**Table of Contents** **Pages**

1. [Introduction 2](#Intro)

2. [Study Freeze Process and Summary 3](#F10Process)

3. [Stool Specimen Collection and Measurement of Vaccine Shedding 5](#ShedStool)

4. [Rotavirus Infection Determination](#Rotadeter)  7

5. [Calculation of Gestational Age 7](#GACal)

6. [Visit Windows 8](#VSTwindow)

7. [Protocol Implementation Dates in the Field](#Protocolversin)  9

8.[Bayley Scales of Infant and Toddler Development assessment](#Bayley) 9

9. [Description on HLA data structure 9](#HAL)

10. [PROVIDE Study Data Files Included in this Release 11](#F10DataFiles)

**Appendices**

[Appendix A: Additional details on HLA data 18](#appendix)

**1.** **Introduction**

Oral polio and rotavirus vaccines are significantly less effective in children living in the developing world. Environmental enteropathy, which is associated with intestinal inflammation, decreased absorption and increased permeability, may contribute substantially to oral vaccine failure in developing country settings. The study sought to determine the association of environmental enteropathy with the efficacy of oral polio and rotavirus vaccines in children in Bangladesh.

The PROVIDE (**P**erformance of **R**otavirus and **O**ral Polio **V**accines in **DE**veloping Countries) study was structured as a 2x2 clinical trial with a prospective birth cohort enrollment of 700 infants and their mothers in the urban slum of Mirpur, Dhaka, Bangladesh. All children received the full series of Bangladesh’s Expanded Program on Immunization vaccines, including oral polio vaccine (OPV) and followed for their first 2 years of life. Half the children were randomized to receive an inactivated polio vaccine (IPV) for their 4th dose at 39 weeks instead of the standard OPV dose. In the other trial, half the infants received the 2-dose rotavirus vaccine (Rotarix) with delayed dosing at 10 and 17 weeks of age.

The database snapshot (Freeze 10) was conducted on February 19, 2016 and contains clinical information with disclosure of randomization treatment arm for Polio and Rotavirus vaccine along with lab assay results for data/specimens collected between May 22, 2011 and February 19, 2016. Additional clinical and lab assay results have been added post data freeze 10. The release contains 38 individual clinical tables designed to capture the study related data along with 49 lab assay results and 22 derived management tables. The details of the tables along with additional information for better understanding of the study procedures and data are provided in this document.

This package contains the following:

PROVIDE\_Data\_Documentation.docx This document

PROVIDE\_CRFs Directory of study case report forms

PROVIDE\_Data Directory of data files

PROVIDE\_DataDefinition Directory of data definition files

PROVIDE\_Protocol Latest version of protocol

PROVIDE Annotated Bibliography List of published papers

**2.** **Study Freeze Process and Summary**

1. In Screening Eligibility form data (bv\_sel\_screening), HHID (household ID) and MSCN (Screening Number) variables are not released. All participants that were entered in the screening form are not included in this CRF table in compliance with HIPPA guidelines.
2. Missing data is retained with values as specified during the design of each individual CRF. Typically missing data will be formatted to contain 9s, eg 9, 99, 999.99, -9, 09/09/1999.
3. The diarrheal surveillance form table bv\_dir\_diarrheal\_surveillance is pre-populated to record the surveillance data from birth through study duration (24 months). We have retained all records in this table, including those that had not been data entered as of study closeout. The non-data entered records will contain wholly missing values.
4. Special consideration to be taken while using following lab assay results tables
   1. **lab\_bv\_pcr**: When rotact value is ≥ 36 the assay is repeated only for rota virus resulting in duplicate records in the table with -9 in other viruses. This necessitates using astro, napo and sopa virus results from first assay and rota virus results from the second assay. To avoid confusion to the user, duplicate records are collapsed to include rota result from second assay and results for remaining viruses from first assay.
   2. **lab\_bv\_rpliga. & lab\_bv\_rpligg**: additional columns are introduced to show the calculation for conversion of Endpoint titer to units per ml (uml) as per algorithm forwarded by University of Vermont (UVM). A second variable uml2 is included using modified UVM algorithm in which the “*normalized positive control*” is derived using geometric mean of positive control A & B and “*average titer*” is derived using geometric mean of titer A & B instead of simple average. Please refer data description files “lab\_bv\_rpliga.xls” & “lab\_bv\_rpliga.xls” for better understanding.
   3. **Iterative visit schedule:** The infant follow-up (FU) visits was originally scheduled using their birth dates and were static once generated but the investigators decided to shift the visit dates based on actual visit dates of vaccination at wk 6 and 10. This was done with the recognition that for maximum efficacy of Pentavalent and OPV vaccines they require a minimum of 4 weeks between doses. The protocol was thus modified in *May 2011* to include the iterative visit schedules based on actual visit at weeks 6 and 10. However, given the 7 days’ window on each of these vaccination visits, the overall clinic visit schedule may be pushed out up to 4 weeks cumulatively from target age for each FU visit. In protocol version 6 modifications were made so the visit schedules will continue to be generated iteratively in this window but they resume clinic visits per their actual age from date of birth starting at age 24 weeks visit. This was implemented in *September 2012* and therefore, there are ~ 88 children whose week 52 occurred beyond their visit window. Effort was made to ensure that all five doses of polio vaccine are received before 52 weeks of life. For purposes of primary outcome data analysis in the Polio arm children that received Polio vaccination outside the window of this iterative visit schedule are considered as not adhering to protocol and thus flagged.
   4. **Reg1b results**: The kit used to assay the first batch containing 239 samples on 80 kids at visit weeks 12, 24, and 40 was not calibrated correctly resulting in very high OD values and new kits were ordered to run all samples except those at 24 and 40 (159) that were in batch1. These records in the data file are coded -9.99999.
5. Additional management tables are included in the dataset which contain **derived data**:
   1. mgmt\_aed: adverse events experienced by infants during first two years of life in long format.
   2. mgmt\_bv\_danth: Anthropometry measurement along with their WHO Z Scores
   3. mgmt\_bv\_ddep: Diarrheal episodes, RUUSKA score and corresponding diarrheal rotavirus status. The scoring categories for fever was updated from 5-point scale (0-4) to 4-point scale (0-3) to match with 20-point maximum scores as defined per original publication (ref: Scan J Infec Dis 22: 259-267 1990).
   4. mgmt\_bv\_dinf: Derived infant data (one record per infant) to include breastfeeding
   5. mgmt\_bv\_lmratio: Derived Lactulose/Mannitol ratio
   6. mgmt\_mib\_mother\_infant\_baseline: Baseline data captured from both mother and infant during enrollment for rota outcome analysis
   7. mgmt\_mgmt\_ppg\_polioprotgp: Polio protocol groups used in polio trial analysis
   8. mgmt\_prn\_polio\_random: Polio randomization group
   9. mgmt\_psi1\_poliosi: PolioS1\_Polio shedding index
   10. mgmt\_psi2\_poliosi: PolioS2\_Polio shedding index
   11. mgmt\_psi3\_poliosi: PolioS3\_Polio shedding index
   12. mgmt\_pso\_poliotrial\_sna: derived table with polio trial SNA data – one row per child wide form data table
   13. mgmt\_pta\_poliotrial\_ae: Retain adverse events that occurred during first year of infants among kids randomized with Polio intervention.
   14. mgmt\_pto\_poliotrial\_outcome: retain polio outcome, pcr and qpcr at week 14, 52 – one row per child
   15. mgmt\_pts\_poliotrial\_sae: Retain serious adverse events that occurred during first year of infants and added additional column for polio intervention arm.
   16. mgmt\_rrn\_rota\_random: Rota randomization group
   17. mgmt\_rta\_rotatrial\_ae: Retain adverse events that occurred during first year of infants
   18. mgmt\_rtf\_rotatrial\_fup: Retain diarrheal episodes during first year of infants. Updated consequent to modifications in mgmt\_bv\_ddep
   19. mgmt\_rtf2\_rotatrial\_fup2: Retain diarrheal episodes during two year of life of the infants. Updated consequent to modifications in mgmt\_bv\_ddep
   20. mgmt\_rto: Diarrheal episode information consolidated by infant; one row per child. Updated consequent to modifications in mgmt\_bv\_ddep and updates to rotaprod variable.
   21. mgmt\_rto2: Diarrheal episode information consolidated by infant during first two years of life; one row per child. Updated consequent to modifications in mgmt\_bv\_ddep and updates to rotaprod variable.
   22. mgmt\_rts: Retain serious adverse events that occurred during first year of infants and added additional column for Rota intervention arm.

*(The exclusive breast feeding (BF) data is assessed to the last day of exclusive BF before any other food was introduced even if the mother adopted exclusive BF subsequent to that date.)*

**3.** **Shedding Stool Specimen collection and measurement of vaccine shedding**

**3a. Calculation of specimen collection window**

In tables, bv\_ps1 or bv\_ps2 or bv\_rss; the shedding stool series are collected as per following schedule, with week 14 visit date (f14dt - OPV) or 17 visit date (f17dt - Rotarix) or 52 visit date (f52dt - Polio) being the date of vaccination/visit:

Example:

Day 0 (-1)

Day 4 (+1)

Day 11 (+1)

Day 18 (+1)

Day 25 (+1)

This means that counting from f14dt, any specimen collected in the following windows is considered valid:

Day 0: Day -1 to Day 0 pre-vax

Day 4: Day 3 to Day 5 post-vax

Day 11: Day 10 to Day 12 post-vax

Day 18: Day 17 to Day 19 post-vax

Day 25: Day 24 to Day 26 post-vax

**3b. Determination of viral shedding of any vaccine**

The presence and absence of viral shedding of any vaccine was done by measurement of cell culture and qPCR at week 14 and week 52 for poliovirus vaccine. While stool is collected at 5 discrete time points on day 0 before vaccination and 4, 11, 18, 25 after the tOPV challenge, only the first positive cell culture was run for PCR. During the initial few months the last positive sample was used to perform PCR but it was changed to first because it was difficult to obtain cell lines in the last sample.

Additional vaccine virus samples were collected at week 6 (for pre and post-polio vaccine) and week 10 (for pre and post Rotavirus vaccine) visits in protocol version 6 (implemented ~June 16, 2012). These stool samples were separated from aliquot collected for assessing fecal cytokines (bv\_fcs table). Initially when bv\_fcs form was created three samples for cytokines were scheduled at week 6 (Pre/Post vaccination) and week 10 (post vaccination). An additional sample at week 10 **pre vaccination** sample was introduced in protocol 6 so that pre/post vaccine virus excretion samples could be separated. Due to this there is slight confusion in the variable name sequence in bv\_fcs table - see the table below.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Protocol** | **Sample requirement** | **Variable** | **Cytokines** | | **Protocol** | **Vaccine Virus Excretion** | |
| **Sample ID** | **Sample Count** | **Sample ID** | **Sample count** |
| V4 | week 06 (-3days) pre vax | fcst1 | BVC-SID-13-01 | 399 | V6 (09/01/12) | BVC-SID-11-14 | 122 |
| V4 | week 06 (+ 1-3days) post vax | fcst2 | BVC-SID-13-02 | 395 | BVC-SID-11-15 | 124 |
| V6 | week 10 (-3days) pre vax | **fcst4** | BVC-SID-13-04 | 151 | BVC-SID-11-16 | 151 |
| V4 | week 10 (+ 1-3days) post vax | fcst3 | BVC-SID-13-03 | 427 | BVC-SID-11-17 | 150 |

**4.** **Rotavirus infection determination**

Presence of rotavirus infection in diarrheal samples was performed by both ELISA and multiplex RT-PCR methods. All samples positive to rotavirus were genotyped using RT-PCR (table: lab\_bv\_dsgene) method and/or TAC array method (table: lab\_bv\_tacd). The original plan was to genotype all positive samples by both ELISA and multiplex RT-PCR methods (n=284) plus those positive only by Multiplex RT\_PCR method (n=198). However, only a total of 263 were genotyped using RT-PCR. In April of 2014, TAC array method was introduced and all positive samples plus diarrheal samples with RUUSKA score >9 (381) were genotyped.

| **TAC Array** | | | |
| --- | --- | --- | --- |
| **ELISA\_rota** | **Multiplex\_rota** | **Count** | **Comment** |
| **+ve** | **+ve** | 284 |  |
| -ve | **+ve** | 198 |  |
| -ve | -ve | 371 | Ruuska>=9 |
| -ve | -ve | 10 | Ruuska<9 |
| Total |  | 863 |  |

In 2017, additional 1334 samples were processed to contain a total of 2197/3788 stool samples collected from infants during first two years of age.

**5.** **Calculation of Gestational Age (GA)**

Form: BV\_GAA: Calculation of Total GA score and estimated gestational age.

1) Total GA score:

gapos +gasqrw+gaarm+gapop+gascrf+gaheel+gaskin+galanu+gaplan+gabrest+gaeye+if (gagenm=99,0,gagenm) + if(gagenf=99,0,gagenf)

Please refer data definition file BV\_GAA\_Ver1.xls for description of each variable used in the algorithm

2) Table of Estimated Gestational Age:

|  |  |  |
| --- | --- | --- |
| **Total Score** | **Values between** | **Gestational Age in Weeks** |
| -10 | >=-10 and <=-8 | 20 |
| -5 | >=-7 and <=-3 | 22 |
| 0 | >=-2 and <=2 | 24 |
| 5 | >=3 and <=7 | 26 |
| 10 | >=8 and <=12 | 28 |
| 15 | >=13 and <=17 | 30 |
| 20 | >=18 and <=22 | 32 |
| 25 | >=23 and <=27 | 34 |
| 30 | >=28 and <=32 | 36 |
| 35 | >=33 and <=37 | 38 |
| 40 | >=38 and <=42 | 40 |
| 45 | >=43 and <=47 | 42 |
| 50 | >=48 and <=54 | 44 |

**6.** **Visit windows of the Study visits**

**Vst\_1**: Birth (+ 7 days) V4 enrollment stool collection window changed to +10 days post enrollment

**Vst\_2**: Age 6 Weeks (-3 days for stool collection; + 7 days for all other procedures)

**Vst\_2**: Age 6 Weeks (week 6 visit +1-3 days) (FCS and Polio virus excretion included)

**Vst\_3**: Age 10 Weeks (+ 7 days)

**Vst\_3**: Age 10 Weeks (week 10 visit +1-3 days) (FCS and Rota virus excretion included)

**Vst\_4**: Age 12 Weeks (+ 7 days)

**Vst\_5**: Age 14 Weeks (-1 day for stool collection added in V4; + 7 days for all other procedures)

**Vst\_6**: Age 17 Weeks (-1 day for stool collection added in V4; + 7 days for all other procedures)

**Vst\_7**: Age 18 Weeks (one week after week 17 visit +/- 2 days)

**Vst\_8**: Age 24 Weeks (+ 7 days)

**Vst\_9**: Age 39 Weeks (+ 7 days)

**Vst\_10**: Age 40 Weeks (one week after week 39 visit +/- 2 days)

**Vst\_11**: Age 52 Weeks (-1 day for stool collection; + 7 days for all other procedures); V3 window was specified as +/- 4 weeks; V4 -1 day for stool was added; V6 visit window changed to +7 days for all other procedures

**Vst\_12**: Age 53 Weeks (one week after week 52 visit +/- 2 days); V3 window was 53-56 weeks (4 weeks’ window for the visit); V4 added stool window within one week +/- 2 days of week 52 OPV dose but still 4 weeks for all other procedures; V6 visit window changed to week 53 + 7 days

**Vst\_14**: Age 65 Weeks (+ 7 days)

**Vst\_15**: Age 78 Weeks (+ 7 days)

**Vst\_16**: Age 91 Weeks (+ 7 days)

**Vst\_13**: Age 104 Weeks (+/- 30 days) V3 window +/- 8 weeks; V6 window changed to +4 weeks

**7.** **PROVIDE study protocol version implementation dates in the field:**

V3 – 22 May 2011

V4 – 09 Nov 2011

V5 – 19 Jan 2012

V6 – 3 Sept 2012

V7 – 23 Oct 2013

V8 – 19 May 2014

**8.** **Bayley Scales of Infant and Toddler Development assessment at week 104**

The infant and mother pair is contacted anytime during **+/-** 30 days’ window to assess the cognitive development using Bayley Scales of Infant and Toddler Development. There are at least 240 kids whose date of assessment is < actual date of week 104 visit.

**9.** **Description on HLA data structure**

Points to note in lab\_bv\_hla data file:

1. 10 samples were sequenced to test allele calling for DRB1 and DRB345.  These alleles/genotypes are identical to calls in the full HLA sequencing run for each sample and therefore, were not included in the lab\_bv\_hla table.
2. Single quote (ex: **'**02:11) that preceded each allele from the original table (as forwarded by CHORI) has been deleted in ext\_lab\_bv\_hla\_f3 table
3. Where high resolution part of the allele is new, it is designated as "15: new" in the data table. These are detected primarily in allele B\_2 (n=30) and DQB1\_2 (n=4)
4. When “no call” is made or if the data is shown as “\_ \_: new?” especially in HLA A and HLA B, code -9 is applied to indicate missing data for which allele call can be expected. In the current file, these will not appear as they were found in duplicate samples that were excluded due to reasons mentioned in point 1.
5. Where the data is missing and no allele call can be expected, code “Absent” is used; observed mostly in DRB345\_1 and DRB345\_2
6. Refer Appendix A enclosed in this document for additional details

**10.** **PROVIDE Study Data Files Included in this Freeze and Release**

|  |  |  |  |
| --- | --- | --- | --- |
| **S. No** | **Table Name** | **Brief Description** | **Primary Key Columns** |
| **Clinical data tables** | | | |
| 1 | bv\_aer | Adverse event table | sid, aerdt, aewho |
| 2 | bv\_bid\_bayley\_inf\_development | Assessment of cognitive development using Bayley Scales of Infant and Toddler Development assessment | sid , vnum, section, itemnum |
| 3 | bv\_bsc\_bayley\_score | Bayley’s raw, scaled and composite scores | sid, vnum |
| 4 | bv\_dep\_diarrheal\_episode | Clinical data on diarrheal episodes for which specimen is collected | sid, deindt, dedt |
| 5 | bv\_dir\_diarrheal\_surveillance | Diarrheal surveillance data collected during bi-weekly home visits by field research team | sid, dov |
| 6 | bv\_dsmicr | Diarrheal stool specimen lab assay results | sid, specdt |
| 7 | bv\_enc\_infant\_enrollment | Demographic information (gender, birth place, first food, vaccination) on infant collected at enrollment | sid |
| 8 | bv\_enm\_mother\_enrollment | Demographic information (age, age at marriage, first pregnancy, delivery place/type, total pregnancies, vaccination) on mother collected at enrollment | sid |
| 9 | bv\_epd\_edinburgh\_postnatal\_depression | Assessment of maternal depression and general intelligence using the Edinburgh Postnatal Depression Scale | sid |
| 10 | bv\_f06\_week\_06\_visit | Clinic visit data includes Anthropometry, vaccination, specimen collection at week 06 | sid,vstnum |
| 11 | bv\_f10\_week10\_visit | Clinic visit data includes Anthropometry, vaccination, specimen collection at week 10 | sid,vstnum |
| 12 | bv\_f14\_week\_14\_visit | Clinic visit data includes Anthropometry, vaccination, specimen collection at week 14 | sid,vstnum |
| 13 | bv\_f17\_week\_17\_visit | Clinic visit data includes Anthropometry, vaccination, specimen collection at week 17 | sid,vstnum |
| 14 | bv\_f18\_week\_18\_visit | Clinic visit data includes Anthropometry, vaccination, specimen collection at week 18 | sid,vstnum |
| 15 | bv\_f39\_week\_39\_visit | Clinic visit data includes Anthropometry, vaccination, specimen collection at week 39 | sid,vstnum |
| 16 | bv\_f40\_week\_40\_visit | Clinic visit data includes Anthropometry, vaccination, specimen collection at week 40 | sid,vstnum |
| 17 | bv\_f52\_week\_52\_visit | Clinic visit data includes Anthropometry, vaccination, specimen collection at week 52 | sid,vstnum |
| 18 | bv\_f53\_week\_53\_visit | Clinic visit data includes Anthropometry, vaccination, specimen collection at week 53 | sid,vstnum |
| 19 | bv\_f104\_week\_104\_visit | Anthropometry vaccination, specimen collection, & cognitive development at week 104 | sid,vstnum |
| 20 | bv\_fan\_week\_65\_78\_91\_visit | Clinic visit data includes Anthropometry, vaccination, specimen collection, & cognitive development at weeks 65, 78 and 91 | sid, fanvn |
| 21 | bv\_fci\_family\_care\_indicator | Infant family care indicators assessment | sid, vstnum |
| 22 | bv\_fcs\_fecal\_cytokine\_shedding | Clinic visit data includes fecal cytokine shedding specimen collection at week 06 | sid,vstnum |
| 23 | bv\_flm\_week\_12\_24\_visit | Clinic visit data includes Anthropometry, vaccination, urine specimen collection at week 12 & 24 | sid, vstnum |
| 24 | bv\_gaa\_gestational\_age\_assessment | Gestational age assessment form | sid,vstnum |
| 25 | bv\_mfu\_mother\_follow\_up | Mother follow-up form with Anthropometry and specimen collection | sid,vstnum |
| 26 | bv\_neb\_Nebulizer | Infant nebulizer usage data |  |
| 27 | bv\_pdv\_protocol\_deviation | Protocol deviations observed during the course of study | sid, pdvrdt, pdvwho; pdvtype; pdvwk |
| 28 | bv\_ps1\_polio\_vaccine\_shedding\_stool\_wk14 | Clinic visit data includes polio shedding stool specimen collection at week 14 | sid |
| 29 | bv\_ps2\_polio\_vaccine\_shedding\_stool\_wk14 | Clinic visit data includes polio shedding stool specimen collection at week 52 | sid |
| 30 | bv\_ptreat | Infant antibiotic use | sid, dov |
| 31 | bv\_ref\_infant\_referral | Infant referral form | sid, refdt, refreas |
| 32 | bv\_rss\_rotavirus\_vaccine\_shedding\_stool\_17 | Clinic visit data includes Rota shedding stool specimen collection at week 17 | sid |
| 33 | bv\_rvn\_raven\_progressive | Assessment of maternal depression and general intelligence using Raven's Progressive Matrices | sid |
| 34 | bv\_sae\_serious\_adverse\_form | Serious adverse events noted on enrolled mother/infant pairs during the course of study | sid, saerdt, saersdt |
| 35 | bv\_sbo\_sibo | Small intestinal bacterial overgrowth data | sid, doc |
| 36 | bv\_sel\_screening | Screening eligibility form applied before enrollment | sid |
| 37 | bv\_ses\_water | Water, sanitation and socioeconomic data gleaned at the time of enrollment | sid |
| 38 | bv\_ter\_termination | Records early termination or termination of infant mother pair at study completion | sid, terdt, terwho |
| **Lab Assay results** | | | |
| 1 | lab\_bv\_abo | Infant Blood Group | sid |
| 2 | lab\_bv\_act | Infant Activin A measure | sid, vstnum |
| 3 | lab\_bv\_ala | Infant Alpha -1- anti trypsin results | sid, vstnum |
| 4 | lab\_bv\_bml | Breast Milk lipid profile | sid |
| 5 | lab\_bv\_calp | Infant Calprotectin results | sid, vstnum |
| 6 | lab\_bv\_cmv | Infant Cytomegalovirus IgG measure results | sid |
| 7 | lab\_bv\_cmvrt | Infant CMV-RT PCR results | sid, vstnum |
| 8 | Lab\_bv\_crem\_prkca | Infant CREM and PRKCA genotype |  |
| 9 | lab\_bv\_crp | Infant CRP results | sid, vstnum |
| 10 | lab\_bv\_ctt | Infant tetanus measure | sid |
| 11 | lab\_bv\_ctu | Infant cytokine assay results | sid, vstnum |
| 12 | lab\_bv\_cytbm | Breast milk Cytokine measure | sid |
| 13 | lab\_bv\_cytcp | Infant Plasma Cytokine measure | sid, vstnum |
| 14 | lab\_bv\_cytmp | Maternal Plasma Cytokine measure | sid, vstnum |
| 15 | lab\_bv\_dsgene | Diarrheal Stool Rota Gene Expression data | sid, specdt |
| 16 | lab\_bv\_egf | Breast milk EGF biomarker measure | sid |
| 17 | lab\_bv\_endotox | Infant Endotoxin results | sid, vstnum |
| 18 | lab\_bv\_fer | Infant Ferritin Measure | sid, vstnum |
| 19 | lab\_bv\_hla | Infant HLA Measure | sid, vstnum |
| 20 | lab\_bv\_lep | Infant Leptin receptor genotype data | sid |
| 21 | lab\_bv\_lps | Infant LPS/Endocab measure | sid, vstnum |
| 22 | lab\_bv\_lm | Infant Lactulose and Mannitol measure | sid, vstnum |
| 23 | lab\_bv\_mfge8 | Breast Milk MFGE8 biomarker measure | sid |
| 24 | lab\_bv\_mpo | Infant Myeloperoxidase measure | sid, vstnum |
| 25 | lab\_bv\_neop | Infant Neopterin measure | sid, vstnum |
| 26 | lab\_bv\_pals | Infant Polio ALS measure data | sid, vstnum |
| 27 | lab\_bv\_pcr | Infant diarrheal stool PCR for Rota, Astro, Noro, Sapo viruses. | sid, specdt |
| 28 | lab\_bv\_pcrproto | Infant diarrheal stool PCR for protozoa panel. | sid, specdt |
| 29 | lab\_bv\_polcul | Infant polio culture | sid, vstnum, vstdetail |
| 30 | lab\_bv\_polpcr | Infant Polio PCR measure data | sid, vstnum, vstdetail |
| 31 | lab\_bv\_polsna | Polio Neutralizing Antibody | sid, vstnum |
| 32 | lab\_bv\_qpcr | Infant Polio QPCR measure | sid, vstnum |
| 33 | lab\_bv\_rals | Infant Rota ALS measure data | sid, vstnum |
| 34 | lab\_bv\_rbmiga | Breast milk Rota IgA measure | sid, vstnum |
| 35 | lab\_bv\_rbp | Infant RBP measure | sid, vstnum |
| 36 | lab\_bv\_reg1b | Infant Reg1b measure | sid, vstnum |
| 37 | lab\_bv\_rotads | Infant Diarrheal stool Rota ELISA results. | sid, specdt |
| 38 | lab\_bv\_rotavs | Infant Viral shedding stool Rota ELISA results. | sid, vstnum, vstdetail |
| 39 | lab\_bv\_rpliga | Infant Rota plasma IgA measure | sid, vstnum |
| 40 | lab\_bv\_rpligg | Infant Rota plasma IgG measure | sid, vstnum |
| 41 | lab\_bv\_rptype | Infant Rota P Type and Lewis Secretor results | sid |
| 42 | lab\_bv\_scd14 | Infant Soluble CD14 measure data | sid, vstnum |
| 43 | lab\_bv\_speclog | Log maintained on specimen collection | sid, spectype, vstnum, vstdetail |
| 44 | lab\_sspabm | Breastmilk Lewis Secretor Status & Phenotyping Assay | sid |
| 45 | lab\_bv\_tac | PCR results on enteropathogens in infant clinic visit stool at wk6 and wk10 visit | sid, vstnum |
| 46 | lab\_bv\_tacd | PCR results on enteropathogens in infant diarrheal stool samples and stool samples | sid, specdt |
| 47 | lab\_bv\_vac | Infant vaccine data | sid |
| 48 | lab\_bv\_vitd | Infant Vitamin D measure data | sid, vstnum |
| 49 | lab\_bv\_zn | Infant Zinc measure data | sid, vstnum |
| **Derived tables** | | | |
| 1 | mgmt\_bv\_aed | Derived data file on adverse events in infants in long format for convenience of analysis | sid, aerepdt; aeresdt |
| 2 | mgmt\_bv\_danth | Derived data table maintained with WHO **(MGRS)** Z scores (WAZ, HAZ, WHZ) calculated using anthropometric information collected during study clinic visits | sid, vstdate |
| 3 | mgmt\_ddep | Derived data table with variables used from ext\_bv\_dir and ext\_bv\_dep tables to calculate duration of diarrhea, severity, presence of fever, & RUUSKA Score | sid, epidate |
| 4 | mgmt\_bv\_dinf | Derived data table with breast feeding days, last FU date and age | sid |
| 5 | mgmt\_bv\_lmratio | Derived data table on lactulose/ mannitol ratio | sid, lmdoc |
| 6 | mgmt\_mib\_mother\_infant\_baseline | Derived table with baseline data on mother and infant pair for Rota outcome analysis | sid |
| 7 | mgmt\_ppg\_polioprotgp | Polio Intervention Group assignment | sid |
| 8 | mgmt\_prn\_polio\_random | Randomization treatment arm for Polio vaccine | sid |
| 9 | mgmt\_psi1\_poliosi | Polio shedding index (SI) for sabin 1, 2 & 3 (Quantitative shedding index using algorithm: Refer J. Clin. Microbiol. January 2015 vol. 53 no. 1 206-211 for details) | sid |
| 10 | mgmt\_psi2\_poliosi |
| 11 | mgmt\_psi3\_poliosi |
| 12 | mgmt\_pso\_poliotrial\_sna | Derived flat table with SNA results at weeks 6, 18, 40, 53 used for polio outcome analysis | sid |
| 13 | mgmt\_pta\_poliotrial\_ae | Derived table with AE and randomization data for polio outcome analysis | sid, aerepdt, aeresdt |
| 14 | mgmt\_pto\_poliotrial\_outcome | Derived flat table with polio virus excretion PCR and qPCR results at week 14, and 52 on day 0, 4, 11, 18 & 25 used for polio outcome analysis | sid |
| 15 | mgmt\_pts\_poliotrial\_sae | Derived table with SAE data and randomization data for polio outcome analysis | sid, saedov |
| 16 | mgmt\_rrn\_rota\_random | Randomization treatment arm for Rota vaccine | sid |
| 17 | mgmt\_rta\_rotatrial\_ae | Derived table on adverse event data with randomization arm for Rota outcome analysis | sid |
| 18 | mgmt\_rtf\_rotatrial\_fup | Derived table on diarrheal episodes in infants during first 1 year of life for Rota outcome analysis | sid, epidate |
| 19 | mgmt\_rtf2\_rotatrial\_fup | Derived table on diarrheal episodes in infants during first 2 year of life for Rota outcome analysis | sid, epidate |
| 20 | mgmt\_rto\_rotatrial\_outcome | Derived table on diarrheal episodes consolidated and organized by individual infant for Rota outcome analysis | sid |
| 21 | mgmt\_rto2\_rotatrial\_outcome | Derived table on diarrheal episodes consolidated during first 2 years of life and organized by individual infant for Rota outcome analysis | sid |
| 22 | mgmt\_rts\_rotatrial\_sae | Derived table on serious adverse event data with randomization arm for Rota outcome analysis | sid, saedov |

Appendix A: Additional details on HLA data

Genotyping notes for PROVIDE shipments

1. 439 distinct samples were sent along with duplicates for 10 of the samples to compare the performance of DNA from different sources in the assays. Complete genotyping on the duplicates was not performed; data were only given for the DR loci. They have been excluded in freeze 3 table because these alleles/genotypes are identical to calls in the full HLA sequencing run for each sample.
2. All samples were given a CHORI ID upon receipt in the Noble Lab. The spreadsheet contains the Sample ID, the barcode, and the CHORI ID.
3. All calls have been reduced to the 4-digit level of resolution. This is sufficient to determine the primary amino acid sequence of the protein. Additional digits reflect silent polymorphisms and intron polymorphisms and have been truncated for clarity.
4. The two alleles in a genotype are labeled “\_1” and “\_2.” The haplotypes have NOT been phased.
5. All samples produced DNA sequence data for all loci tested. The data do not contain any failed genotypings.
6. One sample for HLA-A and one sample for HLA-B in duplicate sample have sequences that do not match the alleles in the database. These were repeated and produced the same sequence twice. We have designated them as “new?” This is not surprising as this population has not been studied extensively. These don’t appear in the file as duplicates are not included.
7. As agreed at the start of the project, for DQB1, only exon two was sequenced for budgetary reasons. Exon 2 contains the sequence that forms the peptide-binding pocket. Polymorphisms that lie in exon 3 were not tested; thus, some of the calls may represent short ambiguity strings, e.g., DQB1\*02:01 cannot be distinguished from DQB1\*02:02; DQB1\*03:01 cannot be distinguished from DQB1\*03:19.
8. The data labeled as “DRB345\_1” and “DRB345\_2” actually represent three different genes, DRB3, DRB4, and DRB5. A detailed explanation of those data follows:

**Explanation of DRB3, 4, 5 genotyping data:**

DRB1 is the major DRB gene and is the one for which HLA genotyping data are generally reported. Many other DRB genes exist (DRB2, DRB3, DRB4...DRB9). Many of these are pseudogenes; only DRB3, 4, and 5 produce a functional product. Unlike the other HLA loci, DRB3, 4, and 5 are not present on every chromosome. In fact, a maximum of one of them will be present on a given chromosome, with anywhere from zero to two in an individual. The DRB1 genes are in LD with the DRB3, 4, and 5 genes such that we know which of the "secondary" DRB (DRB3, 4, or 5) genes should be present on a chromosome with a given DRB1 gene. Here is the general pattern:

* chromosomes with DRB1\*01:xx, DRB1\*08:xx, or DRB1\*10:xx have no DRB3, 4, or 5
* chromosomes with DRB1\*03:xx, DRB1\*11:xx, DRB1\*12:xx, DRB1\*13:xx, or DRB1\*14:xx, also have a copy of DRB3
* chromosomes with DRB1\*04:xx, DRB1\*07:xx, or DRB1\*09:xx also have a copy of DRB4
* chromosomes with DRB1\*15:xx or DRB1\*16:xx also have a copy of DRB5,

A blank cell seen is because the chromosome(s) in that individual have DRB1\*01:xx, DRB1\*08:xx, or DRB1\*10:xx