**ACE/APP Data Audit: 05-01-2020**

**Breaking down missing data into categories**

|  |  |
| --- | --- |
| Category | Description |
| Single-entered data | Data that are not double entered (removed before going to data portal) |
| Duplicate data | Data pairs that have the same id+visit with different age or id+age with different visit numbers (removed before going to data portal) |
| Orphaned data | Data from which the participant id’s are not from tblSubject (master list of participants) |
| Non-primary data | Data that are from participants who are not marked as “primary” (i.e siblings, excluded) |
| Visit incomplete | Data from the participants which haven’t completed the visit within the time pont (scheduled to be collected/entered) |
| Uncertain | Data that are tracked by coordinators as non-missing but not available in ACCESS (need to look for physical copies) |
| Verified missing data | Data that are tracked by coordinators as missing |

*Table 1. Categories of missing data and definitions*

**Single-entered Data**

* Still waiting for double entry for medical data
* Brianna and Alexa interviewed/hired volunteers, but the process would be on hold due to the pandemic situation.
* Currently 120 participants are involved across medical data (GI History, Autoimmune, Physical Exam, and Tanner Staging)

**Orphaned Data**

* Orphaned data is defined as all data which id is not in the list of tblSubject (master list of primary participants).
* Brianna checked the list and figured out most of them where they came from.
* Further on, Soo will weekly put an updated list into a folder where Brianna can check and audit.

**Duplicate Data**

* Duplicate data is defined as either: 1) data pairs with **the same id and visit** (with different ages) OR 2) data pairs with **the same id and age** (with different visits)
* Same as orphaned data list, Soo will weekly put an updated list into a folder where Brianna can check and audit; the list has been small, and Brianna fixes the data quickly.
* Currently 4 participants (0.67%) are involved across ADOS, CDI-WG, and DAS.

**Suspected Missing Data (“Uncertain” category)**

* A major ongoing project is to account for suspected missing data, which is defined as following: The status of the data is **not** tracked as “Missing Data”, but the data is missing in ACCESS database, when the participant completed the visit where the data belongs to.
* Missing data with miscellaneous reasons
* Sameera, a Senior volunteer, took over this piece of data auditing. It is currently working in progress. More detailed update can come from Brianna.
* List has been shrinking! 😊

**User Logs**

* 0 bug report in April!
* Average time of using the data portal is 9.2 mins.
* The results by months are shown below:

*Table 2. Number of Sessions in April, 2020 by Username for dataportal.explore*

|  |  |
| --- | --- |
| User Name | # of Sessions |
| heath | 12 |
| dwyer | 11 |
| park | 8 |
| wunordahl | 6 |
| andrews | 4 |
| lee | 3 |
| hechtman | 1 |

*Table 3. How Many Time Tables Queried in April, 2020?*

|  |  |
| --- | --- |
| Table Name | Freq |
| tbl\_msel | 16 |
| tbl\_vineland | 13 |
| tbl\_ados | 12 |
| tbl\_cbcl | 11 |
| tbl\_srs | 11 |
| tbl\_eowpvt | 10 |
| tbl\_ppvt | 8 |
| tbl\_cdi\_wg | 7 |
| tbl\_demographics | 7 |
| tbl\_cdi\_ws | 6 |
| tbl\_scan\_details | 6 |
| tbl\_adis | 5 |
| tbl\_dm\_status | 5 |
| tbl\_longitudinal\_iq | 5 |
| tbl\_adi | 3 |
| tbl\_das | 3 |
| tbl\_mori\_regions | 3 |
| tbl\_mori\_total | 3 |
| tbl\_rbs | 3 |
| tbl\_cbcl\_preschool\_items | 2 |
| tbl\_cbcl\_school\_age\_items | 2 |
| tbl\_scq | 2 |
| tbl\_tcv | 2 |
| tbl\_dna | 1 |
| tbl\_fps | 1 |
| tbl\_growth\_measurement | 1 |