

## Breaking down missing data into categories

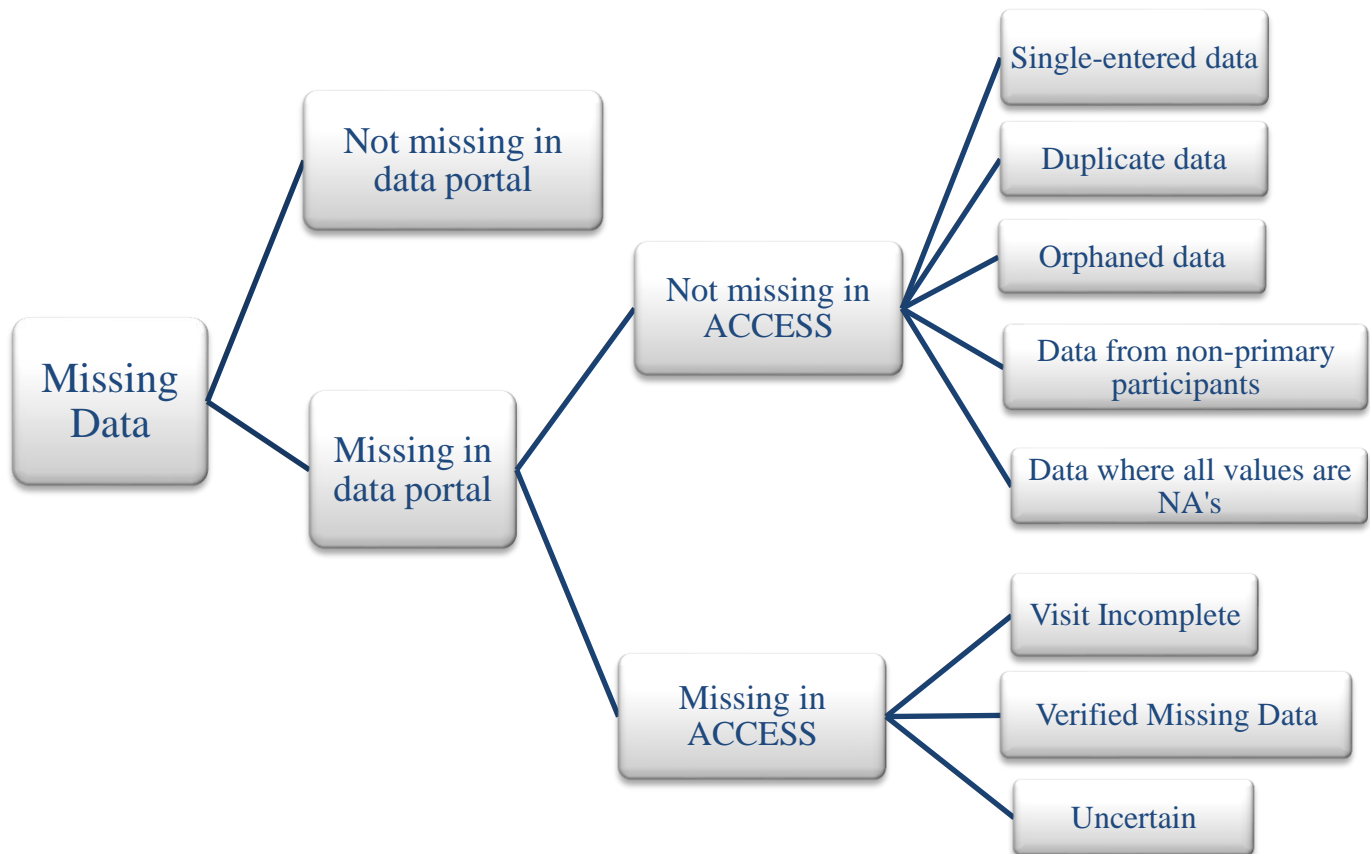


Table 1. Categories of missing data and definitions

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

## Single-entered Data

- Still waiting for double entry for medical data
- Brianna and Alexa have been interviewing/hiring volunteers.
- Currently 107 participants (17.8%) 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.
- Currently only one orphaned data that needs to get fixed with Josh's help

## 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)

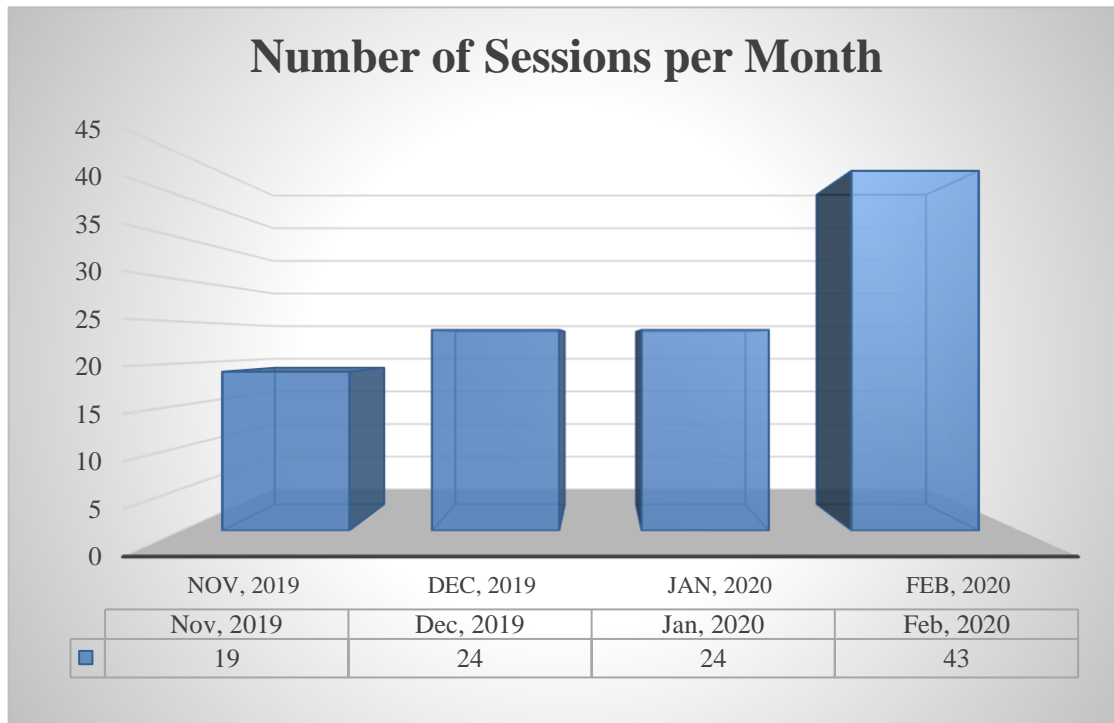
- A major ongoing project is to account for suspected missing data, which is defined as following: The status of the data is **not** tracked as “Missing Data”, but the data is missing in ACCESS database, when the participant completed the visit where the data belongs to.
- Missing data with miscellaneous 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.

## Future Pathway of Data Audit

- So far, we have mainly dealt with the quantity of the data (i.e what and why certain data are missing).
- Next steps would be focusing on the quality of the data (i.e if data we have are accurate and reliable). Brianna has some suggestions on what we can look into.

## User Logs

- 3.7% (4/110) requested bug reports
- Average time of using the data portal is 10.4 mins.
- The results by months are shown below:



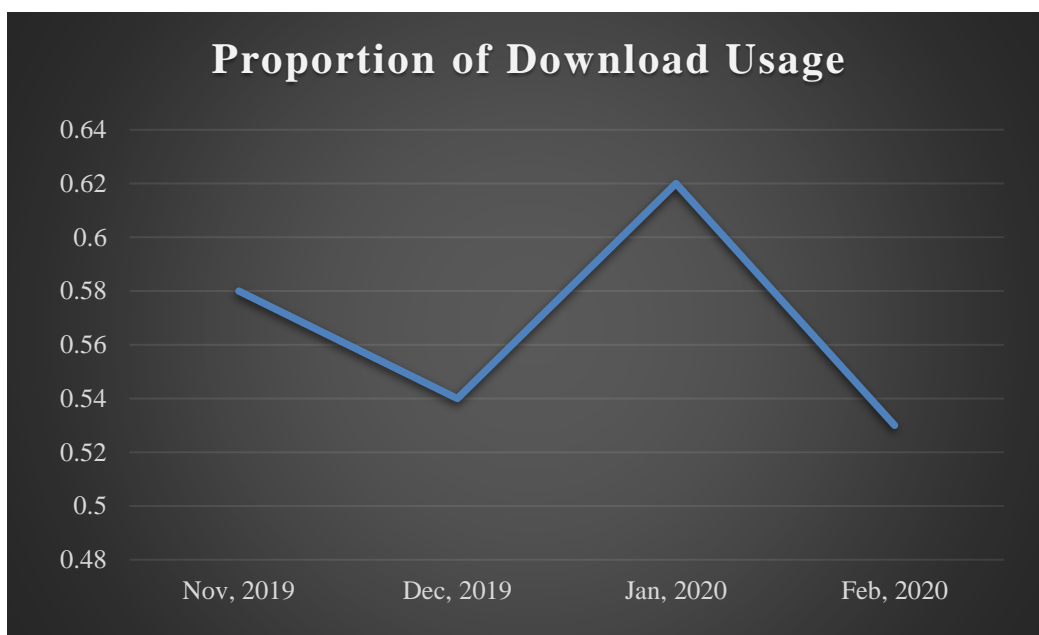
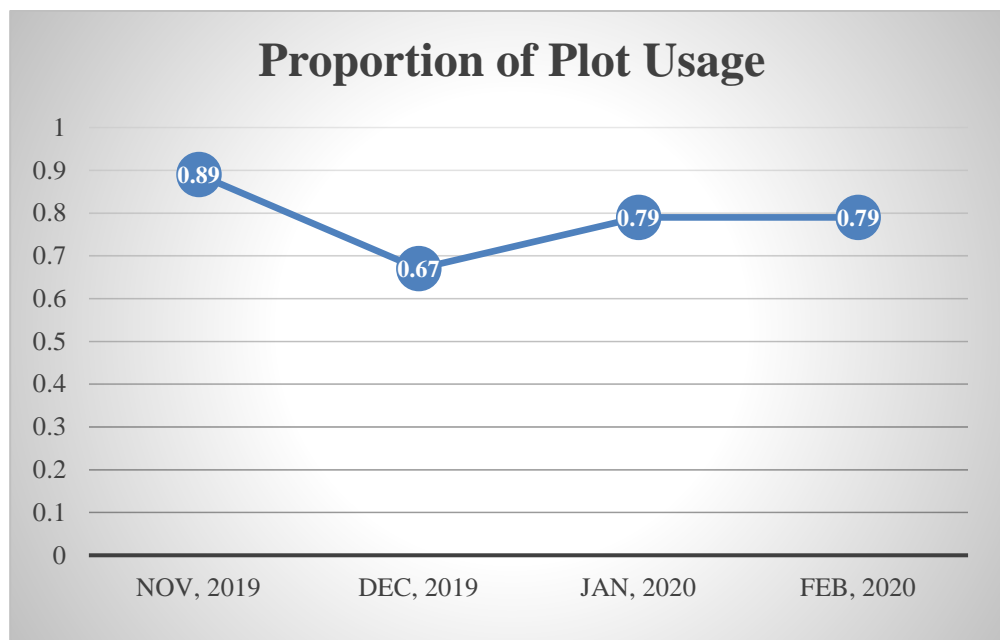


Table 2. Number of Sessions in Feb, 2020 by Username for dataportal.explore

User Name	# of Sessions
<b>Wunordahl</b>	11
<b>Lee</b>	9
<b>Dwyer</b>	7
<b>heath</b>	7
<b>young</b>	3
<b>park</b>	2
<b>andrews</b>	1

Table 3. Number of Sessions in Feb, 2020 by Username for dataportal.proposal

User Name	# of Sessions
<b>wunordahl</b>	2
<b>young</b>	1

Table 4. How Many Time Tables Queried in Feb, 2020?

Table Name	Freq
<b>tbl_ados</b>	20
<b>tbl_dm_status</b>	20
<b>tbl_scan_details</b>	20
<b>tbl_longitudinal_iq</b>	15
<b>tbl_tcv</b>	13
<b>tbl_adi</b>	7
<b>tbl_msel</b>	7
<b>tbl_cbcl</b>	6
<b>tbl_demographics</b>	3
<b>tbl_mori_total</b>	3
<b>tbl_adis</b>	2
<b>tbl_cshq</b>	2
<b>tbl_growth_measurement</b>	2
<b>tbl_mori_regions</b>	2
<b>tbl_rbs</b>	2
<b>tbl_scq</b>	2
<b>tbl_srs</b>	2
<b>tbl_ssp</b>	2
<b>tbl_vineland</b>	2
<b>tbl_das</b>	1

<b>tbl_dsm</b>	1
<b>tbl_masc</b>	1
<b>tbl_masc_parent</b>	1
<b>tbl_nih_toolbox</b>	1
<b>tbl_tanner_staging</b>	1