

# Dataset Integrity Check for the Hepatitis B Research Network (HBRN) Adult Cohort Baseline Dataset Files

**Prepared by Allyson Mateja**

**IMS Inc.**

3901 Calverton Blvd, Suite 200 Calverton MD 20705

**October 2015**

# Table of Contents

1 Standard Disclaimer.....	2
2 Study Background.....	2
3 Archived Datasets.....	3
4 Statistical Methods.....	3
5 Results .....	3
6 Conclusions.....	3
7 References .....	3
Attachment A: SAS Code .....	11
<b>Table A:</b> Variables used to replicate Table 1. <u>Demographic and Clinical Characteristics by Race</u> ..	4
<b>Table B:</b> Comparison of values computed in integrity check to reference article Table 1 values ...	5

## 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) consist of a collection of investigators, the NIDDK, and a Data Coordinating Center (DCC) whose goal is to facilitate research of both acute and chronic hepatitis B in adults and children in North America. The hepatitis B virus (HBV) is a leading cause of death throughout the world, despite the availability of a vaccine and therapy. Therefore, the HBRN wishes to examine the epidemiology, natural history, and clinical virology of the hepatitis B virus.

This report examines the baseline characteristics of the adult cohort. Individuals of at least 18 years of age who are hepatitis B surface antigen (HBsAg) positive are enrolled in the HBRN Adult Cohort study. Baseline data is collected on demographics, medical history, family history of liver disease, and health behaviors. Participants are categorized into various phases of HBeAg-positive and HBeAg-negative HBV infection and stages of HBV disease for monitoring of changes in HBV infection status and quantitative HBsAg levels. The rate of various clinical outcomes—including hepatitis exacerbation marked by alanine aminotransferase (ALT) flare, antigen loss of HBsAg or HBeAg, cirrhosis, development of hepatic decompensation, hepatocellular carcinoma (HCC), death, and liver transplantation—and the factors associated with these outcomes are assessed as primary outcome measures, evaluated at up to 288 weeks. Analyses were done to determine whether a baseline HBsAg below 1,000 IU/mL and HBV DNA below 1,000 IU/mL are accurate predictors of people who are, or who will become, inactive carriers (defined as people who are HBsAg positive, HBeAg negative, have normal ALT and HBV DNA under 1,000 IU/mL on at least two occasions over a period of at least 6 months with HBV DNA under 1,000 IU/mL). Additionally, biospecimens are collected from participants to create a repository of resources for future studies.

Currently only data from the HBRN Adult Baseline Visit is available for request.

### **3 Archived Datasets**

All SAS data files, as provided by the Data Coordinating Center (DCC), are located in the HBRN folder “HBRN Adult Cohort Baseline SAS Datasets” in the data package. For this replication, variables were taken from the “hbrn\_bc”, “hbrn\_bi”, “hbrn\_bp”, “hbrn\_ec”, “hbrn\_uwash\_results”, “hbrn\_screen\_info”, “hbrn\_cdc\_results” and “hbrn\_sl” SAS datasets.

### **4 Statistical Methods**

Analyses were performed to duplicate results for the data published by Ghany, et al [1] in Clinical Gastroenterology and Hepatology in July 2014. To verify the integrity of the dataset, descriptive statistics were computed.

### **5 Results**

For Table 1 in the publication [1], Demographic and Clinical Characteristics by Race, Table A lists the variables that can be used in the replication. Table B compares the results calculated from the archived data file to the results published in Table 1. The results of the replication are similar to the results in publication [1].

### **6 Conclusions**

The NIDDK repository is confident that the HBRN Baseline data files to be distributed are a true copy to the manuscript.

### **7 References**

[1] Ghany M, et al.; The Hepatitis B Research Network (HBRN): Characteristics of the Adult Cohort In North America Reflect the Diversity in Country of Origin and Genotype. Clinical Gastroenterology and Hepatology 2014;13(1):183-192.

**Table A:** Variables used to replicate Table 1 Demographic and Clinical Characteristics by Race

Table Variable	dataset.variable
Age	hbrn_screen_info.age
Sex	hbrn_sl.scrsex
Hispanic	hbrn_bp.hisp
Place of Birth	hbrn_bc.cborn
Presumed Mode of Transmission	hbrn_bi.srcpi
Estimated Duration of HBV	hbrn_bi.hbdur (If this variable is missing, and the mode of transmission was vertical, then hbrn_screen_info.age is used instead)
Ever Received HBV Treatment	hbrn_bc.txhbv
Genotype	hbrn_cdc_results.genotype, hbrn_cdc_results.genotype2, hbrn_bc.bgen
HBeAg	hbrn_bc.hbeag within 2 years of enrollment (Compare hbrn_bc.hbeagy and hbrn_bc.hbeagm to hbrn_ec.erldate)
HBV DNA	hbrn_uwash_results.result_n where cltest=2, hbrn_bc.bdna, hbrn_bc.dnall within 1 year of enrollment if hbrn_uwash_results.result_n is missing (Compare hbrn_bc.bdnay and hbrn_bc.bdnam to hbrn_ec.erldate)
ALT Level	hbrn_bc.alt
Albumin Level	hbrn_bc.alb
Total Bilirubin Level	hbrn_bc.tbili
Platelet Level	hbrn_bc.plat
Race	hbrn_bp.racew, hbrn_bp.raceb, hbrn_bp.racea, hbrn_bp.racei, hbrn_bp.raceh, hbrn_bp.raceo

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

	All			White		
	Paper	DSIC	Difference	Paper	DSIC	Difference
Characteristic	N=1625	N=1625	0	N=184	N=184	0
Age	N=1623	N=1623	0	N=184	N=184	0
	42 (33-52)	42 (33-52)	0 (0)	48 (35-58)	48 (34.5-58)	0 (0)
Age group	N=1623	N=1623	0	N=184	N=184	0
18 to < 30	264 (16.3)	264 (16.3)	0	27 (14.7)	27 (14.7)	0 (0)
30 to <40	452 (27.8)	452 (27.9)	0 (0.1)	37 (20.1)	37 (20.1)	0 (0)
40 to <50	405 (25)	405 (25)	0 (0)	37 (20.1)	37 (20.1)	0 (0)
50 to <60	324 (20)	324 (20)	0 (0)	48 (26.1)	48 (26.1)	0 (0)
> 60	178 (11)	178 (11)	0 (0)	35 (19)	35 (19)	0 (0)
Sex	N=1625	N=1625	0	N=184	N=184	0
Male	829 (51)	829 (51)	0 (0)	111 (60.3)	111 (60.3)	0 (0)
Hispanic	N=1612	N=1612	0	N=183	N=183	0
Yes	18 (1.1)	18 (1.1)	0 (0)	10 (5.5)	10 (5.5)	0 (0)
Place of Birth	N=1615	N=1615	0	N=184	N=184	0
US/Canada	294 (18.2)	294 (18.2)	0 (0)	118 (64.1)	118 (64.1)	0 (0)
Other North America or South America	27 (1.7)	27 (1.7)	0 (0)	10 (5.4)	10 (5.4)	0 (0)
Europe	47 (2.9)	47 (2.9)	0 (0)	44 (23.9)	44 (23.9)	0 (0)
Asia/Australia	1077 (66.7)	1077 (66.7)	0 (0)	10 (5.4)	10 (5.4)	0 (0)
Africa	170 (10.5)	170 (10.5)	0 (0)	2 (1.1)	2 (1.1)	0 (0)
Presumed Mode of Transmission	N=1210	N=1211	1	N=120	N=120	0
Vertical	724 (59.8)	724 (59.8)	0 (0)	27 (22.5)	27 (22.5)	0 (0)
Horizontal/Household	325 (26.9)	322 (26.6)	3 (0.3)	20 (16.7)	20 (16.7)	0 (0)
Sexual	73 (6)	73 (6)	0 (0)	33 (27.5)	33 (27.5)	0 (0)
Medical exposures	62 (5.1)	60 (5)	2 (0.1)	25 (20.8)	25 (20.8)	0 (0)
Drug Use	17 (1.4)	17 (1.4)	0 (0)	13 (10.8)	13 (10.8)	0 (0)
Other	9 (0.7)	15 (1.2)	6 (0.5)	2 (1.7)	2 (1.7)	0 (0)
Estimated Duration of HBV	N=1168	N=1167	1	N=101	N=101	0
	31 (20-43)	31 (20-42)	0 (0-1)	25 (13-35)	25 (13-35)	0 (0)
Ever Received HBV Treatment	N=1623	N=1623	0	N=184	N=184	0
Yes	235 (14.5)	235 (14.5)	0 (0)	24 (13)	24 (13)	0 (0)
Genotype	N=1379	N=1379	0	N=152	N=152	0
A	247 (17.9)	247 (17.9)	0 (0)	84 (55.3)	84 (55.3)	0 (0)
B	531 (38.5)	531 (38.5)	0 (0)	7 (4.6)	7 (4.6)	0 (0)
C	450 (32.6)	450 (32.6)	0 (0)	4 (2.6)	4 (2.6)	0 (0)
D	106 (7.7)	106 (7.7)	0 (0)	50 (32.9)	50 (32.9)	2 (0)
E	35 (2.5)	35 (2.5)	0 (0)	2 (1.3)	2 (1.3)	0 (0)
F/G/H/multiple	10 (0.7)	10 (0.7)	0 (0)	5 (3.3)	5 (3.3)	0 (0)

HBeAg	N=1386	N=1390	4	N=156	N=158	2
Positive	365 (26.3)	366 (26.3)	1 (0)	31 (19.9)	31 (19.6)	0 (0.3)
HBV DNA level	N=1592	N=1593	1	N=178	N=178	0
	3.6 (2.5-5.6)	3.6 (2.5-5.6)	0 (0)	3.0 (2.0-4.2)	3.0 (2.1-4.3)	0 (0.1-0.1)
ALT level, men	N=800	N=796	4	N=109	N=109	0
	38 (27-61)	38 (27-60)	0 (0-1)	46 (30-68)	46 (30-68)	0 (0)
ALT level, women	N=764	N=761	3	N=71	N=71	0
	26 (19-39)	26 (19-39)	0 (0)	29 (22-44)	29 (22-44)	0 (0)
Albumin level	N=1516	N=1516	0	N=175	N=175	0
	4.3 (4.1-4.6)	4.3 (4.1-4.6)	0 (0)	4.3 (4.1-4.6)	4.3 (4.1-4.6)	0 (0)
Total bilirubin level	N=1544	N=1541	3	N=179	N=178	1
	0.6 (0.4-0.8)	0.6 (0.4-0.8)	0 (0)	0.6 (0.4-0.8)	0.6 (0.4-0.8)	0 (0)
Platelet level	N=1389	N=1387	2	N=167	N=167	0
	217 (179-256)	217 (179-256)	0 (0)	208 (170-257)	208 (170-257)	0 (0)

	Black			Asian		
	Paper	DSIC	Difference	Paper	DSIC	Difference
Characteristic	N=237	N=234	3	N=1158	N=1158	0
Age	N=237	N=234	3	N=1157	N=1157	0
	43 (34-54)	43 (34-54)	0 (0)	41 (32-51)	41 (32-51)	0 (0)
Age group	N=237	N=234	3	N=1157	N=1157	0
18 to < 30	31 (13.1)	31 (13.3)	0 (0.2)	194 (16.8)	194 (16.8)	0 (0)
30 to <40	64 (27)	63 (26.9)	1 (0.1)	341 (29.5)	340 (29.4)	1 (0.1)
40 to <50	67 (28.3)	66 (28.2)	1 (0.1)	287 (24.8)	287 (24.8)	0 (0)
50 to <60	44 (18.6)	44 (18.8)	0 (0.2)	227 (19.6)	227 (19.6)	0 (0)
> 60	31 (13.1)	30 (12.8)	1 (0.3)	108 (9.3)	109 (9.4)	1 (0.1)
Sex	N=237	N=234	3	N=1158	N=1157	1
Male	132 (55.7)	131 (56)	1 (0.3)	569 (49.1)	568 (49.1)	1 (0)
Hispanic	N=235	N=234	1	N=1150	N=1154	4
Yes	1 (0.4)	1 (0.4)	0 (0)	4 (0.3)	4 (0.4)	0 (0.1)
Place of Birth	N=237	N=234	3	N=1150	N=1150	0
US/Canada	57 (24.1)	57 (24.4)	0 (0.3)	97 (8.4)	97 (8.4)	0 (0)
Other North America or South America	11 (4.6)	11 (4.7)	0 (0.1)	3 (0.3)	3 (0.3)	0 (0)
Europe	1 (0.4)	1 (0.4)	0 (0)	0 (0.0)	1 (0.1)	1 (0.1)
Asia/Australia	2 (0.8)	2 (0.9)	0 (0.1)	1048 (91.1)	1047 (91)	1 (0.1)
Africa	166 (70)	163 (69.7)	3 (0.1)	2 (0.2)	2 (0.2)	0 (0)
Presumed Mode of Transmission	N=179	N=177	2	N=875	N=875	0
Vertical	40 (22.3)	40 (22.6)	0 (0.3)	626 (71.5)	625 (71.4)	1 (0.1)
Horizontal/Household	97 (54.2)	95 (53.7)	2 (0.5)	205 (23.4)	202 (23.1)	3 (0.3)
Sexual	24 (13.4)	24 (13.6)	0 (0.2)	15 (1.7)	15 (1.7)	0 (0)
Medical exposures	15 (8.4)	15 (8.5)	0 (0.1)	22 (2.5)	20 (2.3)	2 (0.2)
Drug Use	2 (1.1)	2 (1.1)	0 (0)	1 (0.1)	1 (0.1)	0 (0)
Other	1 (0.6)	1 (0.6)	0 (0)	6 (0.7)	12 (1.4)	6 (0.7)
Estimated Duration of HBV	N=154	N=152	2	N=875	N=876	1
	15 (6-32)	15 (6 – 31.5)	0 (0 – 0.5)	34 (23-45)	34 (23 – 45)	0 (0)
Ever Received HBV Treatment	N=237	N=234	3	N=1157	N=1157	0
Yes	20 (8.4)	20 (8.6)	0 (0.2)	183 (15.8)	181 (15.6)	2 (0.2)
Genotype	N=162	N=160	2	N=1023	N=1020	3
A	105 (64.8)	104 (65)	1 (0.2)	50 (4.9)	50 (4.9)	0 (0)
B	7 (4.3)	7 (4.4)	0 (0.1)	510 (49.9)	510 (50)	0 (0.1)
C	3 (1.9)	3 (1.9)	0 (0)	420 (41.1)	416 (40.8)	4 (0.3)
D	12 (7.4)	12 (7.5)	0 (0.1)	40 (3.9)	41 (4)	1 (0.1)
E	31 (19.1)	30 (18.8)	1 (0.3)	2 (0.2)	2 (0.2)	0 (0)
F/G/H/multiple	4 (2.5)	4 (2.5)	0 (0)	1 (0.1)	1 (0.1)	0 (0)
HBeAg	N=189	N=187	2	N=1003	N=1006	3

Positive	19 (10.1)	20 (10.7)	1 (0.6)	305 (30.4)	305 (30.3)	0 (0.1)
HBV DNA level	N=234	N=231	3	N=1136	N=1137	1
	3.0 (2.0-3.9)	3.0 (2.0-3.9)	0 (0)	4.0 (2.7-6.1)	4.0 (2.7-6.1)	0 (0)
ALT level, men	N=127	N=126	1	N=548	N=544	4
	36 (26-63)	36.5 (26-63)	0.5 (0)	37 (26-58)	37 (26-57.5)	0 (0-0.5)
ALT level, women	N=101	N=99	2	N=564	N=562	2
	20 (16-27)	20 (16-27)	0 (0)	27 (19-40)	27 (19-39)	0 (0-1)
Albumin level	N=219	N=217	2	N=1079	N=1079	0
	4.2 (3.9-4.4)	4.2 (3.9-4.4)	0 (0)	4.4 (4.1-4.6)	4.4 (4.1-4.6)	0 (0)
Total bilirubin level	N=225	N=222	3	N=1096	N=1094	2
	0.5 (0.4-0.7)	0.5 (0.4-0.7)	0 (0)	0.6 (0.5-0.9)	0.6 (0.5-0.9)	0 (0)
Platelet level	N=209	N=207	2	N=974	N=971	3
	208 (170-263)	208 (170-264)	0 (0-1)	219 (184-255)	219 (185-255)	0 (1-0)

	Mixed/other		
	Paper	DSIC	Difference
Characteristic	N=38	N=38	0
Age	N=38	N=38	0
	39 (28-47)	39 (28-47)	0 (0)
Age group	N=38	N=38	0
18 to < 30	11 (28.9)	11 (29)	0 (0.1)
30 to <40	8 (21.1)	8 (21.1)	0 (0)
40 to <50	14 (36.8)	14 (36.8)	0 (0)
50 to <60	3 (7.9)	3 (7.9)	0 (0)
> 60	2 (5.3)	2 (5.3)	0 (0)
Sex	N=38	N=38	0
Male	13 (34.2)	13 (34.2)	0 (0)
Hispanic	N=38	N=38	0
Yes	1 (2.6)	1 (2.6)	0 (0)
Place of Birth	N=37	N=37	0
US/Canada	22 (59.5)	22 (59.5)	0 (0)
Other North America or South America	1 (2.7)	1 (2.7)	0 (0)
Europe	1 (2.7)	1 (2.7)	0 (0)
Asia/Australia	13 (35.1)	13 (35.1)	0 (0)
Africa	0 (0.0)	0 (0)	0 (0)
Presumed Mode of Transmission	N=33	N=33	0
Vertical	29 (87.9)	29 (87.9)	0 (0)
Horizontal/Household	2 (6.1)	2 (6.1)	0 (0)
Sexual	1 (3)	1 (3)	0 (0)
Medical exposures	0 (0.0)	0 (0)	0 (0)
Drug Use	1 (3)	1 (3)	0 (0)
Other	0 (0.0)	0 (0)	0 (0)
Estimated Duration of HBV	N=32	N=32	0
	30 (21-42)	29.5 (20.5 – 41.5)	0.5 (0.5 – 0.5)
Ever Received HBV Treatment	N=38	N=38	0
Yes	8 (21.1)	8 (21.1)	0 (0)
Genotype	N=35	N=35	0
A	6 (17.1)	6 (17.1)	0 (0)
B	6 (17.1)	6 (17.1)	0 (0)
C	23 (65.7)	23 (65.7)	0 (0)
D	0 (0.0)	0 (0)	0 (0)
E	0 (0.0)	0 (0)	0 (0)
F/G/H/multiple	0 (0.0)	0 (0)	0 (0)
HBeAg	N=32	N=32	0

Positive	9 (28.1)	9 (28.1)	0 (0)
HBV DNA level	N=38	N=38	0
	3.4 (2.5-4.5)	3.4 (2.5-4.5)	0 (0)
ALT level, men	N=13	N=13	0
	46 (33-91)	46 (33-91)	0 (0)
ALT level, women	N=24	N=24	0
	27 (16-46)	26.5 (15.5-46)	0.5 (0.5-0)
Albumin level	N=36	N=36	0
	4.3 (4.0-4.6)	4.3 (4-4.6)	0 (0)
Total bilirubin level	N=37	N=37	0
	0.5 (0.4-0.7)	0.5 (0.4-0.7)	0 (0)
Platelet level	N=34	N=34	0
	216 (186-263)	215.5 (186-263)	0.5 (0)

```

*****
**;
* Program /prj/niddk/ims_analysis/HBRN/prog_initial_analysis/dsic_hbrn_baseline.sas
* by Allyson Mateja
* modified by
* Date: August 4, 2015
*
* Dataset Integrity Check for "The Hepatitis B Research Network (HBRN): Characteristics of the Adult Cohort in North
* America Reflect the Diversity in Country of Origin and Genotype" (HBRN Cohort Description)
* Ghaney et al.
*
* Modified: October 22, 2015
*****
**;
options nofmterr;

title 'HBRN Baseline DSIC';
title2 ' ';

proc format;
    value targf 1='Consecutive'
                2='Targeted'
                3='Targeted pediatric transfer';

    value agesf 18 -< 30 = '18 to <30'
                30 -< 40 = '30 to <40'
                40 -< 50 = '40 to < 50'
                50 -< 60 = '50 to < 60'
                60 - HIGH = '>60';

    value genderf 1 = 'Male'
                  2 = 'Female';

    value country 1-2 = 'United States/Canada'
                  13,33,49,57,90,94,95,98,109,142,148,170,178,223,237 = 'Other North America and South America'
                  5,30,36,59,69,75,82,85,108,176,181,182,196,200,230,232='Europe'
    3,21,28,39,46,99,102,103,104,110,113,116,117,120,122,133,143,146,151,154,167,174,199,215,218,222,225,235,245,247='Asia/Australia'
    9,37,38,40,41,43,52,55,68,70,80,83,114,124,139,140,161,183,195,203,204,208,217,220,224,241,244='Africa';

    value sourcef 1 = "Vertical"
                  2,3 = "Horizontal/household"
                  4 = "Sexual"
                  5,6,7 = "Medical exposures"
                  8,9 = "Drug use"
                  10="Other";

    value genf 1="A1"
               2="A2"
               3="A3"
               4="A"

```

```

5="B1"
6="B2"
7="B3"
8="B4"
9="B5"
10="B6"
11="B"
12="C1"
13="C2"
14="C3"
15="C"
16="D1"
17="D2"
18="D3"
19="D4"
20="D5"
21="D6"
22="D7"
23="D"
24="E"
25="F1"
26="F2"
27="F3"
28="F4"
29="F"
30="G"
31="H"
32="B/C"

.A=".A"
.C=".C"
.E=".E"
.T=".T";

```

```
libname sas_file '/prj/niddk/ims_analysis/HBRN/private_orig_data/HBRN_Adult_Baseline/HBRN Adult Cohort Baseline SAS Datasets/';
```

```

data screen;
    set sas_file.hbrn_screen_info;

data sl;
    set sas_file.hbrn_sl;
    if consdate ne '.B' then output sl;

data labs;
    set sas_file.hbrn_bc;

data bi;
    set sas_file.hbrn_bi;

data race;
    set sas_file.hbrn_bp;

```

```

data cdc;
    set sas_file.hbrn_cdc_results;

data ec;
    set sas_file.hbrn_ec;

data uwash;
    set sas_file.hbrn_uwash_results;

proc sort data=uwash;
    by id;

proc contents data=screen;

proc contents data=labs;

proc contents data=race;

proc freq data=race;
    tables racew raceb racea raceo;

proc freq data=screen;
    tables scrpreg*antiv enroll notelig screlig cons hcc hdc hiv livtx site_trans*cohortprot;
    format cohortprot targf.;

proc sort data=screen;
    by id;

proc sort data=sl;
    by id;

data enrolled;
    set screen;
    by id;
    if id ne '!B!' then output enrolled;

proc freq data=enrolled;
    tables cons hcc hdc hiv livtx antiv*scrpreg site_trans*cohortprot /list;

data eligible;
    set enrolled;
    by id;
    if cohortprot=1 and site_trans=0 then do;
        if last.id and notelig=0 and enroll ne 0 then output eligible;
    end;

data eligible_v2;
    merge eligible (in=val1)

```

```

        bi          (in=val2);
by id;
if id = '050015YAZ' then do;
    hbdurt=1;
    hbdur=30;
end;
if id = '050086TXR' then do;
    hbdurt=1;
    hbdur=16;
end;
if id = '050087KCR' then do;
    hbdurt=1;
    hbdur=40;
end;
if id = '170002BHS' then do;
    hbdurt=1;
    hbdur=20;
end;
if id = '220080YBU' then do;
    hbdurt = .;
    hbdur = .;
end;
if hbdurt=2 then hbdur = hbdur/12;
if vall and hbphy ne 1 then output eligible_v2;

data all_baseline_participants;
merge eligible_v2 (in=vall)
    race
    sl (keep=id scrsex sracea sraceb sracew sraceh sracei sraceo sraceu sraceos scrhbe scrhbv scrhbvu)
    labs
    cdc
    ec (keep = id erldate);
by id;
if vall and last.id then output all_baseline_participants;

data all_baseline_participants;
set all_baseline_participants;
if hbdur = '.C' and srcpi = 1 then hbdur = age;
hbeag_length = abs(year(erldate) - hbeagy);
if hbeag_length = 2 then do;
    if hbeagy > year(erldate) then do;
        if month(erldate) < hbeagm then hbeag_length = hbeag_length +1;
    end;
    else if hbeagy < year(erldate) then do;
        if month(erldate) > hbeagm then hbeag_length = hbeag_length +1;
    end;
end;

data all_baseline_participants;
set all_baseline_participants;
if year(erldate) = 2011 then erlyr = 11;

```

```

else if year(erldate) = 2012 then erlyr = 12;
else if year(erldate) = 2013 then erlyr = 13;
bdna_length = abs(erlyr - bdnay);
if bdna_length = 1 then do;
    if bdnay > year(erldate) then do;
        if month(erldate) < bdnam then bdna_length = hbeag_length +1;
    end;
    else if bdnay < year(erldate) then do;
        if month(erldate) > bdnam then bdna_length = bdna_length +1;
    end;
end;

end;

proc means data=all_baseline_participants n median q1 q3;
var age;
title3 'All Baseline Participants-age';

proc freq data=all_baseline_participants;
tables age;
format age agesf.;
title3 'All Baseline Participants-Age Range';

proc freq data=all_baseline_participants;
tables scrsex;
format scrsex genderf.;
title3 'All Baseline Participants-Gender';

proc freq data=all_baseline_participants;
tables hisp;
title3 'All Baseline Participants-Hispanic';

proc freq data=all_baseline_participants;
tables cborn;
format cborn country.;
title3 'All Baseline Participants- Place of Birth';

proc freq data=all_baseline_participants;
tables srcpi srcpis;
format srcpi sourcef.;
title3 'All Baseline Participants - Presumed Mode of Transmission';

proc means data=all_baseline_participants n median q1 q3;
var hbdur;
title3 'All Baseline Participants - Estimated Duration of HBV';

proc freq data=all_baseline_participants;
tables txhvb;
title3 'All Baseline Participants - Ever recieved HBV treatment';

proc freq data=all_baseline_participants;
tables bgen genotype genotype2 /list missing;
format bgen genf.;

```

```

        title3 'All Baseline Participants - Genotype';

proc freq data=all_baseline_participants;
    tables genotype*genotype2*bgen /list missing;
    format bgen genf.;
    where genotype in ('ND', '');
    title3 'All Baseline Participants - Genotype';

proc freq data=all_baseline_participants;
    tables hbeag;
    where hbeag_length <= 2;
    title3 'All Baseline Participants - HBeAg';

proc sort data=all_baseline_participants;
    by id;

data hbv_dna;
    set uwash;
    by id;
    if cltest=2 then do;
        if cohort_tmpt = 10 then output hbv_dna;
    end;

data hbv_dna;
    set hbv_dna;
    by id;
    if result_n = '.E' then result_n =0;
    if first.id and last.id then output hbv_dna;

data all_baseline_participants_uwash;
    merge all_baseline_participants (in=val1)
          hbv_dna (in=val2 keep=id result_n cltest cohort_tmpt);
    by id;
    if result_n in (0, .) then do;
        if bunit =1 then do;
            if bdna_length <= 1 then do;
                if bdna = '.F' then bdna = bdnall;
                if bdna = '.G' then bdna = 5000000000;
                result_n=round(log10(bdna), .1);
            end;
        end;
    end;
    if val1 then output all_baseline_participants_uwash;

proc means data=all_baseline_participants_uwash n median q1 q3;
    var result_n ;
    title3 'All Baseline Participants - HBV DNA level';

proc means data=all_baseline_participants n median q1 q3;
    var plat tbili alb;
    title3 'All Baseline Participants - Albumin Level (g/dL), Total bilirubin level (mg/dL), Platelet level (10^3/mm^3)';

```

```

proc sort data=all_baseline_participants;
  by scrsex;

proc means data=all_baseline_participants n median q1 q3;
  var alt;
  by scrsex;
  format scrsex genderf.;
  title3 'All Baseline Participants - ALT Level, U/L by Gender';

proc freq data=all_baseline_participants;
  tables racea*raceb*racew*raceh*racei*raceo*racer raceos /list;
  title3 'Race';

data whites;
  set all_baseline_participants;
  if racew=1 and raceb=0 and racea=0 and raceo=0 and racei=0 and raceh=0 and racer=0 then output whites;

proc means data=whites n median q1 q3;
  var age;
  title3 'Whites - Age';

proc freq data=whites;
  tables age;
  format age agesf.;
  title3 'Whities - Age Range';

proc freq data=whites;
  tables scrsex;
  format scrsex genderf.;
  title3 'Whites - Gender';

proc freq data=whites;
  tables hisp;
  title3 'Whites - Hispanic';

proc freq data=whites;
  tables cborn;
  format cborn country.;
  title3 'Whites - Place of Birth';

proc freq data=whites;
  tables srcpi;
  format srcpi sourcef.;
  title3 'Whites - Presumed Mode of Transmission';

proc means data=whites n median q1 q3;
  var hbdur;
  title3 'Whites - Estimated Duration of HBV';

proc freq data=whites;

```

```

        tables txhbv;
        title3 'Whites - Ever recieved HBV treatment';

proc freq data=whites;
    tables bgen genotype genotype2 /list missing;
    format bgen genf.;
    title3 'Whites - Genotype';

proc freq data=whites;
    tables genotype*genotype2*bgen /list missing;
    format bgen genf.;
    where genotype in ('ND', '');
    title3 'Whites - Genotype';

proc freq data=whites;
    tables hbeag;
    where hbeag_length <= 2;
    title3 'Whites - HBeAg';

proc sort data=whites;
    by id;

data whites_uwash;
    merge whites (in=val1)
           all_baseline_participants_uwash (in=val2 keep = id result_n);
    by id;
    if val1 and val2 then output whites_uwash;

proc means data=whites_uwash n median q1 q3;
    var result_n ;
    title3 'Whites - HBV DNA level';

proc means data=whites n median q1 q3;
    var plat tbili alb;
    title3 'Whites - Albumin Level (g/dL), Total bilirubin level (mg/dL), Platelet level (10^3/mm^3)';

proc sort data=whites;
    by scrsex;

proc means data=whites n median q1 q3;
    var alt;
    by scrsex;
    format scrsex genderf.;
    title3 'Whites - ALT Level, U/L by Gender';

data blacks;
    set all_baseline_participants;
    if racew=0 and raceb=1 and racea=0 and raceo=0 and racei=0 and raceh=0 and racer=0 then output blacks;

```

```

proc means data=blacks n median q1 q3;
    var age;
    title3 'Blacks - Age';

proc freq data=blacks;
    tables age;
    format age agesf.;
    title3 'Blacks - Age Range';

proc freq data=blacks;
    tables scrsex;
    format scrsex genderf.;
    title3 'Blacks - Gender';

proc freq data=blacks;
    tables hisp;
    title3 'Blacks - Hispanic';

proc freq data=blacks;
    tables cborn;
    format cborn country.;
    title3 'Blacks - Place of Birth';

proc freq data=blacks;
    tables srcpi;
    format srcpi sourcef.;
    title3 'Blacks - Presumed Mode of Transmission';

proc means data=blacks n median q1 q3;
    var hbdur;
    title3 'Blacks - Estimated Duration of HBV';

proc freq data=blacks;
    tables txhvb;
    title3 'Blacks - Ever recieved HBV treatment';

proc freq data=blacks;
    tables bgen genotype genotype2 /list missing;
    format bgen genf.;
    title3 'Blacks - Genotype';

proc freq data=blacks;
    tables genotype*genotype2*bgen /list missing;
    format bgen genf.;
    where genotype in ('ND', '');
    title3 'Blacks - Genotype';

proc freq data=blacks;
    tables hbeag;
    where hbeag_length <= 2;
    title3 'Blacks - HBeAg';

```

```

proc sort data=blacks;
  by id;

data blacks_uwash;
  merge blacks (in=val1)
        all_baseline_participants_uwash (in=val2 keep = id result_n);
  by id;
  if val1 and val2 then output blacks_uwash;

proc means data=blacks_uwash n median q1 q3;
  var result_n ;
  title3 'Blacks - HBV DNA level';

proc means data=blacks n median q1 q3;
  var plat tbili alb;
  title3 'Blacks - Albumin Level (g/dL), Total bilirubin level (mg/dL), Platelet level (10^3/mm^3)';

proc sort data=blacks;
  by scrsex;

proc means data=blacks n median q1 q3;
  var alt;
  by scrsex;
  format scrsex genderf.;
  title3 'Blacks - ALT Level, U/L by Gender';

data asians;
  set all_baseline_participants;
  if (racew=0 and raceb=0 and racea=1 and raceo=0 and racei=0 and raceh=0 and racer=0) or
     (racew=0 and raceb=0 and racea=0 and raceo=1 and racei=0 and raceh=0 and racer=0)then output asians;

proc means data=asians n median q1 q3;
  var age;
  title3 'Asians - Age';

proc freq data=asians;
  tables age;
  format age agesf.;
  title3 'Asians - Age Range';

proc freq data=asians;
  tables scrsex;
  format scrsex genderf.;
  title3 'Asians - Gender';

proc freq data=asians;
  tables hisp;
  title3 'Asians - Hispanic';

```

```

proc freq data=asians;
  tables cborn;
  format cborn country.;
  title3 'Asians - Place of Birth';

proc freq data=asians;
  tables srcpi;
  format srcpi sourcef.;
  title3 'Asians - Presumed Mode of Transmission';

proc means data=asians n median q1 q3;
  var hbdur;
  title3 'Asians - Estimated Duration of HBV';

proc freq data=asians;
  tables txhbv;
  title3 'Asians - Ever recieved HBV treatment';

proc freq data=asians;
  tables bgen genotype genotype2 /list missing;
  format bgen genf.;
  title3 'Asians - Genotype';

proc freq data=asians;
  tables genotype*genotype2*bgen /list missing;
  format bgen genf.;
  where genotype in ('ND', '');
  title3 'Asians - Genotype';

proc freq data=asians;
  tables hbeag;
  where hbeag_length <= 2;
  title3 'Asians - HBeAg';

proc sort data=asians;
  by id;

data asians_uwash;
  merge asians (in=val1)
        all_baseline_participants_uwash (in=val2 keep = id result_n);
  by id;
  if val1 and val2 then output asians_uwash;

proc means data=asians_uwash n median q1 q3;
  var result_n ;
  title3 'Asians - HBV DNA level';

proc means data=asians n median q1 q3;
  var plat tbili alb;
  title3 'Asians - Albumin Level (g/dL), Total bilirubin level (mg/dL), Platelet level (10^3/mm^3)';

```

```

proc sort data=asians;
  by scrsex;

proc means data=asians n median q1 q3;
  var alt;
  by scrsex;
  format scrsex genderf.;
  title3 'Asians - ALT Level, U/L by Gender';

data mixed_other;
  set all_baseline_participants;
  if (racew=0 and raceb=1 and racea=1 and raceo=0 and racei=0 and raceh=0 and racer=0) or
    (racew=1 and raceb=0 and racea=1 and raceo=0 and racei=1 and raceh=0 and racer=0) or
    (racew=1 and raceb=0 and racea=1 and raceo=0 and racei=0 and raceh=0 and racer=0) or
    (racew=0 and raceb=0 and racea=1 and raceo=0 and racei=0 and raceh=1 and racer=0) or
    (racew=1 and raceb=0 and racea=0 and raceo=0 and racei=1 and raceh=0 and racer=0) or
    (racew=0 and raceb=0 and racea=0 and raceo=0 and racei=1 and raceh=0 and racer=0) or
    (racew=0 and raceb=0 and racea=0 and raceo=0 and racei=0 and raceh=1 and racer=0) then output mixed_other;

proc means data=mixed_other n median q1 q3;
  var age;
  title3 'Mixed/Other - Age';

proc freq data=mixed_other;
  tables age;
  format age agesf.;
  title3 'Mixed/Other - Age Range';

proc freq data=mixed_other;
  tables scrsex;
  format scrsex genderf.;
  title3 'Mixed/Other - Gender';

proc freq data=mixed_other;
  tables hisp;
  title3 'Mixed/Other - Hispanic';

proc freq data=mixed_other;
  tables cborn;
  format cborn country.;
  title3 'Mixed/Other - Place of Birth';

proc freq data=mixed_other;
  tables srcpi;
  format srcpi sourcef.;
  title3 'Mixed/Other - Presumed Mode of Transmission';

proc means data=mixed_other n median q1 q3;
  var hbdur;
  title3 'Mixed/Other - Estimated Duration of HBV';

```

```

proc freq data=mixed_other;
  tables txhbv;
  title3 'Mixed/Other - Ever recieved HBV treatment';

proc freq data=mixed_other;
  tables bgen genotype genotype2 /list missing;
  format bgen genf.;
  title3 'Mixed/Other - Genotype';

proc freq data=mixed_other;
  tables genotype*genotype2*bgen /list missing;
  format bgen genf.;
  where genotype in ('ND', '');
  title3 'Mixed/Other - Genotype';

proc freq data=mixed_other;
  tables hbeag;
  where hbeag_length <= 2;
  title3 'Mixed/Other - HBeAg';

proc sort data=mixed_other;
  by id;

data mixed_other_uwash;
  merge mixed_other (in=val1)
        all_baseline_participants_uwash (in=val2 keep = id result_n);
  by id;
  if val1 and val2 then output mixed_other_uwash;

proc means data=mixed_other_uwash n median q1 q3;
  var result_n ;
  title3 'Mixed/Other - HBV DNA level';

proc means data=mixed_other n median q1 q3;
  var plat tbili alb;
  title3 'Mixed/Other - Albumin Level (g/dL), Total bilirubin level (mg/dL), Platelet level (10^3/mm^3)';

proc sort data=mixed_other;
  by scrsex;

proc means data=mixed_other n median q1 q3;
  var alt;
  by scrsex;
  format scrsex genderf.;
  title3 'Mixed/Other - ALT Level, U/L by Gender';

```