

Data quality procedures

Raw data were downloaded from the webserver. Raw data were split into two different files:

File 1: 10.856 records from individual participants

File 2: 9561 records on sports

Deleting whole records to preserve high data quality in Individual data

Deleted 2716 records containing no answers at all

Moving 882 records from non-athletes to a different file and deleting records with answers to sociodemographic items led to 841 records from non-athletes remaining. These data can only be used to compare the sample of athletes to the sample of non-athletes in terms of sociodemographic variables

For data from athletes:

Initial Number of records from recreational athletes: 7258

Deleted test records which were entered during the survey phase – 7226 records remaining

Deleted records without any answers – 7146 records remaining

Deleted 9 implausible records and records not entailing information on doping or medication use:

- born 1929, intersexual, did not finish school, answered all RRT questions in less than 2 seconds
- born 1929, did not finish school, did not answer any RRT questions

Identifying tampering with the survey basing on the time to answer the first RRT-question:

- We analyzed the time to answer the first RRT question (see figure 1). With 5 or more seconds, the time seems to increase at a nearly constant rate. Therefore, we set the threshold for detecting tampering with the survey to below 5 seconds to avoid too many false positives. Still these short times can be possible for respondents who have read the instruction page and the related help page thoroughly.

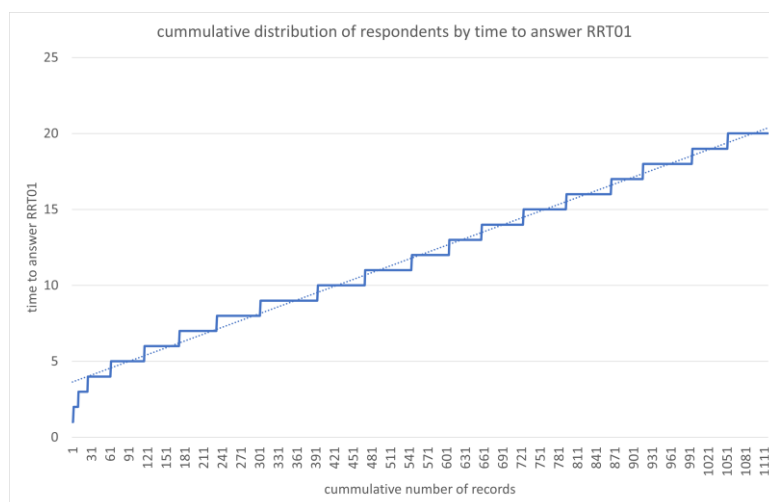


Fig. 1: analysis of the time to answer the first RRT question for answering times below 20 seconds

- Delete records with implausible combinations of age and education and who took less than 5 seconds to answer RRT01 completely and also from sports dataset. Records were only deleted if the time to answer the RRT questions was less than 10 seconds for RRT 01, 02 and 05 and time for RRT 03 (and RRT 04 if applicable) were both less than 5 seconds.
- Basing on this criterion: 3 records deleted. Final sample size: 7143 records.

Deleting single items to preserve high data quality

Delete only items age and education for records with implausible combinations of age and education but who took at least 5 seconds to answer RRT01.

- Born before 1995 and still in school: age and education replaced by missing data codes for 10 records
- Born 2000 or later and holding a doctorate degree: age and education replaced by missing data codes for 3 records
- Born 2004 or later and holding a masters degree: age and education replaced by missing data codes for 2 records

Apply the same logic to equal answers in psycho scales.

- Answered all psycho scales with equal values: 15 records.
- Answered 3 psycho scales with equal values: 395 records.
- Answered 2 psycho scales with equal values: 638 records.
- Answered 1 psycho scale with equal values: 1237 records

For these records we replaced the answers in the psycho-scales with equal values by missing data code.

For the different webpages, the number of respondents per page was:

Individual records with RRT01 answered: 4795

Individual records with RRT02 answered: 4483

Individual records with RRT05 answered: 4225

Individual records with perceived prevalence answered: 3863

Individual records with psy1 answered: 2649

Individual records with psy2 answered: 1598

Individual records with psy3 answered: 2410

Individual records with psy4 answered: 3403

Data quality check, data from recreational athletes

Saved as FPSportData_athletes with data for single sports, including 9561 records

Deleted sport records from individual test records and individuals whose records were deleted because of implausible datasets (see above) – 9431 records remaining

RRT-answers for doping in sports from 5766 different individuals

RRT-answers for doping in more than 1 sport from 1814 individuals

We asked for a shared list of 155 different sports in all questionnaires and additionally for 17 different sports which were seen as relevant in only some participating countries resp. languages (4 in Great Britain, 2 in Denmark, 1 in Germany, 10 in Norway). Beyond that, participants entered 59 different sports which were not listed, raising the number of sports to 218.