoint;
proc freq data=long;
tables sex;
where ittmpt = 1;
run;

*race, using the screening timepoint;
proc freq data=long;
tables race;
where ittmpt = 1;
run;

*age, using the baseline timpoint;
proc means data=long n median min max;
var Age_visit;
where ittmpt = 2;
run;

*BMI, using the baseline timepoint;
proc means data=long n median min max;
var bmi;
where ittmpt = 2;
run;

*HBV genotype;
proc freq data=long;
tables cdcgen_cat;
where ittmpt = 2;
run;
```

```
*hbeag;
proc freq data=long;
tables hbeag;
where ittmpt = 1;
run;

*HBV DNA;
data long_1; set long;
hbvdna_log = log10(hbvdna_c);
run;

proc means data=long_1 n median min max;
var hbvdna_log;
where ittmpt = 2; *change this to 1 for screen, 2 for baseline;
run;

*quant HBeAg;
data long_2; set long_1;
equant_log = log10(equant);
run;

proc means data=long_2 n median min max;
var equant_log;
where ittmpt = 2;
run;

*quant HBsAg;
data long_3; set long_2;
squant_log = log10(squant);
run;

proc means data=long_3 n median min max;
var squant_log;
where ittmpt = 2;
run;

*ALT screening and baseline;
proc sort data=long_3;
by sex;
run;

proc means data=long_3 n median min max;
var alt;
class ittmpt;
by sex;
run;
```

```
*AST, FIB-4, INR, Total Bilirubin, Albumin, Platelets, and WBC: baseline;
proc means data=long_3 n median min max;
var ast Fib4 inr tbili alb plat wbc;
where ittmpt=2;
run;
```