Documentation of weight calculations

The data from the RRT survey were weighted to account for different return rates in different countries as well as in different age groups and different sexes. The objective of weighting is, to approach the population structure better than with an unweighted sample. Nevertheless, this only holds if answering and not answering parts of the population within the subgroups are similar (Elliot, 1991).

As the survey design did not include sample quotas and was constructed to gain as much information from as many respondents in each participating country, design weighting was not feasible. Therefore, we had to calculate "redressment" weights (Arzheimer, 2009) basing on the sampled data. We used data, describing the age and sex structure of the overall national populations as well as data describing the relative frequency of doing in recreational sport by country, sex and age to calculate the cell weights for 7 countries * 5 age categories * 2 sexes (70 cells).

Source of national statistics

Population data were retrieved via the Eurostat data browser (https://ec.europa.eu/eurostat/) for EU-28 at 01. January 2019. Data were selected for males and females in 5 years birth-cohorts starting at 15-19 years old and finishing at 85+ years old.

Source of recreational sport data

Data were taken from Eurobarometer 88.4 (https://www.gesis.org/en/eurobarometer-data-service/home) except data for Norway. This was the most recent wave of Eurobarometer surveys, containing information about sport and physical activities. To calculate the weights, the raw data from the GESIS data archive were analyzed by the researchers.

As the Eurobarometer questionnaire did not contain am item which would unanimously differentiate between recreational athletes and other respondents, we had to approach the vague concept of
"recreational athlete" from the items which were included. Eurobarometer respondents were classified as recreational athletes if their answers to the items

- QB1 How often do you exercise or play sport?
- QB2 And how often do you engage in other physical activity such as cycling from one place to another, dancing, gardening, etc.?

This classification may or may not overlap with different (national) interpretations of the concept "recreational athlete" but it was the best approach to address this issue.

For Norway, data were retrieved through the website of the Norwegian national statistical bureau (https://www.ssb.no/en/). From the database "05782: Training and exercising (per cent)", a table was generated which included the percentage of respondents, answering "I never exercise or train" by age class and by sex.

Weight calculations

Due to the multitude of sub-populations of interest in the RRT study, we calculated weights to the following objectives:

- To correct for age and gender bias on the national level of each participating country for the country analyses.

As from Eurobarometer studies, it is well-known fact that recreational sport participation differs between Northern and Southern Europe, we grouped the participating countries accordingly and calculated additional weights

- to correct for age and gender bias as well as return rates from the 8 participating countries when estimating parameters for multi-national subgroups of the participating countries from Northern, Central/Western and Southern Europe.

Lastly, to calculate for overall results, we also calculated weights

- to correct for age and gender bias as well as return rates from the 8 participating countries when estimating parameters for the overall dataset.

Additionally, we calculated weights which address the need to conclude beyond the scope of the participating countries although representativeness of these weighted samples is questionable:

- To correct for age and gender bias as well as return rates from the 8 participating countries when estimating parameters for multi-national subgroups from Northern, Central/Western and Southern Europe beyond the participating countries.
- To correct for age and gender bias as well as return rates from the 8 participating countries when estimating parameters for North-, Western- and South-Europe.

The last two weights were calculated in case that the results of the prevalence study should also be used in the context of non-participating states. It is important to keep in mind that such use of the data goes beyond the range of evidence which is substantiated by the present survey.

The sample distribution by age classes (as reported by Eurostat) and sex (only male and female) contained cells with extreme low frequencies and partly even zero records. This affected the age class 75 years and older for females from Denmark (1 record) and both sexes from Great Britain, Spain, Germany and Italy in the age classes 65 and above (maximum records per cell: GB: 2, ES: 4, GE: 2, IT: 1). From Greece and Cyprus, there was only one record from both sexes in the age classes 55 and above.

For Norway, age classes were selected as reported by the SSB. There were only three records for males aged 75 and higher and only two for females aged 65 and above.

Based on these data, we decided to calculate country weights only for age classes with at least five records in both sexes, meaning:

- the country weights for Denmark for a recreational sport population aged 15 to 74,
- the country weights for from Great Britain, Spain, Germany and Italy for a recreational sport population aged 15 to 64, and
- the country weights for Greece and Cyprus for a recreational sport population aged 15 to 54,

For the weights for the 8 participating countries as well as the NWS-European countries, we decided on a recreational sport population, aged 15 to 64, taking into account the fact that especially for Southern European countries, the estimate for the upper age class is only based on records from Spain and from Italy. The same limitations hold for weighted estimates for NWS-European countries

For each of the countries resp. the national subgroups, we estimated the age- and gender structure of the recreational sports population based on the Eurobarometer data. Combined with the age- and gender specific population size, the size of the recreational sports population for each country and structured by age and sex was calculated. Based on these values, the relative size of each country-

sex- and age-specific recreational sports population within the overall sports population from this country resp. the relevant multi-national group of countries was calculated.

The weights for individuals in the response dataset was then calculated as this relative size of the (country-sex-and age-specific) recreational sports population, divided by the relative size of this (country-sex-and age-specific) subsample. Weights were normed by dividing them by the mean of the weights in order not to artificially excess the technical sample size in inferential statistics.

Response to the different questions varied widely within the final quality-controlled sample dataset (see table 1). Therefore, a special routine was developed to ensure that correct weights were calculated for each single question separately.

Tab. 1: Number of respondents for different questions. The order of the psychological scales differed between individuals. Numbers for these scales were only counted for complete answers to all items within the scales.

	RRT-questions			Perceived prevalence	psychological scales			
Question	RRT01	RRT02	RRT05		1	2	3	4
N	4795	4483	4225	3863	2649	1593	2410	3403

When (not) to weight

Weighting is important for any descriptive statistics in order to ensure the correct representation of age classes, sexes and countries in the statistic which is calculated. As such, any data which will be presented to an extra-academic public should always be calculated as a weighted statistic.

For inferential statistics, weighting can lead to less biased estimates if the variables which are used to weight the data (here, age, sex and country) correlate with the effect under study. On the other hand, weighted statistics do not per se outperform unweighted statistics, especially if the weighting variables are used (e.g.in a linear model) as additional predictors to reduce the level of error variance. Therefore, the decision to weight or not to weight should be scrutinized in advance. In cases of insecurity, comparing results from weighted and unweighted statistics can provide a hint to which extent the effect under study is stable.

Effects of weighting onto the results

The following tables shows the results from an unweighted and a weighted calculation. Table 2 compares unweighted and weighted subjectively estimated percentages of dopers in the country, the respondent lived in in 2019 including the standard deviation.

Tab. 2: Individually estimated prevalence of doping in the respondent's home country in 2019.

	Dk	en	es	ge	gr+cy	it	no
mean	5.63	10.72	18.09	14.08	21.70	22.02	5.41
sd	11.47	16.77	22.79	18.51	23.71	21.90	10.91
w.mean	5.49	13.77	16.76	14.47	24.50	16.31	5.39
w.sd	10.69	19.71	21.96	18.59	24.13	20.41	11.13

The effect of weighting differs considerably between the included countries. While for the data from Denmark, Norway and Germany, there are only marginal differences (difference between mean and weighted mean < 1), both the size and the direction of the effect differ from the other countries. For Great Britain as well as Greece and Cyprus, weighting increases the mean by approx. 3, it decreases the mean for Spain by slightly more than 1 and for Italy nearly by 4. These effects do not systematically depend on the sample size or on the height of the estimated value.

The effect of weighting procedures on RRT-variables is shown in tables 3 and 4 for the first RRT-question "Did you use OTC medication to enhance your athletic performance in 2019". For Denmark (table 3), all estimates were massively influenced by weighting. For Spain (table 4) there was nearly no influence onto the estimate for the "honest yes" response and much weaker influence for the other estimates.

Tab. 3: Unweighted and weighted estimates of using over the counter medication for performance enhancement in sport in Denmark

DK	unweighted estimates			weighted estimates		
	estim.	low/bca	high/bca	estim.	low/bca	high/bca
honest yes	0.2080	0.1595	0.2615	0.2394	0.1769	0.3128
inc	0.3432	0.2668	0.4196	0.4164	0.3234	0.5095
honest no	0.4488	0.3587	0.5550	0.3441	0.2177	0.4619

Tab. 4: Unweighted and weighted estimates of using over the counter medication for performance enhancement in sport in Spain

ESP	unweighted estimates			weighted estimates		
	estim.	low/bca	high/bca	estim.	low/bca	high/bca
honest yes	0.0441	0.0000	0.1042	0.0422	0.0000	0.1208
inc	0.1738	0.0893	0.2839	0.1988	0.0832	0.3344
honest no	0.7821	0.6523	0.8908	0.7589	0.5899	0.8971

Arzheimer, Kai (2009). Gewichtungsvariation. In: Schoen, Harald; Rattinger, Hans; Gabriel, Oscar (Hrsg.), *Vom Interview zur Analyse*. (S. 361-388) Baden-Baden.

Elliot, D. (1991). *Weighting for non-response.* London: Office of Population Censuses and Surveys.

European Commission, B. (2019). *Eurobarometer 88.4 (2017)*: GESIS Datenarchiv, Köln. ZA6939 Datenfile Version 2.0.0, https://doi.org/10.4232/1.13288.

Häder, Sabine; Gabler, Siegfried (1997): Deviations from the Population and Optimal Weights, in: Saris, W. E., & Kaase, M. (Eds.), Eurobarometer: measurement instruments for opinions in Europe (ZUMA-Nachrichten Spezial, 2, p. 32-44). Mannheim: Zentrum für Umfragen, Methoden und Analysen -ZUMA-. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-49742-1