

<b>Quantitative research - Data processing - Data entry accuracy</b>	Set-up & Conduct- Process & Analyse data	
	VERSION	4.0

## Aim

To explain how to determine the accuracy of data entry.

## Requirements

### Non WMO research

- Prior to data cleaning, researchers should evaluate the data entry error of their data;
- If the amount of errors discovered by data entry evaluation is larger than 3% per (electronic/paper) case report form/(paper) questionnaire, then the form/questionnaire needs to be double-entered entirely.
- As nWMO research does not get monitored by CMC, and may not be checked by a Sponsor's monitor either, it is recommended to include double-data entry for a percentage of all study data and 100% for crucial data points for primary outcomes and Safety evaluation.

### WMO research

- Use study database / electronic data capture (EDC) system that maintains an audit-trail (see 9.2.2 of the NFU Guideline Quality assurance for human research)
- Use the setting 'Confirm Changes' in Castor EDC, or equivalent in a different system, to ensure a reason for change is provided
- Ensure validation checks are built in and documented (including double-data entry procedures if a monitor only performs limited source data verification)

## Documentation

### Non WMO research

- The percentage of inconsequential errors of data entry;
- The number of records re-entered.

### WMO research

- Define automated and manually built-in checks in a Data Validation and Derivation plan
- Ensure the Test documentation includes information on if the automated validations described in this Data Validation and Derivation plan have been tested
- A Study Operations Manual may include relevant sections, cross-references can be made to prevent double administration of all details

## Responsibilities

### Executing researcher:

- To decide whether an extensive (full data entry) or less extensive (sampling) evaluation for data entry errors and interpretation needs to be carried out during the data collection period / prior to start of the data cleaning process;
- To determine the percentage of the data entry error;
- To re-enter (a sample of) registration forms/questionnaires of respondents, if necessary;
- To document the number of records re-entered and the percentage of inconsequential errors found.

### Project leaders:

- In case of sampling, to draw a small sample from the respondents (approx. 5%) and to have these registration forms/questionnaires re-entered by someone other than the individual who completed the first data entry;

# Amsterdam Public Health



Amsterdam UMC  
University Medical Centers



VRIJE  
UNIVERSITEIT  
AMSTERDAM



UNIVERSITY  
OF AMSTERDAM

- To ensure this procedure is carried out (on time).

Research assistant:

- To re-enter (a sample of) registration forms/questionnaires of respondents, if necessary.

## How To

### For paper source documents entered into a spreadsheet and Blaise:

(A database should be used, only with limited variables and a very small patient group, manual entry for non-WMO research is still allowed, but electronic data capture systems always advised). To evaluate the entry errors by sampling, the project leader should draw a small sample from the respondents (approx. 5%), and have the questionnaires or registration forms re-entered into an empty database. This second input should be carried out in principle by someone other than the individual who completed the first data entry. For instance, if the first round of data entry is done by a project assistant the second can be carried out by the researcher. The reliability of the entered data can be assessed by comparing the first and second round of data entry (using an option in SPSS, Data->Compare Datasets).

If the amount of errors discovered is greater than 3% per registration form/questionnaire (in comparison to the total number of variables inputted), then the registration form/questionnaire needs to be double-entered entirely! Subsequently, the first and second input should be compared in the same way. If the second input is carried out by the same person, then the permissible margin of error is smaller than 3%, i.e. 1.5%. This procedure applies to both manual input, for instance by using Blaise and scanned questionnaires. For scanned questionnaires the forms from the sample are re-scanned into a separate file.

For webbased questionnaires: (e.g. Surveys in Castor, Survlyzer or LimeSurvey), this procedure is not applicable, but automated validations should be included to limit data entry errors by participants (e.g. warning if a question is not answered, using standard categories where possible, and defining min-max values).

For electronic case reports forms (e.g. Forms and Reports in Castor EDC), which are completed by study team members, (partial) double-data entry is also advised (by the IGJ inspectorate) as a useful method to improve data quality when not all data is source data verified.

The necessity of full or partial double data entry is determined by issues such as:

- Irregularities observed during data collection;
- The complexity of the registration forms/questionnaires entered (large risk of interpretation errors);
- The required reliability of the data;
- Doubt about the reliability and accuracy of the data entry clerk(s);
- Whether controls have been built in to the data entry program to detect inconsistencies and out-of-range values.

The Research Data Management department offers several courses on Research Data Management to discuss requirements and practical implications, see [Research data management \(amsterdamumc.org\)](https://amsterdamumc.org) or course information on Intranet.

# Amsterdam Public Health



Amsterdam UMC  
University Medical Centers



VRIJE  
UNIVERSITEIT  
AMSTERDAM



UNIVERSITY  
OF AMSTERDAM

## Audit questions

1. Has the data entry been assessed and documented?
  - a. If not, why not?
  - b. If so, how has the assessment been carried out, and what was the result?
2. For poor results (more than 3% errors): Was the entire registration form/questionnaire re-entered, or were other actions taken?

## LINKS

	Link
VUmc Data Management Course	<a href="http://www.datamanagement-vumc.nl/cursussen.aspx">http://www.datamanagement-vumc.nl/cursussen.aspx</a>

## DOCUMENT HISTORY

Version	Status	Date	Name
4.0	Revision	26MAY2021	Elize Vlaminck, Dr. Wouter van Ballegooijen
3.0	Text updated	02DEC2016	EMGO
2.0	Revision format	16JUL2015	EMGO
1.2	Availability of a new comparison programme	31OCT2013	EMGO
1.1	English translation	01JAN2010	EMGO
1.0	-	31MAR2004	EMGO

## DOCUMENT APPROVAL

Role	Name	Date
Project Leader	Dr. Seta Jahfari	27MAY2021