

Data Quality

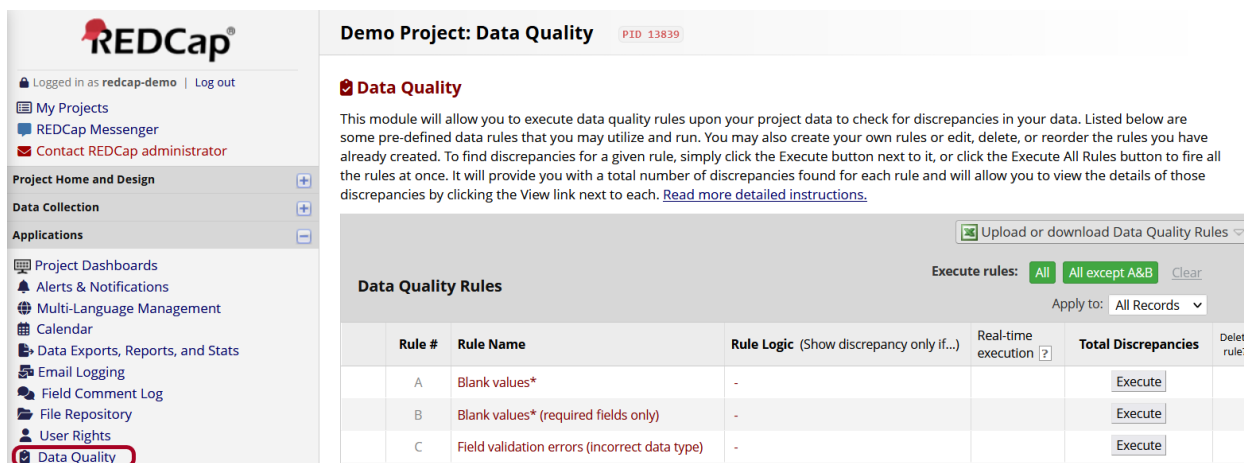
CCTS Biostatistics Core

May 2023

Data Quality Module

The Data Quality module allows you to run pre-defined and custom checks for data errors and discrepancies. These checks can help you identify data entry errors and values that no longer conform to a field's requirements (e.g., field validation or "required" status) if the field has been modified.

Within the module, you can identify values that violate your rules and either correct them or exclude them from future checks.



REDCap
 Logged in as redcap-demo | Log out
 My Projects
 REDCap Messenger
 Contact REDCap administrator

Project Home and Design
 Data Collection
 Applications

Project Dashboards
 Alerts & Notifications
 Multi-Language Management
 Calendar
 Data Exports, Reports, and Stats
 Email Logging
 Field Comment Log
 File Repository
 User Rights
 Data Quality

Demo Project: Data Quality PID 13839

Data Quality

This module will allow you to execute data quality rules upon your project data to check for discrepancies in your data. Listed below are some pre-defined data rules that you may utilize and run. You may also create your own rules or edit, delete, or reorder the rules you have already created. To find discrepancies for a given rule, simply click the Execute button next to it, or click the Execute All Rules button to fire all the rules at once. It will provide you with a total number of discrepancies found for each rule and will allow you to view the details of those discrepancies by clicking the View link next to each. [Read more detailed instructions.](#)

Upload or download Data Quality Rules

Execute rules: All All except A&B Clear

Apply to: All Records

Rule #	Rule Name	Rule Logic (Show discrepancy only if...)	Real-time execution	Total Discrepancies	Delete rule?
A	Blank values*	-		Execute	
B	Blank values* (required fields only)	-		Execute	
C	Field validation errors (incorrect data type)	-		Execute	

Types of Data Quality Rules

Pre-Existing Rules

The Data Quality module comes with nine pre-defined rules to help you identify common issues.

Rule #	Rule Name
A	Blank values*
B	Blank values* (required fields only)
C	Field validation errors (incorrect data type)
D	Field validation errors (out of range)
E	Outliers for numerical fields (numbers, integers, sliders, calc fields)**
F	Hidden fields that contain values***
G	Multiple choice fields with invalid values
H	Incorrect values for calculated fields
I	Fields containing "missing data codes"

- **A. Blank values**
 - Find all fields that have been left blank (empty), provided fields aren't hidden by branching logic.
- **B. Blank values (required fields only)**
 - Find all **required** fields that have been left blank (empty), provided fields aren't hidden by branching logic.
- **C. Field validation errors (incorrect data type)**
 - Identify values that conflict with a field's validation. For example, this rule would flag letters stored in an "integer" field.
- **D. Field validation errors (out of range)**
 - Identify values that exceed the minimum and maximum values for a field. For example, if you have an "age" integer field with lower and upper boundaries of 18 and 120, this rule would flag any values less than 18 or greater than 120.
- **E. Outliers for numerical fields. (numbers, integers, sliders, calc fields)**
 - Identify values in numeric fields that fall more than two standard deviations from the mean.
- **F. Hidden fields that contain values**
 - Fields that are hidden by branching logic are expected to be blank (empty). This rule identifies any fields that should be hidden but contain a value.
- **G. Multiple choice fields with invalid values**
 - **Multiple-choice fields** are expected to contain values that match the defined choice values. This rule identifies values that don't appear in the list of choices.
- **H. Incorrect values for calculated fields**
 - Find values that don't match the expected result of a calculation formula or a @CALCTEXT or @CALCDATE **action tag**.
- **I. Fields containing "missing data codes"**
 - Missing data codes can be used to document why a field is blank. Use this rule to find missing

data codes in your project.

Custom Rules

You may wish to define one or more custom data quality rules. For example, your project may have two versions of a variable, `item_v1` and `item_v2`. The following rule will find records that contain mismatched values in the two fields:

```
[item_v1]<>' and [item_v2]<>' and [item_v1] != [item_v2]
```

To define a custom rule for your project:

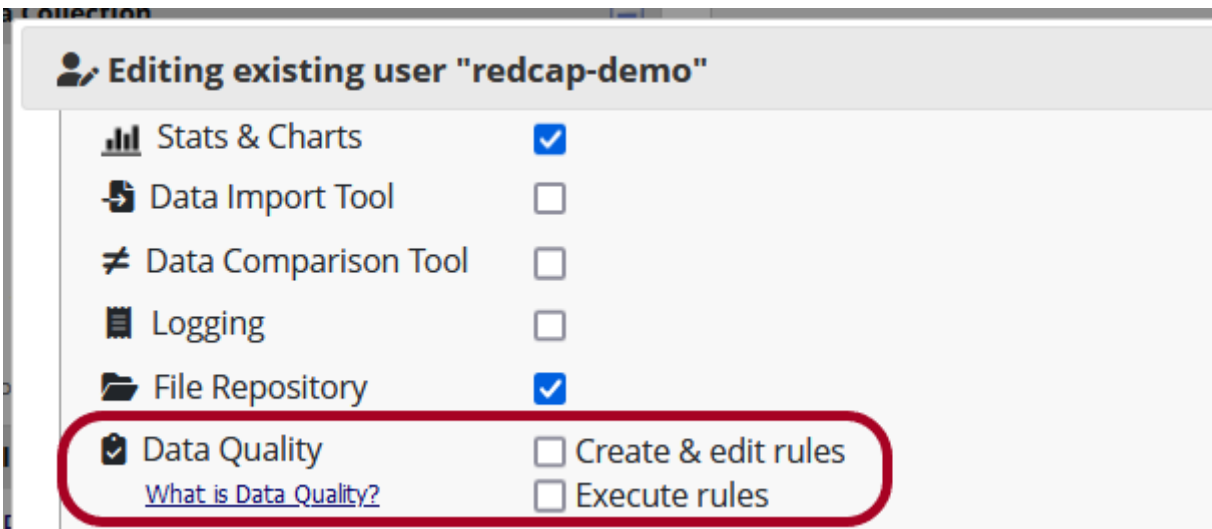
1. Give the rule a name or label in the “Rule Name” box.
2. Type the syntax into the “Rule Logic” box.
3. Decide whether you want to execute the rule in real time on data entry forms. If real-time execution is enabled, users will see a popup if they try to save a form with data that violates the rule.
4. Click Add.

Rule #	Rule Name	Rule Logic (Show discrepancy only if...)	Real-time execution ?
1	Conflicts between Item V1 and Item V2	[item_v1]<>" and [item_v2]<>" and [item_v1] != [item_v2]	<input checked="" type="checkbox"/>
	<div style="border: 1px solid #ccc; height: 60px; width: 100%;"></div> <p>Add</p> <p>Enter descriptive name for new rule (e.g., Participants below age 18)</p>	<div style="border: 1px solid #ccc; height: 60px; width: 100%;"></div> <p>Enter logic for new rule (e.g., [age] < 18)</p>	<input type="checkbox"/> Execute in real time on data entry forms ?

Steps to Run Data Quality Rules

Ensure Access

Use the [User Rights](#) page to define who can create custom rules and who can execute existing rules.



Execute One Rule

Run an individual rule on all records by clicking the `Execute` button beside the rule.

The screenshot shows the 'Data Quality Rules' interface. At the top right, there is a button 'Upload or download Data Quality Rules'. Below it, the 'Execute rules:' section has three buttons: 'All', 'All except A&B', and 'All custom', along with a 'Clear' link. The 'Apply to:' dropdown is set to 'All Records'. The table below has the following structure:

	Rule #	Rule Name	Rule Logic (Show discrepancy only if...)	Real-time execution ?	Total Discrepancies	Delete rule?
	A	Blank values*	-		Execute	

If REDCap finds any records that violate the rule, the number of matches will appear in the Total Discrepancies column. Click “view” to see a list of matches or “export” to download matches as a CSV file.

The screenshot shows the 'Data Quality Rules' interface after processing. A green checkmark and the text 'Processing Complete!' are visible. The 'Execute rules:' section remains the same. The 'Apply to:' dropdown is still 'All Records'. The table now shows:

	Rule #	Rule Name	Rule Logic (Show discrepancy only if...)	Real-time execution ?	Total Discrepancies	Delete rule?
	A	Blank values*	-		52 export view	

Execute All Rules

To execute all rules an once, all rules except the “Blank Values” rules, or all custom rules, click the appropriate button.

The screenshot shows the 'Data Quality Rules' interface. The 'Execute rules:' section is highlighted with a red box, showing the 'All', 'All except A&B', and 'All custom' buttons. The 'Apply to:' dropdown is 'All Records'. The table below has the following structure:

	Rule #	Rule Name	Rule Logic (Show discrepancy only if...)	Real-time execution ?	Total Discrepancies	Delete rule?
	A	Blank values*	-		Execute	
	B	Blank values* (required fields only)	-		Execute	

Limit to One Record

By default, rules are run on all records. To execute rules on a specific record, select the record ID from the dropdown menu.

The screenshot shows the 'Data Quality Rules' interface. The 'Apply to:' dropdown menu is open, showing options: 'All Records', '1', '2', and '3'. The 'Execute rules:' section is 'All'. The table below has the following structure:

	Rule #	Rule Name	Rule Logic (Show discrepancy only if...)	Real-time execution ?	Total	Delete rule?
	A	Blank values*	-			

Fix Incorrect Values

After you execute one or more rules, click “view” to see a list of values that violate a rule. The list will include the record ID and the details of the issue.

Take followup actions to address each issue:

1. Click a link in the “Discrepant fields with their values” column to navigate to the corresponding record.
2. If the discrepancy isn’t a concern and shouldn’t be returned the next time you run the data quality rule, click “exclude”.

Rule: **Blank values***
Discrepancies found: **52** [Export results \(CSV\)](#)

Record	Discrepant fields with their values	Status	Exclude
1	"Street, City, State, ZIP" address = [no data]	Missing value	2 exclude
1	"Phone number" telephone = [no data]	Missing value	exclude
1	"Clinician's Email" clinician_email = [no data]	Missing value	exclude

Rule H will find discrepancies between saved values and the current results of formulas in calculated fields. This rule may be useful if formulas have changed, if calculated fields were added after records were created, if calculated fields appear on instruments that don’t yet have saved data for a record, or if calculations rely on the “today” and “now” smart variables.

When Rule H identifies discrepancies, you will have the option to fix all issues at once by clicking the **Fix calcs now** button. To exclude an individual match from the action, first click “exclude.”

Rule: **Incorrect values for calculated fields**
Discrepancies found: **2** [Export results \(CSV\)](#)

Fix ALL incorrect calculated values? [Fix calcs now](#)

Record	Discrepant fields with their values	Status	Exclude
3	"Yn Both" yn_both = [no data] (Calculated value should be "0")	Incorrect value	exclude
3	"Cumulative" cumulative = [no data] (Calculated value should be "0")	Incorrect value	exclude

Calculated values fixed!

✔ **2 calculated values were successfully fixed.**

(The data quality rule is now automatically re-running to ensure that all calculated field values were updated correctly.)

Close