**Cohort update:**

1. Participants 32\_441040 and 32\_441057 had their cohort changed from “Healthy” to HIV. This was due to these participants taking drugs for HIV (the CBC site confirmed and approved the cohort change)
2. Cohort Definitions were changed to align more with template designs
   1. Cancer Cohort:  Participants specifically enrolled with cancer
   2. HIV Cohort: Participants specifically enrolled with HIV
   3. Transplant Cohort: Participants specifically enrolled who received an organ or cell transplant
   4. IBD Cohort: Participants specifically enrolled who have IBD
   5. All other participants were grouped into the following cohorts
      1. Healthy Cohort: Participants not in the above cohorts and either report “No”, “Unknown” or “Not Reported” for all comorbidities (including cohorts above)
      2. Comorbidity Cohort: Participants not in the above cohorts and report “Yes” for any (at least one) comorbidity (including cohorts above)

Note: Depending on enrollment, a Participant could have cancer and be in either the Cancer Cohort or the Comorbidity Cohort. This determination was made at the collection site. This would also apply to HIV, IBD, and organ transplant participants respectively.

**Cohort Name Changes:**

1. Participants\_With\_Autoimmune\_Condition was removed and was replace with Participants\_in\_IBD\_Cohort

**Participants\_in\_the\_Comorbidity\_Cohort:**

1. This new file contains all the comorbidities for participants in the comorbidity cohort.

**Participants\_in\_IBD\_Cohort**

1. ICD10\_block\_description was removed as this field was redundant and confusing
2. Added Visit\_Info\_ID which was missing from v4

**Visit\_Info\_ID:**

1. Visits that where flagged as F0A, F0B, or F0C are Pre-Baseline Visits. Because these technically are not follow-up they are being changed to PB1, PB2, PB3 where PB = ‘Prior to Baseline’

**Demographics:**

|  |  |
| --- | --- |
|  | VRSS 5.0: |
| in V4.0.0 | 3271 |
| New | 46 |
| Dropped out | 27 |
| Vaccination errors | 6 |
| Total | 3,284 |

1. Changed table to order by Cohort (previously was by CBC)
2. Added Sunday\_Prior\_To\_First\_Visit
3. Added Year of First Visit to the Summary Table (Year of Participants First Visit)
4. Split ASU out into their subsites as they are different locations (was done in foresight for Columbia Data)
5. 6 participants are removed due to vaccine history not able to be determined
   1. 32\_221131: only has booster 1, missing primary series
   2. 32\_221140: only has booster 1, missing primary series
   3. 32\_221185: Dose 2 (day 0) occurs before dose 1 (day 49)
   4. 32\_221186: Dose 2 (day 0) occurs before dose 1 (day 49)
   5. 32\_221172: Booster 1 (day -21) occurs before dose 1 (day 0)
   6. 32\_221173: Booster 2 (day -50) and Booster 3 (-22) occur before dose 1 (day 4)

**Smoking\_Alcohol\_Drug:**

1. Split the follow up table by moving the pre-baseline visits into their own table. So now there are 3 tabs in the file
   1. Baseline (visit = 1)
   2. Pre-baseline (visit is less than 1)
   3. Follow up (visit is greater than 1)

**Assay\_Data:**

1. Added Assay Metadata for Assay\_ID: 14\_030 which was previously missing from V4.0.0
2. Added Assay Target data for Assay\_ID: 14\_030 which was previously missing from V4.0.0
3. Added Assay\_QC Data (new sheet)
4. In the Metadata Assays 41\_487 and 41\_488 had a data entry error that was corrected
   1. 41\_487 is now correctly labeled as COVID-19 Nucleocapsid (which aligns correctly with the assay target and Biospecimen Test Results
   2. 41\_488 is now correctly labeled as COVID-19 Spike (which aligns correctly with the assay target and Biospecimen Test Results
5. In assay Target, assay ID 41\_491 has a Spike target and a subregion of recombinant trimeric SARS-CoV-2 spike protein. This was normalized to Full Length spike and No subregion

**Biospecimen\_Test\_Results:**

1. Updated the summary chart to reflect all the different tests performed
2. Added total assays by visit (0 means no test was done) (Assays by Visit)
3. Added summary of collection to test (Collection\_To\_Test\_Interval) (Test Result Data)
4. Positive\_Control\_Reading had missing values replaced with “Not Reported”
5. Negative\_Control\_Reading had missing values replaced with “Not Reported”
6. Biospecimen\_Test\_Results\_Comments has missing values replaced with “N/A”

**Participant\_Visit\_Info:**

|  |  |
| --- | --- |
|  | VRSS 5.0: |
| in V4.0.0 | 14,336 |
| New | 1,221 |
| Dropped out | 67 |
| Total | 15,490 |

**Vaccination Data**:

1. Summarized complete vaccination history for both baseline and each follow-up visit
2. Example: Primary Series: Moderna with 1 Extra Dose and 1 Booster
3. Added columns to differentiate the various boosters between original, bivalent, XBB1.5
4. Added values for Dose 1, Dose 2, and Dose 3 to differentiate between primary and additional doses that are not boosters
5. Sheet Name “Visit Vaccination and Symptoms” was renamed to “Visit Vacc and Side Effects”

**Smoking\_Alcohol\_Drug\_Use\_5.1.0:**

1. Renamed 'Former drinker' to 'Current non-drinker' as they have the same meaning
2. Renamed ‘Former smoker’ to ‘Current non-smoker’ as they have the same meaning

**Covid\_History:**

1. Under Symptoms Replaced Missing and N/A for positive cases with “No symptoms reported”
2. Under Other\_Symptoms Replaced Missing and N/A for positive cases with “No symptoms reported”
3. Added column for the total number of symptoms

**Treatment:**

1. Changed the Condition table to number of unique conditions as well as rolled up the smaller cases (<5) for space reasons
2. The full list of which conditions a participant has, is on “Conditions by Participant”
3. Removed the tab for summarizing the treatment names as there are too many to justify (299 unique drug names once harmonized)
4. Harmonized dosage units to a standard form (lowercase values and correct units)

**Comorbidities**:

1. Split all the comorbidities into 6 new files
   1. All\_Participant\_Comorbidities and All\_Participant\_Comorbidities\_Summary
   2. All\_Participant\_Chronic\_Conditions and All\_Participant\_Chronic\_Conditions\_Summary
   3. All\_Participant\_Infections and All\_Participant\_Infections\_Summary
2. Comorbidities have ICD10 codes (either provided via EMR from each site or interpreted from the free text by Elaine and her team)
3. Clarified N/A versus not reported responses
   1. N/A: Participant does not have Condition (answered No at baseline)
   2. Not Reported: Participant did not answer the question, Unable to determine if the condition exists
4. Added data dictionary to each of the summary files

**Participants\_in\_the\_Transplant\_Cohort:**

1. Combined the Organ Transplant Comorbidity and Organ Transplant Cohort tables into one file. There was transplant data that was originally missing. Filled the Cohort specific data with “Not Reported” for conditions that only occurred in the baseline templates (148 Participants)
2. Flipped the summary table from By Site to by Cohort instead, to show that not all people who had transplants are in the transplant cohort

**Participants\_in\_the\_HIV\_Cohort:**

1. Added ASU data (EMR update) and more UMN CD4 data that was previously not included in the last release

**Participants\_in\_the\_Comorbidity\_Cohort:**

1. This new file contains all the comorbidities for participants in the comorbidity cohort.

**Participants\_in\_IBD\_Cohort**

1. ICD10\_block\_description was removed as this field was redundant and confusing
2. Added Visit\_Info\_ID which was missing from v4