Dataset Integrity Check for the Hepatitis B Research Network (HBRN) Pediatric Baseline Dataset Files

Prepared by Allyson Mateja
IMS, Inc.
3901 Calverton Blvd, Suite 200, Calverton, MD 20705
June 21, 2016

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) 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 pediatric cohort. Individuals of between 6 months and 18 years of age who are hepatitis B surface antigen (HBsAg) positive are enrolled in the HBRN Pediatric 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 hepative decompensation, heptaocellular 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 Pediatric 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 Data folder in the data package. For this replication, variables were taken from the “hbrn_bcp”, “hbrn_bip”, “hbrn_ecp”, “hbrn_uwash_results”, “hbrn_slp”, “hbrn_cdc_results” and “hbrn_peds_ages” SAS datasets.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Schwarz, et al [1] in the Journal of Pediatrics in December 2015. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For Table 1 in the publication [1], Baseline characteristics by HBeAg status, Table A lists the variables that can be used in the replication, and Table B compares the results calculated from the archived data file to the results published in Table 1. The results of the replication are a close match to the results in publication [1].

For Table 2 in the publication [1], Baseline characteristics by genotype, Table C lists the variables that can be used in the replication, and Table D compares the results calculated from the archived data file to the results published in Table 2. The results of the replication are a close match to the results in publication [1].

For the Supplemental Table in the publication [1], Baseline characteristics across family composition, Table E lists the variables that can be used in the replication, and Table F compares the results calculated from the archived data file to the results published in the Supplemental Table. The results of the replication are a close match to the results in publication [1].

6 Conclusions

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

7 References

[1] Schwarz, K.B., Cloonan, Y.K., Ling, S.C., Murray, K.F., Rodriguez-Baez, N., Schwarzenberg, S.J., Teckman, J., Raeva-Ganova, L., Rosenthal, P., and the Hepatitis B Research Network. Children with Chronic Hepatitis B in the US and Canada in the Hepatitis B Research Network. Journal of Pediatrics 2015; 167(6): 1287-1294.

Table A: Variables used to replicate Table 1: Baseline characteristics by HBeAg status

Table Variable	**dataset.variable**
HBeAg	hbrn_bcp.hbeag, hbrn_bcp.hbeagy, hbrn_bcp.hbeagm, hbrn_ecp.erldate
Age	hbrn_peds_ages.age
Sex	hbrn_slp.scrsex
Race	hbrn_slp.sracew, hbrn_slp.sraceb, hbrn_slp.sracea, hbrn_slp.sracei, hbrn_slp.sraceh, hbrn_slp.sraceo
Adopted	hbrn_bcp.adopt
Ever Received HBV Treatment	hbrn_bcp.txhbv, hbrn_bcp.txb1
Vertical Transmission	hbrn_bip.srcpi
HBV DNA	hbrn_uwash_results.result_n (where hbrn_uwash_results.cltest=1 and hbrn_uwash_results.cohort_tmpt=10), hbrn_bcp.bunit, hbrn_bcp.bdna, hbrn_bcp.bdnall
ALT	hbrn_bcp.alt
ALT x ULN	hbrn_bcp.alt, hbrn_peds_ages.age, hbrn_slp.scrsex
Platelets	hbrn_bcp.plat

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

Characteristic	All Manuscript N=343	All DSIC N=343	Difference N=0	HBeAg(-) Manuscript N=81 (26%)	HBeAg(-) DSIC N=81 (26%)	Difference N=0 (0)
Age (yrs)	N=343	N=343	0	N=81	N=81	0
Mean (SD)	10.4 (4.5)	10.4 (4.5)	0 (0)	12.4 (3.7)	12.4 (3.7)	0 (0)
Sex	N=343	N=343	0	N=81	N=81	0
Female	210 (61%)	210 (61%)	0 (0)	43 (53%)	43 (53%)	0 (0)
Race	N=341	N=341	0	N=80	N=80	0
White	29 (9%)	29 (9%)	0 (0)	19 (24%)	19 (24%)	0 (0)
Black	36 (11%)	36 (11%)	0 (0)	12 (15%)	12 (15%)	0 (0)
Asian	265 (78%)	265 (78%)	0 (0)	47 (59%)	47 (59%)	0 (0)
Mixed/Other	11 (3%)	11 (3%)	0 (0)	2 (3%)	2 (3%)	0 (0)
Adopted	N=343	N=343	0	N=81	N=81	0
	188 (55%)	188 (55%)	0 (0)	40 (49%)	40 (49%)	0 (0)
Ever Received HBV Treatment	N=343	N=343	0	N=81	N=81	0
	43 (13%)	43 (13%)	0 (0)	23 (28%)	23 (28%)	0 (0)
Vertical Transmission	N=277	N=277	0	N=55	N=55	0
	269 (97%)	269 (97%)	0 (0)	53 (96%)	53 (96%)	0 (0)
HBV DNA (log_{10} IU/mL)	N=326	N=326	0	N=78	N=77	1
Median (25:75)	8.1 (4.6:8.6)	8.1 (4.9:8.6)	0 (0.3:0)	2.5 (1.3:3.4)	2.6 (1.7:3.4)	0.1 (0.4:0)
≥6 log_{10} IU/mL	234 (72%)	235 (72%)	1 (0)	2 (3%)	2 (3%)	0 (0)
Alt (U/L), Males	N=130	N=130	0	N=36	N=36	0
Median (25:75)	42.5 (32:59)	42.5 (32:59)	0 (0:0)	38.5 (32.5:43.5)	38.5 (32.5:43.5)	0 (0:0)
ALT (U/L), Females	N = 204	N = 204	0	N=42	N=42	0
Median (25:75)	37 (25:49)	37 (25:49)	0 (0:0)	29 (19:40)	29 (19:40)	0 (0:0)
ALT x ULN	N=334	N=334	0	N=78	N=78	0
≤ 1 x ULN	152 (46%)	152 (46%)	0 (0)	44 (56%)	44 (56%)	0 (0)
>1 to 2 x ULN	132 (40%)	132 (40%)	0 (0)	31 (40%)	31 (40%)	0 (0)
> 2 x ULN	50 (15%)	50 (15%)	0 (0)	3 (4%)	3 (4%)	0 (0)
Platelets ($10^3/mm^3$)	N=296	N=296	0 (0)	N=71	N=71	0
Median (25:75)	267 (233.5: 310)	267 (233.5:310)	0 (0:0)	259 (234:310)	259 (234:310)	0 (0:0)
< 160,000/mm^3	12 (4%)	12 (4%)	0 (0)	5 (7%)	5 (7%)	0 (0)

Characteristic	HBeAg (+) Manuscript N=226 (74%)	HBeAg (+) DSIC N=227 (74%)	Difference N=1 (0)
Age (yrs)	N=226	N=227	1
Mean (SD)	9.7 (4.5)	9.7 (4.5)	0 (0)
Sex	N=226	N=227	1
Female	146 (65%)	147 (65%)	1 (0)
Race	N=225	N=226	1
White	7 (3%)	7 (3%)	0 (0)
Black	21 (9%)	21 (9%)	0 (0)
Asian	189 (84%)	190 (84%)	1 (0)
Mixed/Other	8 (4%)	8 (4%)	0 (0)
Adopted	N=226	N=227	1
	125 (55%)	126 (56%)	1 (1)
Ever Received HBV Treatment	N=226	N=227	1
	14 (6%)	14 (6%)	0 (0)
Vertical Transmission	N=189	N=190	0
	184 (97%)	185 (97%)	1 (0)
HBV DNA (log_{10} IU/mL)	N=217	N=219	2
Median (25:75)	8.3 (8:9)	8.3 (8:8.9)	0 (0:0.1)
≥6 log_{10} IU/mL	206 (95%)	208 (95%)	2 (0)
Alt (U/L), Males	N=80	N=80	0
Median (25:75)	44.5 (34:73)	44.5 (34:73)	0 (0:0)
ALT (U/L), Females	N=144	N=145	1
Median (25:75)	38 (27:59)	38 (27:59)	0 (0:0)
ALT x ULN	N=224	N=225	1
≤ 1 x ULN	95 (42%)	96 (43%)	1 (1)
>1 to 2 x ULN	86 (38%)	86 (38%)	0 (0)
> 2 x ULN	43 (19%)	43 (19%)	0 (0)
Platelets ($10^3/mm^3$)	N=199	N=200	1
Median (25:75)	274 (239:308)	274 (240.5:309)	0 (1.5:1)
< 160,000/mm^3	5 (3%)	5 (3%)	0 (0)

Table C: Variables used to replicate Table 2: Baseline characteristics by genotype

Table Variable	**dataset.variable**
Genotype	hbrn_bcp.bgen, hbrn_cdc_results.genotype, hbrn_cdc_results.genotype2
Age	hbrn_peds_ages.age
Sex	hbrn_slp.scrsex
Race	hbrn_slp.sracew, hbrn_slp.sraceb, hbrn_slp.sracea, hbrn_slp.sracei, hbrn_slp.sraceh, hbrn_slp.sraceo
HBeAg	hbrn_bcp.hbeag, hbrn_bcp.hbeagy, hbrn_bcp.hbeagm, hbrn_ecp.erldate
HBV DNA	hbrn_uwash_results.result_n (where hbrn_uwash_results.cltest=1 and hbrn_uwash_results.cohort_tmpt=10), hbrn_bcp.bunit, hbrn_bcp.bdna, hbrn_bcp.bdnall
ALT	hbrn_bcp.alt
ALT x ULN	hbrn_bcp.alt, hbrn_peds_ages.age, hbrn_slp.scrsex

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

Characteristic N (%)	A Manuscript 11 (5)	A DSIC 11 (5)	Difference 0 (0)	B Manuscript 99 (43)	B DSIC 99 (43)	Difference 0 (0)
Age (yrs)	N=11	N=11	0	N=99	N=99	0
Mean (SD)	11.9 (4.5)	11.9 (4.5)	0 (0)	10.2 (4.6)	10.2 (4.6)	0 (0)
Sex	N=11	N=11	0	N=99	N=99	0
Female	5 (45%)	5 (45%)	0 (0)	71 (72%)	71 (72%)	0 (0)
Race	N=11	N=11	0	N=99	N=99	0
White	1 (9%)	1 (9%)	0 (0)	0 (0%)	0 (0%)	0 (0)
Black	7 (64%)	7 (64%)	0 (0)	1 (1%)	1 (1%)	0 (0)
Asian	3 (27%)	3 (27%)	0 (0)	97 (98%)	97 (98%)	0 (0)
Mixed/Other	0 (0%)	0 (0%)	0 (0)	1 (1%)	1 (1%)	0 (0)
HBeAg	N=9	N=9	0	N=88	N=88	0
HBeAg (+)	8 (89%)	8 (89%)	0	84 (95%)	84 (95%)	0 (0)
HBV DNA (log_{10} IU/mL)	N=11	N=11	0	N=95	N=96	1
Median (25:75)	7.9 (3.7:8.3)	7.9 (3.7:8.3)	0 (0:0)	8.2 (7.9:8.6)	8.2 (7.9:8.6)	0 (0:0)
≥6 log_{10} IU/mL	7 (64%)	7 (64%)	0 (0)	88 (93%)	89 (93%)	1 (0)
ALT (U/L), Males	N=6	N=6	0	N=27	N=27	0
Median (25:75)	53.5 (38:115)	53.5 (38:115)	0 (0:0)	43 (23:55)	43 (23:55)	0 (0:0)
ALT (U/L), Females	N=4	N=4	0	N=70	N=70	0
Median (25:75)	88 (37:747.5)	88 (37:747.5)	0 (0:0)	33 (25:47)	33 (25:47)	0 (0:0)
ALT	N=10	N=10	0	N=97	N=97	0
≤ 1 x ULN	3 (30%)	3 (30%)	0 (0)	55 (57%)	55 (57%)	0 (0)
>1 to 2 x ULN	3 (30%)	3 (30%)	0 (0)	27 (28%)	27 (28%)	0 (0)
> 2 x ULN	4 (40%)	4 (40%)	0 (0)	15 (15%)	15 (15%)	0 (0)

Characteristic N (%)	C Manuscript 74 (32)	C DSIC 74 (32)	Difference 0 (0)	D Manuscript 36 (16)	D DSIC 36 (16)	Difference 0 (0)
Age (yrs)	N=74	N=74	0	N=36	N=36	0
Mean (SD)	9.4 (4.1)	9.4 (4.1)	0 (0)	11.9 (4.2)	11.9 (4.2)	0 (0)
Sex	N=74	N=74	0	N=36	N=36	0
Female	46 (62%)	46 (62%)	0 (0)	22 (61%)	22 (61%)	0 (0)
Race	N=72	N=72	0	N=36	N=36	0
White	0 (0%)	0 (0%)	0 (0)	18 (50%)	18 (50%)	0 (0)
Black	0 (0%)	0 (0%)	0 (0)	0 (0%)	0 (0%)	0 (0)

Characteristic N (%)	C Manuscript 74 (32)	C DSIC 74 (32)	Difference 0 (0)	D Manuscript 36 (16)	D DSIC 36 (16)	Difference 0 (0)
Asian	69 (96%)	69 (96%)	0 (0)	18 (50%)	18 (50%)	0 (0)
Mixed/Other	3 (4%)	3 (4%)	0 (0)	0 (0%)	0 (0%)	0 (0)
HBeAg	N=67	N=68	1	N=35	N=35	0
HBeAg (+)	55 (82%)	56 (82%)	1 (0)	15 (43%)	15 (43%)	0 (0)
HBV DNA (log_{10} IU/mL)	N=72	N=72	0	N=35	N=35	0
Median (25:75)	8.3 (7.7:8.8)	8.2 (7.7:8.7)	0.1(0:0.1)	4.1 (3.1:8.3)	4.1 (3.1:8.3)	0 (0:0)
≥6 log_{10} IU/mL	61 (85%)	61 (85%)	0 (0)	12 (34%)	12 (34%)	0 (0)
ALT (U/L), Males	N=28	N=28	0	N=13	N=13	0
Median (25:75)	43 (34.5:57)	43 (34.5:57)	0 (0:0)	35 (30:41)	35 (30:41)	0 (0:0)
ALT (U/L), Females	N=45	N=45	0	N= 22	N= 22	0
Median (25:75)	36 (24:46)	36 (24:46)	0 (0:0)	40 (34:48)	40 (34:48)	0 (0:0)
ALT	N=73	N=73	0	N=35	N=35	0
≤ 1 x ULN	31 (42%)	31 (42%)	0 (0)	16 (46%)	16 (46%)	0 (0)
>1 to 2 x ULN	32 (44%)	32 (44%)	0 (0)	18 (51%)	18 (51%)	0 (0)
> 2 x ULN	10 (14%)	10 (14%)	0 (0)	1 (3%)	1 (3%)	0 (0)

Characteristic N (%)	E/Multiple Manuscript 10 (4)	E/Multiple DSIC 10 (4)	Difference 0 (0)
Age (yrs)	N=10	N=10	0
Mean (SD)	14.0 (2.9)	14.0 (2.9)	0 (0)
Sex	N=10	N=10	0
Female	4 (40%)	4 (40%)	0 (0)
Race	N=10	N=10	0
White	1 (10%)	1 (10%)	0 (0)
Black	8 (80%)	8 (80%)	0 (0)
Asian	0 (0%)	0 (0%)	0 (0)
Mixed/Other	1 (10%)	1 (10%)	0 (0)
HBeAg	N=10	N=10	0
HBeAg (+)	3 (30%)	3 (30%)	0 (0)
HBV DNA (log_{10} IU/mL)	N=10	N=10	0
Median (25:75)	4.4 (3.1:8.2)	4.5 (3.1:8.2)	0.1 (0:0)
≥6 log_{10} IU/mL	3 (30%)	3 (30%)	0 (0)
ALT (U/L), Males	N=6	N=6	0
Median (25:75)	73 (32:102)	73 (32:102)	0 (0:0)
ALT (U/L), Females	N=4	N=4	0

Characteristic N (%)	E/Multiple Manuscript 10 (4)	E/Multiple DSIC 10 (4)	Difference 0 (0)
Median (25:75)	45.5 (39:55.5)	45.5 (39:55.5)	0 (0:0)
ALT	N=10	N=10	0
≤ 1 x ULN	2 (10%)	2 (10%)	0 (0)
>1 to 2 x ULN	5 (50%)	5 (50%)	0 (0)
> 2 x ULN	3 (30%)	3 (30%)	0 (0)

Table E: Variables used to replicate Supplemental Table: Baseline characteristics across family composition

Table Variable	**dataset.variable**
Adopted	hbrn_bcp.adopt
Immigrant	hbrn_bcp.adopt, hbrn_bcp.cborn
Age	hbrn_peds_ages.age
Sex	hbrn_slp.scrsex
Race	hbrn_slp.sracew, hbrn_slp.sraceb, hbrn_slp.sracea, hbrn_slp.sracei, hbrn_slp.sraceh, hbrn_slp.sraceo
Place of Birth	hbrn_bcp.cborn
Combined Child-Parent Immigration Status	hbrn_bcp.cborn, hbrn_bcp.cbornm, hbrn_bcp.cbornf
Ever Received HBV Treatment	hbrn_bcp.txhbv, hbrn_bcp.txb1
HBeAg	hbrn_bcp.hbeag, hbrn_bcp.hbeagy, hbrn_bcp.hbeagm, hbrn_ecp.erldate
HBV DNA	hbrn_uwash_results.result_n (where hbrn_uwash_results.cltest=1 and hbrn_uwash_results.cohort_tmpt=10), hbrn_bcp.bunit, hbrn_bcp.bdna, hbrn_bcp.bdnall
ALT	hbrn_bcp.alt, hbrn_peds_ages.age, hbrn_slp.scrsex
Genotype	hbrn_bcp.bgen, hbrn_cdc_results.genotype, hbrn_cdc_results.genotype2

Table F: Comparison of values computed in integrity check to reference article Supplemental Table values

Characteristic	Adopted Manuscript N=188	Adopted DSIC N=188	Difference N=0	Living with Biological Parent(s) - Immigrant Manuscript N=73	Living with Biological Parent(s) - Immigrant DSIC N=73	Difference N=0
Age	N=188	N=188	0	N=73	N=73	0
<3 yrs	7 (39)	7 (39)	0 (0)	3 (17)	3 (17)	0 (0)
3 to <8 yrs	59 (63)	60 (64)	1 (1)	11 (12)	11 (12)	0 (0)
8 to <12 yrs	57 (65)	56 (64)	1 (1)	16 (18)	16 (18)	0 (0)
12 to <15 yrs	30 (41)	30 (41)	0 (0)	24 (33)	24 (33)	0 (0)
15 to <18 yrs	35 (50)	35 (50)	0 (0)	19 (27)	19 (27)	0 (0)
Sex	N=188	N=188	0	N=73	N=73	0
Male	47 (36)	47 (36)	0 (0)	39 (30)	39 (30)	0 (0)
Female	141 (67)	141 (67)	0 (0)	34 (16)	34 (16)	0 (0)
Race	N=186	N=186	0	N=73	N=73	0
White	23 (79)	23 (79)	0 (0)	1 (3)	1 (3)	0 (0)
Black	22 (61)	22 (61)	0 (0)	9 (25)	9 (25)	0 (0)
Asian	138 (52)	138 (52)	0 (0)	63 (24)	63 (24)	0 (0)
Mixed/Other	3 (27)	3 (27)	0 (0)	0 (0)	0 (0)	0 (0)
Place of Birth	N=188	N=188	0	N=73	N=73	0
US/Canada	6 (7)	6 (7)	0 (0)	0 (0)	0 (0)	0 (0)
Other North America and South America	5 (100)	5 (100)	0 (0)	0 (0)	0 (0)	0 (0)
Europe	24 (100)	24 (100)	0 (0)	0 (0)	0 (0)	0 (0)
Asia	137 (68)	137 (68)	0 (0)	64 (32)	64 (32)	0 (0)
Africa	16 (64)	16 (64)	0 (0)	9 (36)	9 (36)	0 (0)
Combined Child-Parent Immigration Status	N=185	N=183	0	N=73	N=73	0
Immigrated to US/Canada	180 (71)	178 (71)	2 (0)	73 (29)	73 (29)	0 (0)
Born in US/Canada & at least one parent foreign-born	2 (3)	2 (3)	0 (0)	0 (0)	0 (0)	0 (0)
Participant & both parents born in US/Canada	3 (33)	3 (33)	0 (0)	0 (0)	0 (0)	0 (0)
Ever Received HBV Treatment	N=188	N=188	0	N=73	N=73	0
	33 (77)	33 (77)	0 (0)	6 (14)	6 (14)	0 (0)
HBeAg	N=168	N=169	1	N=70	N=70	0

Characteristic	Adopted Manuscript N=188	Adopted DSIC N=188	Difference N=0	Living with Biological Parent(s) - Immigrant Manuscript N=73	Living with Biological Parent(s) - Immigrant DSIC N=73	Difference N=0
HBeAg(-)	40 (49)	41 (50)	1 (1)	30 (37)	30 (37)	0 (0)
HBeAg(+)	128 (55)	128 (55)	0 (0)	40 (17)	40 (17)	0 (0)
HBV DNA (log_{10} IU/mL)	N=178	N=179	1	N=69	N=69	0
<3 log_{10} IU/mL	25 (50)	24 (50)	1 (0)	16 (32)	16 (33)	0 (1)
3 to <6 log_{10} IU/mL	26 (62)	27 (63)	1 (1)	13 (31)	13 (30)	0 (1)
≥6 log_{10} IU/mL	127 (55)	128 (55)	1 (0)	40 (17)	40 (17)	0 (0)
ALT	N=183	N=183	0	N=72	N=72	0
≤1 x ULN	93 (62)	93 (62)	0 (0)	31 (21)	31 (21)	0 (0)
>1 to 2 x ULN	66 (50)	66 (50)	0 (0)	32 (24)	32 (24)	0 (0)
>2 x ULN	24 (48)	24 (48)	0 (0)	9 (18)	9 (18)	0 (0)
Genotype	N=127	N=127	0	N=50	N=50	0
A	6 (55)	6 (55)	0 (0)	1 (9)	1 (9)	0 (0)
B	62 (63)	62 (63)	0 (0)	17 (17)	17 (17)	0 (0)
C	33 (45)	33 (45)	0 (0)	17 (23)	17 (23)	0 (0)
D	22 (61)	22 (61)	0 (0)	10 (28)	10 (28)	0 (0)
E	4 (44)	4 (44)	0 (0)	5 (56)	5 (56)	0 (0)
Multiple	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

Characteristic	Living with Biological Parent(s) - Non-Immigrant Manuscript N=81	Living with Biological Parent(s) - Non-Immigrant DSIC N=81	Difference N=0
Age	N=81	N=81	0
<3 yrs	8 (44)	8 (44)	0 (0)
3 to <8 yrs	23 (25)	23 (24)	0 (1)
8 to <12 yrs	15 (17)	15 (17)	0 (0)
12 to <15 yrs	19 (26)	19 (26)	0 (0)
15 to <18 yrs	16 (23)	16 (23)	0 (0)
Sex	N=81	N=81	0
Male	46 (35)	46 (35)	0 (0)
Female	35 (17)	35 (17)	0 (0)
Race	N=81	N=81	0

Characteristic	Living with Biological Parent(s) - Non-Immigrant Manuscript N=81	Living with Biological Parent(s) - Non-Immigrant DSIC N=81	Difference N=0
White	5 (17)	5 (17)	0 (0)
Black	5 (14)	5 (14)	0 (0)
Asian	63 (24)	63 (24)	0 (0)
Mixed/Other	8 (73)	8 (73)	0 (0)
Place of Birth	N=81	N=81	0
US/Canada	81 (93)	81 (93)	0 (0)
Other North America and South America	0 (0)	0 (0)	0 (0)
Europe	0 (0)	0 (0)	0 (0)
Asia	0 (0)	0 (0)	0 (0)
Africa	0 (0)	0 (0)	0 (0)
Combined Child-Parent Immigration Status	N=81	N=78	3
Immigrated to US/Canada	0 (0)	0 (0)	0 (0)
Born in US/Canada & at least one parent foreign-born	75 (97)	72 (97)	3 (0)
Participant & both parents born in US/Canada	6 (67)	6 (67)	0 (0)
Ever Received HBV Treatment	N=81	N=81	0
	4 (9)	4 (9)	0 (0)
HBeAg	N=75	N=75	0
HBeAg(-)	11 (14)	11 (13)	1 (0)
HBeAg(+)	64 (28)	64 (28)	0 (0)
HBV DNA (log_{10} IU/mL)	N=78	N=77	1
<3 log_{10} IU/mL	9 (18)	8 (17)	1 (1)
3 to <6 log_{10} IU/mL	3 (7)	3 (7)	0 (0)
≥6 log_{10} IU/mL	66 (28)	66 (28)	0 (0)
ALT	N=78	N=78	0
≤1 x ULN	27 (18)	27 (18)	0 (0)
>1 to 2 x ULN	34 (26)	34 (26)	0 (0)
>2 x ULN	17 (34)	17 (34)	0 (0)

Characteristic	Living with Biological Parent(s) - Non-Immigrant Manuscript N=81	Living with Biological Parent(s) - Non-Immigrant DSIC N=81	Difference N=0
Genotype	N=52	N=52	0
A	4 (36)	4 (36)	0 (0)
B	19 (19)	19 (19)	0 (0)
C	24 (32)	24 (32)	0 (0)
D	4 (11)	4 (11)	0 (0)
E	0 (0)	0 (0)	0 (0)
Multiple	1 (100)	1 (100)	0 (0)

Attachment A: SAS Code

```
*** HBRN Pediatric Baseline DSIC;
*** Programmer: Allyson Mateja;
*** Date: 6/7/2016;

title 'HBRN Pediatric Baseline DSIC';
title2 ' ';

proc format;
        value sexf 1 = 'Male'
                   2 = 'Female';

        value bornf 9, 68, 70, 83, 114, 124, 161, 198, 204, 208 = 'Africa'
                    21, 46, 102, 110, 113, 116, 117, 133, 151, 167, 174, 194, 218, 235, 245 = 'Asia'
                    23, 36, 182, 196, 230 = 'Europe'
                    95 = 'Other North America and South America'
                    1, 2 = 'US/Canada';

libname hbrn '/prj/niddk/ims_analysis/HBRN/private_orig_data/HBRN_Peds_Baseline/HBRN Peds Cohort Baseline SAS Datasets/';

options nofmterr;

data ages;
        set hbrn.hbrn_peds_ages;

data bcp;
        set hbrn.hbrn_bcp;

data bip;
        set hbrn.hbrn_bip;

data slp;
        set hbrn.hbrn_slp;

data ecp;
        set hbrn.hbrn_ecp;

proc contents data = bcp;
proc contents data = ages;

proc sort data = bcp;
        by id;

proc sort data = ages;
        by id;

proc sort data = bip;
```

```
	by id;

proc sort data = ecp;
	by id;

proc sort data = slp;
	by id rescr_seq;

data slp;
	set slp;
	by id;
	if last.id then output;

data uwash;
	set hbrn.hbrn_uwash_results;

data hbv_dna;
	set uwash;
	if cltest = 1 and cohort_tmpt = 10;

data table1;
	length alt_uln_status $10.
	       race $12.;
	merge ages    (in=val1)
	      bcp     (keep=id hbeag hbeagy hbeagm adopt txhbv alt plat bdna bunit bdnall bgen cborn cbornf cbornm txb1)
	      slp     (keep=id scrsex sracew sraceb sracea sracei sraceh sraceo)
	      bip     (keep=id srcpi)
	      ecp     (keep=id erldate)
	      hbv_dna (keep=id result_n);
	by id;
	if srcpi = 1 then vertical_transmission = 1;
	else if srcpi in (2, 3, 6) then vertical_transmission = 0;
	if hbeag = .H then hbeag = 0;
	if txb1 = .C then txhbv = 0;
	if sracew = 1 and sraceb = 0 and sracea = 0 and sracei = 0 and sraceh = 0 and sraceo = 0 then race = 'White';
	else if sracew = 0 and sraceb = 1 and sracea = 0 and sracei = 0 and sraceh = 0 and sraceo = 0 then race = 'Black';
	else if sracew = 0 and sraceb = 0 and sracea = 1 and sracei = 0 and sraceh = 0 and sraceo = 0 then race = 'Asian';
	else if (sracew = 1 or sraceb = 1 or sracea = 1 or sracei = 1 or sraceh = 1 or sraceo = 1) then race = 'Mixed/Other';
	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;
	if bunit =1 and bdna not in (.C, .F, .G) then bdna = round(log10(bdna), .1);
	if result_n not in (.F, .J) then result_n = round(log10(result_n), .1);
	if result_n = . then do;
	       if bunit = 1 then do;
```

```
                    if bdna not in (.C, .F., .G) then result_n = bdna;
                  if bdna = .F then result_n = round(log10(bdnall), .1);
                  if bdna = .G then result_n = round(log10(5000000000), .1);
                end;
        end;
        if result_n = .J then result_n = 0;
        if result_n = .F then result_n = round(log10(bdnall), .1);
        if result_n >= 6 then hbv_greater_6 = 1;
        else if 0 <= result_n < 6 then hbv_greater_6 = 0;
        if scrsex = 1 and age <= 1.5 and 0 <= alt <= 60 then alt_uln_status = '<= 1';
        if scrsex = 2 and age <= 1.5 and 0 <= alt <= 55 then alt_uln_status = '<= 1';
        if scrsex = 1 and age > 1.5 and 0 <= alt <= 40 then alt_uln_status = '<= 1';
        if scrsex = 2 and age > 1.5 and 0 <= alt <= 35 then alt_uln_status = '<= 1';
        if scrsex = 1 and age <= 1.5 and 60 < alt <= 120 then alt_uln_status = '>1 to 2';
        if scrsex = 2 and age <= 1.5 and 55 < alt <= 110 then alt_uln_status = '>1 to 2';
        if scrsex = 1 and age > 1.5 and 40 < alt <= 80 then alt_uln_status = '>1 to 2';
        if scrsex = 2 and age > 1.5 and 35 < alt <= 70 then alt_uln_status = '>1 to 2';
        if scrsex = 1 and age <= 1.5 and alt > 120 then alt_uln_status = '>2';
        if scrsex = 2 and age <= 1.5 and alt > 110 then alt_uln_status = '>2';
        if scrsex = 1 and age > 1.5 and alt > 80 then alt_uln_status = '>2';
        if scrsex = 2 and age > 1.5 and alt > 70 then alt_uln_status = '>2';
        if 0 <= plat < 160 then plat_less_160000 = 1;
        else if plat > 160 then plat_less_160000 = 0;
        else plat_less_160000 = .;
        if val1 then output table1;

proc freq data = table1;
        tables hbeag /list ;
        where hbeag_length <= 2 ;
        title3 'Table 1 - HBeAg Status';

proc sort data = table1;
        by hbeag;

proc means data = table1;
        var age;
        title3 'Table 1 - Age';

proc means data = table1;
        var age;
        class hbeag;
        where hbeag in (0,1) and hbeag_length <= 2;

proc freq data = table1;
        tables scrsex;
        format scrsex sexf.;
        title3 'Table 1 - Sex';

proc freq data = table1;
        tables scrsex;
        format scrsex sexf.;
```

```
        by hbeag;
        where hbeag in (0,1) and hbeag_length <= 2;

proc freq data = table1;
        tables race;
        title3 'Table 1 - Race';

proc freq data = table1;
        tables race;
        by hbeag;
        where hbeag in (0,1) and hbeag_length <= 2;

proc freq data = table1;
        tables adopt;
        title3 'Table 1 - Adpoted';

proc freq data = table1;
        tables adopt;
        by hbeag;
        where hbeag in (0,1) and hbeag_length <= 2;

proc freq data = table1;
        tables txhbv /list missing;
        title3 'Table 1 - Ever Recieved HBV Treatment';

proc freq data = table1;
        tables txhbv;
        by hbeag;
        where hbeag in (0,1) and hbeag_length <= 2;

proc freq data = table1;
        tables vertical_transmission;
        title3 'Table 1 - Vertical Transmission';

proc freq data = table1;
        tables vertical_transmission;
        by hbeag;
        where hbeag in (0,1) and hbeag_length <= 2;

proc means data = table1 n median p25 p75;
        var result_n;
        title3 'Table 1 - HBV DNA';

proc means data = table1 n median p25 p75;
        var result_n;
        class hbeag;
        where hbeag in (0,1) and hbeag_length <= 2;

proc freq data = table1;
        tables hbv_greater_6;
```

```
proc freq data = table1;
       tables hbv_greater_6;
       by hbeag;
       where hbeag in (0,1) and hbeag_length <= 2;

proc means data = table1 n median p25 p75;
       var alt;
       where scrsex = 1;
       title3 'Table 1 - ALT (U/L), Males';

proc means data = table1 n median p25 p75;
       var alt;
       class hbeag;
       where scrsex = 1 and hbeag in (0,1) and hbeag_length <= 2;

proc means data = table1 n median p25 p75;
       var alt;
       where scrsex = 2;
       title3 'Table 1 - ALT (U/L), Females';

proc means data = table1 n median p25 p75;
       var alt;
       class hbeag;
       where scrsex = 2 and hbeag in (0,1) and hbeag_length <= 2;

proc freq data = table1;
       tables alt_uln_status;
       title3 'Table 1 - ALT x ULN';

proc freq data = table1;
       tables alt_uln_status;
       by hbeag;
       where hbeag in (0,1) and hbeag_length <= 2;

proc means data = table1 n median p25 p75;
       var plat;
       title3 'Table 1 - Platelets';

proc means data = table1 n median p25 p75;
       var plat;
       where scrsex = 2;
       class hbeag;
       where hbeag in (0,1) and hbeag_length <= 2;

proc freq data = table1;
       tables plat_less_160000;

proc freq data = table1;
       tables plat_less_160000;
       by hbeag;
       where hbeag in (0,1) and hbeag_length <= 2;
```

```
data cdc;
       set hbrn.hbrn_cdc_results;

proc sort data = cdc;
       by id;

proc sort data = table1;
       by id;

data table2;
       merge table1 (in=val1)
             cdc    (in=val2);
       by id;
       if genotype = 'ND' then genotype = genotype2;
       if bgen = 4 or genotype = 'A' then gen_type = 'A';
       if bgen = 11 or genotype = 'B' then gen_type = 'B';
       if bgen in (13, 15) or genotype = 'C' then gen_type = 'C';
       if bgen = 23 or genotype = 'D' then gen_type = 'D';
       if bgen = 24 or genotype in ('E', 'E,C') then gen_type = 'E';
       if val1 then output table2;

proc freq data = table2;
       tables gen_type;
       title3 'Table 2 - Genotype';

data table2;
       set table2;
       if gen_type not in ('', ' ');

proc sort data = table2;
       by gen_type;

proc means data = table2;
       var age;
       class gen_type;
       title3 'Table 2 - Age';

proc freq data =      table2;
       tables scrsex;
       format scrsex sexf.;
       by gen_type;
       title3 'Table 2 - Sex';

proc freq data = table2;
       tables race;
       by gen_type;
       title3 'Table 2 - Race';

proc freq data = table2;
       tables hbeag;
```

```
        by gen_type;
        where hbeag_length <= 2;
        title3 'Table 2 - HBeAg';

proc means data = table2 n median p25 p75;
        var result_n;
        class gen_type;
        title3 'Table 2 - HBV DNA';

proc freq data = table2;
        tables hbv_greater_6;
        by gen_type;

proc means data = table2 n median p25 p75;
        var alt;
        class gen_type;
        where scrsex = 1;
        title3 'Table 2 - ALT (U/L), Males';

proc means data = table2 n median p25 p75;
        var alt;
        class gen_type;
        where scrsex = 2;
        title3 'Table 2 - ALT (U/L), Females';

proc freq data = table2;
        tables alt_uln_status;
        by gen_type;
        title3 'Table 2 - ALT';

data table3;
        length status age_group $15.
               gen_type $8.
               child_parent_status $55.
               hbv_dna $8.;
        merge table1 (in=val1)
              cdc    (in=val2);
        by id;
        if adopt = 1 then status = 'Adopted';
        if adopt = 0 and cborn ne .C and cborn not in (1,2) then status = 'Immigrant';
        else if adopt = 0 and cborn in (1,2) then status = 'Non-Immigrant';
        if cborn in (1,2) and cbornm in (1,2) and cbornf in (1,2) then child_parent_status = 'Participant & both parents born in
US/Canada';
        else if cborn not in (1,2) and ((cbornm not in (1,2, .C) and cbornf not in (1,2, .C)) or (cbornf = .C and cbornm = .C)) then
child_parent_status = 'Immigrated to US/Canada';
        else if cbornf ne .C and cbornm ne .C then child_parent_status = 'Born in US/Canada & at least one parent foreign-born';
        if age < 3 then age_group = '<3';
        if 3 <= age < 8 then age_group = '3 to < 8';
        if 8 <= age < 12 then age_group = '8 to <12';
        if 12 <= age < 15 then age_group = '12 to <15';
        if 15 <= age < 18 then age_group = '15 to <18';
```

```
        if bgen = 4 or genotype = 'A' then gen_type = 'A';
        if bgen = 11 or genotype = 'B' then gen_type = 'B';
        if bgen in (13, 15) or genotype = 'C' then gen_type = 'C';
        if bgen = 23 or genotype = 'D' then gen_type = 'D';
        if bgen = 24 or genotype = 'E' then gen_type = 'E';
        if genotype = 'E,C' then gen_type = 'Multiple';
        if 0 <= result_n < 3 then hbv_dna = '<3';
        if 3 <= result_n < 6 then hbv_dna = '3 to < 6';
        if result_n >= 6 then hbv_dna = '>=6';
        if val1 then output;

proc freq data = table3;
        tables status ;
        title3 'Supplemental Table - Family Composition';

data table3;
        set table3;
        if status not in ('', ' ');

proc sort data = table3;
        by status;

proc freq data = table3;
        tables age_group*status /nocol nopercent;
        title3 'Supplemental Table - Age';

proc freq data = table3;
        tables scrsex*status /nocol nopercent;
        format scrsex sexf.;
        title3 'Supplemental Table - Sex';

proc freq data = table3;
        tables race*status /nocol nopercent;
        title3 'Supplemental Table - Race';

proc freq data = table3;
        tables cborn*status /nocol nopercent ;
        format cborn bornf.;
        title3 'Supplemental Table - Place of Birth';

proc freq data = table3;
        tables child_parent_status*status /nocol nopercent;
        title3 'Supplemental Table - Combined Child-Parent Immigration Status';

proc freq data = table3;
        tables txhbv*status /nocol nopercent;
        title3 'Supplemental Table - Ever Received HBV Treatment';

proc freq data = table3;
        tables hbeag*status /nocol nopercent;
        title3 'Supplemental Table - HbEAg';
```

```
proc freq data = table3;
       tables hbv_dna*status /nocol nopercent;
       title3 'Supplemental Table - HBV DNA';

proc freq data = table3;
       tables alt_uln_status*status /nocol nopercent;
       title3 'Supplemental Table - ALT';

proc freq data = table3;
       tables gen_type*status /nocol nopercent;
       title3 'Supplemental Table - Genotype';
```