Palestinians' Psychological Conditions Survey 2022

Note on Weights Calculation

Diego Zardetto, Giulio Barcaroli

1. Preamble

In the first quarter of 2022, the Palestinian Central Bureau of Statistics (PCBS), with technical assistance from the World Bank, conducted a sample survey to assess the impact of the May 2021 hostilities with the Government of Israel on mental health, psychological well-being, employment status and other aspects of quality of life in the West Bank and Gaza. Data collection was carried out by PCBS through Computer-Assisted Personal Interviews (CAPI) from March 13th to April 12th, 2022.

This note provides a basic description of the procedural steps that have been undertaken to calculate final weights of the 2022 Palestinians' Psychological Conditions Survey (PPCS 2022). The interested reader is referred to (Haziza and Beaumont, 2017) for a general review about the construction of weights in probability surveys.

2. Background information on the PPCS sample

2.1 Panel structure and representativeness

The PPCS survey is a panel survey. Its sample was designed to be a randomly selected subsample of the respondent sample of the first round of the COVID-19 Rapid Assessment Phone Survey (RAPS 1), which was fielded by PCBS between June and August of 2020. Notably, the sample targeted by RAPS 1 was itself a continuation of the last panel wave of the Socio-Economic and Food Security Survey (SEFSEC), which took place in 2018. Moreover, for the individual module of the PPCS survey, PCBS tried to reinterview the same adult individual who was interviewed for the individual module of the SEFSEC 2018 survey. As a result, the sample of the 2022 PPCS survey can be regarded as a new (partial) wave of the SEFSEC panel, which started in 2013.

Panel information extending over such a long period of time is extremely valuable for impact evaluation and causal inference. At the same time, panel surveys can be considered fully representative of their target population only at round 1, whereas their representativeness decays as the panel ages, as a consequence of panel attrition and structural changes in the target population. Specific weighting techniques have been applied to counteract those effects in the case of the PPCS sample, thereby improving its ability to provide a reliable representation of 2022 West Bank and Gaza. In this respect, it is important to note that both panel ancestors in the lineage of the PPCS sample (namely SEFSEC and RAPS) were considered by PCBS representative at national-level, governorate-level, and rural/urban/camps-level at the time of the first wave.

2.2 Sample size and selection

The planned sample size for PPCS was set to 7,057 households, and the 8,709 respondent households of RAPS 1 served as a sampling frame for PPCS. The selection of PPCS households followed a one-stage cluster sampling design. More precisely, 641 Enumeration Areas (EA) were randomly selected from the 1,824 EAs of

RAPS 1 with Probability Proportional to Size (PPS). All the RAPS 1 households contained in the selected EAs were included in the PPCS sample. To meet the logistic needs of PCBS, variable X = 'number of RAPS 1 respondent households per EA' was used as measure of size (MOS) for the PPS algorithm. This fully preserved the probability nature of the PPCS household sample, but also made the inclusion of EAs containing fewer RAPS 1 households less likely, thus avoiding logistical challenges and associated high data collection costs. The choice of the MOS variable illustrated above clearly explains how a 35% sampling rate at EA-level (35% = 641 / 1,824) resulted in the desired 81% sampling rate at household-level (81% = 7,057 / 8,709).

The individual questionnaire of the PPCS survey was administered to one selected adult member (aged 18 years or above) of each respondent household. For each respondent PPCS household, interviewers attempted to identify and re-interview the same adult individual who responded to the individual module of SEFSEC 2018. Only if the attempt was unsuccessful, the interviewer used a Kish grid (which accompanied the questionnaire) to randomly select, with equal probability, one adult from among all adult members of the household.

2.3 Sample distribution and household-level response rates

Table 1 reports the composition of the planned and realized PPCS samples in terms of EAs and households by region (Gaza and West Bank) and governorate (16 governorates). The table also shows the attained EA-level completion rate (i.e. the percentage of planned EAs for which interviews were actually collected during fieldwork) and household-level response rate (i.e. the percentage of planned households that responded to the survey).

EA-level completion rate was 94% for West Bank and Gaza as a whole, with slight differences by region (93% Gaza, 94% West Bank). Fieldwork operations resulted in sizably lower than average completion rates for some governorates, notably Deir Al-Balah (90%) in Gaza, and Nablus (89%), Tulkarm (91%), and Jerusalem (92%) in West Bank.

Of 7,057 planned households, 917 did not respond, yielding an overall household nonresponse rate of 13%[1]. Household-level response rate was higher in Gaza (90%) than in West Bank (85%), resulting in 87% for West Bank and Gaza as a whole. In the Gaza region, lower than average response rates were recorded for North Gaza (88%) and Rafah (89%). In the West Bank region, response rates turned out to be significantly below the regional average for the governorates of Jerusalem (71%) and Ramallah & Al-Bireh (73%).

[1] Of those who did not respond, in 84% of cases the household was not found or had moved abroad, in 12% of cases the household refused to be interviewed again, and the remaining households only partially completed the interview.

Table 1: Distribution of the planned and realized PPCS sample by region and governorate. EA-level completion rates and household-level response rates are also shown.

		Enumeration Areas			Households		
Region	Governorate	Planned Sample	Realized Sample	Completion Rate (Realized / Planned)%	Planned Sample	Realized Sample	Response Rate (Realized / Planned)%
Gaza	Deir Al-Balah	39	35	90%	454	410	90%
Gaza	Gaza	88	81	92%	878	788	90%
Gaza	Khan Yunis	59	56	95%	678	618	91%
Gaza	North Gaza	57	55	96%	625	550	88%
Gaza	Rafah	41	38	93%	391	347	89%
Gaza		**284**	**265**	**93%**	**3,026**	**2,713**	**90%**
West Bank	Bethlehem	28	27	96%	323	293	91%
West Bank	Hebron	67	62	93%	843	750	89%
West Bank	Jenin	43	41	95%	449	414	92%
West Bank	Jericho & Al Aghwar	13	13	100%	161	134	83%
West Bank	Jerusalem	36	33	92%	321	229	71%
West Bank	Nablus	57	51	89%	595	504	85%
West Bank	Qalqiliya	15	15	100%	191	171	90%
West Bank	Ramallah & Al-Bireh	43	42	98%	464	339	73%
West Bank	Salfit	19	19	100%	179	146	82%
West Bank	Tubas & Northern Valleys	13	13	100%	192	170	89%
West Bank	Tulkarm	23	21	91%	313	277	88%
West Bank		**357**	**337**	**94%**	**4,031**	**3,427**	**85%**
West Bank & Gaza		**641**	**602**	**94%**	**7,057**	**6,140**	**87%**

Table 2 reports the composition of the planned and realized PPCS samples in terms of EAs and households by region (Gaza and West Bank) and locality type (rural, urban, camps), along with completion rates and response rates.

EA-level completion rates were lower in urban areas (93%) than in rural areas and camps (95% and 96%, respectively), with very small differences between the two regions.

Household-level response rates did not vary significantly by locality type for West Bank and Gaza as a whole but were consistently lower in West Bank than in Gaza for the same locality type (84% vs 89% for camps and 84% vs 90% in urban areas, respectively).

Table 2: Distribution of the planned and realized PPCS sample by region and locality type. EA-level completion rates and household-level response rates are also shown.

		Enumeration Areas			Households		
Region	Locality Type	Planned Sample	Realized Sample	Completion Rate (Realized / Planned)%	Planned Sample	Realized Sample	Response Rate (Realized / Planned)%
Gaza	Camps	45	43	96%	459	410	89%
Gaza	Urban	239	222	93%	2,567	2,303	90%
Gaza		**284**	**265**	**93%**	**3,026**	**2,713**	**90%**
West Bank	Camps	23	22	96%	285	239	84%
West Bank	Rural	84	80	95%	943	827	88%
West Bank	Urban	250	235	94%	2,803	2,361	84%
West Bank		**357**	**337**	**94%**	**4,031**	**3,427**	**85%**
WB&G	Camps	68	65	96%	744	649	87%
WB&G	Rural	84	80	95%	943	827	88%
WB&G	Urban	489	457	93%	5,370	4,664	87%
West Bank & Gaza		**641**	**602**	**94%**	**7,057**	**6,140**	**87%**

2.4 Data structure of the realized sample and individual-level response rate

After data collection and editing, three datasets were made available by PCBS for the calculation of weights of the PPCS sample.

- The first dataset, denoted here as S_{hh}, stores household-level information available for each household belonging to the planned PPCS sample, whether respondent or non-respondent. This dataset has 7,057 rows.
- The second dataset, denoted here as S_{roster}, stores individual-level information available for each person which was found in the residence of each respondent household of the PPCS sample. This dataset has 36,621 rows.
- The third dataset, denoted here as S_{ind18+}, stores individual-level information collected through the individual questionnaire of the PPCS survey, which was administered to one selected adult member (aged 18 years or above) of each respondent household. This dataset has 6,138 rows.

By linking datasets S_{hh} and S_{roster}, it was possible to build a binary variable 'RESP' identifying respondent (RESP = 1) and non-respondent (RESP = 0) households belonging to S_{hh}. The number of households with RESP = 1 was found to be 6,140, in accordance with the figures shown in Tables 1 and 2.

Out of the 36,621 roster individuals contained in dataset S_{roster}, only 32,276 (88% for the roster dataset) were considered eligible for interview. The remaining 4,345 individuals in S_{roster} were not considered eligible because they did not reside in the residence of the household (they had left the household since SEFSEC 2018 or were only temporarily present there at the time of the interview). Structural information (e.g. sex, age, …) was successfully collected by PCBS for all the 32,276 eligible roster individuals.

Finally, out of the 6,138 adult individuals contained in dataset S_{ind18+}, 5,877 responded, yielding an individual-level response rate of 96% for the adult individuals' sample.[2]

3. Main steps of the weights calculation procedure

The fundamental objectives of the weights calculation procedure were (i) mitigation of bias risks and (ii) improvement of estimation efficiency. This section lists the main procedural steps that led to the final PPCS weights. Subsequent sections will elaborate on each of these steps.

[S1] Derive initial weights for PPCS households and roster individuals (see Section 3.1).

[S2] Adjust the weights of PPCS households and roster individuals for household-level nonresponse (see Section 3.2).

[S3] Calibrate the weights obtained at step S2, using as calibration benchmarks suitable household-level and individual-level aggregates provided by PCBS. Note that this step generates *integrated* household-level and individual-level calibration weights (see Section 3.3).

[S4] Suitably trim the calibration weights obtained at step S3 (see Section 3.4).

[S5] Derive initial weights for the sample of adult individuals (one per household) who responded to the individual module of the questionnaire (see Section 3.5).

[S6] Calibrate the weights obtained at step S5, using as calibration benchmarks suitable household-level and individual-level aggregates provided by PCBS (see Section 3.6).

Step S6 ends the procedure. At that stage, the weights of all the sampling units encompassed by the PPCS respondent sample (households, roster individuals, and adult individuals interviewed through the individual module) are ready for dissemination and to be used for general purpose statistical analysis.

3.1 S1 – Calculation of initial weights of households and roster individuals

As illustrated in Section 2.2, the PPCS sample is a probability subsample selected from the RAPS 1 respondent sample through PPS selection of EAs. Therefore, the initial weights of PPCS households must be obtained multiplying the final household weights of RAPS 1 (provided by PCBS) by the reciprocals of the inclusion probabilities generated by the PPS sampling algorithm:

$$d_{ij} = w_{ij}^{\mathrm{RAPS1}} \times \frac{1}{\pi_i} \qquad (1)$$

In equation (1), d_{ij} denotes the initial weight of household *j* belonging to EA *i*, w_{ij}^{RAPS1} denotes the final RAPS1 weight of the same household, and π_i is the PPS inclusion probability of EA *i*. The inclusion

[2] Of those who did not complete the individual interview, the majority (70 percent) were not located and there was no suitable replacement, some partially started the interview but could not or did not want to complete it, and even fewer refused to take part from the outset.

probabilities under PPS sampling without replacement, using variable X as MOS, can be expressed symbolically[3] as follows:

$$\pi_i = m \times \frac{X_i}{\sum_{i \epsilon \mathrm{RAPS1}} X_i} \qquad (2)$$

where m is the PPCS sample size in terms of EAs, X_i is the MOS of EA *i*, and the sum at denominator extends to all the EAs of the sampling frame (i.e. the RAPS 1 respondent sample).

It is important to note that all the household members of each sampled household were included in the roster PPCS sample. Therefore, all roster individuals within any PPCS household share the same inclusion probability, which equals the inclusion probability of the household they belong to. The same holds true for the weights. As a result, equation (1) also gives the initial weight of each roster member *k* belonging to household *j* located in EA *i*, namely:

$$d_{kij} = d_{ij} = w_{ij}^{\mathrm{RAPS1}} \times \frac{1}{\pi_i} \quad \forall k \qquad (3)$$

In what follows, unnecessary subscripts will be dropped for notational convenience and the PPCS initial weight of respondent sample unit *k* (be it a household of S_{hh} or a roster individual of S_{roster}) will be simply denoted as d_k.

3.2 S2 – Adjustment of weights for household-level nonresponse

Total nonresponse occurs when a sampled unit, for whatever reason, either does not respond at all to a survey, or fails to provide enough information for its data to be usable in the estimation phase. Total nonresponse results in estimation efficiency loss and increased risks of bias. In an effort to mitigate the risk of bias, survey weights need to be adjusted for total nonresponse (Särndal and Lundstrom, 2005). To this end, response propensity modeling and calibration are commonly applied alternatives, the choice between the two being mainly driven by the available auxiliary information.

As shown in Section 2.3, the household-level nonresponse rate was non-negligible for the PPCS sample. Out of 7,057 planned households, 917 did not respond, yielding an overall nonresponse rate of 13%. Given the origin of the PPCS sample, i.e. its provenance from the RAPS and SEFSEC panels, rich information was available on both respondent and non-respondent households. Moreover, non-respondent households were enough to enable a response propensity modeling approach to nonresponse. This approach, often called the propensity score method (Haziza and Beaumont, 2017), entailed several steps.

- First, a logistic model was developed to estimate household-level response probabilities, using variable 'RESP' (introduced in Section 2.4) as dependent variable and suitable variables derived from the RAPS and/or SEFSEC surveys as predictors. Potential candidate variables to be used as predictors spanned different domains, e.g. territory, socio-demographics, housing, wellbeing, and consumptions. After

[3] In practice, formula (2) can result in $\pi_i > 1$ for some units characterized by large values of the MOS variable X. Suppose this happens for a subset of *k* units out of the *m* desired ones. In this case, the inclusion probabilities of those *k* units are set to 1, whereas the inclusion probabilities of the remaining *m' = m – k* units are re-calculated through formula (2), using m' instead of m and a restricted sampling frame that excludes the first *k* units. If probabilities greater than 1 are generated again, the above steps are iterated until $\pi_i \leq 1$holds true for all the units. This algorithm is automatically applied by R package sampling (Tillé and Matei, 2021).

careful exploration (see, for instance, (Valliant, Dever, and Kreuter, 2013), section 13.5.2), only a parsimonious subset of the available variables was selected[4].

- Second, the logistic model was fit and used to predict response probabilities, $\hat{p}_k$, for all the respondent PPCS households.
- Third, deciles of predicted response probabilities were calculated and used to cluster the respondent PPCS households into ten reasonably homogeneous and non-overlapping classes.
- Fourth, the ten response propensity classes were treated as Response Homogeneity Groups (RHG) (see Särndal, Swensson, and Wretman, 1992), and household-level nonresponse adjustment factors, $g\_nonresp_k$, were computed as reciprocals of weighted response rates[5] within propensity classes.

Note that the propensity score method summarized here only exploits the fitted logistic model to define the Response Homogeneity Groups. In other words, it does *not* use the predicted response probabilities, $\hat{p}_k$, to compute the nonresponse adjustment factors, $g\_nonresp_k$. More precisely, in the usual propensity modeling approach, nonresponse adjustment factors would be calculated as the inverse of the average $\hat{p}_k$ within each class. Instead, the propensity score method calculates them as reciprocals of weighted response rates within each class (RHG). The latter choice is arguably more robust to model misspecification than the former, and less prone to generate very large weight adjustments that may result in unstable estimates.

The nonresponse adjusted weights of PPCS households and roster individuals, w_k^{NRA}, were obtained by multiplying the initial weights, d_k, of equation (3) by the nonresponse adjustment factors, $g\_nonresp_k$, calculated using the propensity score method:

$$w_k^{NRA} = g\_nonresp_k \times d_k \tag{4}$$

3.3 S3 – Calibration of nonresponse adjusted weights of households and roster individuals

Calibration minimally adjusts survey weights so that survey estimates exactly match population parameters that are known from sources external to the survey (Särndal, 2007). These known population parameters are called ‘calibration benchmarks’ or ‘calibration controls’ and usually take the form of population totals. The survey variables for which calibration benchmarks are available are called ‘auxiliary variables’.

Calibration typically increases estimation efficiency: the stronger is the correlation between the interest variable(s) and the auxiliary variables, the larger will be the efficiency gain. Moreover, depending on how the auxiliary variables are chosen[6], calibration can also provide an additional layer of protection against nonresponse and/or frame under-coverage bias (Särndal and Lundstrom, 2005).

[4] These variables, selected through an AIC (Akaike Information Criterion) minimization procedure, are: ‘governorate’, ‘locality type’ (rural, urban, camps), ‘household size’, ‘average monthly expenditure on food’, ‘assistance’ (i.e. whether any member of the household receives any type of assistance), and ‘wellbeing’ (i.e. a 0-100 score derived from SEFSEC 2018).

[5] Note that these are response rates in terms of households.

[6] Nonresponse bias reduction can be achieved by calibration if the auxiliary variables: (i) are correlated to response propensity; (ii) are correlated to the interest variable(s); (iii) do identify important estimation domains. Powerful auxiliary variables should ideally have all the above properties (i), (ii) and (iii). However, any of those properties is beneficial in its own right.

Calibration of PPCS weights of households and roster individuals was performed using as calibration benchmarks household-level and individual-level population totals for 2022 West Bank and Gaza. These totals were provided by PCBS[7] and are reported in Tables 3 and 4, respectively. The first set of calibration benchmarks, reported in Table 3, encompasses 43 totals, representing counts of Palestinian households by governorate and locality type. The second set of calibration benchmarks, reported in Table 4, encompasses 68 totals, representing counts of Palestinian persons by region, five-year age class, and sex. Note that a single set of calibration weights was sought that *simultaneously* fulfills all the 111 calibration constraints induced by the household-level and individual-level benchmarks of Tables 3 and 4 (111 = 43 + 68).

To solve the calibration problem, the R software ReGenesees (Zardetto, 2015 and 2023) was used. Owing to the simultaneous presence of household-level and individual-level population benchmarks, the calibration task had to be undertaken at individual-level. However, ReGenesees facilities for *cluster-level* weights adjustments made it possible to produce identical calibration weights across members of the same household. Calibration weights with this property are known as *integrated* individual-household weights, see (Lemaitre and Dufour, 1987) and (Heldal, 1992). This property is desirable for calibration weights since *design* weights are inherently constant within each household in the PPCS survey (see the end of Section 3.1). Note, in addition, that a *range-restricted* calibration algorithm was applied, so as to prevent negative or exceedingly high calibration weights. More specifically, calibration adjustment factors g_cal_k (namely the so-called 'calibration g-weights') were constrained to the minimum bounding interval $g\_cal_k \in [0.38, 11.54]$. Exact convergence of the calibration algorithm was obtained: all the 111 calibration benchmarks were matched. This resulted in (i) perfect elimination of any estimation bias affecting the auxiliary variables and (ii) mitigation of possible residual bias in any variable which happens to be correlated with the auxiliary ones.

After calibration, the integrated PPCS weight of sampling unit *k* (household or individual) can be expressed in terms of the nonresponse adjusted weights in equation (4) as follows:

$$w_k^{CAL} = g\_cal_k \times w_k^{NRA} \tag{5}$$

where the calibration g-weights, g_cal_k, are the same for all roster individuals which are members of any same household and are equal to the g-weight of the household as well. As usual for complex calibration tasks, the calibration g-weights g_cal_k of equation (5) cannot be expressed in analytic closed-form.

[7] These population totals were obtained by PCBS as the result of a demographic estimation exercise.

Table 3: First set of population totals used as calibration benchmarks for PPCS households and roster individuals. These totals represent counts of Palestinian households by governorate and locality type in 2022.

Region	Governorate	Locality Type	Number of Households
West Bank	Jenin	Urban	45,013
West Bank	Jenin	Rural	26,062
West Bank	Jenin	Camps	2,432
West Bank	Tubas & Northern Valleys	Urban	10,051
West Bank	Tubas & Northern Valleys	Rural	2,940
West Bank	Tubas & Northern Valleys	Camps	1,321
West Bank	Tulkarm	Urban	31,601
West Bank	Tulkarm	Rural	7,647
West Bank	Tulkarm	Camps	3,767
West Bank	Nablus	Urban	53,739
West Bank	Nablus	Rural	29,436
West Bank	Nablus	Camps	6,845
West Bank	Qalqiliya	Urban	18,124
West Bank	Qalqiliya	Rural	8,300
West Bank	Salfit	Urban	12,816
West Bank	Salfit	Rural	5,037
West Bank	Ramallah & Al-Bireh	Urban	38,125
West Bank	Ramallah & Al-Bireh	Rural	35,144
West Bank	Ramallah & Al-Bireh	Camps	3,794
West Bank	Jericho & Al Aghwar	Urban	6,030
West Bank	Jericho & Al Aghwar	Rural	2,429
West Bank	Jericho & Al Aghwar	Camps	3,079
West Bank	Jerusalem	Urban	23,938
West Bank	Jerusalem	Rural	10,448
West Bank	Jerusalem	Camps	1,960
West Bank	Jerusalem	Urban	63,394
West Bank	Jerusalem	Camps	2,711
West Bank	Bethlehem	Urban	34,085
West Bank	Bethlehem	Rural	13,783
West Bank	Bethlehem	Camps	3,082
West Bank	Hebron	Urban	144,787
West Bank	Hebron	Rural	21,721
West Bank	Hebron	Camps	3,975
Gaza	North Gaza	Urban	66,033
Gaza	North Gaza	Camps	10,222
Gaza	Gaza	Urban	121,424
Gaza	Gaza	Camps	8,084
Gaza	Deir Al-Balah	Urban	37,940
Gaza	Deir Al-Balah	Camps	17,110
Gaza	Khan Yunis	Urban	67,076
Gaza	Khan Yunis	Camps	8,384
Gaza	Rafah	Urban	39,994
Gaza	Rafah	Camps	7,408

Table 4: Second set of population totals used as calibration benchmarks for PPCS households and roster individuals. These totals represent counts of Palestinian persons by region, five-year age class, and sex in 2022.

Region	Age Class	Female	Male
West Bank	0-4	198,740.8	206,492.2
West Bank	5-9	181,417.7	191,880.6
West Bank	10-14	169,318.5	177,518.2
West Bank	15-19	157,096.2	163,449.3
West Bank	20-24	145,231.2	151,722.2
West Bank	25-29	135,666.1	144,403.4
West Bank	30-34	116,646.0	122,928.0
West Bank	35-39	91,210.4	92,156.4
West Bank	40-44	78,294.1	79,120.6
West Bank	45-49	69,905.2	72,588.0
West Bank	50-54	58,964.4	61,796.2
West Bank	55-59	47,935.3	50,846.4
West Bank	60-64	36,461.2	38,101.1
West Bank	65-69	23,527.2	23,537.0
West Bank	70-74	16,228.5	15,053.0
West Bank	75-79	10,665.5	8,611.4
West Bank	80+	11,948.6	7,787.8
Gaza	0-4	159,597.0	165,673.4
Gaza	5-9	136,751.8	143,130.3
Gaza	10-14	131,332.3	137,404.3
Gaza	15-19	108,596.1	113,594.2
Gaza	20-24	91,960.1	95,910.8
Gaza	25-29	94,643.5	96,879.1
Gaza	30-34	80,582.3	81,822.4
Gaza	35-39	59,062.9	58,506.0
Gaza	40-44	47,417.9	46,756.8
Gaza	45-49	38,662.9	39,037.9
Gaza	50-54	29,193.4	29,910.3
Gaza	55-59	26,087.6	27,921.4
Gaza	60-64	18,099.5	18,478.1
Gaza	65-69	13,320.2	12,502.2
Gaza	70-74	9,515.6	8,750.3
Gaza	75-79	5,144.0	4,354.4
Gaza	80+	5,220.8	3,167.4

3.4 S4 – Consistent trimming of calibration weights of households and roster individuals

Unduly large calibration weights might lead to unstable estimates and inflate standard errors and confidence intervals. At the same time, negative calibration weights, or calibration weights whose value is less than one, may challenge the interpretation of end-users and therefore be perceived as undesirable. For these reasons, calibration weights may be trimmed using a suitable procedure. However, trimming calibration weights can result in introducing bias in survey estimators. Therefore, it is advisable to apply trimming procedures sparingly and carefully.

In the light of these considerations, PPCS calibration weights were trimmed to avoid weights larger than the maximum nonresponse adjusted weight, $\max\left(w_k^{NRA}\right)$, derived from equation (4). In fact, no weights smaller than one had been produced in step S3 (and negative calibration weights had been prevented by construction, given the calibration bounds described in Section 3.3). To tackle the trimming task, ReGenesees was used. The software made it possible to trim calibration weights to the desired interval[8] while simultaneously *preserving* (i) all the *calibration constraints* discussed in Section 3.3 and (ii) the individual-household *integration* property. In other words, after trimming, PPCS weights are still integrated and still able to reproduce, in estimation, all the population benchmarks reported in Tables 3 and 4.

In terms of the calibration weights in equation (5), these trimmed calibration weights can be written as:

$$w_k^{TRIM} = g\_trim_k \times w_k^{CAL} \tag{6}$$

Note that the weights in equation (6) above are the *final* PPCS weights for respondent households and roster individuals, namely the weights that should be used for general purpose analyses of PPCS survey variables that were *not* collected through the individual module of the questionnaire.

Table 5 summarizes the sample distribution of the weights of PPCS roster individuals as obtained along steps S1-S4 of the weights calculation procedure. Kish Unequal Weighting Effect (UWE) is also reported for each set of weights. Following Kish's definition (Kish, 1992), the UWE is calculated as 1 plus the relative sample variance of the weights. It can be regarded as a measure of how far the weights at hand are from the case of a self-weighting sample (UWE = 1).

Table 5: Summary of the sample distribution of roster individuals' weights along steps S1-S4 of the weight calculation procedure (symbol 'Q' stands for quartile). Kish Unequal Weighting Effect (UWE) is also reported.

Weights Type	**Min**	**1st Q**	**Median**	**Mean**	**3rd Q**	**Max**	**UWE**
Design [eq. (3)]	6.49	20.70	28.12	122.98	42.81	10,767.94	23.39
Nonresponse adjusted [eq. (4)]	7.42	26.38	36.31	158.41	56.46	12,000.64	22.85
Calibrated [eq. (5)]	2.95	26.69	57.62	164.09	104.20	14,831.44	15.35
Trimmed [eq. (6)]	2.95	26.74	57.94	164.09	105.69	12,000.64	14.43

It can be noted that the design weights were far from being constant from the outset, as expected from Section 3.1. Indeed, those design weights had embodied both the intrinsic variability of the final weights inherited from the RAPS 1 survey and the additional variability brought about by the PPS subsampling algorithm. Importantly, however, all the undertaken adjustment steps S2-S4 resulted in a reduction of the variability of the weights. This reduction was especially significant for the calibration step (-33% in terms of previous UWE) and was not negligible for the trimming step as well (-6% in terms of previous UWE). Overall, it is safe to conclude that the weights adjustment pipeline S2-S4 resulted in an increase of the estimation efficiency of the PPCS sample of respondent households and roster individuals.

[8] The trimming interval was set to $w_k^{TRIM} \in [2.95,\ 12000.64]$, see Table 5.

3.5 S5 – Calculation of initial weights of the adult individuals' sample

As anticipated in Section 2.4, the S_{ind18+} dataset stores individual-level information collected through the individual questionnaire of the PPCS survey, which was administered to one selected adult member (aged 18 years or above) of each respondent household. The selection mechanism was controlled by PCBS and applied during fieldwork according to the following protocol. For each respondent PPCS household, PCBS enumerators made an attempt to identify and re-interview the *same* adult individual who responded to the individual module of SEFSEC 2018. Only if the attempt was unsuccessful, they used a Kish grid (which accompanied the questionnaire) to randomly selected – with equal probability – one adult from the adult members of the household. Importantly, the re-interview attempt was mostly successful: among PPCS adult respondents, 91% had already responded to the individual module of SEFSEC 2018.

Concerning the described follow-up protocol, an important observation is in order. The re-interview attempt performed by PCBS enumerators is clearly an uncontrolled sample selection step, meaning that it is impossible to calculate the probability of its outcomes (i.e. success/failure) in a design-based sampling approach. Note that these probabilities would be needed to calculate the actual inclusion probability of *any* respondent adult in the S_{ind18+} sample, even those who were actually selected using the Kish grid. Since, under the adopted follow-up protocol and selection mechanism, it was impossible to calculate actual inclusion probabilities, the decision was made to *approximate* those probabilities based on (i) the already computed weights of PPCS respondent households and roster individuals, and (ii) the composition of each respondent household. To this end, the initial weight of each adult individual *k* belonging to the S_{ind18+} sample was computed as follows:

$$\tilde{d}_k^{18+} = w_k^{TRIM} \times \left(\frac{1}{\mathrm{n\_adults}_k} \right)^{-1} \tag{7}$$

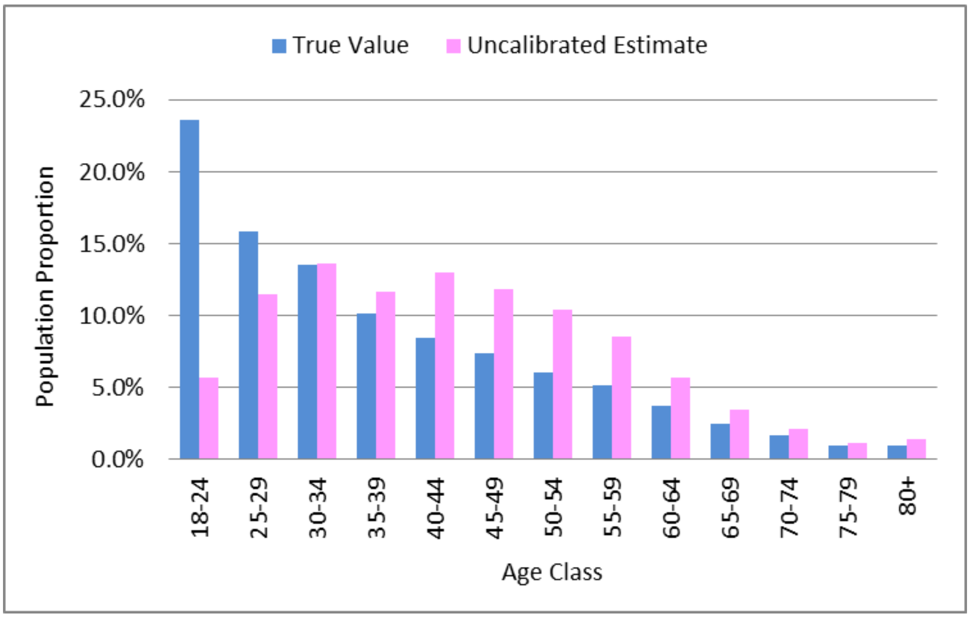
where w_k^{TRIM} is the final roster-level weight of individual *k* from equation (6), and $\mathrm{n\_adults}_k$ gives the number of individuals aged 18 years or above within the household individual *k* belongs to.

Note that equation (7) would only be appropriate in case the adult interviewed through the individual module of the questionnaire had actually been selected at random with equal probability from the $\mathrm{n\_adults}_k$ adult members of the household. For this reason, at odds with equation (3), the *approximate* initial weight *defined* by equation (7) was denoted with a tilde hat.

3.6 S6 – Calibration of weights of the adult individuals' sample

The follow-up protocol adopted by PCBS and illustrated in Section 3.5 was manifestly intended to maximize the overlap between the samples of adult individuals observed by SEFSEC 2018 and PPCS 2022, thereby boosting the statistical power of PPCS 2022 in terms of longitudinal analysis. However, in accordance with the observations of Section 2.1, that same protocol inevitably reduced the cross-sectional representativeness of the PPCS individual sample. For instance, the S_{ind18+} sample turned out to under-represent young adults and over-represent middle-aged and elderly adults, as can be recognized in Figure 1. Clearly, this is because the respondents to the individual module of SEFSEC 2018 had already to be adult in 2018, thus being at least 22 years old when responding to PPCS in 2022. Recall that former respondents to SEFSEC 2018 constitute the overwhelming majority (91%) of the S_{ind18+} respondent sample. Therefore, individuals below age 22 could enter the PPCS adult sample only very unfrequently, as a result of random selection through the Kish grid upon failure of the re-interview attempt.

Figure 1: Population proportion by age class. The blue bar chart shows true values derived from the calibration benchmarks provided by PCBS and reported in Table 6 below. The pink bar chart shows the uncalibrated estimates derived from the PPCS adult individuals' sample using the initial weights of equation (7). Note that the apparent under-estimation of younger adults (aged from 18 to 29 years old) and over-estimation of older adults (aged 35 years old and above) are an inherent feature of the PPCS adult individuals' sample and do not depend on the approximation adopted to compute the initial weights through equation (7). Indeed, the same features would be obtained using unweighted estimates.



The calibration step of individual adults' weights described in this section was designed to counteract those undesired effects and mitigate as much as possible any potential bias of PPCS cross-sectional estimates, including from nonresponse. As shown in Section 2.4, the individual-level nonresponse rate was almost negligible for the S_{ind18+} sample. Out of 6,138 adult individuals, only 261 did not respond, yielding an overall nonresponse rate of 4%.

Calibration of PPCS weights of adult individuals was performed using as calibration benchmarks household-level and individual-level population totals for 2022 West Bank and Gaza. These totals were provided by PCBS and are reported in Tables 3 and 6, respectively. The first set of calibration benchmarks, reported in Table 3 and already introduced in Section 3.3, encompasses 43 totals, representing counts of Palestinian households by governorate and locality type. The second set of calibration benchmarks, reported in Table 6, encompasses 52 totals, representing counts of Palestinian *adult* persons by region, age class, and sex. Note that a single set of calibration weights was sought that *simultaneously* fulfills all the 95 calibration constraints induced by the household-level and individual-level benchmarks of Tables 3 and 6 (95 = 43 + 52).

Table 6: Second set of population totals used as calibration benchmarks for PPCS adult individuals. These totals represent counts of Palestinian *adult* persons by region, age class, and sex in 2022.

Region	Age Class	Female	Male
West Bank	18-24	208,069.7	217,101.9
West Bank	25-29	135,666.1	144,403.4
West Bank	30-34	116,646.0	122,928.0
West Bank	35-39	91,210.4	92,156.4
West Bank	40-44	78,294.1	79,120.6
West Bank	45-49	69,905.2	72,588.0
West Bank	50-54	58,964.4	61,796.2
West Bank	55-59	47,935.3	50,846.4
West Bank	60-64	36,461.2	38,101.1
West Bank	65-69	23,527.2	23,537.0
West Bank	70-74	16,228.5	15,053.0
West Bank	75-79	10,665.5	8,611.4
West Bank	80+	11,948.6	7,787.8
Gaza	18-24	135,398.5	141,348.4
Gaza	25-29	94,643.5	96,879.1
Gaza	30-34	80,582.3	81,822.4
Gaza	35-39	59,062.9	58,506.0
Gaza	40-44	47,417.9	46,756.8
Gaza	45-49	38,662.9	39,037.9
Gaza	50-54	29,193.4	29,910.3
Gaza	55-59	26,087.6	27,921.4
Gaza	60-64	18,099.5	18,478.1
Gaza	65-69	13,320.2	12,502.2
Gaza	70-74	9,515.6	8,750.3
Gaza	75-79	5,144.0	4,354.4
Gaza	80+	5,220.8	3,167.4

Once more, the R software ReGenesees was used to solve the calibration problem. As before, a *range-restricted* calibration algorithm was applied, so as to prevent negative or exceedingly high calibration weights. More specifically, calibration adjustment factors $g\_cal_k^{18+}$ (namely the so-called 'calibration g-weights') were constrained to the minimum bounding interval $g\_cal_k^{18+} \in [0.325, 5.512]$. Exact convergence of the calibration algorithm was obtained: all the 95 calibration benchmarks were matched. This resulted in (i) perfect elimination of any estimation bias affecting the auxiliary variables (e.g. the bias originally affecting the age distribution, as strikingly highlighted by Figure 1) and (ii) mitigation of possible residual bias in any variable which happens to be correlated with the auxiliary ones.

After calibration, the PPCS weight of each respondent adult individual *k* belonging to the S_{ind18+} sample can be expressed in terms of the initial weights in equation (7) as follows:

$$w_k^{CAL18+} = g\_cal_k^{18+} \times \tilde{d}_k^{18+} \quad (8)$$

As usual for complex calibration tasks, the calibration g-weights $g\_cal_k^{18+}$ of equation (8) cannot be expressed in analytic closed-form.

Table 7 summarizes the sample distribution of the weights of PPCS adult individuals as obtained along steps S5-S6 of the weights calculation procedure. Kish UWE is also reported for each set of weights.

Table 7: Summary of the sample distribution of adult individuals' weights along steps S5-S6 of the weight calculation procedure (symbol 'Q' stands for quartile). Kish Unequal Weighting Effect (UWE) is also reported.

Weights Type	Min	1st Q	Median	Mean	3rd Q	Max	UWE
Design [approximate, eq. (7)]	5.35	72.74	156.04	501.39	330.56	61,005.98	14.47
Calibrated [eq. (8)]	3.30	65.12	142.18	505.59	331.10	28,663.21	10.22

Unsurprisingly, the approximate design weights of equation (7) exhibit a UWE which is close, but slightly larger, than the UWE of the trimmed weights of roster individuals (see Table 5). In line with what was already observed for the roster individuals, the calibration step significantly reduced the UWE also for the adult individuals' sample (-29%). Moreover, Table 7 shows that the calibration step did not produce exceedingly small or large weights. Even more, calibration resulted in weights characterized by a considerably shorter range than the approximate design weights. In particular, the calibration step S6 made the right-tail of the weights' distribution noticeably shorter, thus ruling out any need of an additional trimming step. Overall, it is safe to conclude that the weights adjustment pipeline S5-S6 resulted in an increase of the estimation efficiency of the PPCS sample of respondent adult individuals.

In the light of the above considerations, the calibration weights in equation (8) were approved as *final* PPCS weights for respondent adult individuals, namely the weights that should be used for general purpose analyses of PPCS survey variables that were collected through the individual module of the questionnaire.

4. Final remark

This note described the calculation of final weights for the Palestinians' Psychological Conditions Survey (PPCS 2022). The calculation entailed 6 main procedural steps, which have been concisely illustrated in dedicated sections of the note. The intermediate output of each procedural step has been symbolically summarized by the equation appearing at the end of the corresponding section. The final weights that will be used in practice by most end-users are those reported in equations (6) and (8).

5. References

Haziza, D., and Beaumont, J. F. (2017). Construction of weights in surveys: A review. *Statistical Science*, *32*(2), 206-226.

Heldal, J. (1992). A method for calibration of weights in sample surveys. *Working Papers from the Central Bureau of Statistics of Norway*, SSB.

Kish, L. (1992). Weighting for unequal Pi. *Journal of Official Statistics*, 8, 183-200.

Lemaitre, G., and Dufour, J. (1987). An Integrated Method for Weighting Persons and Families. *Survey Methodology,* 13(2), 199-207.

Särndal, C.E. (2007). The Calibration Approach in Survey Theory and Practice. *Survey Methodology*, 33 (2), 99-119.

Särndal, C.E., and Lundstrom, S. (2005). Estimation in Surveys with Nonresponse. John Wiley & Sons.

Särndal, C.E., Swensson, B., and Wretman, J. (1992). Model Assisted Survey Sampling*,* Springer Verlag.

Tillé, Y. and Matei, A. (2021). sampling: Survey Sampling. R package version 2.9. URL: https://cran.r-project.org/package=sampling

Valliant, R., Dever, J. A., and Kreuter, F. (2013). Practical tools for designing and weighting survey samples (Vol. 1). New York: Springer.

Zardetto, D. (2015). ReGenesees: An advanced R system for calibration, estimation and sampling error assessment in complex sample surveys. *Journal of Official Statistics*, *31*(2), 177.

Zardetto, D. (2023). ReGenesees: R Evolved Generalized Software for Sampling Estimates and Errors in Surveys. R package version 2.2. URL: https://diegozardetto.github.io/ReGenesees/