Swiss Salt Study 2: Explanation of Weighting Factors for the Research Dataset

(Version 1.0, 2024)

Calculation of sampling weights

Marginal totals for each sampling list were obtained from the Swiss Federal Office of Statistics, and an average sampling frame was calculated to integrate both samplings. Sampling weights were defined as the inverse of inclusion probability and were then adjusted for non-response. The predicted probability of response to urine collection was estimated using a logistic regression model that considered age group, sex, nationality, household size, marital status, and region as predictors. Subsequently, a k-means clustering algorithm was employed to group subjects with similar predicted response probabilities. The number of clusters was determined based on the ratio of the sum of squares between clusters to the total sum of squares. To correct for non-response, each participant's sampling weight was divided by the average probability of response within their cluster. Weights corrected for non-response were then calibrated to ensure that some characteristics of the sample of respondants match those in the population within different subgroups. The subgroups were defined according to strata (combinations of sex and center), age groups, nationality, household size, marital status and region. Additionally, weights were calibrated on the season of urine collection (in consideration of variations in nutritional habits across seasons) by assuming an equal repartition of the population across the 4 seasons. Finally, weight trimming was implemented based on the approach proposed by Potter et al. (1990). An appropriate value denoted as w0 was determined based on the 99^{th} quantile of the theoretical (Beta) distribution of the reciprocal of scaled weights. All weights exceeding w0 were set to w0, and the remaining weights were adjusted so as to force the sum of weights to equal the size of the sampling frame. This process was repeated iteratively until no further weight trimming was necessary. After applying calibrated weights, both trimmed and untrimmed, the sample characteristics of participants became similar to those of the population within different subgroups. Despite potentially introducing a small bias, trimmed weights were applied due to their slightly smaller variability and range compared to untrimmed weights, in order to obtain more precise estimation. The final point estimate (e.g. average sodium concentration) was obtained by calculating a weighted average of the outcome measured on the participants. A confidence intervals for the weighted mean was constructed using the with-replacement bootstrap procedure

described in Bessonneau et al. (2021). A total of 200 bootstraps was constructed in this way, with each boostrap providing calibrated and trimmed weights for the participants. The standard error of the weighted mean was determined by calculating the standard deviation of the weighted means obtained across the 200 bootstrap resamples.

Compared with the final report (Swiss Salt Study 2, second survey on salt consumption in Switzerland: Main results**), there are 2 slight modifications to the protocol for calculating the weights:**

1/ The same predictors were used in the non-response model, with the exception of the region, which was replaced by the centre. In this way, the model is a little closer to the definition of the strata (which combine sex and centre). On the other hand, when calibrating, the region was also taken into account (as before).

2/ For “weight trimming”, the statistician used a simpler approach which consists of not allowing weights greater than the median + 5 times the inter-quartile range.

Excel file with weights

It contains 4 tabs:

- w1_urine: sampling weights
- w2_urine: weights corrected for non-response
- w3_urine: calibrated weights
- w4_urine: calibrated and « trimmed » weights

In practice, only weights w3 or w4 should be used. Each line corresponds to the identifier of an invited person, but only weights >0 indicate participants. The « Estimate » column contains the weights to be used in the desired estimate. The « Bootstrap_ » columns contain the bootstrap weights used to calculate the confidence intervals.