



# RADx-UP CDCC Data Submission Guidance

*Updated May 11, 2023*

## Table of Contents

Part 1: Start Up Information.....	4
Introduction .....	4
myRADx-UP Home .....	4
Data Managers.....	4
Upload Frequency .....	4
File Type/Delimiters .....	4
File Scope and Deleting Files.....	4
Data Sharing Concerns.....	5
Consent Variables: .....	6
Participant & Event Identifiers .....	9
• Record_ID:.....	9
• Longitudinal Data - Event Identifiers: .....	<b>Error! Bookmark not defined.</b>
• Key Uniqueness: .....	<b>Error! Bookmark not defined.</b>
CDE Conformance.....	11
• Overall Conformance:.....	11
• Dropdown/Radio Fields: .....	11
• Multi-Select (Checkbox Elements):.....	11
• Dates/Timestamps: .....	12
• Text Fields with Other Validation: .....	12
• Text Fields: .....	12
• Height Fields:.....	12
Mapping Guidance .....	12
• Exact Matches: .....	12
• Modified Language, Non-Modified Meaning: .....	12
• Modified Code Lists Mapped Back to CDE: .....	12

- Examples of Mapping: Adding Answer Options ..... 12
- Examples of Mapping: Adding Answer Options ..... 13
- CDE Versions ..... 13
- FOR Version 1.0 CDEs: SAS Synonyms..... 14
- Data Quality Report ..... 14
- Part 2: Uploading Data & Data Quality.....17
- How to Submit Data.....17
- Submission Status Definitions..... 19
- Data Quality Report Definitions ..... 20
  - Table of Contents: ..... 22
  - Data Quality Summary:..... 22
  - CDE Conformance Summary:..... 23
  - CDE Values Outside of Conformant Values: ..... 23
  - CDEs Expected but not Received:..... 24
  - CDEs Received but not Expected:..... 27
  - CDEs with Failed Branching Logic: ..... 27
  - Summary Statistics: Categorical Variables: ..... 28
  - Summary Statistics: Continuous Variables: ..... 29
- Data Cleaning Guidance: values Outside Expected Range: ..... 29
  1. Fixing Data:..... 29
  2. Remove Data:..... 30
  3. Valid Data: ..... 30
- Data Cleaning Guidance: CDE Expectations ..... 30
- Data Cleaning Guidance: Branching Logic Failures..... 30
- Upload Troubleshooting Guide ..... 32
  - If Rejected: ..... 32
  - If Incomplete: ..... 32
- Part 3: Data Quality Resolution Requests..... 33
  - Overview ..... 33
  - Pagination..... 34
  - Description of Columns: Conformance Table ..... 34
  - Description of Columns: Missingness ..... 36
  - Description of Columns: Branching Logic..... 37

Examples ..... 39

    True and Valid Data: A Participants Age is 114. .... 39

    Submitted Exception: household\_congregate\_2..... 41

    Address and Fix Issue: age\_yrs is 1413. .... 43

    Example: Fixing a Missing CDE..... 44

## Part 1: Start Up Information

### INTRODUCTION

This document is intended to provide guidance to RADx-UP projects on the study data upload process and specify transfer requirements for NIH RADx-UP Tier1 and Tier2 Common Data Elements (CDEs). Collecting harmonized data across the consortium is a key requirement of the RADx-UP project, both to ingest the data properly and consistently understand the meaning of the data collected.

The data transfer process is designed to be as straightforward as possible. Given that the majority of RADx-UP projects are using REDCap, most guidance has been aligned with REDCap processes. Users will find that many of the requirements are met by default when exporting data from a REDCap project.

### MYRADX-UP HOME

*myRADx-UPhome* is a secure web portal that all RADx-UP projects will have access to. You will be able to upload study data to the CDCC through this portal. **Please do not use email or other means to share any participant-level study data with the CDCC.**

### DATA MANAGERS

A few members of your project team, chosen by the project principal investigator (PI), will be given access to upload study data via *myRADx-UPhome*. These users will also receive data upload notifications, reminders, and associated feedback. Please contact your Engagement Impact Team (EIT) to make changes to the list of users who can upload data from your project.

### UPLOAD FREQUENCY

Projects are encouraged to upload data to *myRADx-UPhome* on a monthly basis. Uploading frequently allows your data to be included in NIH dashboards, RADx-UP Data Dashboard and in the data sets for RADxUP Consortial Publications. Additionally, frequent uploads give you ample time to resolve the identified data quality issues.

If it has been more than 90 days from your last upload it could result in a meeting with your NIH PO. Please contact our team if you are not able to upload for more than 90 days. Your project may request a limited data upload pause for reasons related to project design or unexpected circumstances such as staffing shortages, natural disasters, etc.

### FILE TYPE/DELIMITERS

Projects are currently expected to upload a single comma separated value file (.csv) with column headers in the first row. Please save the file with UTF-8 encoding to ensure the upload is successful.

If you are using REDCap, you will be able to select both the file type (.csv) and the delimiter (comma or pipe) as part of the export process. During the import to the portal, you will be able to select the delimiter matching your file type as part of the upload process. **Be sure to export as a "raw" file, not as "labeled".**

### FILE SCOPE AND DELETING FILES

All study data files uploaded to *myRADx-UPhome* are expected to be cumulative. This will help reduce complexity for the projects by eliminating the need to determine changes since the last upload.

In addition, the CDCC will request data managers enter start and end dates of data collection during the data upload process. Start date is the first date of collection; end date is the last date of active data collection. While the study data collection is ongoing, end date corresponds to the date on which the data was extracted from the source data system.

You have the option now to delete prior uploads. Please only delete files that were uploaded in error such as non-consented data or variables outside your DTA or IRB approval. We store file history of previous uploads unless they are deleted. **You do not need to delete uploads that were Incomplete/Rejected/Technical Errors.**

### DATA SHARING CONCERNS

**Please upload only data elements that are covered under a participant’s consent and allowed by the data transfer agreement your organization has established with Duke.** If certain participants did not consent to share their data with the CDCC, please exclude their data from the study data files uploaded to *myRADx-UPhome*. If your study design allows optional sharing of additional identifying information with the CDCC, please design your process so that the optional shared data matches the individual consent.

If you are anonymizing your participant identifiers (record\_id column in the file), please ensure that the identifiers remain consistent between data uploads. This helps us identify any potential quality and data integrity issues over time.

You are responsible for ensuring the de-identification standards of your DTA or IRB are met. The table below is the CDCC guidelines based on the [HHS guidelines](#) on which CDEs are not allowed by type of DTA. **Please check your own DTA and contact your IRB or other officials to ensure your upload file meets the required standards for de-identification.**

DTA security level	Data NOT Allowed	CDEs NOT Allowed
PHI	all data allowed	all allowed
Limited	Names, street address (city, county, state, and zip code allowed), Telephone / Fax numbers, email, ssn, Medical record numbers, Health plan beneficiary numbers.	first_name, last_name, current_street, current_street2, mobile_phone, home_phone, other_phone, personal_email, other_email, ssn, mrn
Deidentified	Names, any geographic location smaller than state, all elements of date except year, all ages over 89 and any dates indicative of that age, telephone numbers, facsimile numbers, email address, ssn, medical record numbers, certificate/license numbers, vehicle identifiers, device identifiers, URLs, IP addresses, biometric identifiers, full face photographic images, any other unique identifying number, characteristic, or code	first_name, last_name, current_street, current_street2, mobile_phone, home_phone, other_phone, personal_email, other_email, ssn, mrn, dob_mdy, age_yrs > 89, zip_code, current_county, current_city, consentdt_mdy, sociodem_date_mdy, housing_date_mdy, work_ppe_date_mdy, med_hx_date_mdy, hlthstat_date_mdy, vacc_date_mdy, test_date_mdy, positivemonth_covidtest,

		recentmonth_covidtest, covid_test_date_mdy, covid_test_collect_datetime, covid_test_result_datetime, covid_test_result_sent_datetime, sym_date_mdy, alcohol_date_mdy, iden_date_mdy, med_date_mdy, drg_date_mdy, disability_date_mdy
--	--	---

### Consent Variables:

Please note that the “identifiable information” specified on the consent form is “name, address, contact information, and date of birth.”

You must make sure these variables align with your consent form. The below tables are scenarios where participant consents could potentially conflict, and so the CDCC is removing data to be certain nothing is shared beyond what is consented. You should review with your PI and IRB to determine what data you deem shareable in each scenario. If you deem the data is sharable, you must adjust the variable responses to allow the data through. We apologize for this inconvenience, but we are committed to ensuring all data is consented.

Before uploading, you should always review your consent variables and ensure you are not sharing data the participant did not consent to share. Please either adjust the branching logic for these values or remove the data specified as not allowed. **You should not be asking participants consent\_zip without also asking consent\_ident.**

The identifiable CDEs are: *zip\_code, current\_county, age\_yrs > 89, first\_name, last\_name, current\_street, current\_street2, current\_city, mobile\_phone, home\_phone, other\_phone, personal\_email, other\_email, ssn, mrn, dob\_mdy*

*If you are using versions later than 1.5:* There are also participant level consent variables within the code book (*consent\_ident, consent\_zip, consent\_ssn*). The CDE *consent\_ident* reads “I agree to let The Duke Clinical Research Institute to collect all identifiable information” while *consent\_zip* reads “I agree to let The Duke Clinical Research Institute to collect only my zip code and no other identifiable information”. This could lead to conflicts depending on the answers given by the participant.

The CDCC advises that *consent\_zip* should only be answered if *consent\_ident* is no. IF your project team agrees that is consistent with the language in your consent form, you can add this as branching logic (*show consent\_zip ONLY IF consent\_ident = o*) to your surveys to avoid these issues in the future. You may version up to the CDE “*consent\_zip\_2*” which reads “I agree to let the Duke Clinical Research Institute collect my zip code” and removes the issue.

*If you are using versions after 1:5:* There are also participant level consent variables within the code book (*consent\_ident, consent\_zip\_2*). If participants say “No” to sharing data, you must remove the corresponding identifiable data.

IF you are asking participants all three consents (versions earlier than 1.5), refer to the following table:

<b>consent_ident</b>	<b>consent_zip</b>	<b>consent_ssn</b>		
I agree to let The Duke Clinical Research Institute collect all identifiable information.	I agree to let The Duke Clinical Research Institute to collect only my zip code and no other identifiable information.	I agree to let The Duke Clinical Research Institute collect my Social Security number.	<b>CDEs Allowed in Upload</b>	<b>Defer to your IRB</b>
Yes	Yes	Yes	<i>Zip code only</i>	Yes
Yes	Yes	No	<i>Zip code only</i>	Yes
Yes	No	No	<i>All identifiers except SSN</i>	No
Yes	No	Yes	<i>All identifiers</i>	No
No	No	No	<i>No identifiers</i>	No
No	No	Yes	<i>SSN only</i>	No
No	Yes	No	<i>Zip code only</i>	No
No	Yes	Yes	<i>Zip code only</i>	Yes

If you are asking participants only consent\_ident and consent\_zip (versions earlier than 1.5) :

<b>consent_ident</b>	<b>consent_zip</b>		
I agree to let The Duke Clinical Research Institute collect all identifiable information.	I agree to let The Duke Clinical Research Institute collect only my zip code and no other identifiable information.	<b>CDEs Allowed in Upload</b>	<b>Defer to your IRB</b>
Yes	Yes	<i>Only zip code</i>	Yes
Yes	No	<i>All identifiers</i>	No
No	Yes	<i>Zip code only</i>	No
No	No	<i>No identifiers</i>	No

If you are asking participants only consent\_ident and consent\_zip\_2 (versions later than 1.5) :

<b>consent_ident</b>	<b>consent_zip_2</b>		
I agree to let The Duke Clinical Research Institute collect all identifiable information.	I agree to let The Duke Clinical Research Institute collect my zip code.	<b>CDEs Allowed in Upload</b>	<b>Defer to your IRB</b>
Yes	Yes	<i>All identifiers</i>	Yes
Yes	No	<i>All identifiers except Zip code</i>	No
No	Yes	<i>Zip code only</i>	No
No	No	<i>No identifiers</i>	No

If you are asking participants only consent\_ident and consent\_ssn (versions earlier than 1.5) :

<b>consent_ident</b>	<b>consent_ssn</b>		
I agree to let The Duke Clinical Research Institute collect all identifiable information.	I agree to let The Duke Clinical Research Institute collect my Social Security number.	<b>CDEs Allowed in Upload</b>	<b>Defer to your IRB</b>
Yes	Yes	<i>All</i>	No
Yes	No	<i>All except SSN</i>	No
No	Yes	<i>SSN only</i>	No
No	No	<i>No identifiers</i>	No

If you are asking participants only consent\_ident :

<b>consent_ident</b>		
I agree to let The Duke Clinical Research Institute collect all identifiable information.	<b>CDEs Allowed in Upload</b>	<b>Defer to your IRB</b>



Yes	<i>All identifiers</i>	No
No	<i>No identifiers</i>	No

## PARTICIPANT & EVENT IDENTIFIERS

- **Record\_ID:** record\_id is always expected to be present for each row of data and should contain a unique participant ID. A record\_id should always uniquely identify a participant within a study upload file. Lack of this uniqueness could lead to wrong enrollment numbers being reported to NIH and cause other issues.

Record\_id should be anonymized but the de-identified id should be consistent across data uploads for a given participant.

Record\_ids can be alphanumeric but should not contain any spaces or other special characters apart from dashes and underscores.

*For Supplemental Projects:* Your record IDs should be consistent between your original projects and your new project. If Participant 1 is in the original project, they should also be Participant 1 in the new project. Record IDs should only overlap if it is the same participant in your new study, otherwise they should be unique numbers.

If you are collecting longitudinal data or cross-sectional data, we will ask you to fill out a spreadsheet explaining your data structure. Your informaticist will provide this spreadsheet and provide guidance on your structure. **Please discuss with the CDCC before you begin collecting data.**

- **Longitudinal Data: You are collecting longitudinal data if you are collecting a CDE more than once per participants. Your data would contain multiple records per participant, where each record represents a distinct event/visit.** Your data file would then have multiple rows of data for each participant, for example one row/record for a baseline visit where you collect sociodemographic and perform test 1, another row/record for follow-up tests. If you are collecting longitudinal data, please include redcap\_event\_name, redcap\_repeat\_instrument and redcap\_repeat\_instance columns as needed in your study data file.
  - We ask that any CDEs (specifically sociodemographics) that are being collected only once per participant be collapsed into the first visit performed. For example, if you collect baseline information as redcap\_event\_name = 'baseline' and then perform the first test as redcap\_event\_name = 'test1', you should combine these two rows into the 'baseline' redcap\_event\_name row. New rows should only be used when CDEs are collected another time.
  - We ask that any CDEs (specifically sociodemographics) that do not change between visits be collected only once.

- Please refer to Table 1 below to understand which columns are required based on the REDCap functionality (Data Collection Type) you are using to collect longitudinal data.

Table 1

	Columns in the file			
Data Collection Type	record_id	redcap_event_name	redcap_repeat_instrument	redcap_repeat_instance
Single event	Required			
Multiple events + No repeating forms	Required	Required		
Single event + Repeating forms	Required		Required	Required
Multiple events + Repeating forms	Required	Required	Required	Required

- If you are using a non-REDCap data system, please work with the Data Core Informaticists to develop your data file.
- **Cross-Sectional Data:** You are collecting cross-sectional data if all CDEs are only captured once per participant and each record contains information associated with one participant. This applies to data structure where CDEs are collected only once (e.g. demographic information). Your data should contain each CDE once per participant and your data file should have only one row for that participant. Any CDEs that are collected only once may be combined into one row. Additional rows/events should not be used to distinguish sections.
  - For example, do not use one redcap\_event\_name for demographics, one for medical history, etc. If no CDEs are collected multiple times, this should be one row of data.
- **Key Uniqueness:** In addition, for repeating and multiple event data, depending on the REDCap functionality you are using, the columns marked as "Required" in the table above must be unique for each row of data in your file.
  - For example, for multiple event data with no repeating forms, the record\_id and redcap\_event\_name together are unique and occur only once within a file.

## CDE CONFORMANCE

- Overall Conformance:** When a column name/header in your study file matches a CDE name (column "Variable" in CDE codebook), the data values in the column should match data type and/or the code list in the CDE codebook. To achieve this conformance, please ensure that your code lists and data types exactly match the ones specified in the CDE guidance. If they do not, please work with your EIT and primary informaticist(s) to resolve this.
- Dropdown/Radio Fields:** Values in the columns that are dropdown or radio button fields should be the coded values or "raw" values. Additionally, the raw values in each of the columns should match the options specified in the CDE guidance codebook. To achieve this conformance, please ensure that your code lists exactly match the code lists in the CDE guidance. If they do not, please work with your EIT and primary informaticist(s) to resolve this.

Example: 1, 0, 98 and 99 are the expected values in the column "pregnancy\_status".

Additionally, 1, 0, 98 and 99 in the column "pregnancy\_status" should always map to the options "pregnant", "Not Pregnant", "Don't know" and "Prefer not to answer" respectively as specified in the CDE guidance.

- Multi-Select (Checkbox Elements):** CDEs which are setup as a (multiselect) checkbox should be configured with the REDCap modification of appending \_\_\_X (three underscores plus the numbered value of the option).

Data within each column is either 1 or 0, with 1 indicating that the option was selected and 0 indicating the option was not selected.

Example: In the CDE race\_ethn\_asian\_detail, columns are actually named race\_ethn\_asian\_detail\_\_\_1, race\_ethn\_asian\_detail\_\_\_2, race\_ethn\_asian\_detail\_\_\_3, race\_ethn\_asian\_detail\_\_\_4 and race\_ethn\_asian\_detail\_\_\_5. Each of these columns has an expected value of either 1 or 0.

record_id	race_ethn_race__1	race_ethn_race__2	race_ethn_race__3
1234	0	1	0
1235	0	0	1
1236	0	1	1

Additionally, race\_ethn\_asian\_detail\_\_\_1, race\_ethn\_asian\_detail\_\_\_2, race\_ethn\_asian\_detail\_\_\_3, race\_ethn\_asian\_detail\_\_\_4, race\_ethn\_asian\_detail\_\_\_5 should map to Japanese, Filipino, Chinese, Korean and Other Asian respectively as specified in the CDE guidance.

When using a multiselect CDE, **all** columns must be submitted to the CDCC in your upload file for the CDE to count as submitted. If even one column is missing from the CDE, the CDE will be counted as missing. For example, if your upload file contains race\_ethn\_asian\_detail\_\_\_2 through race\_ethn\_asian\_detail\_\_\_5 but does not contain race\_ethn\_asian\_detail\_\_\_1, the CDE will show as missing in your Data Quality Report.

- **Dates/Timestamps:** All date fields should be in the REDCap YYYY-MM-DD format. For example, 2:28pm on June 1st 2021 would be passed as "2021-06-01 14:28".
- **Text Fields with Other Validation:** Data collected in texts fields with validation (zip codes, phone number) should contain only valid expected values.  
 Example: Values in the column zip\_code are expected to be valid US Zip Code values.  
 Values in the column age\_yrs are expected to be integers ranging from 0 to 110.
- **Text Fields:** Data collected in text fields without validation should be enclosed in quotation marks (") if they contain comma characters (,). If you are using REDCap, quotation marks are automatically added when data is downloaded using Firefox. Please ensure your file has fields with commas enclosed in quotation marks before uploading the file.
- **Height Fields:** We expect height fields to be split across two CDEs: *self\_reported\_height\_feet* and *self\_reported\_height\_inches* or *self\_reported\_height\_meters* and *self\_reported\_height\_centimeters*.

#### MAPPING GUIDANCE

- **Exact Matches:** CDEs which match exactly in language and code list values should be labeled with the original CDE variable name.
- **Modified Language, Non-Modified Meaning:** CDEs with modified language that does not change intent or scope of the question should still use the original CDE variable name.
- **Modified Code Lists Mapped Back to CDE:** CDEs with modified code lists, but all answers were mapped to the original CDE options should still use the original CDE variable name. Examples of this include inserting additional options into the 'Other' category and populating the specify field if available.
- **Examples of Mapping: Adding Answer Options** Project is collecting greater detail on individual Hispanic ethnicities than specified in the CDE codebook, and the participant has selected 'Ecuadorian' from an expanded list (see figure 1 below).

Please populate the term 'Ecuadorian' in the race\_ethn\_hispanic\_other field, but also populate race\_ethn\_hispanic with a '1' and race\_ethn\_hispanic\_detail with a '4'. These data are also known

based on the participant's answer.

17	race_ethn_hispanic	Are you of Hispanic, Latino, or Spanish origin?	<table border="1"> <tr><td>radio</td></tr> <tr><td>0 No, not of Hispanic, Latino, or Spanish origin</td></tr> <tr><td>1 Yes, of Hispanic, Latino, or Spanish origin</td></tr> <tr><td>99 Prefer not to answer</td></tr> </table> <p>Field Annotation: PX011901   <a href="https://www.phenxtoolkit.org/protocols/view/11901">https://www.phenxtoolkit.org/protocols/view/11901</a>   U.S. Census Bureau, Census 2020, Questionnaire   Reduced to OMB, Added 99, Prefer not to answer</p>	radio	0 No, not of Hispanic, Latino, or Spanish origin	1 Yes, of Hispanic, Latino, or Spanish origin	99 Prefer not to answer					
radio												
0 No, not of Hispanic, Latino, or Spanish origin												
1 Yes, of Hispanic, Latino, or Spanish origin												
99 Prefer not to answer												
18	race_ethn_hispanic_detail_2 Show the field ONLY if: [race_ethn_hispanic] = '1'	Please specify your origin	<table border="1"> <tr><td>checkbox</td></tr> <tr><td>1 race_ethn_hispanic_detail_2__1 Mexican, Mexican Am., Chicano</td></tr> <tr><td>2 race_ethn_hispanic_detail_2__2 Puerto Rican</td></tr> <tr><td>3 race_ethn_hispanic_detail_2__3 Cuban</td></tr> <tr><td>4 race_ethn_hispanic_detail_2__4 Another Hispanic, Latino, or Spanish origin</td></tr> <tr><td>5 race_ethn_hispanic_detail_2__5 Salvadoran</td></tr> <tr><td>6 race_ethn_hispanic_detail_2__6 Dominican</td></tr> <tr><td>7 race_ethn_hispanic_detail_2__7 Colombian</td></tr> <tr><td>99 race_ethn_hispanic_detail_2__99 Prefer not to answer</td></tr> </table> <p>Field Annotation: @NONEOFTHEABOVE = '99'   PX011901</p>	checkbox	1 race_ethn_hispanic_detail_2__1 Mexican, Mexican Am., Chicano	2 race_ethn_hispanic_detail_2__2 Puerto Rican	3 race_ethn_hispanic_detail_2__3 Cuban	4 race_ethn_hispanic_detail_2__4 Another Hispanic, Latino, or Spanish origin	5 race_ethn_hispanic_detail_2__5 Salvadoran	6 race_ethn_hispanic_detail_2__6 Dominican	7 race_ethn_hispanic_detail_2__7 Colombian	99 race_ethn_hispanic_detail_2__99 Prefer not to answer
checkbox												
1 race_ethn_hispanic_detail_2__1 Mexican, Mexican Am., Chicano												
2 race_ethn_hispanic_detail_2__2 Puerto Rican												
3 race_ethn_hispanic_detail_2__3 Cuban												
4 race_ethn_hispanic_detail_2__4 Another Hispanic, Latino, or Spanish origin												
5 race_ethn_hispanic_detail_2__5 Salvadoran												
6 race_ethn_hispanic_detail_2__6 Dominican												
7 race_ethn_hispanic_detail_2__7 Colombian												
99 race_ethn_hispanic_detail_2__99 Prefer not to answer												
19	race_ethn_hispanic_other Show the field ONLY if: [race_ethn_hispanic_detail_2(4)] = '1'	Please specify other Hispanic, Latino, or Spanish origin. For example, Salvadoran, Dominican, Colombian, Guatemalan, Spaniard, Ecuadorian, etc.	<table border="1"> <tr><td>text</td></tr> <tr><td>Field Annotation: PX011901</td></tr> </table>	text	Field Annotation: PX011901							
text												
Field Annotation: PX011901												

Ecuadorian ↑

- Examples of Mapping: Adding Answer Options** Project added an answer option to the question household\_famgen: "7 = Living with grandparents". In this case, the new answer option can be mapped back into CDE approved "None of These". Before submitting data, the project can change all values of "7" to "90".

household_famgen_2	What best describes the people at your home:	<table border="1"> <tr><td>radio</td></tr> <tr><td>1 Just me</td></tr> <tr><td>2 Living with spouse, no kids</td></tr> <tr><td>3 Family including kids</td></tr> <tr><td>4 Family with 3 generations (parents, children, grandchildren)</td></tr> <tr><td>5 Family with 4 generations</td></tr> <tr><td>6 Living with roommates</td></tr> <tr><td>90 None of these</td></tr> </table> <p>Field Annotation: @NONEOFTHEABOVE = '90'     Recommendation from RADx-UP projects. Update, added roommates as an option</p>	radio	1 Just me	2 Living with spouse, no kids	3 Family including kids	4 Family with 3 generations (parents, children, grandchildren)	5 Family with 4 generations	6 Living with roommates	90 None of these
radio										
1 Just me										
2 Living with spouse, no kids										
3 Family including kids										
4 Family with 3 generations (parents, children, grandchildren)										
5 Family with 4 generations										
6 Living with roommates										
90 None of these										

### CDE VERSIONS

There are multiple versions of CDEs released, with each new version reflecting project feedback and requests. You can find details on the CDE versions, including change logs and codebooks, here: <https://radx-up.org/learning-resources/cdes/#files>. Incorporating new versions of the CDEs is **not required** for any RADx-UP project already collecting data, but incoming projects should use the most recent version available.

For projects that are already uploading data and are considering changing from one version of CDEs to another, you do not have to incorporate every new version change. You can choose to limit your changes to a few CDEs that you like in the newer version. For example, maybe you choose to change to “race\_ethn\_hispanic\_detail\_2” but do not use any of the other Version 1.1 CDEs – that is allowed. Feel free to mix and match the new versions to find the CDE version best suited for your project.

Please note that you can submit multiple versions of the same CDE. If the first half of your project uses 1.0, but the second half uses 1.1, you can submit both versions. However, a single participant may only have data for one version of the CDEs, or it will show up as an error in your Data Quality Report.

Version 1.5 can be uploaded today, and the CDCC will be ready to accept Versions 1.6 and 1.7 uploads in spring of 2023.

### FOR VERSION 1.0 CDES: SAS SYNONYMS

The CDE synonyms described in this section are completely optional. If you are using SAS to create the file and are encountering issues with the length of the CDE variable names, you may want to use these synonyms.

For CDEs over 32 characters, SAS truncates their variable names, so the following synonyms are shorter versions of the CDEs that meet the SAS length requirement. You can now use these synonyms to replace the original longer CDE names in your database and the upload file. Note that synonym data will appear in reports under the original longer CDE name.

You should NOT pass both the abbreviated CDE and the original CDE. If your file contains both the abbreviated synonym CDE and the original CDE, it will fail to upload.

New shortened CDE synonym	Original longer CDE (still valid to submit)
cov_pan_chlng_trans	covid_pandemic_challenges_transportation
self_rpt_hlth_stat_asses	self_reported_health_status_assessment
cov_pan_chal_med	covid_pandemic_challenges_medications
cov_pan_chal_hlth	covid_pandemic_challenges_healthcare
cov_tst_col_set_oth	covid_test_collection_setting_other
cov_tst_perf_loc_oth	covid_test_performed_location_other
cov_tst_spec_col_oth	covid_test_specimen_collector_other

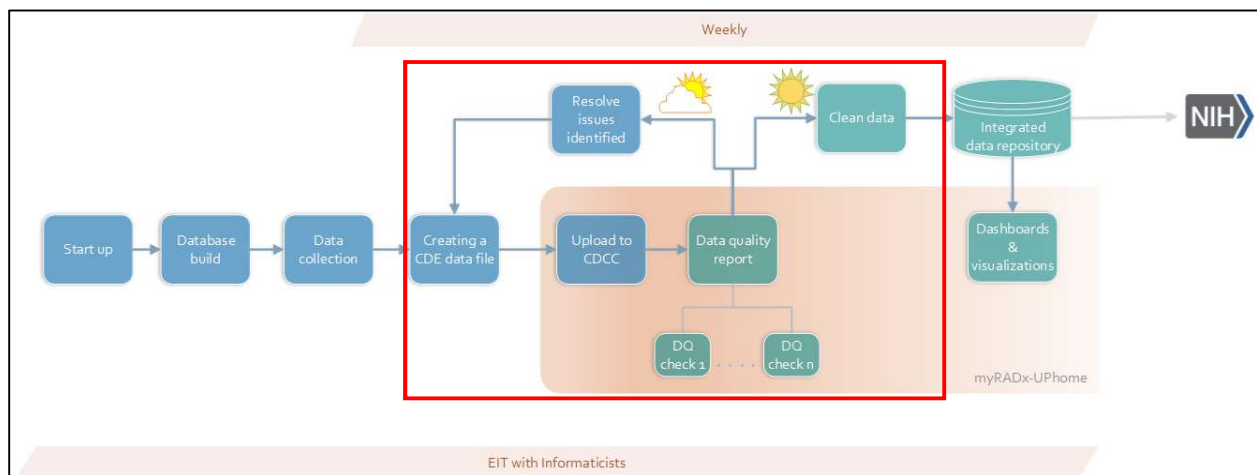
### DATA QUALITY REPORT

After uploading a file to the CDCC, a data quality report will be generated in real time in the “Study Data File History” table. It will be a hyperlink reading “Data Quality Report”. This report contains the results of data quality checks, with more coming in the future.

- **Adherence to Data Submission Guidance:** These checks verify the file’s adherence to the type and structure requirements listed in this document and in the RADx-UP Data Dictionary.
  - Seen in Study Data File History Status
  - Example: Do you have any characters in a column that should be integer?
- **CDE Conformance:** variable names, response values
  - Embedded in Data Quality Report

- Example: The CDE "age\_yrs" has allowed values 0-110. A participant with age\_yrs = 200 will trigger this.
- **CDE exception process:** ensure projects are sharing all the expected CDEs per the REDCap NIH Exception tracking form
  - Embedded in Data Quality Report
  - Example: Does your file contain all the CDEs your tracking form says you are submitting?
- **Logical data checks:** ensure data follows the branching logic described in the codebook
  - Embedded in Data Quality Report
  - Example: did a participant select "Yes" to "Have you have received a flu vaccine in the last six months" but also selected "No" to "Have you ever received a flu vaccination?"
- **CDE summary statistics:** Report summary statistics (Frequency distributions and eight-number summaries) on Tier 1 CDEs
  - Embedded in Data Quality Report

Below is a diagram of the process of readying and uploading data to the CDCC and eventually to NIH. The goal of the data quality report is for you to review the report, clean your data, and resubmit a clean file.



**You should do this within four weeks (28 days) of receiving your report. You will receive formal email notifications about this timeline. You must respond to these emails or use the Data Quality Resolution Table to avoid further escalation.**

Using the Data Quality Resolution Table (found under the Data Management tab), data managers can view issues shown in the Data Quality and address these issues. You can use this feature to inform the CDCC if anything was incorrectly flagged as a conformance issue, if it is valid data, if you added an answer option via a Wording Request Change form, or if this will not be corrected until a future aim. You can also choose a timeline for fixing any issues if you foresee needing over a month to fix these items.

If you have any questions on errors reported, reach out to your EIT/primary informaticist and we will get to back you through email or schedule a time to meet. We are here to help you with this process!



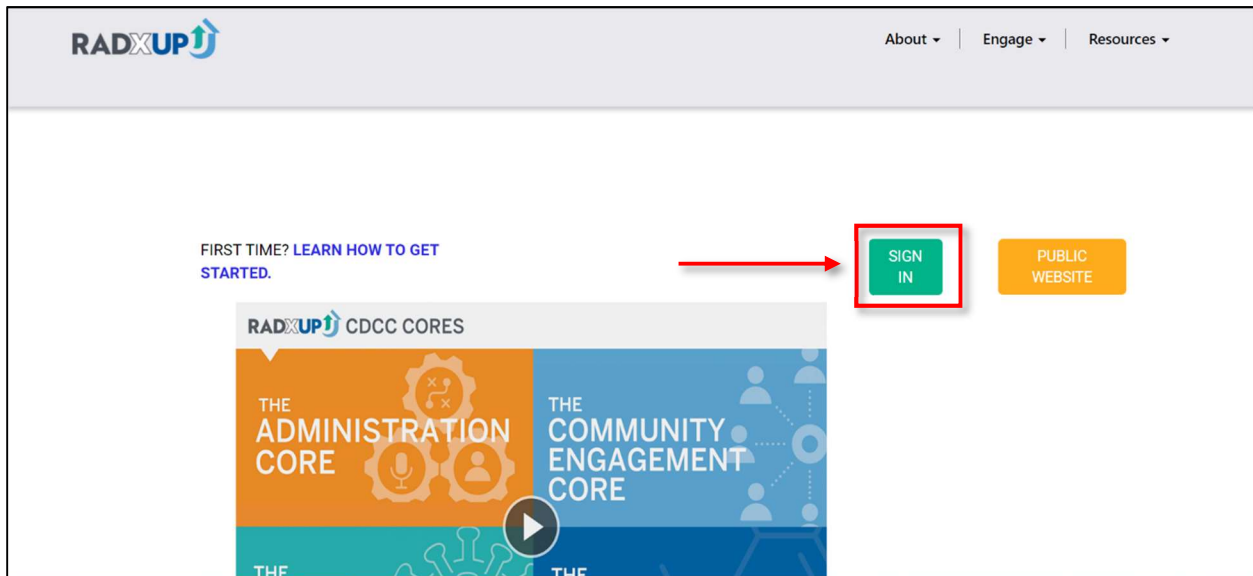


## Part 2: Uploading Data & Data Quality

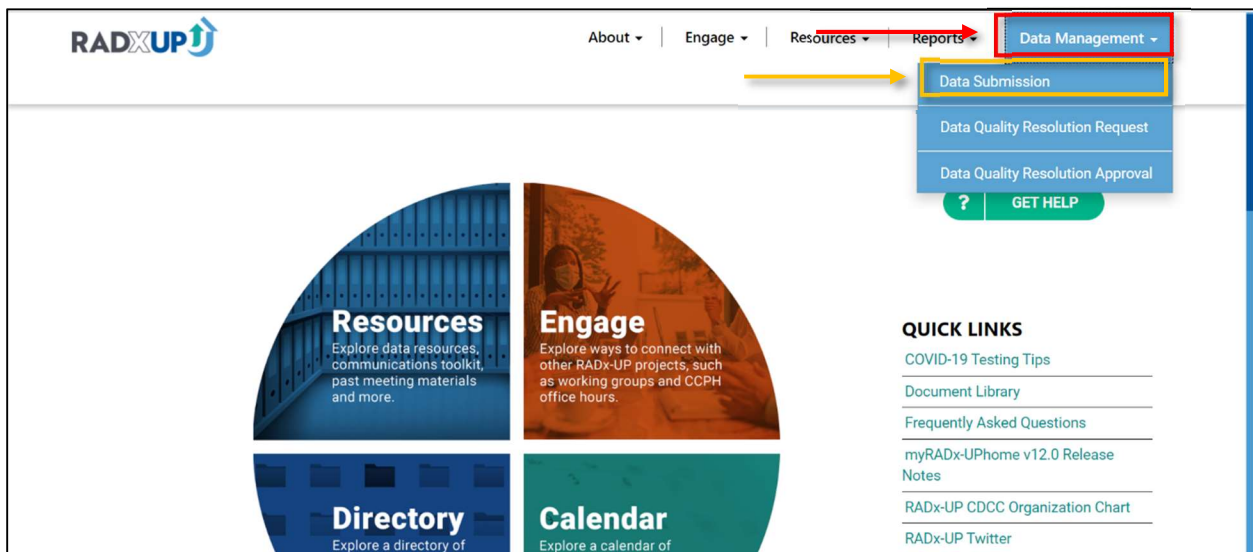
**Note:** The screenshots in this section may be out of date, but the information will remain the same. This section was last updated 2/22/2022.

### HOW TO SUBMIT DATA

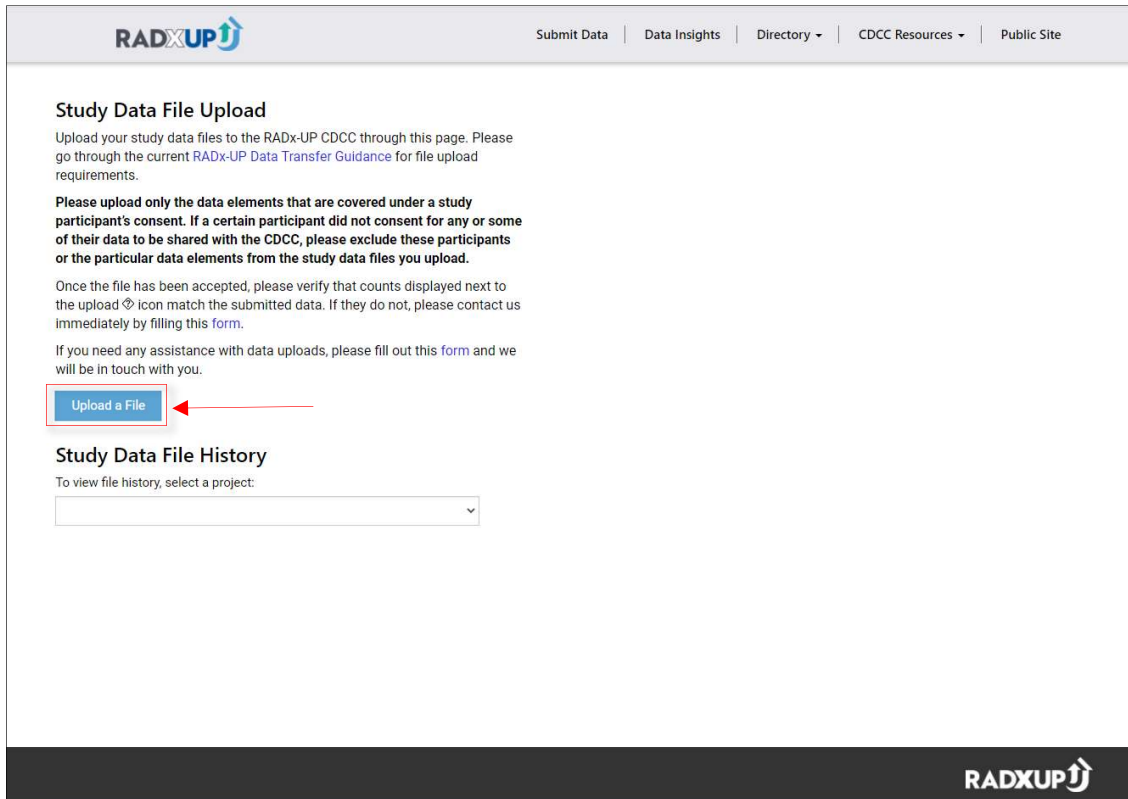
1. Go to [myhome.radx-up.org](https://myhome.radx-up.org) and click **Sign In**. Sign in through your InCommon Account or see the section "Signing in with Sponsored Accounts".



2. Under the "Data Management" tab, select "Data Submission"



3. Click 'Upload a File'



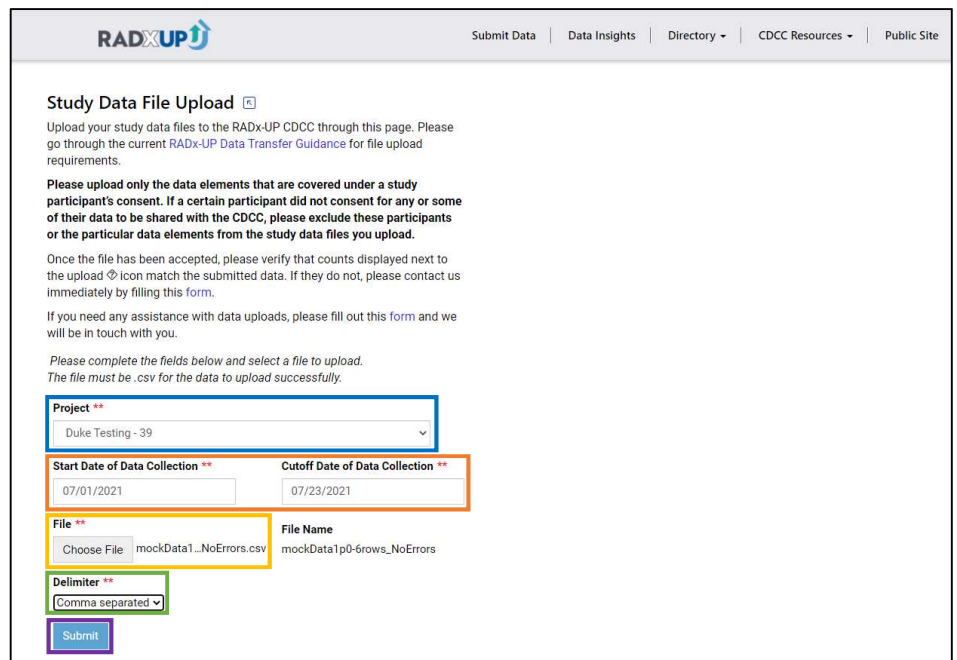
4. Select your project from the drop-down menu (most people will only have one project)

5. Add start date and cutoff date for the file you are submitting. Remember the file is cumulative and the start date should be the same for every upload.

6. 'Choose File'

7. Select Delimiter

8. 'Submit'





After this, your file will appear in the Study File Data History. Click the green "Refresh" to view progress as the status changes and you wait on the DQ report. Submission statuses are shown beside

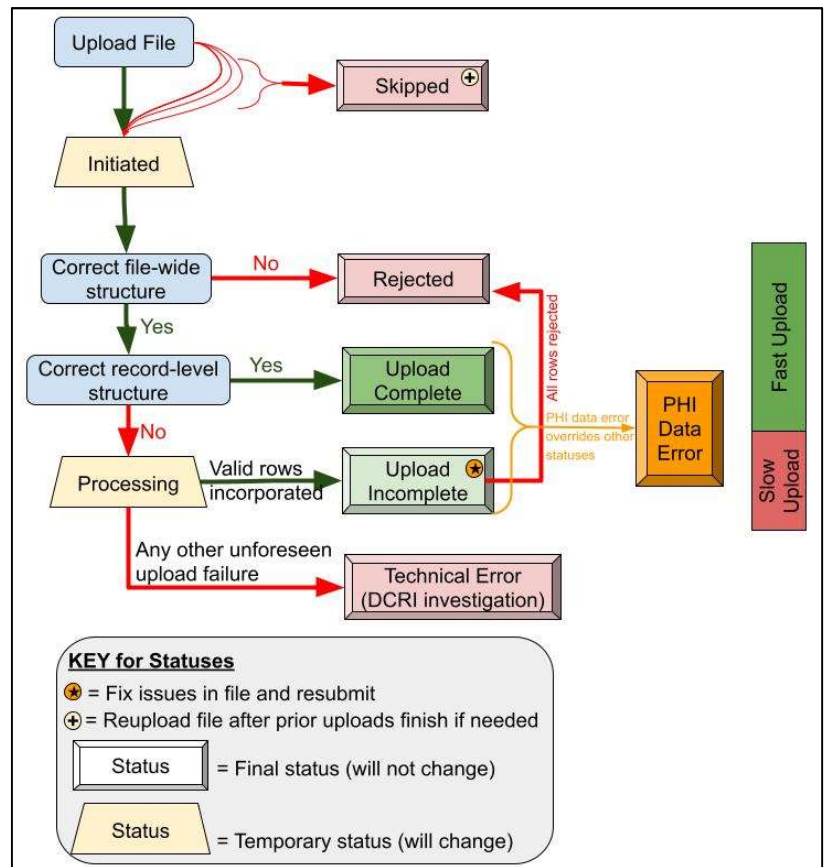
“Status”. Clicking the [green question mark](#) beside “Upload Incomplete” will show the error message that causes this status.

File Name	Status	Data Start Date	Data Cutoff Date	Upload Date	Upload ID	Uploaded By
mockData1p4-2badEml-3ageEr-12rows-noRowFails.csv	Upload Complete Data Quality Report	07/05/2021	07/08/2021	07/23/2021 04:16:37	471	Sage Arbor
mockData1p0-6rows.csv	Upload Incomplete Data Quality Report	07/22/2021	07/22/2021	07/22/2021 16:09:38	470	Jonah Schiestle

### SUBMISSION STATUS DEFINITIONS

Once a file is uploaded, there are a number of statuses your file can take on. After clicking “Submit”, the job will be shown in the job log table at the bottom of the page with the temporary status “Initiated”. If you click the fresh arrow , you will see an updated status.

- **“Skipped”**: Attempting to upload files in quick succession will cause the latter files to fail. If you see “Skipped”, you need to wait until your first file upload reaches a final status before trying to upload again. “Skipped” files will never be incorporated into the RADx-UP database.
- **“Rejected”**: If the structure of the file is significantly malformed and no data can be ingested the upload attempt will end with a final status of “Rejected”. Click the question mark next to “Rejected”  to see the specific error message and reach out to your EITs if you need help understanding the message. You can also use the Upload Troubleshooting Guide contained within this document.
- **“Technical Error”**: Something went wrong processing the file, likely due to a software error. Reach out to your EITs to let us know if this error occurs.
- **“PHI Data Error”**: If you upload CDEs that are considered identifiable and are not allowed per your specific DTA, they will be removed from your upload. If you upload identifiers on a row where “consent\_ident” or “consent\_zip” are No, then the identifiers will be removed. See the section on consent variables above for more information.
- **“Upload Complete”**: all rows have been uploaded to myRADxUP.



- **“Processing”**: The file has some rows of data which cannot be incorporated. A slower row by row check of the data will proceed. If you click on the question mark next to “Processing” you will get an update of how many rows in your file have been processed. Click the green Refresh to keep updating the status until processing is finished.
- **“Upload Incomplete”**: This status comes after the “Processing” status. The question mark next to “Upload Incomplete” will give you information on which rows had issues so you can fix the file before your next upload. Below is an example of the message you will see with an Uploaded Incomplete status. You can use the Upload Troubleshooting Guide contained within this document to help understand the error messages. Reach out to your EITs if you have any additional questions.

File Name	Status	Data Start Date	Data Cutoff Date	Upload Date	Upload ID	Uploaded By
mockData1p4-2badEmI-3ageEr-12rows-noRowFails.csv	Upload Complete <a href="#">Data Quality Report</a>	07/05/2021	07/08/2021	07/23/2021 04:16:37	471	Sage Arbor
mockData1p0-6rows.csv	Upload Incomplete <a href="#">Data Quality Report</a>	07/22/2021	07/22/2021	07/22/2021 16:09:38	470	Jonah Schiestle
mockData1p0-6rows.csv	Upload Complete <a href="#">Data Quality Report</a>			07/21 16:06:07	469	Jonah Schiestle
mockData1p0-1000rows - 1 bad text in date.csv	Upload Incomplete <a href="#">Data Quality Report</a>			07/21 11:17:58	468	Sage Arbor
mockData1p0-100rows - 1 bad text in date.csv	Upload Incomplete <a href="#">Data Quality Report</a>			07/21 11:13:57	467	Sage Arbor
test_upload.csv	Upload Complete <a href="#">Data Quality Report</a>			07/21 10:13:28	466	Bhargav Srinivas Adagarla
tt02.csv	Upload Incomplete			07/21 09:31:38	465	Sage Arbor
tt01.csv	Rejected	07/05/2021	07/08/2021	07/22/2021 09:29:44	464	Sage Arbor

**Result Message**

Some rows have failed to load. Row detail is included here. If you can not find errors in your upload file please contact your Engagement Impact Team (EIT) and share any errors below for them to investigate the problem.; Error on record 39.1869541.testingRcRinstr.29837.baseline\_arm\_ Conversion failed when converting date and/or time from character string.

A summary of the statuses can be seen in the following table.

Status	Description
<b>Initiated</b>	The file has been uploaded, processing has begun
<b>Upload Complete</b>	The job execution has completed successfully.
<b>Rejected</b>	The file failed entirely upon submission
<b>Processing</b>	The file contains at least one row that is not in the correct format. It will be processed row by row and take quite a bit longer. “Processing ” status will be updated in real time.
<b>Upload Incomplete</b>	File contained malformed rows but was processable
<b>Skipped</b>	When two submissions were attempted at the same time to the same awardee.
<b>Technical Error</b>	Processing failed. The reason is unknown and is likely due to a software failure
<b>PHI Data Error</b>	The file contains data outside of your DTA/The file contains conflicting consent variables.

### DATA QUALITY REPORT DEFINITIONS

Once you submit data, you will see your upload appear in the Study Data File History section. Your file can go through a number of statuses, but once you have the status “Upload Complete” or “Upload Incomplete”, your data quality report will be generated. Depending on the size of your file, this may

take a few seconds or few minutes. Click the **green refresh arrow** and wait on the **Data Quality Report** to appear. If you have no issues, no report will appear.

**Study Data File History**

To view file history, select a project:

Duke Testing - 39

File Name	Status	Data Start Date	Data Cutoff Date	Upload Date	Upload ID	Uploaded By
mockData1p4-2badEml-3ageEr-12rows-noRowFails.csv	Upload Complete <a href="#">Data Quality Report</a>	07/05/2021	07/08/2021	07/23/2021 04:16:37	471	Sage Arbor
mockData1p0-6rows.csv	Upload Incomplete <a href="#">Data Quality Report</a>	07/22/2021	07/22/2021	07/22/2021 16:09:38	470	Jonah Schiestle

If you click on the link, you'll be taken to the Data Quality Report. This report contains 12 pages:

1. Cover Page (shown to the right)
2. Table of Contents
3. Data Quality Summary
4. Data Conformance Summary
5. CDE Values Outside of Expected Values
6. CDEs Expected but not Received
7. CDEs Received but not Expected
8. CDE Branching Logic Summary
9. CDEs with Failed Branching Logic
10. Summary Statistics – Categorical Variables
11. Summary Statistics – Continuous Variables
12. CDE Missingness by Tracking Status

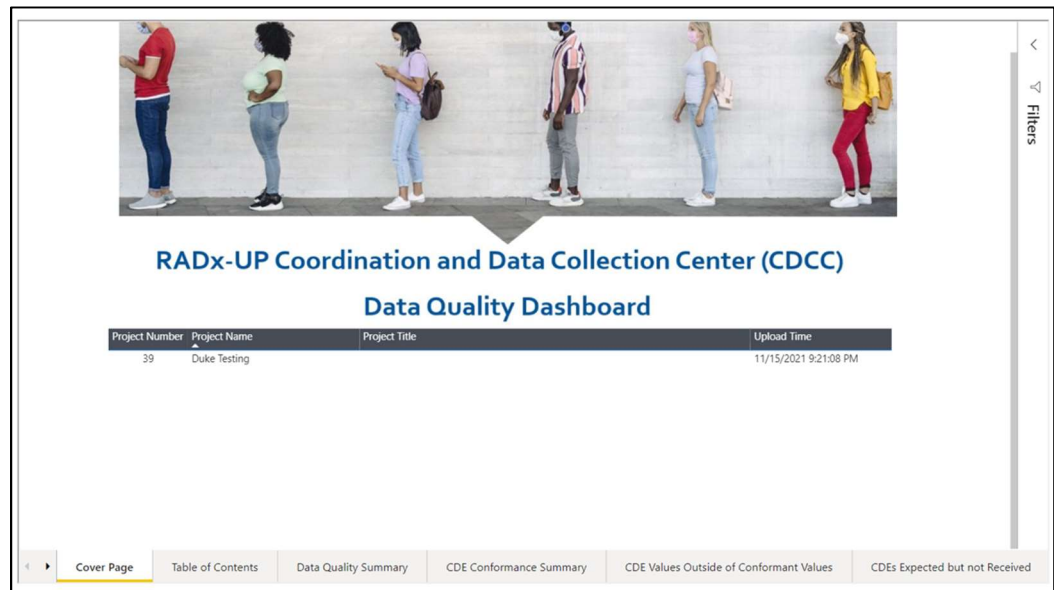


Table of Contents	
1. Cover Page	
2. Table of Contents	
3. Data Quality Summary	
4. CDE Conformance Summary	
5. CDE Values Outside of Conformant Values	
6. CDEs Expected but not Received	
7. CDEs Received but not Expected	
8. CDEs Branching Logic Summary: A CDE-level summary of data that does not meet the allowed branching logic specified in the Data Dictionary Codebook	
9. CDEs with Failed Branching Logic	
10. Summary Statistics - Categorical CDEs	
11. Summary Statistics - Continuous CDEs	

**Table of Contents:** The table of contents page contains links to all other pages. Click on a page title to be taken to the corresponding page. Hovering over the page titles shows a definition of the page. Clicking on the home button in the upper left corner of the content pages returns you to the table of contents. *Note: Pages beyond page 7 may not display on the bottom bar unless you click [the right arrow](#) to view more tabs.*

**Data Quality Summary:** High-level upload and data quality summary. This page shows the metadata for your data upload. You can hover over the blue question mark at the top left corner of the page to see descriptions of the columns.

- *Upload ID:* the unique ID for this upload
- *File Name:* The name of the file you uploaded
- *Rows Rejected:* The number of rows rejected due to file type and data structure issues
- *Rows:* The number of rows of data present in upload file
- *Nonconformant Data Points:* The number of data points outside of accepted values specified in the Data Dictionary Codebook
- *Columns:* The number of columns present in upload file
- *Unique Participants:* The number of unique participants present in upload file
- *Expected CDEs:* The number of Tier 1 CDEs expected to be submitted per the NIH CDE Tracking Form in REDCap
- *Submitted CDEs:* The number of Tier 1 CDEs submitted in this file and expected per the NIH CDE Tracking Form in REDCap
- *Missing CDEs:* The number of Tier 1 CDEs not submitted in this file that are expected to be submitted per the NIH CDE Tracking Form in REDCap
- *Extra CDEs:* The number of Tier 1 CDEs submitted in this file that are not expected to be submitted per the NIH CDE Tracking Form in REDCap

Data Quality Summary												
Project Number	Upload ID	File Name	Rows Rejected	Rows	Columns	Unique Participants	Non-Conformant Data Points	Expected CDEs	Submitted CDEs	Missing CDEs	Extra CDEs	Upload Time
39	745	mockData1p0-1001rows	0	1,001	49	1,001	224	147	21	128	2	11/15/2021 9:21:08 PM
<b>Total</b>			<b>0</b>	<b>1,001</b>	<b>49</b>	<b>1,001</b>	<b>224</b>	<b>147</b>	<b>21</b>	<b>128</b>	<b>2</b>	

- *Total Logical Errors*: The number of data points with branching logic failures as specified in the Data Dictionary Codebook
- *Upload Time*: When the file was uploaded

**CDE Conformance Summary:** CDE-level summary of data outside of accepted values specified in the Data Dictionary Codebook. Click on a field name to filter the next page to that specific CDE. You can hover over the blue question mark at the top left corner of the page to see descriptions of the columns.

- *Upload ID*: The unique upload ID for this upload
- *CDE Name*: The CDE name irrespective of version as specified in the Data Dictionary Codebook
- *Column Name*: The specific CDE name corresponding to the version submitted by your project
- *Percent Missing*: The percent of unique participants with null/blank values for this CDE irrespective of version
- *Collecting Status*: CDE collection information marked in the project’s most recent and valid NIH CDE Tracking Form
- *All Null*: Will be marked “Yes” if the column was submitted in the upload file but is entirely blank
- *Tracking Form Type*: Type of tracking form instance (Initial Project Request, NIH Prior Approval Decision, and Study Closed Default Decision)
- *Tracking Form Date*: Date of tracking form submission
- *Upload Time*: The file upload time

CDE Conformance Summary			
Project Number	Upload ID	CDE Name	Issue Count
39	634	age_yrs	6
39	634	self_reported_height_centimeters	1
39	634	self_reported_height_feet	1
39	634	self_reported_height_inches	1
39	634	self_reported_height_meters	1
39	634	self_reported_weight_kgs	1
39	634	self_reported_weight_lbs	1
<b>Total</b>			<b>12</b>

**CDE Values Outside of Conformant Values:** Datum-level summary of data outside of accepted values specified in the Data Dictionary Codebook. You can hover over the blue question mark at the top left corner of the page to see descriptions of the columns.

- *Upload ID*: The upload ID for this upload
- *Participant ID*: Unique participant record ID
- *CDE Name*: Common Data Element as specified in the Data Dictionary Codebook
- *Conformant Values*: The accepted values specified in the Data Dictionary Codebook

- **CDE Value:** The nonconformant value present in upload file
- **Upload Time:** When this file was uploaded

**CDEs Expected but not Received:** Summary of Tier 1 CDEs expected per the NIH CDE Tracking Form in REDCap that are missing in upload file. You can hover over the blue question mark at the top left corner of the page to see descriptions of the columns. In addition, you can use the “Percent Missing” filter to select CDEs that are 100% missing versus partially missing.

- **Upload ID:** The unique ID for this upload
- **CDE Name:** Common Data Element as specified in the Data Dictionary Codebook
- **Column Name:** The specific column missing (may differ from CDE Name if it is a multiselect CDE)

CDEs Expected but not Received									
A flow diagram of the CDE Tracking Form process can be found <a href="#">here</a>									
Percent Missing									
<input type="checkbox"/> 100% Missing <input type="checkbox"/> Partially Missing									
Project Number	Upload ID	CDE Name	Column Name	Percent Missing	Collecting Status	All Null	Tracking Form Type	Tracking Form Date	Upload Time
24	2505	age_yrs		25.00	Collecting and sharing with the CDCC	N	Prior Approval Decision	08/31/2022	5/3/2023 9:15:4
24	2505	alcohol_daysperweek	alcohol_daysperweek	100.00	Collecting and sharing with the CDCC	N	Prior Approval Decision	08/31/2022	5/3/2023 9:15:4
24	2505	bio_sex_birth	Indeterminate- Consult Data Dictionary Codebook	100.00	Collecting and sharing with the CDCC	N	Prior Approval Decision	08/31/2022	5/3/2023 9:15:4
24	2505	cc_asthma	Indeterminate- Consult Data Dictionary Codebook	100.00	Collecting and sharing with the CDCC	N	Prior Approval Decision	08/31/2022	5/3/2023 9:15:4
24	2505	cc_asud	Indeterminate- Consult Data Dictionary Codebook	100.00	Collecting and sharing with the CDCC	N	Prior Approval Decision	08/31/2022	5/3/2023 9:15:4
24	2505	cc_autoimm	Indeterminate- Consult Data Dictionary Codebook	100.00	Collecting and sharing with the CDCC	N	Prior Approval Decision	08/31/2022	5/3/2023 9:15:4
24	2505	cc_cancer	Indeterminate- Consult Data Dictionary Codebook	100.00	Collecting and sharing with the CDCC	N	Prior Approval Decision	08/31/2022	5/3/2023 9:15:4
24	2505	cc_chronickd	Indeterminate- Consult Data Dictionary Codebook	100.00	Collecting and sharing with the CDCC	N	Prior Approval Decision	08/31/2022	5/3/2023 9:15:4
24	2505	cc_clung	Indeterminate- Consult Data Dictionary Codebook	100.00	Collecting and sharing with the CDCC	N	Prior Approval Decision	08/31/2022	5/3/2023 9:15:4
24	2505	cc_copd	Indeterminate- Consult Data Dictionary Codebook	100.00	Collecting and sharing with the CDCC	N	Prior Approval Decision	08/31/2022	5/3/2023 9:15:4
24	2505	cc_cvd	Indeterminate- Consult Data Dictionary Codebook	100.00	Collecting and sharing with the CDCC	N	Prior Approval Decision	08/31/2022	5/3/2023 9:15:4
24	2505	cc_depression	Indeterminate- Consult Data Dictionary Codebook	100.00	Collecting and sharing with the CDCC	N	Prior Approval Decision	08/31/2022	5/3/2023 9:15:4
24	2505	cc_diabetes	Indeterminate- Consult Data Dictionary Codebook	100.00	Collecting and sharing with the CDCC	N	Prior Approval Decision	08/31/2022	5/3/2023 9:15:4
24	2505	cc_hypertension	Indeterminate- Consult Data Dictionary Codebook	100.00	Collecting and sharing with the CDCC	N	Prior Approval Decision	08/31/2022	5/3/2023 9:15:4

If the Column Name says “Indeterminate – consult Data Dictionary Codebook”, there are multiple versions of this CDE that could have differing column names. Please consult the codebook for the right column name for the version you are using.

CDE Values Outside of Conformant Values						
Project Number	Upload ID	Participant ID	CDE Name	Conformant Values	CDE Value	Upload Time
39	745	1551135	bio_sex_birth	0,1,2,96,99	55	11/15/2021 9:21:08 PM
39	745	715055	consent_ident	0,1	3	11/15/2021 9:21:08 PM
39	745	1037494	household_congregate	1,2,3,4,5,6,7,8,9,90	0	11/15/2021 9:21:08 PM
39	745	1187003	household_congregate	1,2,3,4,5,6,7,8,9,90	0	11/15/2021 9:21:08 PM
39	745	1341006	household_congregate	1,2,3,4,5,6,7,8,9,90	0	11/15/2021 9:21:08 PM



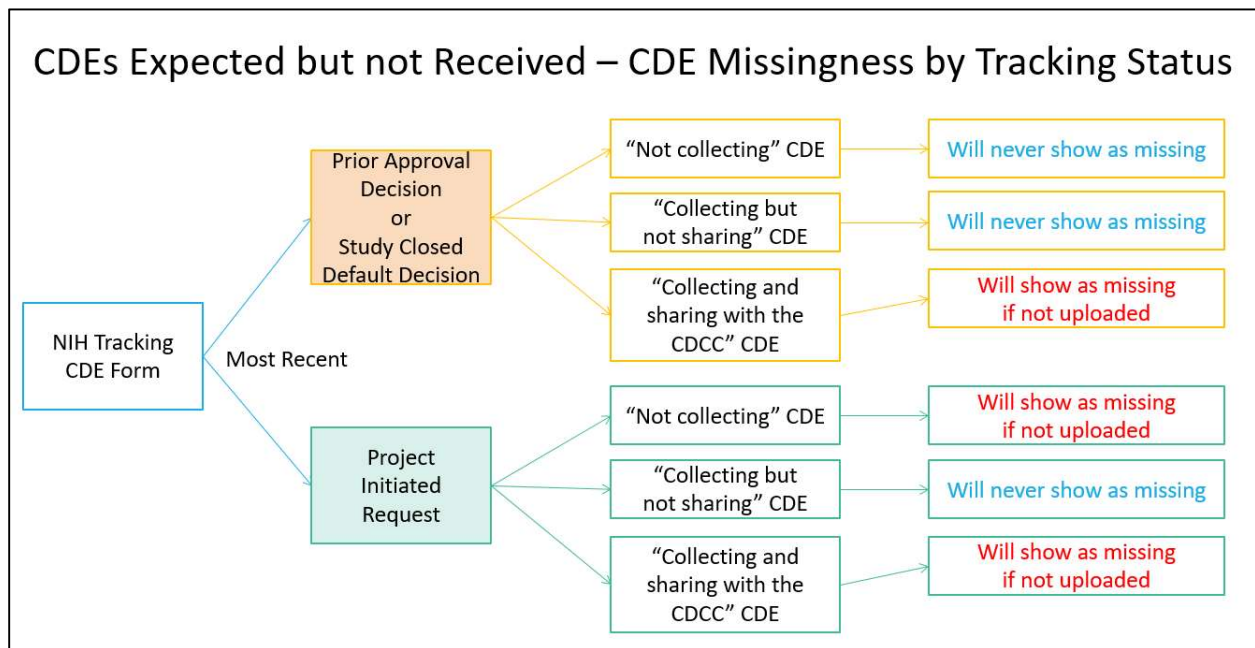
Different actions are required depending on the extent to which a CDE is missing. First, the following explains the meaning and necessary next steps regarding “100% Missing CDEs”.

### 100% MISSING CDES

Per the change in the NIH Terms and Conditions (effective November 2021), all RADx-UP project teams are expected to collect all Tier 1 CDEs (excluding the consent variables previously referenced) unless they receive an NIH Prior Approval Decision indicating otherwise.

If a project’s most recent tracking form does not have NIH prior approval, all Tier 1 CDEs (excluding consent variables) marked as “Not collecting” or “Collecting and sharing with the CDCC” will show as 100%missing on the Data Quality Dashboard if not submitted. If the project’s most recent tracking form does have NIH approval—or if the project has closed without yet receiving NIH prior approval—then only CDEs marked as “Collecting and sharing with the CDCC” will show as missing if not submitted.

For more information, please refer to the diagram on the next page that displays CDE missingness by tracking status.



Although we encourage project teams to provide consent information when possible, Tier 1 Consent CDEs are not required by the NIH for data submissions where they do not apply to your study design.

For example, if you are conducting a quality improvement study that does not require consent or are only using secondary data, you do not have to request a prior approval in order to exclude these variables from your data set.

If you are conducting a study that requires consent, you should include the variables listed below as applicable to your study:

- consent\_given (recommended for all studies, indicates at the study level whether consent was required for the study, may be autofilled)

- consentdt\_mdy (to be used when consent is conducted with participants, date of consent)
- *consent\_ident, consent\_zip\_2, consent\_recontact* (used only when consent process is conducted and includes the *opt-in/opt-out* variables for data sharing)
- consent\_given (recommended for all studies, indicates at the study level whether consent was required for the study, may be autofilled)
- consent\_complete (used only when consent is conducted or completion status other than date is available)

If your project team is excluding any of these consent variables, please ensure that their exclusion is correctly represented as “Not collecting” or “Collecting but not sharing” on your NIH CDE Tracking Form.

Since late October 2022, missing consent CDEs marked “Not collecting” in REDCap will not show as expected but not received on your Data Quality Dashboard, regardless of whether the associated tracking form instance is a project request or NIH Prior Approval Decision. Unlike the remaining Tier 1 CDEs, exclusion of consent CDEs does not require a prior approval request to the NIH Office of Grants Management.

In addition, the NIH has specified certain CDE groups within which not all CDEs need to be submitted:

- *zip\_code/zip\_code\_3digit/current\_county*
- *dob\_mdy/lage\_yrs*
- *self\_reported\_weight\_kgs/self\_reported\_weight\_lbs*
- *self\_reported\_height\_meters/self\_reported\_height\_feet*
- *self\_reported\_height\_centimeters/self\_reported\_height\_inches*

Within these groups, only 1 CDE needs to be submitted. For example, if a project indicates on their tracking form that they are collecting and sharing *zip\_code\_3digit* but not *zip\_code* or *current\_county*, then only *zip\_code\_3digit* will be expected and show as missing if not submitted. However, if all CDEs are marked as “Collecting and sharing with the CDCC” or “Not collecting”, then they will all be expected and show as missing if not submitted. Please ensure your tracking form correctly reflects which CDEs you intend to submit for a particular group.

### **PARTIALLY MISSING CDES**

These CDEs will show up on this page if they are over 20% missing, meaning more than 20% of participants did not answer the survey question.

Please use the Data Quality Resolution Request page (under the Data Manager tab) to let us know why this data is missing. We provide the following possible reasons a CDE may be over 20% missing:

- Aims/Phase related
- Added to survey after data collection started
- Secondary data set did not include
- Participants left blank
- Branching logic not applied
- Other - please specify in Data Manager Notes

You are encouraged to add details to the Data Manager Notes field. You can select multiple reasons if multiple may apply. Once you've filled out requests, you click "Submit" at the bottom of the page. The CDCC will then review these reasons and approve the requests. Once approved, they will be removed from your Data Quality Report "CDEs Expected but Not Received" page.

**CDEs Received but not Expected:** Summary of Tier 1 CDEs not expected per the NIH CDE Tracking Form in REDCap that were received in upload file. You can hover over the blue question mark at the top left corner of the page to see descriptions of the columns.

- *Upload ID:* The unique ID for this upload
- *CDE Name:* The extra CDE
- *All Null:* Are all data points null for this CDE?

**CDEs with Failed Branching Logic:** Summary of Tier 1 CDEs with branching logic errors. Branching logic is described in the Data Dictionary. You may need to click the right arrow in the bottom left corner to view the page. You can hover over the blue question mark at the top left corner of the page to see descriptions of the columns. You can scroll across the table to see all columns.

- *Dependent CDE Column Name:* The Common Data Element being acted on by the branching logic as specified in the Data Dictionary Codebook
- *Dependent CDE Name:* The CDE name irrespective of version or multi-select as specified in the Data Dictionary Codebook
- *Dependent CDE Question:* The question corresponding to this CDE as specified in the Data Dictionary Codebook

Project Number		Upload ID	CDE Name	All Null	Upload Time
39	745	consent_given	No	11/15/2021 9:21:08 PM	
39	745	consentdt_mdy	No	11/15/2021 9:21:08 PM	

- *Dependent CDE Value:* The encoded value of the CDE
- *Dependent CDE Answer:* The meaning of the numerical encoding for this CDE as specified in the Data Dictionary Codebook
- *Allowed Logic:* Branching logic acting on the CDE as specified in the Data Dictionary Codebook
- *Logic CDE Name:* The CDE within the branching logic that caused the error
- *Logic CDE Question:* The question corresponding to the CDE within the branching logic as specified in the Data Dictionary Codebook

- **Logic CDE Value:** User-entered value that fails the branching logic
- **Logic CDE Answer:** The meaning on the numerical encoding for the logic CDE value as specified in the Data Dictionary Codebook
- **Upload Time:** File upload time

CDEs with Failed Branching Logic								RADXUP
Upload ID	Participant ID	Dependent CDE Name	Dependent CDE Column	Dependent CDE Question	Dependent CDE Value	Dependent CDE Answer	Allowed	
910	1	race_ethn_hispanic_detail	race_ethn_hispanic_detail	Please specify your origin	0	N/A	(race_ethn_hispanic = 1)	
910	10	race_ethn_hispanic_detail	race_ethn_hispanic_detail	Please specify your origin	0	N/A	(race_ethn_hispanic = 1)	
910	10	race_ethn_islander_detail	race_ethn_islander_detail__3		1	Samoan	(race_ethn_race__4 = 1)	
910	11	consent_ident	consent_ident	I agree to let The Duke Clinical Research Institute to collect all identifiable information.	0	No	(consent_given = 1)	
910	11	consent_recontact	consent_recontact	I agree to be contacted for future research.	0	No	(consent_given = 1)	
910	11	consent_ssn	consent_ssn	I agree to let The Duke Clinical Research Institute to collect my Social Security number.	1	Yes	(consent_given = 1)	
910	11	consent_zip	consent_zip	I agree to let The Duke Clinical Research Institute to collect only my zip code and no other identifiable information.	1	Yes	(consent_given = 1)	
910	11	race_ethn_hispanic_detail	race_ethn_hispanic_detail	Please specify your origin	0	N/A	(race_ethn_hispanic = 1)	
910	11	race_ethn_islander_detail	race_ethn_islander_detail__4		1	Tongan	(race_ethn_race__4 = 1)	
910	12	consent_ident	consent_ident	I agree to let The Duke Clinical Research Institute to collect all identifiable information.	0	No	(consent_given = 1)	
910	12	consent_recontact	consent_recontact	I agree to be contacted for future research.	1	Yes	(consent_given = 1)	

Navigation: Data Quality Summary | CDE Conformance Summary | CDE Values Outside of Conformant Values | CDEs Expected but not Received | CDEs Received but not Expected | **CDEs with Failed Branching Logic**

**Summary Statistics: Categorical Variables:** Frequency Distributions for Tier 1 categorical CDEs. You may need to click the right arrow in the bottom left corner to vie the page. You can hover over the blue question mark at the top left corner of the page to see descriptions of the columns. You can scroll across the table to see all columns.

- **Total Number of Records N:** The number of unique rows in most recent incorporated data submission, including longitudinal data when applicable
- **Percent Non-Conformance Removed:** The percentage of non-null data points in expected CDEs submitted to the CDC that are non-conformant
- **Upload ID:** The unique ID for this upload
- **Section Characteristic:** Survey section title as specified in Data Dictionary Codebook
- **CDE Question:** The CDE question label as specified in the Data Dictionary Codebook
- **CDE Name:** The CDE name irrespective of version as specified in the Data Dictionary Codebook
- **Column Name:** Specific column name of most recent CDE version as specified in the Data Dictionary Codebook
- **CDE Value:** The encoded value of the CDE
- **CDE Answer:** The meaning of the numerical encoding for the CDE value as specified in the Data Dictionary Codebook
- **Rows n:** The number of unique rows for which this answer choice was selected

- **Percentage:** The percentage of the CDE Question that corresponds to the CDE answer, calculated using the formula  $(n / M) * 100$

Upload ID	Section Characteristic	CDE Question	CDE Name	Column Name	CDE Value	CDE Answer	Rows	Percentage
1256	Consent	I agree to let the RADx-UP CDCC collect all identifiable information.	consent_ident	consent_ident	1	Yes	1	20.00
1256	Demographics	What is your race?Mark one or more boxes AND print origins.	race_ethn_race	race_ethn_race_2	1	Black or African American	1	20.00
1256	Demographics	What is your race?Mark one or more boxes AND print origins.	race_ethn_race	race_ethn_race_3	1	Asian	1	20.00
1256	Demographics	What is your race?Mark one or more boxes AND print origins.	race_ethn_race	race_ethn_race_99	1	Prefer not to answer	1	20.00

**Summary Statistics: Continuous Variables:** Number summaries and missingness for Tier 1 continuous response. You may need to click the right arrow in the bottom left corner to view the page. You can hover over the blue question mark at the top left corner of the page to see descriptions of the columns. You can scroll across the table to see all columns.

- **Total Number of Records N:** The number of unique rows in most recent incorporated data submission, including longitudinal data when applicable
- **Percent Non-Conformance Removed:** The percentage of non-null data points in expected CDEs submitted to the CDCC that are non-conformant
- **Unique ID:** The unique ID for this upload
- **Section Characteristic:** Survey section title as specified in Data Dictionary Codebook
- **CDE Question:** The CDE question label as specified in the Data Dictionary Codebook
- **CDE Name:** The CDE name irrespective of version as specified in the Data Dictionary Codebook
- **Column Name:** Specific column name of most recent CDE version as specified in the Data Dictionary Codebook
- **Rows n:** The number of unique rows for which this answer choice was populated
- **Missing:** The number of rows for which this CDE was not answered, calculated using  $N - n$
- **Mean:** The arithmetic mean for this CDE among those who answered
- **Standard Deviation:** The standard deviation for this CDE among those who answered
- **Minimum:** The minimum for this CDE among those who answered
- **First Quartile:** The first quartile for this CDE among those who answered
- **Median:** The median for this CDE among those who answered
- **Third Quartile:** The third quartile for this CDE among those who answered
- **Maximum:** The maximum for this CDE among those who answered

### DATA CLEANING GUIDANCE: VALUES OUTSIDE EXPECTED RANGE:

Once you've viewed your report, you are responsible for cleaning issues and resubmitting. These examples are for illustration, but you should follow your study protocols and work with your statistician/PI to make data cleaning decisions. If you have any questions about your Data Quality Report, you can reach out to your EIT or primary informaticist(s).

**1. Fixing Data:** If it's a reasonable assumption, you can "fix" the data.

- a. **Example 1:** self\_reported\_height\_inches was entered as 64. The allowed values were 0-12.
  - i. When checking this row, you see self\_reported\_height\_feet is 5.333333.
  - ii. It's reasonable to assume 64 is their total height in inches, and they are 5 feet 4 inches tall.
- b. **Example 2:** age\_yrs was entered as 1993. The allowed values are 0-110.
  - i. If you have date of birth and date of data collection, you can use that to infer their age.
  - ii. If you don't have a way to confirm date of birth, see #2 below.

**2. Remove Data:** If it's a nonsensical answer and you can't make a reasonable assumption, you can null out the data.

- a. **Example 1:** age\_yrs was entered as 1538. The allowed values are 0-110.
  - i. If you do not have date of birth and date of data collection, you should null this value out.
- b. **Example 2:** smoker\_number was entered as 44,848. The allowed values are 0-80.
  - i. There's no way to understand what was meant by this answer. You should null this value out.

**3. Valid Data:** If you have a flagged error, but you double check data and it is correct, you should use the Data Quality Resolution page to flag there.

- a. **Example 1:** age\_yrs was entered as 114. The allowed values are 0-110.
  - i. You know you did have a participant who was 114. Do not null out this value! See the Data Quality Resolution section of this document.
- b. **Example 2:** you have a value of '98' in a categorical variable only allowing 0 or 1.
  - i. You got this approved by the CDCC before you began uploading by uploading a Wording Request Change to your NIH CDE Tracking Form. See the Data Quality Resolution section of this document for how to flag this as valid data.

#### DATA CLEANING GUIDANCE: CDE EXPECTATIONS

You can resubmit a new file with any extra columns removed or new columns added. If needed, you can work with your EITs to update your NIH CDE Tracking Survey in REDCap. **Before adding any CDEs to your file, be sure to check your DTA/IRB consent forms and ensure you are allowed to share this data.**

If Column Name for CDEs Expected but Not Received says "Indeterminate – consult Data Dictionary Codebook", there are multiple versions of this CDE that could have differing column names. Please consult the codebook for the right column name for the version you are using.

#### DATA CLEANING GUIDANCE: BRANCHING LOGIC FAILURES

Some logic failures may be allowable and you would need to fill out the Data Quality Resolution Table to let us know that this is true and valid data. For example, if looking at these two questions about employment status, "cur\_employ\_stat\_specify" is only supposed to be filled out if a participant selects "Other (Specify)" from current\_employment\_status. However, if a participant chose another option in

“current\_employment\_status” and chose to add more context in the “cur\_employ\_stat\_specify”, that is not conflicting data.

33	current_employment_status	We would like to know about what you do -- are you working now, looking for work, retired, keeping house, a student, or something else?	<table border="1"> <tr><td>1</td><td>Working now</td></tr> <tr><td>2</td><td>Only temporarily laid off, sick leave or maternity leave</td></tr> <tr><td>3</td><td>Looking for work, unemployed</td></tr> <tr><td>4</td><td>Retired</td></tr> <tr><td>5</td><td>Disabled, permanently or temporarily</td></tr> <tr><td>6</td><td>Keeping house</td></tr> <tr><td>7</td><td>Student</td></tr> <tr><td>96</td><td>Other (Specify)</td></tr> <tr><td>98</td><td>Prefer not to answer</td></tr> <tr><td>99</td><td>Don't know</td></tr> </table> <p>Field Annotation: PX011301   <a href="https://www.phenxtoolkit.org/protocols/view/11301">https://www.phenxtoolkit.org/protocols/view/11301</a> Study of Income Dynamics (PSID), 2007   Added 99, Prefer not to answer</p>	1	Working now	2	Only temporarily laid off, sick leave or maternity leave	3	Looking for work, unemployed	4	Retired	5	Disabled, permanently or temporarily	6	Keeping house	7	Student	96	Other (Specify)	98	Prefer not to answer	99	Don't know
1	Working now																						
2	Only temporarily laid off, sick leave or maternity leave																						
3	Looking for work, unemployed																						
4	Retired																						
5	Disabled, permanently or temporarily																						
6	Keeping house																						
7	Student																						
96	Other (Specify)																						
98	Prefer not to answer																						
99	Don't know																						
34	cur_employ_stat_specify	Current employment status, Other - specify	<p>Show the field ONLY if: [current_employment_status] = '96'</p> <p>Field Annotation: PX011301   <a href="https://www.phenxtoolkit.org/protocols/view/11301">https://www.phenxtoolkit.org/protocols/view/11301</a> Study of Income Dynamics (PSID), 2007</p>																				

There may be cases where data is conflicting. For example, a participant must say “Yes” to flu\_vaccinehistind “Have you ever received a flu vaccination?” to be asked flu\_vaccinseason “Have you received a flu vaccine in the last 6 months?” If a participant says “No” to flu\_vaccinehistind, but “Yes” to flu\_vaccinseason, the answers conflict.

88	flu_vaccinehistind	Section Header: <i>Vaccination</i> Have you ever received a flu vaccination?	<table border="1"> <tr><td>1</td><td>Yes</td></tr> <tr><td>0</td><td>No</td></tr> <tr><td>98</td><td>Don't know</td></tr> </table> <p>Field Annotation: C08096</p>	1	Yes	0	No	98	Don't know
1	Yes								
0	No								
98	Don't know								
89	flu_vaccine_season_2	Have you received a flu vaccine in the last 12 months?	<table border="1"> <tr><td>1</td><td>Yes</td></tr> <tr><td>0</td><td>No</td></tr> <tr><td>98</td><td>Don't know</td></tr> </table> <p>Field Annotation: UPENN SURVEY</p>	1	Yes	0	No	98	Don't know
1	Yes								
0	No								
98	Don't know								

If re-contacting the project or confirming the data is not possible, our general advice for fixing logic issues is to accept the first response answer (the independent CDE contained within the branching logic) provided and remove the second answer (the dependent CDE). In this flu vaccine example, you would take the answer from flu\_vaccinehistind as the independent CDE and remove the response to flu\_vaccinseason as it should have depended on the answer to flu\_vaccinehistind.

Upload ID	Section Characteristic	CDE Question	CDE Name	Column Name	Rows	Missing	Mean	Standard Deviation	Minimum	First Quartile	Median	Third Quartile	Maximum
1256	Alcohol and Tobacco/Nicotine Use	If you smoke every day, on average, how many cigarettes per day do you smoke?	smoker_number	smoker_number	0	5							
1256	Demographics	Age For babies less than 1 year old, do not write the age in months. Write 0 as the age.	age yrs	age yrs	3	2	49.00	37.80	6	35	64.00	70	77

## UPLOAD TROUBLESHOOTING GUIDE

If your file was rejected or some rows were rejected, you can use the following guide to help trouble shoot. This is not an exhaustive list, but a place to begin. If you cannot identify the error, reach out to your EIT and your primary informaticist(s) can meet with you to help out.

### If Rejected:

- **Ensure each record has a unique 'record\_id'** in a column labeled 'record\_id'. If 'record\_id' is missing, the file will be rejected.
- **Ensure the column headers** in your file match the column headers in the RADx-UP Codebook. At least some columns must have expected headers, spelled exactly as they are listed in the Codebook.
- **Special / Hidden Characters:** If column headers contain special characters (i.e., Ñ, ĩ, á), the file may be rejected. If you suspect hidden characters, try submitting the first few columns only or rewrite the header into a new file and copy in one or two rows of data.

### If Incomplete:

You can use the uid\_rc from the error message pop up (shown in Submission Statuses Section) to identify the row causing the error.

- **String in Date/Time:** check every date column to be sure they are correctly formatted yyyy-mm-dd (2021-08-06 9:00) Records with other date formats (i.e., 2021-AUG-06 9:00) will not load.
- **Large Int in Small Int Column:** Be sure for columns where years are coded as numbers, you aren't putting the year in the column. For example, for positiveyear\_covidtest and recentyear\_covidtest, 1=2019, 2=2020, 3=2021.
- **Look for text in integer fields** i.e., 'positive' or 'negative' instead of the expected coded values of 1, 2 in 'recentresult\_covidtest'
- Ensure that response options for each variable in your file align with response options in the Codebook. Records with unexpected values will not be loaded.



## Part 3: Data Quality Resolution Requests

### OVERVIEW

There is a new function for Data Managers. This feature allows data managers to view conformance issues shown in the Data Quality Report on the page “CDE Value Outside of Accepted Values” and address these issues. You can use this feature to inform the CDCC if anything was incorrectly flagged as a conformance issue, if it is valid data, if you added an answer option via a Wording Request Change form, or if this will not be corrected until a future aim. You can also choose a timeline for fixing any issues if you foresee needing over a month to fix these items.

**You do not have to use the table for every issue. If you plan to address and fix issues within four weeks of your upload, you do not need to fill out a request.**

Under the “Data Management” header on *myRADx-UPhome*, select “Data Quality Resolution Request” to view your table. Select your project from the drop-down menu. There will be multiple tables for each type of issue in the Data Quality Report – “CDEs Outside of Allowed Values” (Conformance) and “CDEs Expected but not Received” (Missing). You can apply any filters you’d like. The filter options include filtering by approval status or resolution request reason. *Note that “Select” means that no option has been selected on Approval Status or Resolution Request Reason.*

### Data Quality Resolution Request

Select Project \*\*

Duke Testing - 39
▼
↻

Please see below data quality issues. Please review and add feedback around resolution. The informaticists will review and discuss with your project team if needed.

Show the following Approval Status(es):  Select  Requested  Approved  Denied

Resolution Request Reasons(es):  Select  Verified and true outlier in the submitted data  An exception was submitted for this CDE

We will address this and fix the issue  We are unable to address this issue and will remove the data from our dataset

#### Conformance Issues

Initial Upload ID	Most Recent Upload ID	Participant ID	CDE Name	Conformant Values	CDE Value	Quality Resolution Type	Resolution Request Reason
761	761	10	recentresult_covidtest	1,2,3,4,98,99	66	Conformance Issue ▼	Select ▼
761	761	11	recentresult_covidtest	1,2,3,4,98,99	67	Conformance Issue ▼	Select ▼
761	761	12	recentresult_covidtest	1,2,3,4,98,99	68	Conformance Issue ▼	Select ▼

## PAGINATION

If you have a large number of issues, they will be separated into pages within the table. The CDEs are sorted alphabetically, so to locate a specific CDE you need to find the page containing that letter.

Be sure to finish editing and click “Submit” before changing pages within the table. Additionally, the filters only apply to the page currently showing.

## DESCRIPTION OF COLUMNS: CONFORMANCE TABLE

The un-editable columns on the table are:

- **Initial Upload ID:** The first upload to contain this issue
  - *you can see which file names and dates correspond to an upload ID on your data submission page*
- **Most Recent Upload ID:** The most recent upload to contain this issue
  - *you can see which file names and dates correspond to an upload ID on your data submission page*
- **Participant ID:** the unique record ID that corresponds to this issue
- **CDE Name:** Common Data Element as specified in the Data Dictionary Codebook
- **Conformant Values:** The accepted values specified in the Data Dictionary Codebook
- **CDE Value:** The nonconformant value present in upload file
- **Quality Resolution Type:** whether this is an issue of conformance, missingness, or a future issue type
  - *Right now, this table is only displaying conformance issues and all values will say “Conformance Issue”*
- **CDCC Response:** the response from the CDCC on the status of this request. It will be “Requested” after you have filled out a request, and then it will change to either “Approved” or “Denied”.

Quality Resolution Type	Resolution Request Reason	Does this apply to every CDE with this value?	Please include the label of this field option	Expected Resolution Date 1	Expected Resolution Date 2	Expected Resolution Date 3	Describe Why the Data Cannot be Resolved	Data Manager Notes	CDCC Response
Conformance Issue	Select	Select		mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy			Select
Conformance Issue	Select	Select		mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy			Select
Conformance Issue	Select	Select		mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy			Select
Conformance Issue	Select	Select		mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy			Select
Conformance Issue	Select	Select		mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy			Select

Submit

The columns you can edit are:

- **Resolution Request Reason:** the reason you are submitting a request to the CDCC concerning this data point.
  - Verified and true outlier in the submitted data: You have verified that this outlier is a true data point and it needs to be included in RADx-UP reporting.

- *Example: a participant's age was truly 114.*
  - An exception was submitted for this CDE: You submitted a Wording Request Change to add an answer option to a categorical variable and it should be recorded.
    - *Example: you added "99", "Prefer Not to Answer" to the variable jobless\_covid and submitted that change via the NIH CDE Tracking Survey. It is currently being reviewed by the NIH or has been approved.*
  - We will address this and fix the issue: You know this variable is wrong and are planning to fix this and are entering the time needed to address it.
    - *Example: Someone put their birth date in the age\_yrs column. You will confirm that is their birthdate and replace the value with their actual age.*
  - We are unable to address this issue and will remove the data from our dataset: You know this variable is wrong and cannot fix it, but are planning to remove it from your upload and are entering the time needed.
    - *Example: Someone put 1432 in the age\_yrs column. This is nonsensical, you do not have their birthdate recorded, and you are going to remove this from your upload file.*
- **Does this apply to every CDE with this value?:** is this an issue that applies to all CDEs of this value? Yes or No answer.
  - *If this was a wording request change that added an answer option to this CDE, it will apply to every instance of that value.*
  - *Example: you added "99", "Prefer Not to Answer" to the variable jobless\_covid and submitted that change via the NIH CDE Tracking Survey. All instances of "99" in "jobless\_covid" are the same issue defined here, and you do not need to fill out a new row for every time this appears.*
  - *Note that this applies to the specific CDE and VALUE. If someone enters "100" in "jobless\_covid", you will need to address that separately.*
- **Please include the label of this field option:** If you added an answer option, please fill out here the meaning of the value.
  - *Example: you added "99" to "jobless\_covid", which means "Prefer Not to Answer". In this box, you would type "Prefer Not to Answer"*
- **Expected Resolution Date 1:** the date you expect to have this issue resolved in your upload file, either by fixing the value or removing it from the upload.
- **Expected Resolution Date 2:** If you miss Date 1 and need more time, you can fill out another resolution date.
- **Expected Resolution Date 3:** If you miss Dates 1 and 2, you can fill out another resolution date.
- **Describe Why the Data Cannot be Removed:** If you selected answer option "We are unable to address this issue and will remove the data from our dataset" for Resolution Request Reason, please explain the details here on why it cannot be corrected.
- **Data Manager Notes:** This is a space to type any relevant notes to the issues and timeline you are proposing.
  - *If you are requesting over eight weeks of time to fix issues, please explain here why your project needs extra time.*

After you are done editing, click the blue Submit button in the bottom left beneath the table. Your primary informaticist will review the request and change it to either "Approved" or "Denied". You can always reach out to your EITs about any issues you have.

## DESCRIPTION OF COLUMNS: MISSINGNESS

The un-editable columns on the table are:

- **Initial Upload ID:** The first upload to contain this issue
  - *you can see which file names and dates correspond to an upload ID on your data submission page*
- **Most Recent Upload ID:** The most recent upload to contain this issue
  - *you can see which file names and dates correspond to an upload ID on your data submission page*
- **CDE Name:** Common Data Element as specified in the Data Dictionary Codebook
- **Column Name:** Version and checkbox-specific CDE as specified in the Data Dictionary Codebook
- **Quality Resolution Type:** whether this is an issue of conformance, missingness, or a future issue type
  - *This table will only display "CDE Expected but not Submitted"*
- **CDCC Response:** the response from the CDCC on the status of this request. It will be "Requested" after you have filled out a request, and then it will change to either "Approved" or "Denied".

The columns you can edit are:

### MISSINGNESS / EXPECTED ISSUES

Initial Upload ID	Most Recent Upload ID	CDE Name	Column Name	Percent Missing	Quality Resolution Type	100% Missing Resolution Request Reason	Partially Missing Resolution Request Reason	Expected Resolution Date 1	Expected Resolution Date 2	Expected Resolution Date 3	Describe Why the Data Cannot be Resolved	Data Manager Notes	CDCC Response
2007	2010	alcohol_days_perweek	alcohol_days_perweek	100	Missingness Issue	Select ▼	None selected -	mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy			Select ▼
2007	2010	cc_asthma	cc_asthma	100	Missingness Issue	Select ▼	None selected -	mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy			Select ▼
2007	2010	cc_asud	cc_asud	100	Missingness Issue	Select ▼	None selected -	mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy			Select ▼
2007	2010	cc_autoimm	cc_autoimm	100	Missingness Issue	Select ▼	None selected -	mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy			Select ▼
2007	2010	cc_cancer	cc_cancer	100	Missingness Issue	Select ▼	None selected -	mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy			Select ▼
2007	2010	cc_chronicid	cc_chronicid	100	Missingness Issue	Select ▼	None selected -	mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy			Select ▼

- **100% Missing Resolution Request Reason:** the reason you are submitting a request to the CDCC concerning this 100% missing CDE.
  - An exception was submitted for this CDE: You submitted a Wording Request Change to add an answer option to a categorical variable and it should be recorded.
    - *This will not be relevant to missing CDEs*
    - *If this is approved by an informaticist, this CDE will no longer show as missing on the Data Quality Dashboard*
  - We will address this and fix the issue: You know this variable is wrong and are planning to fix this and are entering the time needed to address it.
    - *Example: You are adding this CDE to your survey but implementation will take some time.*
  - No participants answered this question: You are collecting and sharing this CDE, but there is no participant data for it (data is null for all participants).
    - *If this is approved by an informaticist, this CDE will no longer show as missing on the Data Quality Dashboard*
- **Partially Missing Resolution Request Reason:** the reason you are submitting a request to this CDCC concerning this partially missing CDE

- Aims/Phase related
    - *Example: Your project only collected bio\_sex\_birth for the first phase of your study.*
  - Added to survey after data collection started
    - *Example: Your project issued different CDEs at different points of data collection*
  - Secondary data set did not include
  - Participants left blank
  - Branching logic not applied
    - *Example: A participant answers race\_hispanic\_detail\_other without answering race\_ethn\_hispanic\_detail*
  - Other - please specify in Data Manager Notes
- **Expected Resolution Date 1:** the date you expect to have this issue resolved in your upload file, either by fixing the value or removing it from the upload.
  - **Expected Resolution Date 2:** If you miss Date 1 and need more time, you can fill out another resolution date.
  - **Expected Resolution Date 3:** If you miss Dates 1 and 2, you can fill out another resolution date.
  - **Describe Why the Data Cannot be Removed:** If you selected answer option “We are unable to address this issue and will remove the data from our dataset” for Resolution Request Reason, please explain the details here on why it cannot be corrected.
  - **Data Manager Notes:** This is a space to type any relevant notes to the issues and timeline you are proposing.
    - *If you are requesting over eight weeks of time to fix issues, please explain here why your project needs extra time.*

After you are done editing, click the blue Submit button in the bottom left beneath the table. Your primary informaticist will review the request and change it to either “Approved” or “Denied”. You can always reach out to your EITs about any issues you have.

#### DESCRIPTION OF COLUMNS: BRANCHING LOGIC

- **Initial Upload ID:** The first upload to contain this issue
  - *you can see which file names and dates correspond to an upload ID on your data submission page*
- **Most Recent Upload ID:** The most recent upload to contain this issue
  - *you can see which file names and dates correspond to an upload ID on your data submission page*
- **Participant ID:** the unique record ID that corresponds to this issue
- **CDE Name:** Common Data Element as specified in the Data Dictionary Codebook
- **Conformant Values:** The accepted values specified in the Data Dictionary Codebook
- **CDE Value:** The nonconformant value present in upload file
- **Quality Resolution Type:** whether this is an issue of conformance, missingness, or a future issue type
  - *Right now, this table is only displaying conformance issues and all values will say “Conformance Issue”*
- **CDCC Response:** the response from the CDCC on the status of this request. It will be

“Requested” after you have filled out a request, and then it will change to either “Approved” or “Denied”.

The columns you can edit are:

- **Resolution Request Reason:** the reason you are submitting a request to the CDCC concerning this data point.
  - Verified and true outlier in the submitted data: You have verified that this outlier is a true data point and it needs to be included in RADx-UP reporting.
    - *Example: a participant has true data that did not follow the branching logic*
  - An exception was submitted for this CDE: You submitted a Wording Request Change to add an answer option to a categorical variable and it should be recorded.
    - *Example: you are not collecting the CDE used in the branching language*
  - We will address this and fix the issue: You know this variable is wrong and are planning to fix this and are entering the time needed to address it.
    - *Example: Someone said they got a flu vaccine in the last six months, but never in their life. You are going to figure out which variable is correct.*
  - We are unable to address this issue and will remove the data from our dataset: You know this variable is wrong and cannot fix it, but are planning to remove it from your upload and are entering the time needed.
    - *Example: Someone said they got a flu vaccine in the last six months, but never in their life. You have no way of verifying so you are removing data that does not align.*
- **Does this apply to every CDE with this value?:** is this an issue that applies to all CDEs of this value? Yes or No answer.
  - *If this was a wording request change that added an answer option to this CDE, it will apply to every instance of that value.*
  - *Example: you asked all participants the question "pregnancy\_status" without regarding bio\_sex\_birth first and need all cases of this to be allowed through.*
  - *Note that this applies to the specific CDE and VALUE. If you apply this to all cases of pregnancy\_status = 'No', it does not also apply to cases where pregnancy\_status = 'Yes'.*
- **Expected Resolution Date 1:** the date you expect to have this issue resolved in your upload file, either by fixing the value or removing it from the upload.
- **Expected Resolution Date 2:** If you miss Date 1 and need more time, you can fill out another resolution date.
- **Expected Resolution Date 3:** If you miss Dates 1 and 2, you can fill out another resolution date.
- **Describe Why the Data Cannot be Removed:** If you selected answer option “We are unable to address this issue and will remove the data from our dataset” for Resolution Request Reason, please explain the details here on why it cannot be corrected.
- **Data Manager Notes:** This is a space to type any relevant notes to the issues and timeline you are proposing.
  - *If you are requesting over eight weeks of time to fix issues, please explain here why your project needs extra time.*

## EXAMPLES

### True and Valid Data: A Participant's Age is 114.

If you have an outlier in your continuous variables that you have verified is a true value, you can use this table to inform the CDCC. You can go to the resolution table in MRUH and fill out this information.

The table shows all issues in your file and you will need to locate the CDE you want to check. They are in alphabetical order, so age\_yrs is at the top in this file.

Please see below data quality issues. Please review and add feedback around resolution. The informaticists will review and discuss with your project team if needed.

Show the following Approval Status(es):  Select  Requested  Approved  Denied

Resolution Request Reasons(es):  Select  Verified and true outlier in the submitted data  An exception was submitted for this CDE

We will address this and fix the issue  We are unable to address this issue and will remove the data from our dataset

#### Conformance Issues

Initial Upload ID	Most Recent Upload ID	Participant ID	CDE Name	Conformant Values	CDE Value	Quality Resolution Type	Resolution Request Reason	Does this apply to every CDE with this value?
788	788	7025597.0	age_yrs	0-110	114	Conformance Issue	Select	Select
788	788	5054638.0	age_yrs	0-110	1413	Conformance Issue	Select	Select
788	788	7025597.0	household_congr egate_2	1,10,2,3,4,5,6, 7,8,9,90	98	Conformance Issue	Select	Select
788	788	1869541.0	household_congr egate_2	1,10,2,3,4,5,6, 7,8,9,90	99	Conformance Issue	Select	Select
788	788	2719007.0	household_congr egate_2	1,10,2,3,4,5,6, 7,8,9,90	99	Conformance Issue	Select	Select

Under "Resolution Request Reason", you would select "Verified and true outlier in the submitted date". For "Does this apply to every CDE with this value?", you could select "Yes" if you're expecting even more truly 114-year-old participants – note that this won't apply to all participants with an age above 110, but only to those that are 114. You can select "No" if this is an outlier and you don't expect it to happen again.

Please see below data quality issues. Please review and add feedback around resolution. The informaticists will review and discuss with your project team if needed.

Show the following Approval Status(es):  Select  Requested  Approved  Denied

Resolution Request Reasons(es):  Select  Verified and true outlier in the submitted data  An exception was submitted for this CDE

We will address this and fix the issue  We are unable to address this issue and will remove the data from our dataset

**Conformance Issues**

Initial Upload ID	Most Recent Upload ID	Participant ID	CDE Name	Conformant Values	CDE Value	Quality Resolution Type	Resolution Request Reason	Does this apply to every CDE with this value?
788	788	7025597.0	age_yrs	0-110	114	Conformance Issue	Select	Select
788	788	5054638.0	age_yrs	0-110	1413	Conformance Issue	Select <b>Verified and true outlier in the submitted data</b> An exception was submitted for this CDE We will address this and fix the issue We are unable to address this issue and will remove the data from our dataset	Select
788	788	7025597.0	household_congr egate_2	1,10,2,3,4,5,6,7,8,9,90	98	Conformance Issue	Select	Select
788	788	1869541.0	household_congr egate_2	1,10,2,3,4,5,6,7,8,9,90	99	Conformance Issue	Select	Select
788	788	2719007.0	household_congr	1,10,2,3,4,5,6,	99	Conformance Issue	Select	Select

You don't need to input any Resolution Dates because this is a true data point. You can scroll over to Data Manager notes and input any information you'd like the CDCC to know about this data point.

Please see below data quality issues. Please review and add feedback around resolution. The informaticists will review and discuss with your project team if needed.

Show the following Approval Status(es):  Select  Requested  Approved  Denied

Resolution Request Reasons(es):  Select  Verified and true outlier in the submitted data  An exception was submitted for this CDE

We will address this and fix the issue  We are unable to address this issue and will remove the data from our dataset

**Conformance Issues**

Include the label of option	Expected Resolution Date 1	Expected Resolution Date 2	Expected Resolution Date 3	Describe Why the Data Cannot be Resolved	Data Manager Notes	CDCC Response
<input type="checkbox"/>	mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy		I verified the birthdate of this participant and they are 114 years old.	Select
<input type="checkbox"/>	mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy			Select
<input type="checkbox"/>	mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy			Select
<input type="checkbox"/>	mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy			Select

If this is all the editing you want to do, you can click the "Submit" button in the bottom left beneath the table. You can edit multiple CDEs on a page before submitting the page.



Show the following Approval Status(es):  Select  Requested  Approved  Denied

Resolution Request Reasons(es):  Select  Verified and true outlier in the submitted data  An exception was submitted

We will address this and fix the issue  We are unable to address this issue and will remove the data from our dataset

**Conformance Issues**

Initial Upload ID	Most Recent Upload ID	Participant ID	CDE Name	Conformant Values	CDE Value	Quality Resolution Type
788	788	5054638.0	age_yrs	0-110	1413	Conformance Issue
788	788	7025597.0	household_congregate_2	1,10,2,3,4,5,6,7,8,9,90	98	Conformance Issue
788	788	1869541.0	household_congregate_2	1,10,2,3,4,5,6,7,8,9,90	99	Conformance Issue
788	788	2719007.0	household_congregate_2	1,10,2,3,4,5,6,7,8,9,90	99	Conformance Issue
788	788	5732613.0	household_congr	1,10,2,3,4,5,6	99	Conformance Issue

A pop-up will appear that says “Data Successfully Inserted/Updated” when the submission is complete.

### Submitted Exception: household\_congregate\_2

If you submitted a Wording Request Exception through your project’s CDE Tracking Survey located in REDCap, you may have extra answer options that are showing as non-conformant in your data quality report. With this table, you can inform the CDCC that this is a valid answer option being reviewed by the NIH and you’d like it to no longer show up in the Data Quality Report.

You may have multiple rows of the same issue due to your exception request. You only need to fill out one row and use the column “Does this apply to every CDE with this value?”. Below you see three rows with “99” for household\_congregate\_2. On any of the rows, you can fill out the “Resolution Request Reason” with “An exception was submitted”. It will automatically fill in the “Does this apply to every CDE with this value?” with a “Yes”. In the text field “Please include the label of this field option”, type the meaning of “99” for household\_congregate\_2 within your study. In this example, “99” means “Living with a sibling”.

Please see below data quality issues. Please review and add feedback around resolution. The informaticists will review and discuss with your project team if needed.

Show the following Approval Status(es):  Select  Requested  Approved  Denied

Resolution Request Reasons(es):  Select  Verified and true outlier in the submitted data  An exception was submitted for this CDE

We will address this and fix the issue  We are unable to address this issue and will remove the data from our dataset

### Conformance Issues

CDE Name	Conformant Values	CDE Value	Quality Resolution Type	Resolution Request Reason	Does this apply to every CDE with this value?	Please include the label of this field option	Expected Date 1
age_yrs	0-110	1413	Conformance Issue	Select	Select		mm/dd/yy
household_congregate_2	1,10,2,3,4,5,6,7,8,9,90	98	Conformance Issue	Select	Select		mm/dd/yy
household_congregate_2	1,10,2,3,4,5,6,7,8,9,90	99	Conformance Issue	An exception was submitted	Yes	Living with a sibling	mm/dd/yy
household_congregate_2	1,10,2,3,4,5,6,7,8,9,90	99	Conformance Issue	Select	Select		mm/dd/yy
household_congregate_2	1,10,2,3,4,5,6,7,8,9,90	99	Conformance Issue	Select	Select		mm/dd/yy

You do not need any expected Resolution Dates in this case because the issue is resolved. You can type notes in the Data Manager Notes box to explain the situation, and then click Submit.

Please see below data quality issues. Please review and add feedback around resolution. The informaticists will review and discuss with your project team if needed.

Show the following Approval Status(es):  Select  Requested  Approved  Denied

Resolution Request Reasons(es):  Select  Verified and true outlier in the submitted data  An exception was submitted for this CDE

We will address this and fix the issue  We are unable to address this issue and will remove the data from our dataset

### Conformance Issues

Include the label of option	Expected Resolution Date 1	Expected Resolution Date 2	Expected Resolution Date 3	Describe Why the Data Cannot be Resolved	Data Manager Notes	CDC Response
	mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy			Select
	mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy			Select
Living with a sibling	mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy		We added this answer option and requested it in our CDE Tracking Survey	Select
	mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy			Select
	mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy			Select

Once you've submitted, you'll see that the answers you filled in for one row are not filled into every row where "household\_congregate\_2" was "99". Notice that the row with "household\_congregate\_2" equaling "98" does not have anything filled out because while it is the same CDE, it has a different value.

Please see below data quality issues. Please review and add feedback around resolution. The informaticists will review and discuss with your project team if needed.

Show the following Approval Status(es):  Select  Requested  Approved  Denied

Resolution Request Reasons(es):  Select  Verified and true outlier in the submitted data  An exception was submitted for this CDE

We will address this and fix the issue  We are unable to address this issue and will remove the data from our dataset

**Conformance Issues**

CDE Name	Conformant Values	CDE Value	Quality Resolution Type	Resolution Request Reason	Does this apply to every CDE with this value?	Please include the label of this field option	Expected Resolution Date 1
age_yrs	0-110	1413	Conformance Issue	Select	Select		mm/dd/yy
household_congr egate_2	1,10,2,3,4,5,6,7,8,9,90	98	Conformance Issue	Select	Select		mm/dd/yy
household_congr egate_2	1,10,2,3,4,5,6,7,8,9,90	99	Conformance Issue	An exception was submitted for this CDE	Yes	Living with a sibling	mm/dd/yy
household_congr egate_2	1,10,2,3,4,5,6,7,8,9,90	99	Conformance Issue	An exception was submitted for this CDE	Yes	Living with a sibling	mm/dd/yy
household_congr egate_2	1,10,2,3,4,5,6,7,8,9,90	99	Conformance Issue	An exception was submitted for this CDE	Yes	Living with a sibling	mm/dd/yy

Submit

### Address and Fix Issue: age\_yrs is 1413.

If you have an issue that you need time to fix and address, you can inform the CDCC using the table. For example, you have an age of 1413. You can consult with your study protocol and statisticians on if you can “fix” the issue or if you need to remove it from your upload.

Show the following Approval Status(es):  Select  Requested  Approved  Denied

Resolution Request Reasons(es):  Select  Verified and true outlier in the submitted data  An exception was submitted for this CDE

We will address this and fix the issue  We are unable to address this issue and will remove the data from our dataset

**Conformance Issues**

CDE Name	Conformant Values	CDE Value	Quality Resolution Type	Resolution Request Reason	Does this apply to every CDE with this value?	Please include the label of this field option	Expected Resolution Date 1
age_yrs	0-110	114	Conformance Issue	Verified and true outlier in the submitted data	Select		mm/dd/yy
age_yrs	0-110	1413	Conformance Issue	Select	Select		mm/dd/yy
household_congr egate_2	1,10,2,3,4,5,6,7,8,9,90	98	Conformance Issue	An exception was submitted for this CDE	Yes	Living with a sibling	mm/dd/yy
household_congr egate_2	1,10,2,3,4,5,6,7,8,9,90	99	Conformance Issue	An exception was submitted for this CDE	Yes	Living with a sibling	12/09/20
household_congr egate_2	1,10,2,3,4,5,6,7,8,9,90	99	Conformance Issue	An exception was submitted for this CDE	Yes	Living with a sibling	12/09/20

Submit

Now you can enter a date you will fix this by. You can enter only one date or all three at once. If it is more than eight weeks overdue, you should provide an explanation for why you need more time.

Show the following Approval Status(es):  Select  Requested  Approved  Denied

Resolution Request Reasons(es):  Select  Verified and true outlier in the submitted data  An exception was submitted for this CDE

We will address this and fix the issue  We are unable to address this issue and will remove the data from our dataset

**Conformance Issues**

Include the label of option	Expected Resolution Date 1	Expected Resolution Date 2	Expected Resolution Date 3	Describe Why the Data Cannot be Resolved	Data Manager Notes	CDCC Response
	mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy		patient actually 114	Select <input type="button" value="v"/>
	01/22/2022 <input type="button" value="calendar"/>	mm/dd/yyyy <input type="button" value="calendar"/>	mm/dd/yyyy		We are working on this.	Select <input type="button" value="v"/>
	mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy			Select <input type="button" value="v"/>
with a sibling	12/09/2021	12/10/2021	12/11/2021		sageTest01	Select <input type="button" value="v"/>
with a sibling	12/09/2021	12/10/2021	12/11/2021		sageTest01	Select <input type="button" value="v"/>

Then you can click submit.

### Example: Fixing a Missing CDE

If you need time to add a CDE to your project survey, you can note this in the resolution table. You can select "We will address and fix this issue" for your Resolution Request Reason and input the date you expect to upload a fixed file by in Expected Resolution Date 1. Provide context for how much time you need in the Data Manager Notes box and click submit.

If you will never be collecting a CDE, you still need to address and fix the issue by submitting a new CDE Tracking Form REDCap Survey. If you need time to coordinate with your EITs and project leader on which CDEs you are collecting or not, then you can follow the same process described here to note the time you need.