TFS-Info-13 Information on correction weights

Version: 1 from 9 April 2024
Authors: Anke Macdonald

Introduction

Since March 2024, the correction weights have been available for application through the NAKO TransferHub.

Special gratitude is extended to Dr. Matthias Sand (Leibniz Institute for the Social Sciences GESIS, Department 'Survey Design and Methodology') for his expert guidance and active support in determining the correction weights.

Below, you will find some information on the development and application of correction weights. Contact person for questions on this topic is Stefan Rach (rach@leibniz-bips.de).

1. Weighting

The correction weights are provided as total weights, which account for both the influence of varying inclusion probabilities due to the sample design (design weights) and the impact of non-response as well as the distribution of specific characteristics in the population (calibration weights).

The correction weights may **only be used with the data from the baseline examination**; separate correction weights may have to be determined for later examination waves and/or the GEFU surveys.

You will find the total weights in the NAKO data dictionary under Basic data->Correction weights->Total weights. (Please note: Currently still under Basic data->Demography -> For the NAKO dataset... -> Correction weights -> Total weights. Will be corrected soon.)

As a rule, you will apply for the variable pairs

- "Total weight normalised" (wgt_total_18sc) and "Geographical centre allocation normalised" (wgt_geoSC_18sc)

and/or

- "Total weight normalised" (wgt_total_16sc) and "Geographical centre allocation normalised" (wgt_geoSC_16sc)

(For the difference between **_18sc** and **_16sc** see section Standardisation)

If only design weights are required for certain analyses, these can also be provided if this is justified accordingly when applying for the data. You will find the design weights in the NAKO data dictionary under Basic data->Correction weights->Quality Assurance.

2. Representativeness

The weighting was not based on the population distribution of Germany as a whole but rather on the **population distribution within the study regions covered by the NAKO study centres**. Consequently, any statements regarding the representativeness of weighted analysis results (if they can or should be made at all) are strictly limited to these specific study regions.

3. Standardisation / renormalisation

The total weights were standardised to the size of the sub-samples of the study centres, i.e. for each study centre the sum of the total weights equals exactly the number of cases from this study centre contained in the data set. It should be noted here that the **geographical** study centre affiliation according to the address **at the time of invitation** (or time of sampling) is taken into account. For a small number of cases (centre changers/movers), this differs from the study centre in which the participants were ultimately examined. The geographical study centre affiliation is provided in the variables wgt_geoSC_18sc and wgt_geoSC_16sc.

The correction weights are provided in two standardisations:

- Standardisation to 18 study centres: The total weights are contained in the variable wgt_total_18sc; the assignment to the study centre is contained in the variable wgt_geoSC_18sc.

- Standardisation by merging the three Berlin study centres (51, 52, 53) into one study region (50): The total weights are contained in the variable wgt_total_16sc; the allocation to the study centre is contained in the variable wgt_geoSC_16sc.

(The differences between the variables wgt_total_18sc and wgt_total_16sc are limited to the 3 Berlin study centres; the remaining values are identical.)

4. Renormalisation of the weights before use

The size of the NAKO sample is constantly decreasing due to incoming contradictions. At the time the correction weights were created, there were 204,754 data sets in the sample. Data users who received data at an earlier date could therefore have cases in their datasets for which no correction weights are available. These would then have to be excluded for analyses using the correction weights.

Data users whose data is provided at a later date may receive fewer than these 204,754 data sets. Furthermore, cases are also regularly excluded as part of the analyses, so that the number of cases is also reduced here.

If the analysis dataset does not include all 204,754 cases with correction weights for the reasons mentioned above, data users are advised to perform a renormalisation before using the weights. This process ensures that the sum of the total weights for each study centre precisely matches the number of cases from this study centre included in the dataset. Renormalisation

is performed by dividing the individual weights for each study centre by the sum of the weights and multiplying by the corresponding number of cases.

5. Subsequent application for correction weights

It is possible for all use and access projects to apply for the correction weights retrospectively, even if the complete data transfer has already taken place.

*If only the **total weights** are required, an informal request by e-mail to transferstelle@nako.de with the following wording is sufficient:*

Within the framework of the NAKO-xxx use and access project and after the data has already been transferred, I hereby request the total weights consisting of the variable pairs

- "Total weight normalised" (wgt_total_18sc) and "Geographical centre allocation normalised" (wgt_geoSC_18sc)

and/or

- "Total weight normalised" (wgt_total_16sc) and "Geographical centre allocation normalised" (wgt_geoSC_16sc)

The variables are then uploaded by data management with the same pseudonymisation as the original variable set in the use and access application.

*If the **design weights** are required, an informal application should also be sent by e-mail with the following wording to transferstelle@nako.de:*

I hereby apply for the design weights as part of the NAKO-xxx use and access project after the data has already been transferred.

Justification: xxx

The UAC checks the justification and, if it recommends approval, the variables are uploaded by the data management with the same pseudonymisation as the original variable set for the use and access application.