Annex 5 Weighting procedure

1. Introduction

The size of the SILC sample is limited in comparison to the population to be analysed. To compensate for these imperfections, an adequate weighting has to be used during the analyses of survey data. In fact, this weighting takes into account the probability of a household being included in the sample and therefore also the fact that we are dealing with a stratified sample. Although the samples are representative of the population living in Switzerland, the same cannot be said for the people who actually respond to the survey. Certain types of households and persons are more willing to respond than others. If data were not weighted, the results would therefore be biased towards the characteristics of the households who are more willing to respond.

The extent of variability in sampling can be estimated by variance, the coefficient of variation or by the confidence intervals of an estimator. The analytic procedures of statistical software are based on the assumption that the sample was the result of simple random sampling. This is not the case for the SILC survey (proportional stratified sample). Furthermore, these procedures take into account neither the corrections made for non-response nor the marginal calibration, giving them the tendency to under-estimate variance and thereby also underestimating confidence intervals. For this reason weights must be used in order to correctly estimate confidence intervals during analysis.

The probability of responding cannot be separated from the characteristics of households or individuals. There is also a sizeable risk that this probability is not independent of the variables of interest (income, risk of poverty). The whole point of weightings, therefore, is to determine the characteristics of persons/households that do not respond and to increase the weights of persons/households who most resemble them. The same approach is taken by reducing the weights of households/persons most like the profiles that are most willing to respond. The last stage, called calibration, enables the sample's weight to be balanced so that the totals of the socio-economic characteristics of the weighted sample coincide as far as possible with the known totals of these characteristics in Switzerland's permanent resident population.

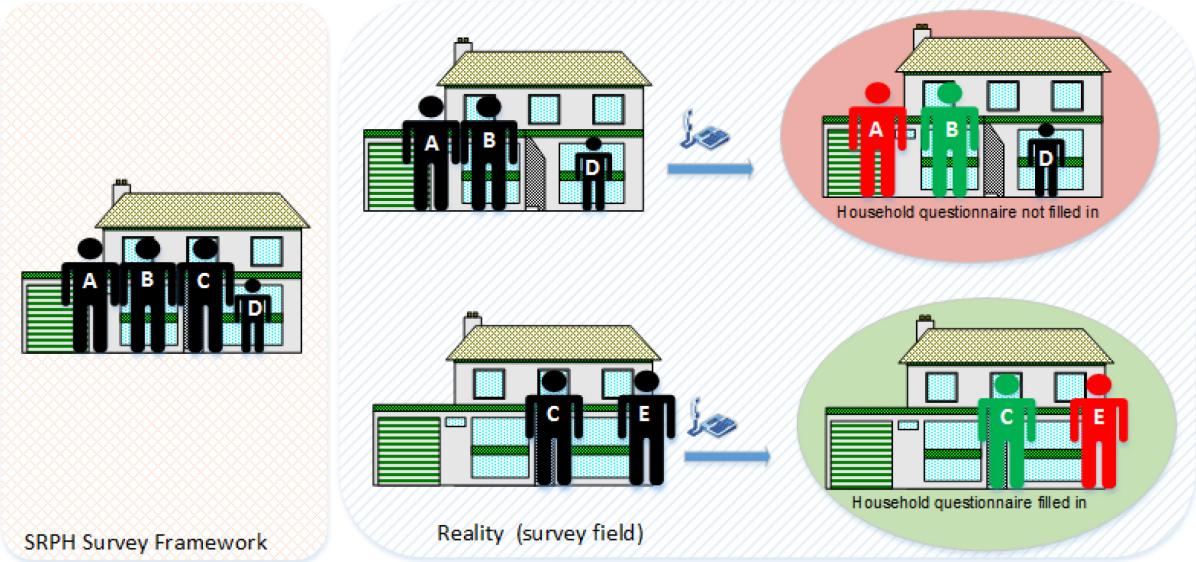
2. Cross-sectional weightings

2.1. Correction for non-response

As the SRPH survey framework is register-based, there may be a slight discrepancy between the structure of households at the time the sample is drawn for wave 1 and the real situation at the moment the survey institute calls households (see diagram 1). This can cause problems, for example if a household member drawn on the basis of SRPH has recently moved and at the time of the first telephone contact is living in a new household (person C in the example below). As the sampling units are persons living in households that have been drawn from the sample, both "field" households have to be followed and their members are all considered as being longitudinal. In this example, if person C lives with a new person (E), not included in the initial sample because they were not present in the household ABCD that was drawn, this person is called "cohabitant of wave 1". This person is treated in the weighting process as longitudinal.

This sampling is followed by a phase of non-response after the initial telephone contact. After this stage, households that are complete in their "field" composition are considered as responding households. A household is complete when it has replied to the grid questionnaire, to the household questionnaire (house on a green background below) and at least one (longitudinal) member has replied to the individual questionnaire (green person). In the above example, household CE is complete but household ABD is not because it did not respond to the household questionnaire (house on red background).

Diagram 1 Illustration of non-response in survey-framework and field households



One problem is that we are departing from the classic schema of two-phase direct sampling because the composition of the households observed in the field does not necessarily correspond to that of the framework. It is difficult in such cases to establish a clear-cut response status for certain drawn households, as in the example above. Some individuals can also belong, in the field, to a responding household although they are not a member of one of the households selected in the framework (person E). Such individuals are selected indirectly.

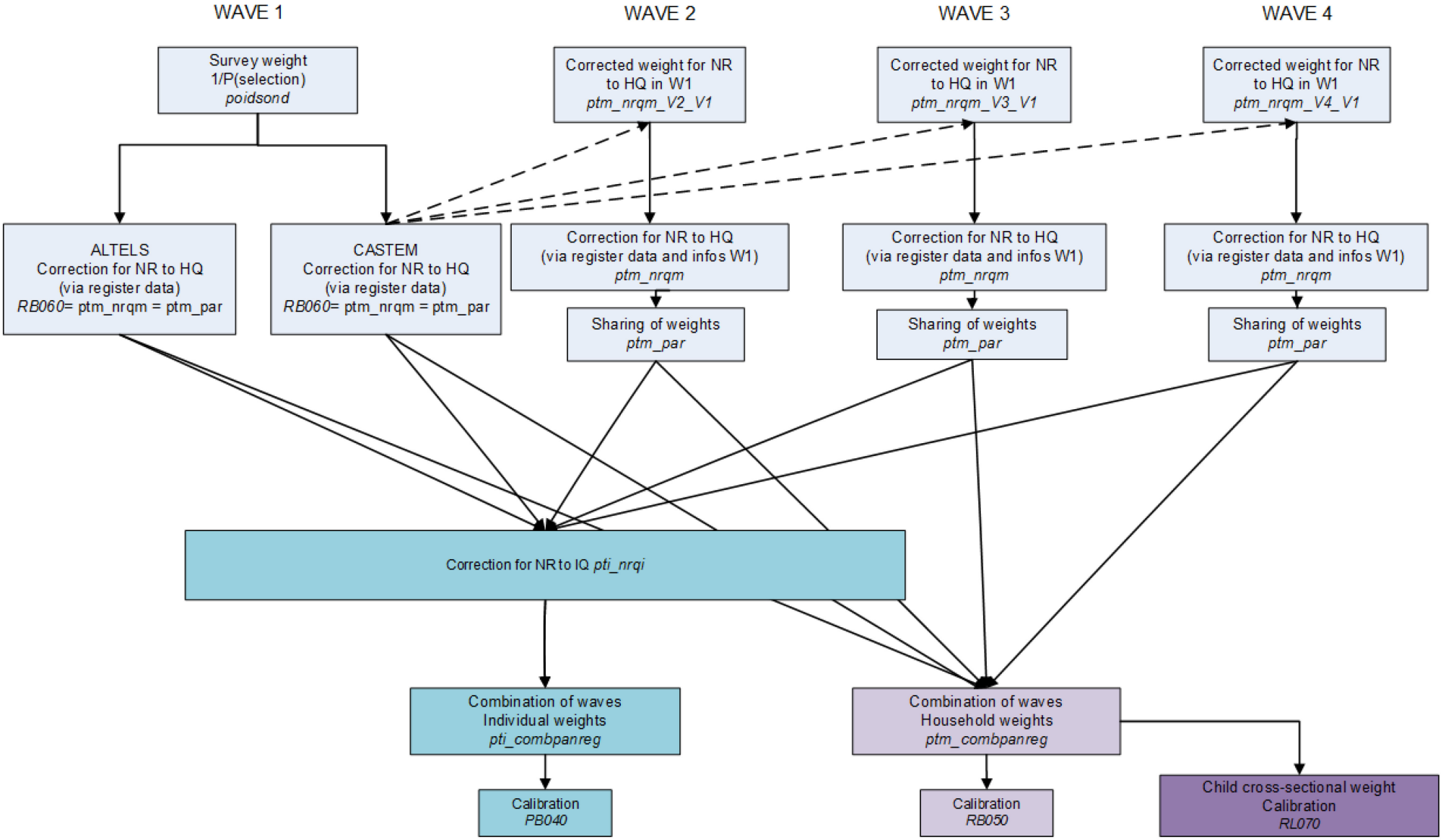
Although the notion of household may vary between the framework and the field, this is not the case for individuals. For individuals who belong to one of the households selected in the framework, it is possible to establish an unequivocal response status (respondent or not). This prompted us to formalise the selection procedure into a direct, two-phase survey of individuals belonging to one of the households in the framework, through which (indirect) access could be obtained to households in their “field” composition.

Initially the weighting process will correct non-responses to the household questionnaire. A model is created to establish whether or not an individual belongs to a complete household. This does not take into account the fact that an individual may have replied or not to the individual questionnaire (this would be done in the correction for individual non-response). Modelling at individual level enables the use of individual and household variables, leading to more precise results than could be obtained with modelling at household level. We should point out that although the response status is defined at individual level, we ensure that when estimating response probabilities, individuals belonging to the same “field” household have the same estimated response probability (i.e. the same ptm_nrqm weight), since all individuals from the same “field” household always have the same response status to the household questionnaire.

The correction process for household non-response is different for the sample of wave 1 than for waves 2, 3 and 4 . The initial weight or the survey weight from wave 1 has to be corrected for non-response in the current wave, but does not require weight sharing with cohabitants (2.1.1.2). For waves 2 to 4, the survey weight has already been corrected for non-response in wave 1 but has to be corrected for non-response in the current wave, then shared with the new cohabitants. The weighting process is shown in diagram 2 below.

Subsequently, the probability of response to the individual questionnaire will have to be modelled on the basis of this household weight for persons living in complete households.

Diagram 2. Schema of weighting process IQ = individual questionnaire, HQ = household questionnaire, NR non-response



2.1.1. Correction process for household non-response in wave 1

The solution chosen here is to calibrate the net sample (respondents) to the drawn gross sample (SRPH) and then to inverse the g-weights (ratio between the calibrated weight and the survey weight) as an estimate of the probabilities of response. So that the probabilities of response are the same for individuals in the same "field" household, the calibration is included at the level "field" household. The survey weight is the opposite of the probability of an individual being selected, which is given to all individuals in the sampled households.

To do this, we start by establishing explanatory auxiliary variables for non-response. As non-response is highly dependent upon whether a household has a known phone number or not, the sample was divided into two and different variables were retained for ALTELS (no known phone number) and CASTEMS (with known phone number). More details on the differences between ALTELS and CASTEM can be found under 6.3.3 *Appendix Minimizing non-response error* or in the 2014 Quality Report.

For reasons of stability in the weighting process, the variables selected in 2014 are used to correct for non-response, both individual and household. These are presented in table 1. Selection process is detailed in Appendix Weighting, ESQRS 2014 or in the 2014 Quality Report.

Once the variables were selected the integrated calibration was carried out. The probability of responding was the opposite of the g-weight, which has to be greater than 1 for the response rate to be less than 1. The calibration was therefore, conducted using the logit method, enabling limits to the g-weight to be selected in order to avoid this problem. For wave 1, the limits used for response probabilities were 0.1 and 0.95.

CASTEM
When choosing the explanatory variables by segmentation, we decided to freeze the first two branches of the segmentation tree and to keep a fixed intersection of two variables (Married family with child(ren)* and Family with child(ren)* composition of household by nationality). The calibration converged during 7 repetitions with a concordance rate of 71.4%.

ALTEL
The calibration converged during 5 repetitions with a concordance rate of 62.1%.

The concordance rate can assess the model by measuring the degree of similarity between the ranked values observed and those predicted by the model. It enables assessment of the model by indicating the percentage of cases in which the response status it predicted is the same as the actual response status. The concordance rate was calculated using the SAS procedure "Rank Correlation of Observed Responses and Predicted Probabilities ".

Table 1. Variables retained to correct for non-response. Information used to fill in the variable is indicated in Source (CATI, SRPH, CCO[1] or GWS[2] registers). Household is abbreviated HH.

					Household			Ind.
			Source	Reference period	W1		W234	
Variable	Codes used				castem	altel		
Age group of the oldest person in hh	1 : Up to age 34 2 : Age 35 - 44 3 : Age 45 - 54	4: Age 55 - 64 5 : Age 65 - 74 6 : Age 75 and more	SRPH	09.T-4 09.T-3 09.T-2 09.T-1	x		x	x
Age groups	1 : Age 0-15 2 : Age 16-20 3 : Age 21-34 4 : Age 35-44	5 : age 45-54 6 : age 55-64 7 : age 65-74 8 : age 75 and more	SRPH	09.T-1				x
Age groups	1 : Age 0-15 2 : Age 16-20 3 : Age 21-34 4 : Age 35-44	5: age 45-54 6 : age 55-64 7 : age 65-74 8 : age 75 and more	SRPH	09.T-4 09.T-3 09.T-2 09.T-1	x		x	
Civil status	1 : Single 2 : Married	3: Widowed 4 : Divorced	SRPH	09.T-4 09.T-3 09.T-2 09.T-1	x		x	x
Nationality group (2 groups)	1 : Switzerland / Northern and Western Europe	2 : Southern Europe / Other countries	SRPH	09.T-1		x		
Nationality group (4 groups)	1 : Switzerland 2 : Northern and Western Europe	3 : Southern Europe 4 : Other countries	SRPH	09.T-4 09.T-3 09.T-2 09.T-1	x		x	x
hh composition by nationality	0 : Only Swiss 1 : Foreign and Swiss mixed	2 : Only foreigners but at least one neighbouring country 3 : Only foreigners but no-one from neighbouring country	SRPH	09.T-4 09.T-3 09.T-2 09.T-1	x		x	x
hh composition by sex	0 : Male and female 1 : Only male	2 : Only female	SRPH	09.T-4 09.T-3 09.T-2 09.T-1		x	x	x
Type of family	110 : Single person aged under 65 130 : Single person aged over 65 210 : 2 adults aged under 65, without other hh members 230 : 2 adults aged under 65, without other hh members	300 : Single adult with at least one child (previously 400) 400 : 2 adults with one child or more (<18) or 18-24 (min 20y. age diff. with adult) (previously 300) 910 : Other households (previously 500 & 900)	SRPH	09.T-4 09.T-3 09.T-2 09.T-1	x		x	x
Type of family and number of children	401 : Family hh with 1 child (previously 301) 402 : Family hh with 2 children (previously 302)	403 : Family hh with 3 children or more (previously 303) 910 : Other hh (previously 900)	SRPH	09.T-4 09.T-3 09.T-2 09.T-1			x	x
hh size according to SRPH	1 : 1 person hh 2 : 2 person hh 3 : 3 person hh	4 : 4 person hh 5 : 5-person hh and more	SRPH	09.T-1	x			
Maximum h level of education	0 : Other 1 : Professional education and training		CATI	T-3 T-2 T-1 T			x	x
Maximum hh level of education	0 : Other 1 : Advanced professional education and training		CATI	T-3 T-2 T-1 T			x	x
Maximum hh level of education	0 : Other 1 : No training		CATI	T-3 T-2 T-1 T			x	x
Interested in politics	1: 0-6 with (0= no interest / 10 = maximum interest)	2: 7-10 with (0= no interest / 10 = maximum interest)	CATI	T-3 T-2 T-1 T			x	x
Employment status in 4 groups	1 : full-time paid work (min. 37 hours/week) / unpaid work in family business/farm 2 : part-time paid work (1 - 36 hours/week) / work in sheltered workshop 3 : In education / domestic tasks (max. age 64/65) / other pensioner / unemployed / other situation (continuing education, unpaid leave, ...) / Military or civil compulsory service / Apprentice 4 : retired (pensioner) AVS/AHV, pension fund, early retirment) / child not at school		CATI	T-3 T-2 T-1 T			x	x
Presence of supplementary benefits in hh	0 : No supplementary benefits	1 : One or more in hh	CCO	09.T-1	x			
Presence of unemployment allowances in hh	0 : No unemployment allowance	1 : One or more unemployment allowances	CCO	09.T-1	x		x	x
Presence of disability pensions in hh	0 : No AI/IA pension	1 : One or more AI/IA pensions	CCO	09.T-1	x		x	x
Number of old-age pensions in the hh	0 : No old-age pension 1 : One old-age pension	2 : Several old-age pensions	CCO	09.T-1	x			
Number of incomes from employment in hh	0 : No income from employment 1 : 1 income from employment 2 : 2 incomes from employment	3 : 3 incomes from employment 4 : 4 or more incomes from employment	CCO	09.T-1	x		x	x
Group of total equivalent incomes CCO	1 : <=P50	2 : >P50	CCO	09.T-1		x		
Group of total equivalent incomes CCO	1 : [P0-P20] 2 :]P20 - P40] 3 :]P40-P60]	4 :]P60-P80] 5 :]P80-P100]	CCO	09.T-1	x		x	x
Material deprivation 3 out of 9 items	0: No deprivation 1: Deprivation		CATI	T-3 T-2 T-1 T			x	x
At risk of poverty status at 60% of median total equivalent hh income CCO	0: Not at risk of poverty 1: At risk of poverty		CATI	T-3 T-2 T-1			x	
At risk of poverty status at 60% of median total equivalent hh income CCO	0: Not at risk of poverty 1: At risk of poverty		CCO	09.T-1				x
Major region	1 : Lake Geneva Region VD, VS, GE 2 : Espace Mitteland BE, FR, SO, NE, JU 3 : North-West Switzerland, BS, BL, AG 4 : Zurich ZH	5 : Eastern Switzerland, GL, SH, AR, AI, SG, GR, TG 6 : Central Switzerland, LU, UR, SZ, OW, NW, ZG 7 : Ticino TI	SRPH	09.T-4 09.T-3 09.T-2 09.T-1			x	x
8 categories of commune typology	1 : Centres 2 : Suburban municipalities 3 : High income municipalities 4 : Semi-urban municipalities	5 : Tourist municipalities 6 : Industrial and tertiary municipalities 7 : Rural commuter municipalities 8 : Mixed rural municipalities or Agricultural municipalities	SRPH	09.T-4 09.T-3 09.T-2 09.T-1	x		x	x
Size of hh's commune	1 : 100 000 inhabitants and more 2 : 50 000 - 99 999 inhabitants 3 : 20 000 - 49 999 inhabitants 4 : 10 000 - 19 999 inhabitants	5 : 5 000 - 9 999 inhabitants 6 : 2 000 - 4 999 inhabitants 7 : 1 000 - 1 999 inhabitants 8 : < 1'000 inhabitants	SRPH	T-3 T-2 T-1 T			x	x
Moved house (change of building) in past 2 years	0: No move	1: Change of building	GWS	09.T-1 12.T-1	x		x	x
Rent and accommodation costs	1: 0-1000 2 : 1000 - 1500	3: 1500 and more	CATI	T				x
Living space per hh member	1: <= 20m2/pers	2: > 20m2/pers	GWS	09.T-4 09.T-3 09.T-2 09.T-1	x		x	x

[1] Central Compensation Office register. See 2.4 *Statistical concepts and definitions* of the main part.
[2] GWS : Buildings and Dwellings statistic

2.1.2. Correction process for household non-response in waves 2, 3 and 4

The principle is the same as for correction for non-response in wave 1, except that the initial population is persons drawn from the CASTEM framework and whose household was complete in wave 1. The initial weights are those that their household had after correction for non-response to the household questionnaire in wave 1 (see diagram 2).

The auxiliary variables retained for correcting non-response between wave 1 and the current wave are not the same as those retained to correct non-response in wave 1, because the two response mechanisms are different, especially due to the fact that the balance between refusal and unreachable is very different in wave 1 to that in waves 2, 3 and 4. Firstly, from this point on we have more auxiliary information since the information collected in wave 1 can be used in addition to the information gathered from the registers. The set of variables under consideration as auxiliary variables is thus composed of variables from registers that were already available for the correction of non-response in wave 1 (with the same groups of levels), as well as information collected in wave 1 such as interest in politics, level of education or the at-risk-of-poverty-status. Secondly, there is no longer a distinction between altel / non-altel for the correction of non-response between wave 1 and the current wave, since we consider that as persons have already been contacted in wave 1 (i.e. there is no altel) that they have accepted to participate.

The same auxiliary variables are retained for the correction of non-response in waves 2,3 and 4. These were selected in 2014 (see Table 1.).

The reference population and consequently the relevant information from registers vary depending on the wave that we are addressing. For example

- Wave 2 we use the SPRH of September T-2
- Wave 3 we use the SPRH of September T-3
- Wave 4 we use the SPRH of September T-4

with T the year of the survey (DB010). The sum of each of these weights before correction for non-response is therefore representative of various targeted reference populations. In wave 3 for example we begin with the weight ptm_nrqmW3_W1 (calculated in T-2) and we use the SRPH of September T-3 (date sample was drawn in w3).

Modelling for non-response (calibration) then took place wave by wave, but with the same variables. The limits used were 0.3 and 0.999999999 for respondents. For non-respondents, the lower limit was 0.25 for wave 2 (no upper limit). In fact, the limit for wave 2 had to be lowered so that the calibration could converge with the result. It should be noted that the estimate of response probabilities for non-respondents is only useful for calculating the rate of matches/non-matches which can be used as diagnostic. Modifying the limits for non-respondents had no influence on the response probabilities of respondents.

The following results were obtained:

- Wave 2: 5 repetitions, 69.48% concordance (see definition p. 4)
- Wave 3: 7 repetitions, 68.78% concordance
- Wave 4: 6 repetitions, 71.66% concordance

Weight sharing:

For wave 1, weight sharing is not necessary as the cohabitants of wave 1 inherited the weight of their household during the correction for non-response.

The weight ptm_nrqm of waves 2 to 4 obtained by modelling (calibration) above were given to all household longitudinals. The weight sharing stage allows a part of these weights to be attributed to cohabitants who joined longitudinal households since the first wave. The weight after sharing is the shared transversal household weight: ptm_par.

The shared household weight is the same as the sum of the weight of household longitudinals divided by the number of longitudinals and initially present cohabitants in the household. In our case, we considered that all cohabitants were initially present in the survey base.

$$\text{ptm_par} = \frac{\sum_{i=1}^{L} \text{ptm_nrqm}_i}{L + P}$$

With i: individual
L: number of longitudinals in the household
P: number of initially present cohabitants in the household

2.1.3. Correction for non-response to the individual questionnaire

As mentioned above, individual weight is calculated from household weight. Another correction for non-response with regard to the individual questionnaire this time, has to be carried out first.

To correct for non-response to the individual questionnaire, the initial weight is the household weight adjusted for complete /incomplete non-response and shared between cohabitants, ptm_par. It is positive for both longitudinals and cohabitants who responded to the individual questionnaire and who belong to a complete household. This weight represents the populations from wave 1 from each of the panels before being merged. Their sum total among all individuals from the current wave therefore represents roughly 4 times the size of the population.

The correction for non-response to the individual questionnaire is carried out in the same way as that for the household questionnaire, but the entire process is done for all waves at the same time.

The variables (table 1) were tested in 2014 and then selected by segmentation and by logistic regression. These are used for the correction for individual non-response in subsequent years in order to guarantee continuity in the process. The calibration is then conducted on the basis of the variables selected.

2.2. Combination of waves

As the household weights were corrected for non-response to the household questionnaire and shared between the cohabitants of waves 2 to 4, the next stage, at household level, will be to merge the waves. The process of wave merging is the same as in previous years. The weights adjusted for non-response to both the individual and household level, are combined for all waves by major geographic region.

The weights used for merging waves were all calculated during wave 1 of each of the panels. They therefore represent the population in wave 1 of each of the panels and their sum corresponded therefore to the number of individuals in Switzerland in wave 1. By grouping together the 4 waves and therefore the 4 corresponding sets of weights, the weights' total is close to 32.2 million, i.e. four times the Swiss population. The adjustment factors enable us to reduce this number, by major region, to approximately 8.1 million individuals. The value of these factors is calculated using the method developed by Merkouris (see « Estimation transversale dans le cas des enquêtes auprès des ménages à panels multiples »). This approximate amount will be adjusted later in the final calibration to correspond to the exact total of the Swiss population.

For individual weights, adjustment is done simply by counting the number of individuals per major region and per wave and by looking at the percentage this number represents in relation to the total number of individuals in the major region concerned for all waves. This percentage (allocation factor) defined by wave and by region is then multiplied by the weight adjusted for the non-response (ptm_par) of each individual in the wave and region concerned. The number of individuals concerned in total and per wave is presented in table 2.

Table 2. Allocation factors (AF) when merging waves

Major regon	n	W1		W2		W3		W4	
		n	AF	n	AF	n	AF	n	AF
Lake Geneva region	3567	1175	0.33	902	0.25	816	0.23	674	0.19
Espace Mittelland	4510	1434	0.32	1212	0.27	988	0.22	876	0.19
Northwest Switzerland	2506	805	0.32	670	0.27	546	0.22	485	0.19
Zurich	3310	1094	0.33	887	0.27	727	0.22	602	0.18
Eastern Switzerland	2546	827	0.32	631	0.25	617	0.24	471	0.18
Central Switzerland	1919	572	0.30	529	0.28	465	0.24	353	0.18
Ticino	871	288	0.33	190	0.22	238	0.27	155	0.18

For households, the merging of four waves is done according to the same method, and by replacing individuals by households.

2.3. Final calibration

The aim of this calibration is to adjust the weights so that the totals according to certain calibration variables are identical in the sample to those of the population of reference (December 2021).

The individuals retained in the calibration framework are selected from the framework of 31.12.2021. The calibration framework contains all individuals (even those belonging to the non-permanent population) from households that could potentially be selected in the sample (at least one person in the permanent resident population in a private household at their main place of residence).

The target population is all persons who belong, in the field, on the day of the survey, to one of the households from the selection framework. Therefore, children born after 31.12.2021, as well as foreigners who arrived after this date, or non-permanent residents, belong to the target population if they belong to a framework household.

This definition enables us to ensure consistency between the household composition in the field and at the time of interviews and income recorded in the field during the survey. It does however lead to technical problems during the calibration, as certain responding individuals are not linked with the calibration framework meaning that some of the information necessary for the calibration is not available for them.

When this is the case, the auxiliary calibration variables are imputed to such persons and the weights are then calibrated as if all respondents were actually a sample from the calibration framework. The calibration is integrated so that all persons from the same field household have the same final weight RB050. By calibrating in this way to the calibration framework we can ensure that the total of the final weights is the same as that of the 31.12.2021 framework and that it is consistent from one year to the next.

2.3.1. Selection of calibration variables

Variable selection was made in 2014. The list of variables retained for the household and individual calibration can be found in table 3. Reference period is December T-1 (2021). Further explanations are available in 2014 Quality Report .

Table 3. Variables used for the household and individual calibrations. Information used to fill in the variable is indicated in Source (SRPH or CCO[3] registers).

Variable	Codes used		Source	Household	Individual
At risk of poverty status at 60% of median total equivalent household income	0 : Not at risk of poverty 1 : At risk of poverty		CCO	X	X
Intersection between ARP60 and age group for calibration	0 : if ARP60 = 0 1 : if ARP60 = 1 and age group = 1 2 : if ARP60 = 1 and age group = 2	3 : if ARP60 = 1 and age group = 3 4 : if ARP60 = 1 and age group = 4 5 : if ARP60 = 1 and age group = 5	CCO	X	
Age groups	1: age 0-17 2 : age 18-24 3 : age 25-49	4 : age 50-64 5 : age 65 and more	SRPH	X	X
Intersection between S20 and median total equivalent household	Continue			X	
Type of family	110 : Single person aged under 65 130 : Single person aged over 65 210 : Couples aged over 65, no other household members 230 : Couples aged under 65, no other household	300 : Single parent with at least one child (previously 400) 400 : Family with at least one child (previously 300) 910 : Other type of household (previously 500 & 900)	SRPH	X	X
Major region	1: Lake Geneva Region VD, VS, GE 2 : Espace Mitteland BE, FR, SO, NE, JU 3 : North-West Switzerland, BS, BL, AG 4 : Zurich ZH	5: Eastern Switzerland, GL, SH, AR, AI, SG, GR, TG 6 : Central Switzerland, LU, UR, SZ, OW, NW, ZG 7 : Ticino TI	SRPH	X	X
Household size according to SRPH	1 : 1 person household 2 : 2 person household 3 : 3 person household	4 : 4 person household 5 : 5-person household and more	SRPH	X	X
Linearised with Gini framework index	Continue		CCO	X	
Civil status	1: Single 2: Married	3 : Widowed 4 : Divorced	SRPH	X	X
Nationality group	1: Switzerland 2 : Northern and western Europe	3: Southern Europe 4: Other countries	SRPH	X	X
Intersection between S80 and total equivalent household income	Continue			X	X
Indicative that the total equivalent household income by cco source is < P10	0 : no 1 : yes		CCO	X	X
Indicative that the total equivalent household income by cco source is < P50	0 : no 1 : yes		CCO	X	X
Indicative that the total equivalent household income by cco source is < P20	0 : no 1 : yes		CCO	X	X
Indicative that the total equivalent household income by cco source is < P80	0 : no 1 : yes		CCO	X	X
Sex	1 : man 2 : women		SRPH	X	X

In order to establish the relevance of the variables chosen, comparisons were made of the variance obtained for the principle SILC indicators between a complete model using the entire set of variables available, and the proposed selection. The results obtained allowed us to approve the selection of calibration variables. As is the case for the correction for non-response, the variables selected in 2014 are used in the calibration process over the following years.

2.4. Cross-sectional weighting for children

The weight for children is calculated so that specific variables such as childcare in structures in or outside of school can be weighted.

It is calculated on the basis of the household weight. Correction for total individual non-response to the child proxy is not necessary as the non-response rate is zero. Children born after 31.12.2021 must be excluded from the weightings. Children aged 13 to 16 are not processed.

Eurostat recommends starting with variable RB050 and calibrating it on the total of individuals for each age between 0 and 12 inclusive (age at the time of interview). Nevertheless, weights adjusted for non-response to the household questionnaire, ptm_combpan_reg, have already been calibrated to the age groups by means of the integrated calibration to obtain RB050, which Eurostat does not recommend for the household weight. A second calibration to the age groups alone does not appear appropriate. Therefore the sex * age are used for the calibration, using the raking ratio method of CALMAR2, as previously.

[3] Central Compensation Office register. See 2.4. Statistical concepts and definitions of the main part.

Table 4. Frequency of children by age and by sex

	Sample								Sampling frame			
	unweighted				weighted with ptm_combpan_reg							
Age	**Boys**		**Girls**		**Boys**		**Girls**		**Boys**		**Girls**	
0	64	2.6%	65	2.7%	22 459	2.0%	29 406	2.7%	44 295	3.9%	42 188	3.7%
1	80	3.3%	90	3.7%	46 721	4.2%	41 433	3.8%	44 101	3.9%	41 503	3.6%
2	84	3.5%	84	3.5%	38 912	3.5%	41 193	3.7%	44 327	3.9%	42 325	3.7%
3	91	3.7%	88	3.6%	42 204	3.8%	41 246	3.7%	45 425	4.0%	43 095	3.8%
4	85	3.5%	87	3.6%	47 486	4.3%	41 739	3.8%	45 381	4.0%	43 069	3.8%
5	87	3.6%	92	3.8%	47 146	4.3%	39 221	3.6%	45 980	4.0%	43 942	3.8%
6	103	4.2%	96	3.9%	52 499	4.8%	37 169	3.4%	46 053	4.0%	43 259	3.8%
7	95	3.9%	91	3.7%	44 094	4.0%	44 295	4.0%	45 801	4.0%	43 248	3.8%
8	91	3.7%	117	4.8%	34 111	3.1%	55 595	5.0%	44 866	3.9%	42 376	3.7%
9	102	4.2%	100	4.1%	39 936	3.6%	42 113	3.8%	45 444	4.0%	42 626	3.7%
10	117	4.8%	92	3.8%	58 767	5.3%	36 098	3.3%	44 766	3.9%	42 322	3.7%
11	102	4.2%	96	3.9%	40 846	3.7%	40 981	3.7%	45 039	3.9%	43 308	3.8%
12	124	5.1%	109	4.5%	52 292	4.7%	46 009	4.2%	44 669	3.9%	42 262	3.7%
Total	**1 225**	50.4%	**1 207**	49.6%	**567 472**	51.4%	**536 499**	48.6%	**586 147**	51.3%	**555 523**	48.7%
	2 432				**1 103 972**				**1 141 670**			

3. Longitudinal weightings

Since 2014, the sample has been drawn in the FSO's SRPH new sampling frame. This change led to a revision of the cross-sectional weighting on the SILC14 data (see 2014 Quality Report). The calculation method for the longitudinal weightings enabling an analysis of the transitions over two, three or four years was not affected by this revision. Data are only composed of four panels drawn from the SRPH as of SILC2017 onwards, i.e. the panels 2014 to 2017.

It has thus been possible to revise the longitudinal weighting model while integrating new data from administration registers as well as other methodological improvements. This revision was applied as of the SILC17 longitudinal data (2014 to 2017). These changes imply a break in series between the longitudinal results prior to and following the revision, as was the case between the cross-sectional results SILC13 and SILC14. More details on the change between the previous and the new weighting method are explained in the SILC18 quality report.

3.1. Construction of the weightings RB062, RB063 and RB064

Regardless of weight, only persons living in a complete household from start to finish of the transition will be awarded a longitudinal weight. The main steps are the following:

3.1.1. Correction for non-response

The aim is to correct for non-response in incomplete households at the end of transition (NRT, died or moved out of scope) although the household was complete at the start of transition. Non-response is modelled by a calibration algorithm (CALMAR2[4]), with a constraint at household level in order to obtain the same non-response correction weightings for the members of a single household. The respondents at the start and end of transition (net sample, including cohabitants) are calibrated to the gross sample (respondents at the start of transition regardless of the response status at the end of transition). A second step aims to model the response status of the gross sample and to calculate a concordance rate assessing the quality of the model. The adjustments provide g-weights (inverse of the response probability). This g-weight is multiplied by the cross-sectional weights ptm_par (cross-sectional weight of the start of the transition after NR correction and weight sharing, see 2.1 of each wave and each year concerned. Since the ptm_par weight has already being shared to include the cohabitants, and the correction is made at household level and all the individuals it contains, a new sharing of weights is not necessary.

Rotational groups are dealt with separately for each of the various weights.

The auxiliary variables are the same as those selected for the NR correction of waves 2-3-4 at cross-sectional level, plus the following 3 variables (see table 5 below):

- Household composition by sex of the SRPH;
- Presence of supplementary benefits of the Central Compensation Office (CCO);
- Household size according to SRPH.

The reference periods are detailed in table 5 and are also very similar to those of the cross-sectional weighting. The main difference concerns the SRPH data, taken from September T-x for the cross-sectional NR correction and December T-x for the longitudinal NR correction. This difference is explained by the fact that the wave 1 survey framework dates from September T-1.

Before using auxiliary variables in the calibration model for NR correction, missing values have to be treated because the CALMAR2 macro rejects them. Part of the missing values in the SRPH variables are imputed in a deterministic way using the CATI variables. The rest is imputed randomly.

The results of the models for NR correction are presented below for the various weights.

[4]The SAS macro CALMAR2 (CALMage sur MARges) is used to adjust a sample by reweighting individuals using auxiliary information available on a number of variables, called calibration variables.

RB062

Wave 1: The algorithm converged during 6 repetitions with a concordance rate of 69.34.
Wave 2: The algorithm converged during 6 repetitions with a concordance rate of 73.09
Wave 3: The algorithm converged during 7 repetitions with a concordance rate of 74.82

RB063

Wave 1: The algorithm converged during 5 repetitions with a concordance rate of 69.58
Wave 2: The algorithm converged during 6 repetitions with a concordance rate of 73.23

RB064

Wave 1: The algorithm converged during 6 repetitions with a concordance rate of 71.64

The concordance rate can assess the model by measuring the degree of similarity between the ranked values observed and those predicted by the model. It enables assessment of the model by indicating the percentage of cases in which the response status it predicted is the same as the actual response status.

Table 5 List of variables used to correct for non-response. Information used to fill in the variable is indicated in Source. Household is abbreviated HH.

Variable	Codelist	Source	Reference period T = 2021
Age group of the oldest person in hh	1 : Up to age 34 2 : age 35 - 44 3 : age 45 - 54 4 : age 55 - 64 5 : age 65 - 74 6 : 75 and more	SRPH	12.T-4 12.T-3 12.T-2
At risk of poverty status at 60% of median total equivalent hh income	0: Not at risk of poverty 1: At risk of poverty	CATI	T-3 T-2 T-1
hh composition by nationality	0 : Only Swiss 1 : Foreign and Swiss mixed 2 : Only foreigners but at least one neighbouring country 3 : Only foreigners but no-one from neighbouring country	SRPH	12.T-4 12.T-3 12.T-2
Presence of supplementary benefits in household	0 : No 1 : At least one in household	CdC	12.T-1
Age group	1 : age 0-15 2 : age 16-20 3 : age 21-34 4 : age 35-44 5 : age 45-54 6 : age 55-64 7 : age 65-74 8 : age 75 and more	SRPH	09.T-1 12.T-1
Rent and accommodation costs	1 : 0-1000.-[2 : [1000.- - 1500.-] 3 :]1500.- and more	CATI	T-3 T-2 T-1
Material deprivation 3 out of 9 items	0: Not deprived 1: Deprived	CATI	T-3 T-2 T-1
Type of family	110 : Single person aged under 65 130 : Single person aged over 65 210 : 2 adults aged under 65, without other hh members 230 : 2 adults aged under 65, without other hh members 300 : Single adult with at least one child* 400 : 2 adults with one child* or more 910 : Other households *(<18) or 18-24 (min 20y. age diff. with adult)	SRPH	12.T-4 12.T-3 12.T-2
Type of family and number of children	401 : 2 adults with 1 child * 402 : 2 adults with 2 children* 403 : 2 adults with 3 children* or more 910 : Other households *(<18) or 18-24 (min 20y. age diff. with adult)	SRPH	12.T-4 12.T-3 12.T-2
Maximum hh level of education	1: Professional education and training 2: Advanced professional education and training 3: no training	CATI	T-3 T-2 T-1
hh composition by sex	0 : Male and female 1 : Only male 2: Only female	SRPH	12.T-4 12.T-3 12.T-2
8 categories of commune typology	1: Centres (CEN) 2 : Suburban municipalities (SUB) 3 : High income municipalities (RE) 4 : Semi-urban municipalities (PERI) 5 : Tourist municipalities (TOUR) 6: Industrial and tertiary municipalities (IND) 7 : Rural commuter municipalities (PEND) 8 : Mixed rural municipalities (MIX) and Agricultural municipalities (AGR)	CATI	T-3 T-2 T-1
Major region	1: Lake Geneva Region VD, VS, GE 2 : Espace Mitteland BE, FR, SO, NE, JU 3 : North-West Switzerland, BS, BL, AG 4 : Zurich ZH 5: Eastern Switzerland, GL, SH, AR, AI, SG, GR, TG 6 : Central Switzerland, LU, UR, SZ, OW, NW, ZG 7 : Ticino TI	CATI	T-3 T-2 T-1
hh size according to SRPH	1 : 1 person household 2 : 2 person household 3 : 3 person household 4 : 4 person household 5 : 5-person household and more	SRPH	12.T-4 12.T-3 12.T-2
Interested in politics	1: 0-6 with (0= no interest / 10 = maximum interest) 2: 7-10 with (0= no interest / 10 = maximum interest)	CATI	T-3 T-2 T-1
Civil status	1: Single 2: Married 3: Widowed 4 : Divorced -9 : No information	SRPH	12.T-4 12.T-3 12.T-2
Groupe de nationalité - 4 classes	1: Switzerland 2 : Northern and Western Europe 3 : Southern Europe 4 : Other countries	SRPH	12.T-4 12.T-3 12.T-2
Presence of unemployment allowances in hh	0 : No unemployment allowance 1 : One or more unemployment allowances	CdC	12.T-1
Presence of disability pensions in hh	0 : No AI/IA pension 1 : One or more AI/IA pensions	CdC	12.T-1
Number of old-age pensions in the hh	0 : No old-age pension 1 : One old-age pension 2 : Several old-age pension	CdC	12.T-1
Number of incomes from employment in hh	0: No income from employment 1: 1 income from employment 3: 3 incomes from employment 4: 4 or more incomes from employment	CdC	12.T-1
Employment status in 4 groups	1: occupa in (1,5) full-time paid work (min. 37 hours/week) / unpaid work in family business/farm 2 : occupa in (2, 3, 6) part-time paid work (1 - 36 hours/week) / work in sheltered workshop 3 : occupa in (4, 7, 9, 10, 11, 12, 13) In education / domestic tasks (max. age 64/65) / other pensioner / unemployed / other situation (continuing education, unpaid leave, ...) / Military or civil compulsory service / Apprentice	CATI	T-3 T-2 T-1
Group of total equivalent incomes	1 : [P0-P20] 2 :]P20 - P40] 3 :]P40-P60] 4 :]P60-P80] 5 :]P80-P100]	CdC	12.T-1
Size of hh's commune	1: 100 000 inhabitants and more 2 : 50 000 - 99 999 inhabitants 3 : 20 000 - 49 999 inhabitants 4 : 10 000 - 19 999 inhabitants 5 : 5 000 - 9 999 inhabitants 6 : 2 000 - 4 999 inhabitants 7 : 1 000 - 1 999 inhabitants 8 : < 1 000 inhabitants	CATI	T-3 T-2 T-1
Moved house (change of building) in past 2 years	0 : No move 1 : Change of building	GWS	09.T-1 12.T-1
Living space per hh member	1 : <= 20m2/pers 2 : > 20m2/pers	GWS	12.T-4 12.T-3 12.T-2

3.1.2. Combination of waves

The benefit of this was explained in 2.2. It is only applied for the longitudinal if several rotational groups are concerned, for either RB062 or RB063. Longitudinal waves are combined in the same way as cross-sectional waves.

3.1.3. Final calibration

The final calibration for longitudinal weights is carried out in the same way as for cross-sectional weights with slightly fewer variables (table 6) and slightly different reference periods[5]. This is an individual calibration, without household constraint, carried out on the initial population of the transition and making it possible to obtain different weights between individuals of the same household. The longitudinal weights thus represent the population at the start of the transition. They therefore allow analyses determining the evolution of the population from T-x to T.

Table 6 List of variables for the final calibration

Nom	Label	Codelist	Source	Reference period T = 2021
ARP60_cl1_CAL **ARP60_cl2_CAL** **ARP60_cl3_CAL** **ARP60_cl4_CAL** **ARP60_cl5_CAL**	Intersection between ARP60 and age group for calibration	0 : if ARP60 = 0 1 : if ARP60 = 1 and age group = 1 2 : if ARP60 = 1 and age group = 2 3 : if ARP60 = 1 and age group = 3 4 : if ARP60 = 1 and age group = 4 5 : if ARP60 = 1 and age group = 5	CdC/ SRPH CATI	12.T-4 12.T-3 12.T-2 T-3 T-2 T-1
cl_age_pondmen_CAL	Age groups for calibration	1 : age 0-17 2 : age 18-24 3 : age 25-49 4 : age 50-64 5 : age 65 and more	SRPH CATI	12.T-4 12.T-3 12.T-2 T-3 T-2 T-1
FamTyp19_CAL	Type of family for calibration	110 : Single person aged under 65 130 : Single person aged over 65 210 : 2 adults aged under 65, without other hh members 230 : 2 adults aged under 65, without other hh members 300 : Single adult with at least one child* 400 : 2 adults with one child* or more 910 : Other households *(<18) or 18-24 (min 20y. age diff. with adult)	SRPH CATI	12.T-4 12.T-3 12.T-2 T-3 T-2 T-1
HH_RES_REGCH_2011_2_CAL	Major region	1 : Lake Geneva Region VD, VS, GE 2 : Espace Mitteland BE, FR, SO, NE, JU 3 : North-West Switzerland, BS, BL, AG 4 : Zurich ZH 5 : Eastern Switzerland, GL, SH, AR, AI, SG, GR, TG 6 : Central Switzerland, LU, UR, SZ, OW, NW, ZG 7 : Ticino TI	SRPH CATI	12.T-4 12.T-3 12.T-2 T-3 T-2 T-1
HHsize_cl5_CAL	Household size according to SRPH for calibration	1: 1 person household 2 : 2 person household 3 : 3 person household 4: 4 person household 5 : 5-person household and more	SRPH CATI	12.T-4 12.T-3 12.T-2 T-3 T-2 T-1
MaritalSt5_CAL	Civil status for calibration	1: Single 2: Married 3: Widowed 4 : Divorced	SRPH CATI	12.T-4 12.T-3 12.T-2 T-3 T-2 T-1
Natio4_CAL	Nationality group (4 groups) for calibration	1 : Switzerland 2 : Northern and Western Europe 3 : Southern Europe 4 :Other countries (regroupement avec les < 0)	SRPH CATI	12.T-4 12.T-3 12.T-2 T-3 T-2 T-1
sex_CAL	Sex for calibration	1 : man 2 : woman	SRPH CATI	12.T-4 12.T-3 12.T-2 T-3 T-2 T-1

[5] The reference periods are indicated by the month of year T (DB010 survey year) minus 1, 2, or 3 depending on the weight concerned (RB062, RB063 or RB064 respectively).

3.2. Construction of the weight RB060

The RB060 weight, an intermediate weight not intended for analyses, is constructed on the basis of the longitudinal weights after correction for non-response (see 3.1.1). This weight does not exist for all individuals, but only those living in a complete household at the start and end of the transition. Newborns and cohabitants who arrived in the household after the start of the transition do not have one. Cohabitants do not need to have an RB060 unlike newborns. Part of the weight must therefore be allocated to them via a weight sharing process.
The initial weights - after correction for non-response - Plm_nr_RB062, Plm_nr_RB063 and Plm_nr_RB064, each of which were based on the gross sample of the arrival transition, were therefore taken over, prioritising the weight RB062, then RB063, and then RB064. The sum of the weights was indeed much greater in this order than in the reverse order. For wave 1, the weight is naturally the ptm_par (see 2.1.1).

if filter = 1 ==> RB060_avpart = ptm_par

Otherwise if a person has a weighting RB062 ==> RB060_avpart = plm_nr_RB062

Otherwise if a person has a weighting RB063 ==> RB060_avpart = plm_nr_RB063

Otherwise if a person has a weighting RB064 ==> RB060_avpart = plm_nr_RB064

According to Eurostat guidelines, the weight of the mother is attributed to newborns. If the mother is absent from the household, the weight of the oldest woman in the household or the oldest man is used, in this order.

3.3. Construction of the weighting DB095

This weight, which is not intended for analysis, should be assigned to complete households (DB135 = 1), which are not in w1. Unlike the other longitudinal weights, which exist only for the end of transition year (SILC22 here), the DB095 exists for each of the four years considered in which the household was complete.

The starting weight used is DB090. However, the sum of the DB095 weights per year must correspond to the total households in the population, which is not the case since, for example for SILC18 (w1), only one panel must have a DB095 weight, while 4 panels have a DB090 weight. A corrective factor must therefore be applied to DB090 so that the sum of the weights for each of the panels is equal to that of DB090, representing the population of private households.

It is then necessary to correct DB095_int, and divide it by the number of rotational groups comprising this weight, i.e. 1 for t-1, 2 for t-3 and 3 for t-1 and t, to obtain a final weight DB090.

4. Analysis of weight differentials in RB050

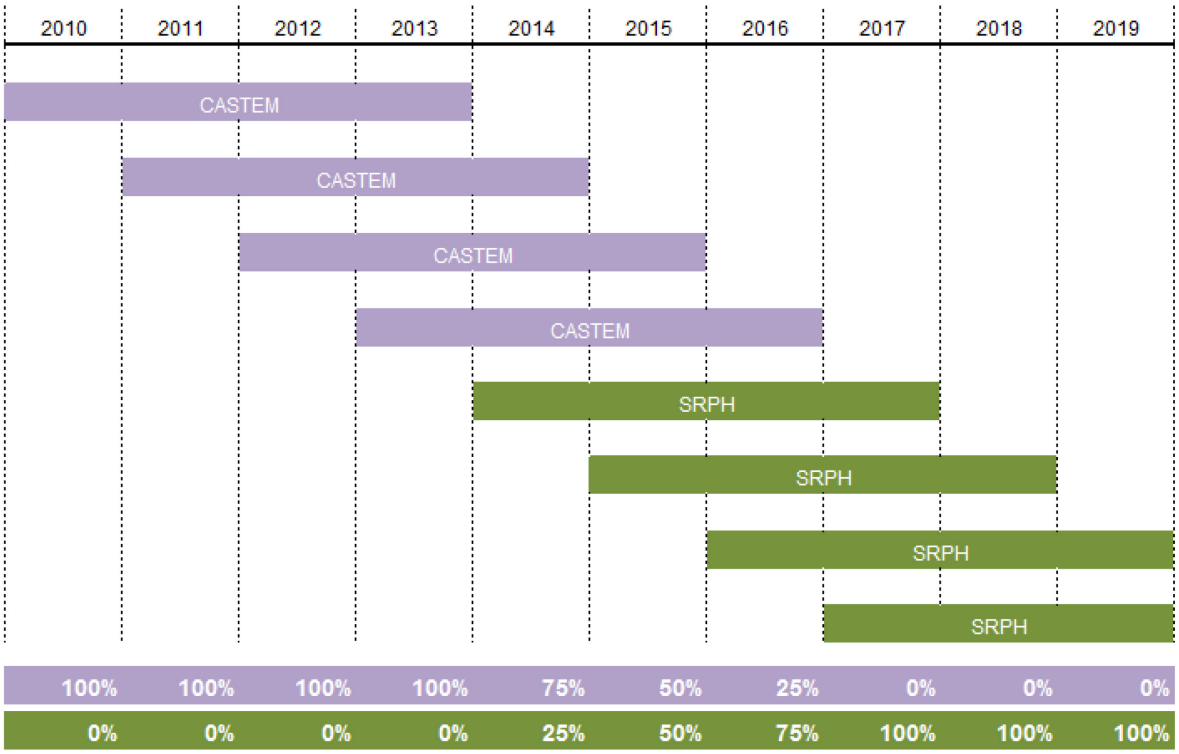
The final household weights RB050 were analysed over several years in order to test the continuity of the response process modelled by weighting, to detect any errors and to assess the impact of changing the weighting method.

The graphics below show the differential of the probability of responding modelled by weighting (1/(average weight of sub-group/total average weight)) for some socio-demographic categories. Values above 1 indicate that the sub-group's probability of responding, as modelled by the weights, is higher than average and that therefore this sub-group tends to respond better than average. In contrast, values below 1 indicate that the sub-group's probability of responding, as modelled by the weights, is lower than average and that therefore this sub-group tends to participate in the survey less well than average.

As well as identifying the sub-groups that have a greater/smaller probability of responding, these values also allow us to observe changes in the weights of sub-groups with the gradual introduction of ALTELS. In fact, the introduction of the ALTEL population increases the coverage of profiles of persons in this group (detailed below). We expected therefore, to see a decrease in the variations in weight, with a slight increase in the weight (decrease of the differential of response probability) of elderly people, who are very rarely ALTELS, and a decrease in the weight (increase of the differential of response probability, diagram 4) of profiles of the ALTEL type, since their representation among respondents was then similar to their proportion in the population. This transition phase is observed until SILC2017, and should then stabilise, as the entire sample is drawn from SRPH (see diagram 3 below).

As ALTELS are not in fact representative of the population, some analyses were conducted during the first sampling in SRPH in 2014 to establish their profile and to measure the impact of their introduction on the main indicators. Some of the results calculated in 2014 showed that the ALTELS represented a population that tended to be young (50% are younger than 44), who often lived alone or in a couple and was more likely to contain foreigners from Southern Europe and other countries (see 2014 Quality Report for more details).

Diagram 3 Share of both sampling frameworks represented in the total sample, by survey year.



Elderly persons are more willing to respond than working-age adults and 25-49 year-olds were the least well-represented. Until change in survey framework, the latter were also less well-represented in the sample as they were most often ALTELS. However, the weight of persons aged over 65 has regularly

declined since 2014 and then increased again since 2017, whereas the opposite was observed for younger people.

Diagram 4 Change in the differential of response probability by age, as modelled by the weighting method, with the introduction of ALTELS and changes to the weighting (from 2014)

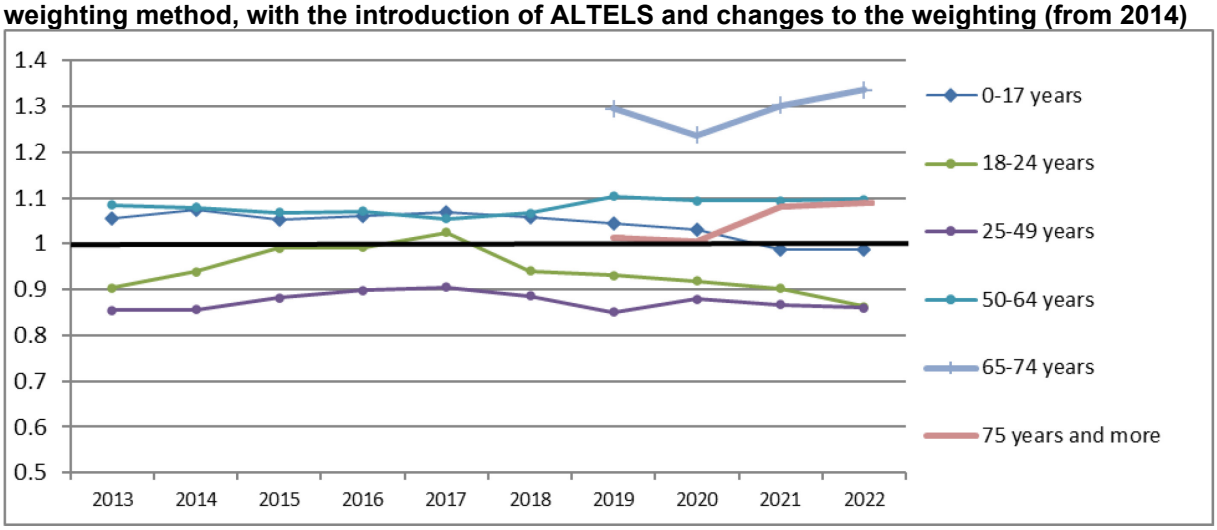


Diagram 5 Change in the differential of response probability by nationality type, as modelled by the weighting method, with the introduction of ALTELS and changes to the weighting (from 2014)

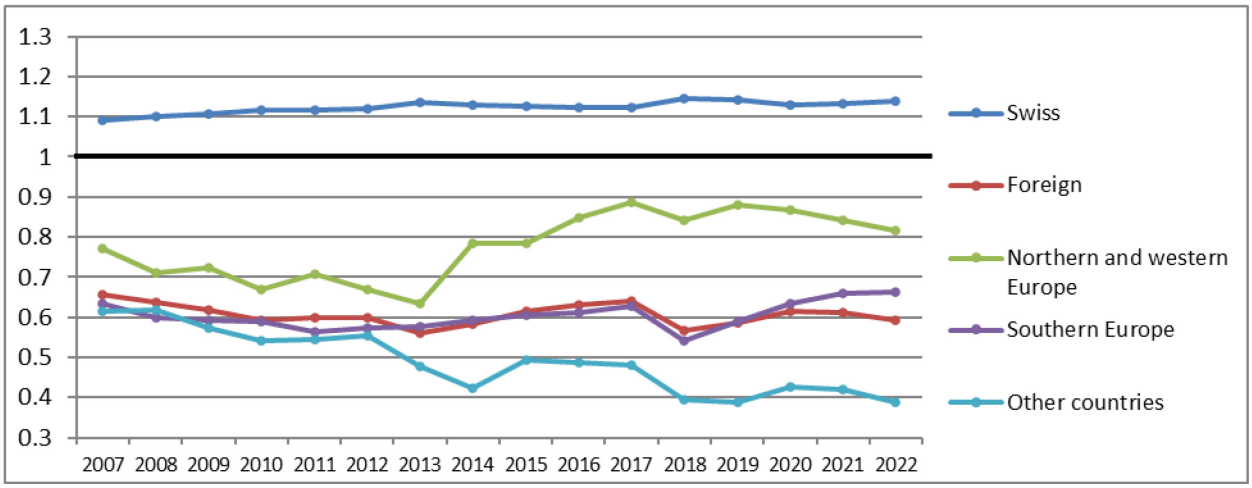
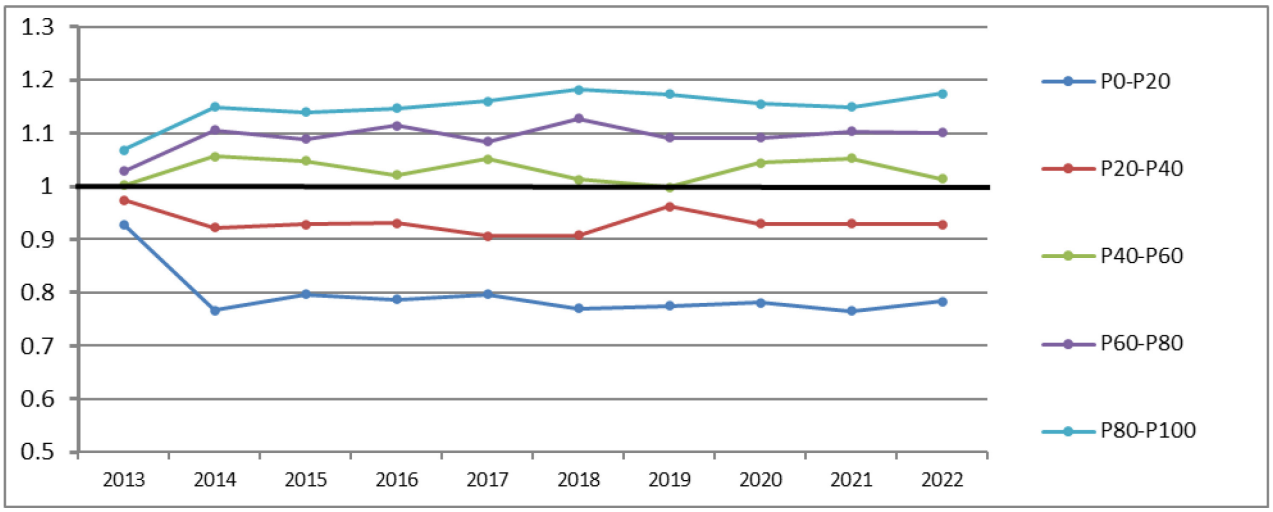


Diagram 5 shows that the response differential for Swiss nationals remains quite stable, whereas that of foreign nationals is much less so. The response probability of foreign nationals from Northern and Western Europe, modelled by weighting, increased relative to the average from 2014 to 2017 and has remained fairly stable since 2017. The response probability differential for foreign nationals from Southern Europe has remained fairly stable since the introduction of ALTELS. The probability of response from nationals of other countries (Eastern European and non-European countries), modelled by weighting, has fallen more sharply than the average since 2017. It should also be noted that until 2013, correction for non-response and the calibration were made without distinguishing between the sub-groups of foreign nationals. The availability of register data enabled us, from 2014 onward, to reuse these same categories of foreign nationals in the correction for non-response and the final calibration.

Whereas the change-over to SRPH and the new weighting reduced weight dispersion by age categories, the reverse effect can be seen by quintiles of equivalent disposable income (dia.6), although the order of weights in relation to the quintile remained the same. Since 2014, the response probability modelled by weights according to quintile have remained relatively stable over time. Thus, from 2014, the new

weighting method attributes a bigger weight to small incomes and a proportionately smaller weight to large incomes. In other words, the new weighting method models a smaller response probability for smaller incomes and a larger one for big incomes. This could be due to the introduction of income variables taken from registers in the weighting process. This means that financially better-off households are more likely to respond, with a gradual decrease by income quintile.

Diagram 6 Change in the differential of response probability by equivalent disposable income, as modelled by the weighting method with the introduction of ALTELS and changes to the weighting (from 2014)



5. Weight variance of RB050

Certain intermediate weights were analysed in relation to their final weights in order to establish the share of the variation coefficient of the final weight created at each stage of weighting. This entailed evaluating which proportion of the non-response process, as modelled by the weights, could be attributed to the stage of correction for non-response (ptm_nrqm) or to the stage of final calibration (RB050). With a proportional stratified sample, a low coefficient of variation of the survey weights is normal (initial weight for wave1). Still in wave 1, a large part of the total variance in the final weights is created at the stage of correction for non-response (ptm_nrqm), and to a lesser extent, during the final calibration (*RB050*). For subsequent waves, the initial variance is already substantial with weight having been gained after correction for non-response in wave 1 (*ptm_nrqmvX_w1*) and accounts for almost half of the total variance. The remainder is due to correction for non-response in the current wave and during the final calibration. For waves 2 to 4, the variability added by the final calibration is smaller than that added by correction for non-response. The latter usually decreases gradually from wave 2 to wave 4, when the non-response corrected weight *ptm_nrqm* increases.

It should be noted that not all stages are represented in the graphic below and that some percentages are missing from the total of waves 2 to 4, in particular those of weight sharing and the merging of waves.

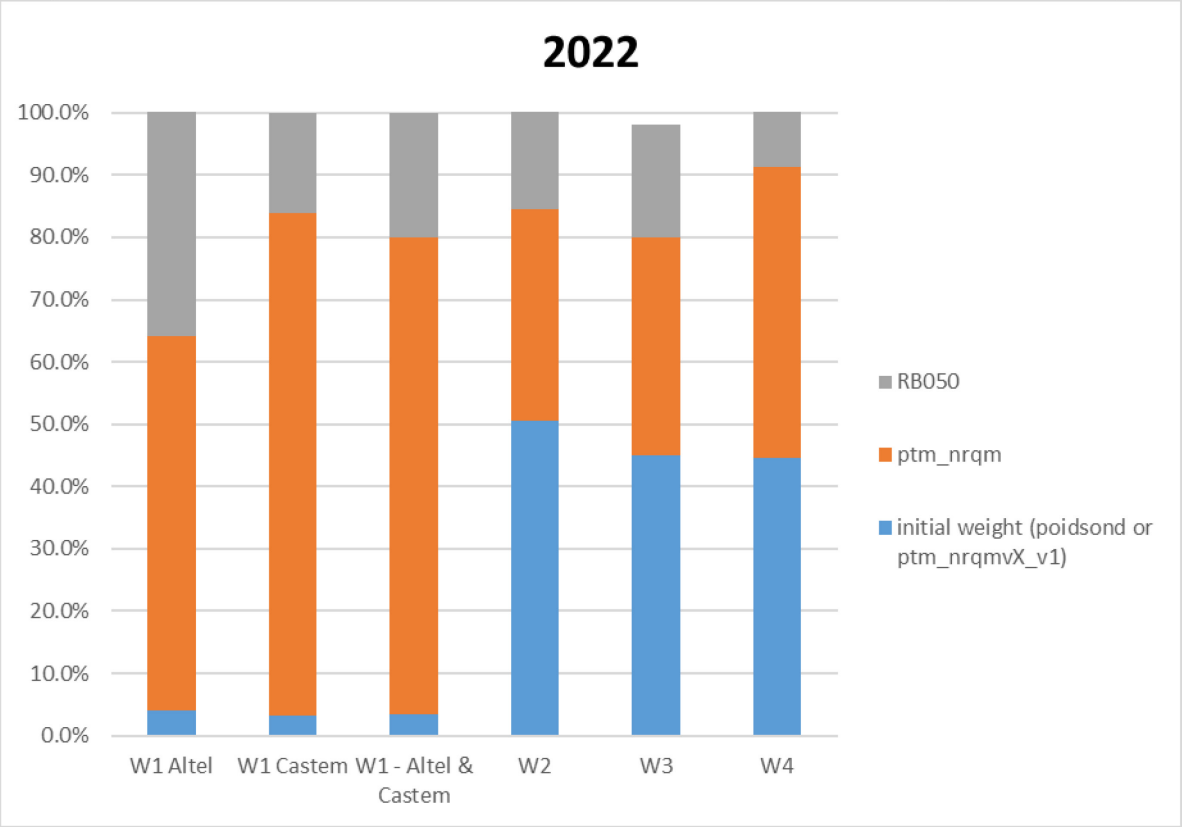



Diagram 8. Share of final coefficient of variation attributed to the main stages of the weighting