WEIGHTS - General Target Variable Report (GVR)

1. General Information

The target variable T_WEIGHT_L2U provides weighting factors for design weights, post-stratification weights, or their combination at a country (L2) level. To identify which type of weight the factors in T_WEIGHT_L2U correspond to, it is necessary to use the methodological indicator QR_WGHTTYPE_SVY (see Table 1.1. and Section 3.4.1).

In SDR2, by country level we mean *de iure* or *de facto* independent countries, even if originally providers of the source data constructed the sample and the weighting factors at a different (L1) level, which, compared to L2, may be lower (e.g. East or West Germany constituting Germany at L2) or higher (e.g. former Czechoslovakia, which is split into Czech Republic and Slovakia at L2). In other words, L1 corresponds to the source sample level and L2 is the harmonized national level (for details, see SDR2 MASTER File Overview document). Since some characteristics of the source level weights may differ for subsamples of a national survey, in this GVR we report counts on the source, L1, sample level.

The aim of sample weighting is to correct distortions between a survey sample and the population it was drawn from (i.e., target population). Distortions can come from (i) a specific sample design, or (ii) the fact that not all sample units were reachable at the time of data collection. Design weights and post-stratification weights, respectively, seek to correct for these kinds of distortions. Both were in the scope of our harmonization and are provided in the MASTER file.

First, *design weights* aim to correct deviations between a target population and a sample whose units had an unequal probability to be included in that sample due to the sampling design used. This is done by equalizing probabilities of inclusion. This includes correcting for the unequal probability of selection in household/address samples, when the sample unit consists of households within which an individual is selected with a probability always depending on the number of household members (and for this reason unequal).

The second type of weights available in SDR2 are *post-stratification weights*. The need for this type of weighting comes from the fact that some sample units do not take part – for any reason – in a survey. Without post-stratification weighting, the realized sample would systematically under- or overrepresent persons with certain socio-demographic characteristics in the target population.

In some cases, weights provided by data producers combine both design and poststratification corrections. We call them *combined weights*. Whenever more than one source weighting factor is provided (either separate – design or poststratification – or combined weights) in the harmonized data file the combined weights are used.

Weights which aim is not correction but adaptation to the population of interest, like *population size weights*, or ones moving the level of analysis to the sub- or supra-populations, like in the case of continents or other grouping factors, are in general not used in the harmonized data file, even if produced by authors and/or provided in original data files at cross-country level (for exceptions see Section 3.2).

Table 1.1 Description of the Target Variables on **T_WEIGHT_L2U**

	Variable description	Variable name	Variable values
Target variable	Weighting factor	T_WEIGHT_L2U	Continuous variable consisting of values > 0.
Source variables			See T_WEIGHT_L2U_CWT_SDR2.xlsx
Quality control variables for data records	Type of source weight	QR_WGHTTYPE_SVY	0 = No weight 1 = Design weight 2 = Post-stratification weight 3 = Combined (design and post-stratification) weight 4 = No information 5 = Constant weight
	Status of source weight variable	QR_WGHTDESC_SVY	0 = Source variable not available, target value set to 1 3 = Source variable not available, target value set to a proportional coefficient 10 = Source value is treated as the target 15 = Source value used after rescaling for sample size 20 = Sample shares a single source value (1) used as the target 25 = Sample shares a single source value (1) used as the target after rescaling 30 = Sample shares a single value, but using a proportional coefficient 35 = Sample shares a single value, but using a proportional coefficient, rescaled 40 = Multiple source values transformed by a proportional coefficient 45 = Multiple source values transformed by a proportional coefficient, rescaled
	Sample size change when source weights applied	QR_WGHTDISTORT_SVY	Interval variable ranging from -978 to 2746

2. Survey Projects (Source Data)

The target variable T_WEIGHT_L2U is provided for all 3474 source samples. However, only 2022 of these provided information about the type of the weighting factor (design, poststratification or combined), no information was available for 337 samples, while for the

remaining 1111 samples, the source weighing factor was either constant for all cases in a sample, or was missing and therefore equal 1 for all cases.

3. General Rules and Procedures

3.1. Source data description

In general, weight values of T_WEIGHT_L2U are directly derived from the source values (just rounded to 5 decimal places) with some exceptions described below.

3.2. Rules of transformation of source variables into target variable

1. When no source weighting variable was provided, we set the value of T_WEIGHT_L2U to 1. This situation occurs in AMB/2006 for Colombia, ARB/1 (all samples), ASES (all samples), CB/2008-2013, 2015 (all samples), CDCEE/1 (all samples) CDCEE/2 for Belarus, Czech Republic, Hungary, Romania, Slovenia, Slovakia, Ukraine, CNEP/3 for Mexico, Mozambique, EB/0.1 (all samples) EB/0.2 (all samples), EB/3 (all samples), EB/5 (all samples), ISSP/1986 for Hungary, ISSP/1992, 1994, 1997 for Sweden, PA1 (all samples), PA2 (all samples), PPE7N for Japan, Nigeria, United States of America.
2. In the case of United Kingdom surveys in EB/3 and EB/5, L1 samples for Great Britain and Northern Ireland shared the same (value =1) constant weighting factor. We calculated the target L2 weights to reflect the differences in the sizes of the two populations. Hence, values on T_WEIGHT_L2U differ for Great Britain and N Ireland samples stemming from in EB/3 and EB/5.
3. In some cases, data providers calculated source weights for standardized sample sizes (of 1000 or 1500 persons). T_WEIGHT_L2U provides recalculated values that reflect the true sample sizes. This situation occurs in: AMB/2008, 2010 (all samples), LITS/2 (all samples), and NEB/1-7 (all samples).
4. When a source weight variable was provided at the sub-country level, we transformed it by a multiplier, to reflect the share of the sub-country population in the country's population. This is the case for United Kingdom (in EB/13, 15, EVS/1-4, ISSP/1989-1991, 1993, 1994, 1998-2002, 2008), Germany (in ISSP/1990-2011 and WVS/3), Belgium (in ISSP/2011), and Israel (in ISSP/1996, 1997, 2001, 2003, 2006-2008, 2010, 2013-2015).
5. We applied a reversed procedure to the one described above (3.2.4) when the source weights were provided at the higher-then-country aggregation level. This occurred in the former Czechoslovakia in ISSP/1992 (where we divided the source sample into the L2 samples corresponding to the Czech Republic and Slovakia), the former Yugoslavia in NEB/5 (where we split the source sample into the L2 samples corresponding to Serbia and Montenegro), and Cyprus (treated as a whole island) in WVS/5-6 (where we divided the original sample into two separate samples of Republic of Cyprus and Turkish Cypriot Community). We scaled all these weights to the country-level sample totals.

3.4. Methodological variables that accompany T_WEIGHT_L2U

In PLUG_SURVEY we provide three quality control variables for source data records.

1. QR_WGHTTYPE_SVY provides information on the type of source weighting factors. Design weights were applied to 254 L1 samples, post-stratification weights – to 1411 samples, and combined weights – to 358 samples. The authors of 339 samples did not provide information on the type of weighting, and 995 samples provide a constant weight, which means that the data were practically not weighted. For the remaining 117 samples we could not identify a weight variable, therefore we used the value 1 as the target weighting factor.
2. QR_WGHTDESC_SVY indicates the basic characteristic of source weight variables and the type of operation, if we performed one, to achieve the L2 weighting factor. For the vast majority of source samples (3095) no operation was required: weights provided in the harmonized data file replicate those from the source data files (2253 samples, QR_WGHTDESC_SVY = 10) or they replicate a single source value, 1 (842 samples, QR_WGHTDESC_SVY = 20). Other situations include:
    - 138 samples in which we rescaled source weights to the sample size (QR_WGHTDESC_SVY = 15);
    - 72 samples where all respondents in the source sample share a single value, which we multiplied by a proportional coefficient to reflect a country level at L2(QR_WGHTDESC_SVY = 30);
    - 32 surveys where multiple source weighting values were provided at L1, which we multiplied by a proportional coefficient to reflect a country level at L2 (QR_WGHTDESC_SVY = 40);
    - 12 samples with a single source weight value 1, which we rescaled to reflect the sample size (QR_WGHTDESC_SVY = 25);
    - 6 samples where multiple source weighting values were provided at L1, which we first multiplied by a proportional coefficient to reflect a country level at L2, and then rescaled (QR_WGHTDESC_SVY = 45);
    - 2 samples where a single source weighting value was provided at L1, which we multiplied by a proportional coefficient to reflect a country level at L2 and then rescaled (QR_WGHTDESC_SVY = 35).

    For the remaining source samples, we could not identify any source weight variable. Hence, we either used the value 1 at L2 (113 samples, QR_WGHTDESC_SVY = 0) or a proportional coefficient, to reflect a country level (4 samples, QR_WGHTDESC_SVY = 3).
3. QR_WGHTDISTORT_SVY indicates the level of distortion, either deflation or inflation, of the sample size after applying source weights to the data. In general, weights should be rescaled before being used, so that they do not change the original sample sizes. Using unscaled weights may artificially increase or decrease the standard errors, and therefore, by influencing the confidence intervals, the statistical analysis might be trapped by Type I or Type II errors. Well-scaled weighting variables guarantee that the sample size remains the same before and after applying weights. QR_WGHTDISTORT_SVY provides information about the number of cases by which the sample will be increased or reduced when the unscaled (original) source weights

are applied. Usually the source sample remains the same (3149 samples). However in 135 surveys, the sample will be inflated and in 190 – deflated.

Note: Users are encouraged to check how the weights are treated in the statistical software they use. While STATA automatically rescales the weights before applying them to analysis, SPSS does not.