

Mobile Data Collection Toolkit - SurveyCTO

Automated quality checks on SurveyCTO

This tutorial is made available under the terms of the
[Creative Commons Attribution-Share Alike 4.0
International License](https://creativecommons.org/licenses/by-sa/4.0/).





What is the benefit of automated quality checks?

Setting up **automated checks** for some (or all) of the server's forms and datasets allows for example to:



- Alert on **outliers** in your data
- Alert on the **frequency of certain values** for a specific variable
- Detect and alert on **potential enumerator effects** (bias)

→ Automated checks result in **warnings**, but do **not prevent the data from being used**.

Activate automated quality checks

Automated data quality checks are activated in the "Monitor" tab:

The screenshot displays the SurveyCTO web interface. At the top left is the SurveyCTO logo. The navigation bar includes tabs for '1. Design', '2. Collect', '3. Monitor', '4. Export', and 'Configure'. The '3. Monitor' tab is highlighted with an orange circle. In the top right corner, the user profile 'renforcement-osc@cartong.org' is visible. Below the navigation bar, a sub-header 'Automated quality checks' is shown. The main content area features a toggle switch labeled 'Automated quality checks', which is currently turned 'ON' and circled in orange. Below this, there are icons for 'Refresh', 'Search', and 'Help'. The main data area shows a list of survey groups under the heading 'Group Ghana'. Two groups are listed: 'Ghana Baseline' and 'Ghana Endline'. Each group has a row of action icons: 'Options', 'Checks', 'Run now', and 'Report'. Below each group name, the status is displayed as 'Quality checks: 0, Nightly? No, Reports to: (nobody)'.

Manage automated quality checks

In this section, you can, for each form:

Automated quality checks

Refresh Search Help

Group Ghana

Ghana Baseline

Options Checks Run now Report

Quality checks: 0, Nightly? No, Reports to: (nobody)

- 1 Configure the **modalities** (options) of the quality checks (frequency, sending a report by email)
- 2 **Configure** automated **quality checks**
- 3 **Launch quality checks** instantly
- 4 **Download the report** following the execution of the quality checks

Creation of an automated quality check

Create an automated quality check (1/4)

To create a new automated quality check (**2**) - example with the *Ghana Baseline* form:

Ghana Baseline

Options Checks Run now Report

Quality checks: 0, Nightly? No, Reports to: (nobody)

Quality checks

Configure automated quality checks to quickly identify potential issues as submissions arrive. Identifying issues early on can help you resolve them before it's too late, and your project is over. You may also configure your workflow to use the results of these checks to intelligently flag certain submissions for review.

[Click here to learn more about configuring quality checks.](#)

You do not yet have any quality checks configured. Click the button below to add your first quality check.

+ Create quality check

Close

Create an automated quality check (2/4)

Select the quality check you wish to apply:

Create a new quality check

What type of check do you want to run?

- Value is too low**
Use this to check every record and warn for each one that has a field value below the threshold you specify.
- Value is too high**
Use this to check every record and warn for each one that has a field value above the threshold you specify.
- Value is an outlier**
Use this to check every record and warn for each one that has a field value more than x times outside the interquartile range (IQR).
- Mean is too low**
Use this to warn for each field that has a mean below the threshold you specify.
- Mean is too high**
Use this to warn for each field that has a mean above the threshold you specify.
- Value is too frequent**
Use this to check for a particular value's frequency and warn whenever it is above the threshold you specify.
- Value is too infrequent**
Use this to check for a particular value's frequency and warn whenever it is below the threshold you specify.
- Group mean is different**
Use this check to test for equality of means across groups using an ANOVA test.
- Group distribution is different**
Use this check to test for equality of distributions across groups using a chi-squared test.

Create an automated quality check (3/4)

Select the field(s)/variable(s) on which you wish to apply the selected quality check:

Create a new quality check

Type of quality check: **Value is too low**

On which field(s) do you want to perform this check?

Search for a specific field...

Filter this list for me ?

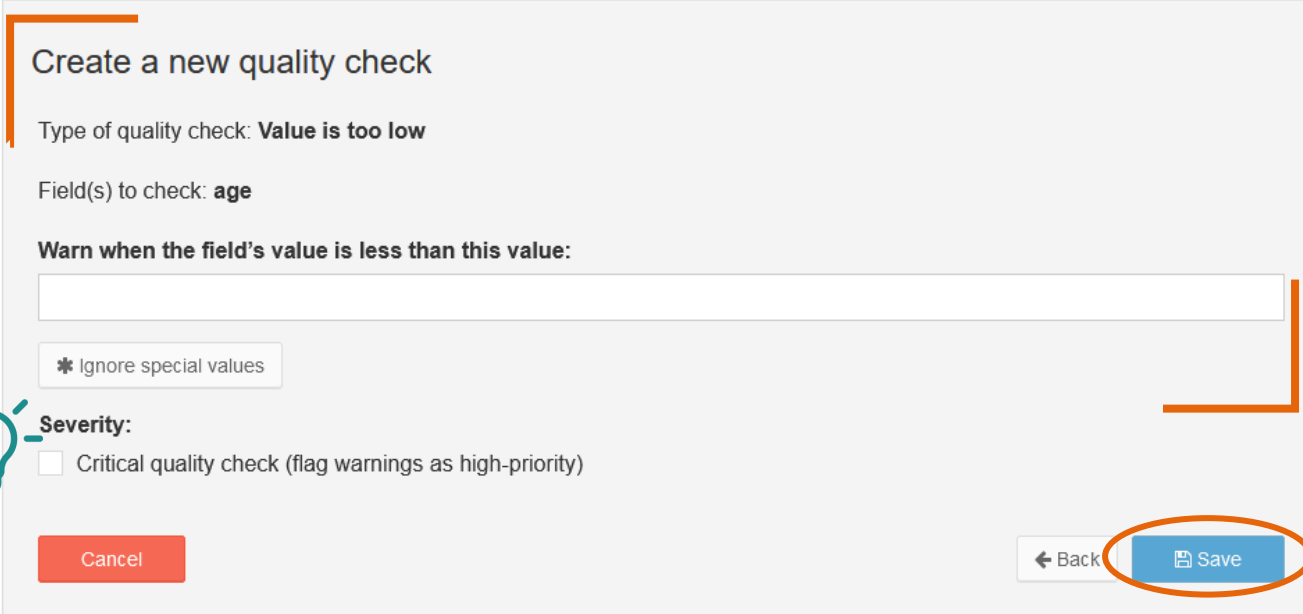
- date_jour
- nom
- age
- age_groupe
- espece_autre
- espece
- label
- formids
- id



You can set up as many quality checks as you want, and each can cover as many fields/variables as you want.

Create an automated quality check (4/4)

Enter the details for the quality check you have selected (e.g. maximum value, % of values that will be considered as extreme, maximum frequency, etc.):



Create a new quality check

Type of quality check: **Value is too low**

Field(s) to check: **age**

Warn when the field's value is less than this value:

* Ignore special values

Severity:

 Critical quality check (flag warnings as high-priority)

Cancel Back Save



Allows you to specify whether or not warnings (in case of non-compliance with the constraints specified above) should be classified as "critical" in the quality check reports.

Example of automated quality checks

Example of automated quality checks (1/8)

Let's apply automated quality checks to the following question (variable name associated with this question: *age*):

How old is Steve's pet?

12

So let's set up the following quality checks:

- 1) If the **age** is **higher than 15**, you want to receive a **warning**.
- 2) If the **age** is **greater than 20**, you want to receive a **critical warning**.

VALUE IS GREATER THAN 20

Fields: age

Edit

Delete

VALUE IS GREATER THAN 15

Fields: age

Edit

Delete

Example of automated quality checks (2/8)

The 2nd quality check must result in a **critical warning**, so it is configured as follows:

Edit a quality check

Type of quality check: **Value is too high**

Field(s) to check: **age**

Warn when the field's value is greater than this value:

* Ignore special values

Severity:

Critical quality check (flag warnings as high-priority)

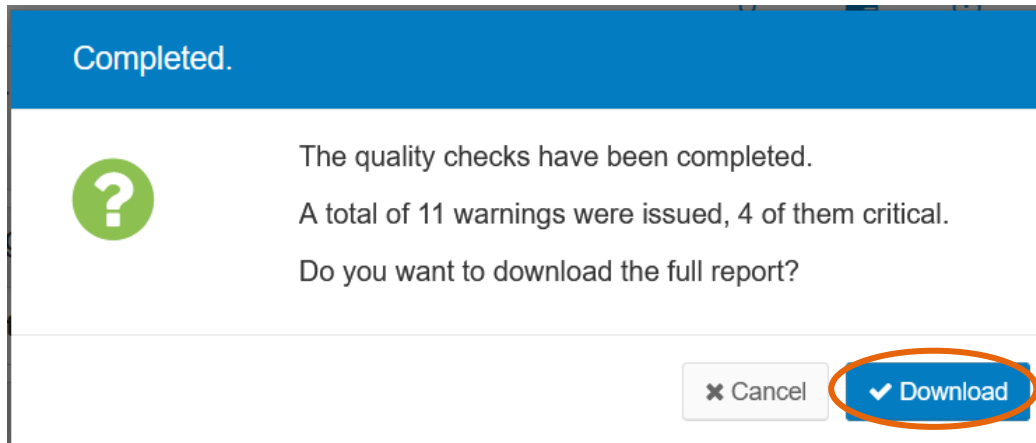
Example of automated quality checks (3/8)

We then launch the automated quality checks:

The screenshot displays the SurveyCTO web interface. At the top left is the SurveyCTO logo. The navigation bar includes '1. Design', '2. Collect', '3. Monitor' (circled in orange), '4. Export', and 'Configure'. On the right, there is a user profile for 'renforcement-osc@cartong.org' and a section for 'Automated quality checks'. Below this, a toggle switch for 'Automated quality checks' is shown in the 'ON' position (circled in orange). A toolbar contains 'Refresh', 'Search', and 'Help' icons. The main content area shows a survey titled 'Steve_TEST_registration' with a 'Run now' button (circled in orange) and 'Options', 'Checks', and 'Report' sub-buttons. Below the survey title, it states 'Quality checks: 2, Nightly? No, Reports to: (nobody)'.

Example of automated quality checks (4/8)

In this case, there are 11 warnings, including 4 critical ones:



→ It is possible to download the automated quality check report (csv file).



In this example, an age greater than 20 would result in both:

- A non-critical warning (age > 15)
- A critical warning (age > 20).

Example of automated quality checks (5/8)

Here is an example of an automated quality check report.

	A	B	C	D	E	F	G	H	I	J	K
1	id	last-reported	dataset-id	warning-id	row-id	field	group-field	group-id	value	warning	critical
11	Steve_example_1- VAL_TOO_HIGH- uuid:6ef9a5a3-2f36-4baa-967c-239e26f793b2-age-20-crit	2023-04-04 07:49:43 UTC	formid.Steve_example_1	VAL_TOO_HIGH	uuid:6ef9a5a3-2f36-4baa-967c-239e26f793b2	age			20	The field 'age' was found to be '22' for the submission with ID 'uuid:6ef9a5a3-2f36-4baa-967c-239e26f793b2'.	1
12	Steve_example_1- VAL_TOO_HIGH- uuid:ee7c6283-d33e-4e2e-8b04-9ff7d6e623eb-age-20-crit	2023-04-04 07:49:43 UTC	formid.Steve_example_1	VAL_TOO_HIGH	uuid:ee7c6283-d33e-4e2e-8b04-9ff7d6e623eb	age			20	The field 'age' was found to be '31' for the submission with ID 'uuid:ee7c6283-d33e-4e2e-8b04-9ff7d6e623eb'.	1
13	Steve_example_1- VAL_TOO_HIGH- uuid:ae02786d-c9cd-4075-8de1-a8d8897c6481-age-15	2023-04-04 07:49:43 UTC	formid.Steve_example_1	VAL_TOO_HIGH	uuid:ae02786d-c9cd-4075-8de1-a8d8897c6481	age			15	The field 'age' was found to be '31' for the submission with ID 'uuid:ae02786d-c9cd-4075-8de1-a8d8897c6481'.	0
14	Steve_example_1- VAL_TOO_HIGH- uuid:d649d59e-f0f3-4333-b2be-ba296140ff4b-age-15	2023-04-04 07:49:43 UTC	formid.Steve_example_1	VAL_TOO_HIGH	uuid:d649d59e-f0f3-4333-b2be-ba296140ff4b	age			15	The field 'age' was found to be '35' for the submission with ID 'uuid:d649d59e-f0f3-4333-b2be-ba296140ff4b'.	0

→ *Warning* column: details of the reasons why a warning was issued.

→ *Critical* column: specifies whether it is a critical warning (1) or not (0).



What happens if a warning is classified as "critical"?



There is **no difference in treatment** (by SurveyCTO) between warnings classified as "critical" and others that are not.

→ **SurveyCTO** simply **mentions which warnings are critical**.

→ It is then up to **the data quality check person to pay closer attention to critical warnings**.

Example of automated quality checks (6/8)

It is also possible to view the warnings in the following way:

The screenshot displays the SurveyCTO web interface. At the top, the navigation menu includes '1. Design', '2. Collect', '3. Monitor', '4. Export', and 'Configure'. The '3. Monitor' tab is highlighted with an orange circle. Below the navigation, the main content area is titled 'Form submissions and dataset data', which is also circled in orange. A toolbar below the title contains 'Refresh', 'Search', and 'Help' icons. The main content area shows the form name 'Steve_TEST_registration' and a 'Monitor form data' button, which is circled in orange. To the right of this button are other options: 'External viewers', 'Review workflow', 'Purge form data', 'Look up by key', and 'Advanced mode'. Below the form name, the text reads 'Form ID: Steve_example_1, Complete submissions: 16 (latest Apr. 3, 2023 at 12:36:36PM)'. The 'Monitor form data' section includes a heading, a descriptive paragraph, and a 'Submissions to include:' section with three radio button options: 'All submissions' (selected), 'A random subset of submissions (8 out of 16)', and 'Recent submissions only (from the last 7 days)'. At the bottom of this section are 'Start Data Explorer' and 'Close' buttons.

Example of automated quality checks (7/8)

The warnings are reported here:

SurveyCTO Data Explorer - Monitor

Steve_TEST_registration

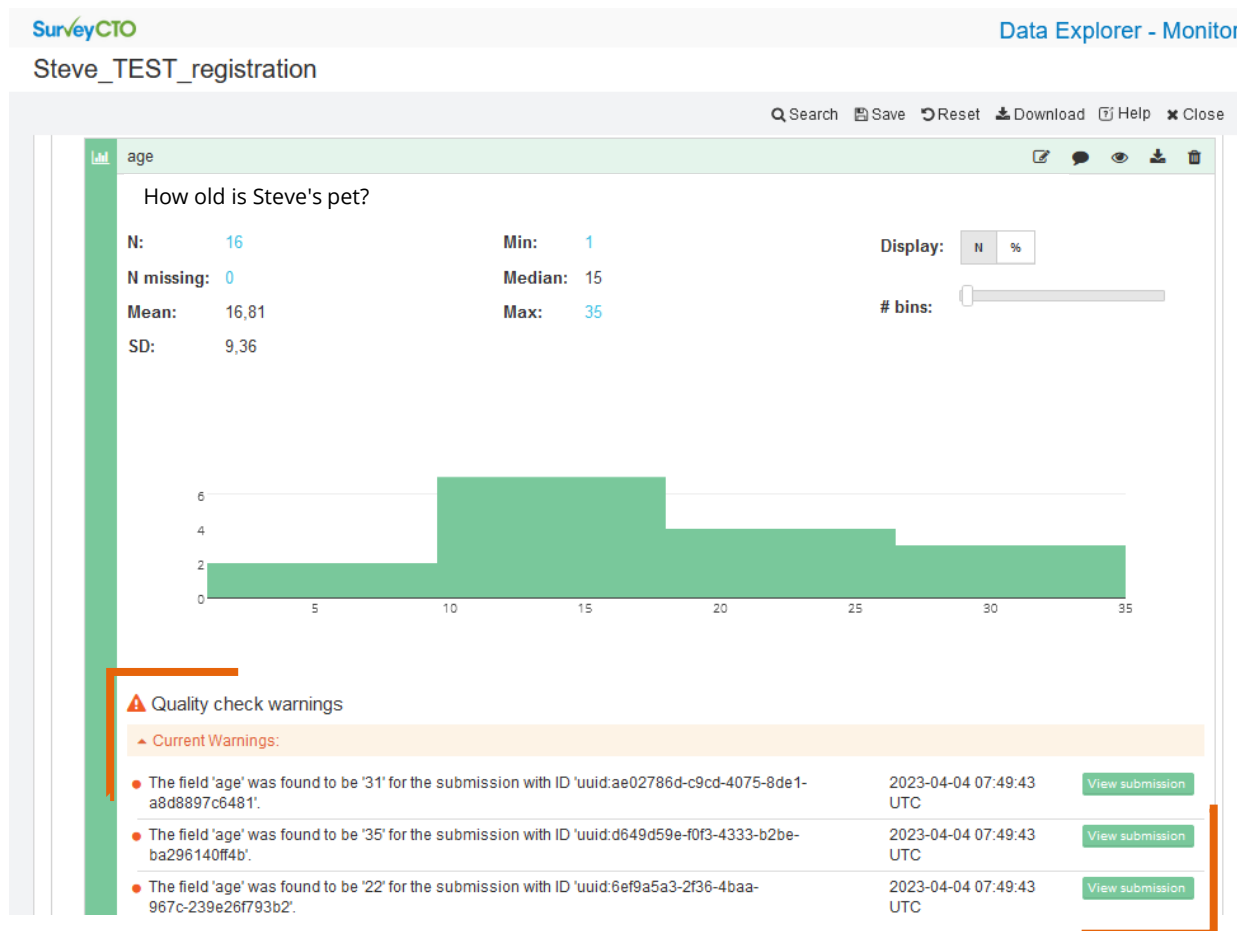
Q Search Save Reset Download Help Close

SUBMISSIONS total: 16 not loaded: 0 excluded: 0 available: 16	FIELDS total: 14 not loaded: 0 note fields: 1 available: 13	DURATION total: 4m 53s average: 18s	CHECKS configured: 3 warnings: 18
--	--	--	--

→ It is also possible to download the automated quality check report (csv file) at this location.

Example of automated quality checks (8/8)

Finally, in the "Data Explorer", the details of the warnings can be found for each question to which the automated quality checks apply.



Limitations of automated quality checks

Limitations of automated quality checks

However, there are some **limitations** regarding automated quality checks:



- x **Unable to** create quality checks for **encrypted fields*** .
- x **Unable to** create quality checks for **repeated fields** (i.e., part of a group of repeated questions - *begin_repeat / end_repeat*, on an XLSForm).

Other SurveyCTO tutorials available

In this toolbox, you will also find the following tutorials:

- **Groups, subgroups** and workspaces
- Centralized management of **roles, users** and **teams**
→ **Scenario** on the use of teams, groups and roles
- **Encryption of** data collected on SurveyCTO
- **Security settings** for SurveyCTO Collect
- **Submission types** on SurveyCTO
- **Case Management** on SurveyCTO
→ **Scenario** on case management and user-based access control
- **Importing** data into SurveyCTO
- **Workflows** of a SurveyCTO form
- Data **filtering** before downloading on the SurveyCTO server
- **Bulk data deletion** on the SurveyCTO server

Acknowledgements

This publication is supported by Terre des hommes (Tdh) and the French Development Agency (AFD). Nevertheless, the ideas and opinions presented in this toolkit do not necessarily represent those of Tdh and AFD.

This presentation was designed using resources from [Flaticon](#).

This tutorial is made available under the terms of the [Creative Commons Attribution-Share Alike 4.0 International License](#).

