# Weighting

The goal of statistical weighting of survey data is to adjust the distribution of the sample to reflect the population’s distribution accounting for the sampling plan, to adjust for changes needed while the survey was in the field, to incorporate differential response rates, and to calibrate the weights to known population characteristics. Weighting achieves these goals. This section provides a sample weighting plan to create statistical weights based on the sample design and adjustments to deal with circumstances faced by the execution of the survey. The following Table W1 summarizes the weighting, adjustment, and planned use for analytical purposes, and it guides the pursuant discussion of the sample weighting plan for the MSMe Covid-19 study.

**Background to Weights**

Sampling weights must account for the sample design and adjustments made during the execution of the study in the field. The MSMe Covid-19 study’s sample design revolves around the following characteristics:

* National stratified probability sample with regional strata, substrata within regions, and either a multistage cluster area probability sample selected within a stratum or substratum or a random sample of businesses from a listed business registry. In either case, probabilities of selection and inclusion can be computed to create design weights.
* Post-field adjustments to the weights may include:
  + Mode of interviewing,
  + Non-Response,
  + Over-sampling of key population subgroups,
  + Handling of Ineligible businesses encountered in the field,
  + Incorporating reliable national statistics of businesses to calibrate weights, and
  + Adjust weights to facilitate combined reporting across countries in the study.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table W1: Summary of Components of Weighting** | | | |
| **#** | **Weighting Effect** | **What it is** | **Assumptions, Reasoning, and Recommendation** |
| 1 | Probability of selection | Design weight | Build F2F weights based on the multistage selection probabilities with documented assumptions  Build CATI weights based on size of list within strata   * Incorporating design weights into the sample weighting plan |
| 2 | Non-response | Accounting for differential response rates | Assumptions include differential response rates by urban/rural, region of country, size of business, and mode.  F2F-only countries: Nonresponse (NR) adjustment needed at cluster level  CATI-only: NR adjustment needed at regional strata and size of business levels  Mixed F2F and CATI countries: Sample design splits F2F and CATI for medium-sized businesses and regional strata.   * Incorporating NR adjustment into sample weighting plan |
| 3 | Eligibility | Accounting for ineligibility of businesses such as screening out of businesses with 250 or more employees | Study will terminate interviews with ineligible businesses with more than 249 employees. The lists can assist in estimating population numbers for medium-sized businesses.  An eligibility adjustment factor can adjust population sizes for the number of medium-sized businesses within regional strata.  The adjustment factor will be calculated using the information captured on the electronic contact form.   * Incorporating Eligibility adjustment into the sample weighting plan. |
| 4 | Quota | Oversample adjustments to bring to account for sample boosts to achieve minimum sample sizes for women-owned businesses | Specific countries require oversamples population subgroups such as women-owned businesses. These segments require oversampling with screening to achieve minimum sample sizes targets.   * Incorporating oversample adjustment in countries where oversampling takes place targeted population segments |
| 5 | Mode | Weighting to adjust for differences by mode (CATI / F2F) in countries where both are conducted | Sample design creates CATI /F2F sub-strata based on geography and size of business.  Response rate differences captured in the NR adjustment and the study’s sample design.   * Not incorporating in sample weighting plan |
| 6 | National Post-stratification weights (within country) | Weighting by business attributes can help calibrate the weights. Candidate attributes are business size and business sector within a country. | Statistical weights are not anchored to known values. Often used in post-stratification and adjustment of weights for the purpose of adjusting weights, method requires on reliable statistics. In most countries, reliable statistics are not available post-Covid 19.   * Not incorporating in sample weighting plan until after a review of the weights and results can be made |
| 7 | Cross-National (across countries) | Weighting each country to enable global comparisons | No weighting / no global figure will be reported / not relevant   * Not incorporating in sample weighting plan |

**Generalized Approach**

In most countries, the study uses a stratified multistage area probability clustered design for all businesses or the micro and small businesses. Countries are first stratified by region, and primary sampling units (PSUs) consisting of large portions of the country are selected at random from each stratum. The staged cluster sample then selects 1st-level secondary sampling units (SSU1s) from within sampled PSUs, and then local areas or 2nd-level SSUs (SSU2s) are selected from the sampled SSU1s. The geographical definitions that make up SSU1s and SSU2s differ by country. The SSU1s and SSU2s are based on each country’s administrative division structure. Where possible, SSU1s are selected from PSUs and SSU2s from SSU1s with probability proportional to their sizes.

Seven countries strictly use a stratified multistage area probability clustered sample design. In these countries, the administrative structure of the country serves as the sampling frame. This provides random sampling to take place without an exhaustive and authoritative list of businesses to serve as a sampling frame. In four other countries, the in-country teams recommended a dual-frame design where micro and small business are sampled using a stratified area probability multistage clustered sample design for micro and small businesses, and a stratified sample from a list sample frame of medium-sized businesses based on registry of businesses. Efforts are taken so the samples from the two frames do not overlap. The final country uses a stratified sample design from list sample frame.

The sample weighting plan accommodates each type of sample design: a) an exclusive stratified multistage area probability clustered design; b) a dual-frame stratified area probability sample design for micro and small businesses and a listed sample frame stratified design for medium-sized businesses, and; c) a listed sample frame stratified sample design. Plus, the sampling weights must explicitly accommodate each component of the weights in Table W1.

*Weighting Effect #1 - Design Weights*

Design weights account for the sample design selection of businesses. This involve the selections of PSUs, SSU1s and SSU2s. This section defines the probabilities of selection for each stage of the study’s stratified area probability multistage cluster sample design and stratified list sample designs.

The design weights are based on the probabilities of selection determined by the sampling plan. Many of components of the probabilities of selection are created prior to execution of the survey. The design weight is based directly on the probabilities of selection for each step of the sample design. The weights need to reflect the two separate sampling plan possibilities, the sample selection from a) an area probability multistage cluster sample frame, or b) the sample selection from listed or business registry sample frame.

1. Design weights under an area probability multistage cluster sample frame

The sample weighting plan accounts for the clustered design within the regional strata. Notation is defined to allow for the derivations of the probabilities of selection.

Notation:

Moss (Measure of Size) – This is an available measure to provide relative sizing for probability proportional to size sampling. The last subscript in the probability of selection expression denotes the level of the sampling, such as h sampling at the stratum level, i is the PSU level, etc. The subscript box to the right provides the legend. The choice of statistics for the measure of size may change depending on the availability of statistics for strata, PSUs, SSU1s and SSU2s.

Subscripts:

h - stratum

i – PSU

j – SSU1

k – SSU2

m - Business

ms – This is the number of units selected at each stage of sampling. The last subscript denoting level of the sampling, e.g., h-stratum, i -PSU, etc. For Stage 1, mh is the number of PSUs to be selected in stratum h. It is the number of SSU1s selected in its PSU in Stage 2, the number of SSU2s selected in its SSU1 in Stage 3, and the number of businesses selected in their SSU2 in Stage 4.

ns – Number of businesses interviewed. Please note the difference between businesses selected and interviewed. “Selected” indicates that interviewers attempted an interview where some cooperated and some did not, while “interviewed” refers to a completed survey.

Ps(S) – is the probability of selection. Probabilities of selection can be computed for each PSU, SSU1, SSU2 and business.

Design weights are the reciprocal of the probability of selection - . The final design weights are the product of the probabilities of selection for each stage of the sample. The probabilities of selection are:

* Probability of PSU i in stratum h is

where is the number of PSUs selected in stratum h.

* Probability of selection for SSU1 j within PSU i for stratum h is

where is the number of SSU1s selected in PSU i in stratum h.

* Probability of selection for SSU2 k within SSU1 j in PSU i for stratum h is

where is the number of SSU2s selected within SSU1 for PSU i in stratum h.

The overall probability of selection for SSU2k is

.

If the measures of size are consistent – from the same data source and definition - for each stage of sampling, the overall probability of selection reduces to

.

In addition, the sample designs try to keep the number of SSU1s and SSU2s selected at each stage the constant within a stratum, so the product of should be a constant value within a stratum. This is not possible for all countries.

An explicit statement for the overall probability of selection of business cannot be stated prior to field in all countries. The number of businesses to be selected is not known until interviewers conduct the random walk to select businesses in a SSU2. The structure for the probability of selection for a business is

where the last term has the expected average number of businesses selected in SSU2 k serves as a place-holder actual amount. After the field is completed, the actual overall probability of selection can be calculated.

1. Design weights under a stratified listed sample frame

The sample weighting plan for stratified listed sample accounts for the number of businesses available in the sampling frame , i.e., the business registry or list. The following notation is used in the deriving the probabilities of selection.

Notation:

* Moss (Measure of Size) – The measure of size is the number of businesses listed in the frame where the subscript denotes the stratum h or PSU i. The notation is kept flexible if individual businesses are sampled at the stratum level or the PSU level.
* ms – Number of businesses selected within level of the sampling denoted by the last subscript, e.g., h-stratum or i -PSU.
* ns – Number of businesses interviewed with last subscript denoting level of the sampling, e.g., h-stratum or i -PSU.
* Ps(S) – Probability of selection with last subscript denoting level of the sampling, e.g., h-stratum or i -PSU.

The overall probability of selection for a business if sample was selected from the listed sample frame for a set of PSUs is

.

The overall probability of selection for a business if the sample was selected across the entire stratum is

.

*Weighting Effect #2 – Nonresponse Adjustments to Design Weights*

Weighting effect #2 accounts for differential response rates across elements of the sample. Response rates are not known at the time of the sample design, so it is an adjustment to the design weights. This section reports the plan to incorporate a nonresponse (NR) adjustment.

The design weights do not reflect the final probabilities of inclusion. The number of businesses that cooperate is known only at the end of fieldwork. The sample design allows for non-cooperating businesses to be replaced in the field. For an area probability multistage cluster sample design, the random walk continues in a SSU2 once callbacks take place and until the designated number of interviews are completed. The listed sample design selects additional businesses once interviewers finish with contacting and following up with businesses in the initial sample. Additional samples are selected such that the targeted number of completed surveys is met.

Notation:

– number of businesses selected in SSU2 j for multistage area cluster sample design where it is the number of businesses with an attempted interview

- number of businesses interviewed in SSU2 j for multistage area cluster sample design

– number of businesses selected in SSU2 j for multistage area cluster sample design where it is the number of businesses with an attempted interview

- number of businesses interviewed in SSU2 j for multistage area cluster sample design

The adjustment factor for response rate is calculated at the level of interviewing, the SSU2. The NR adjustment for each SSU2 in a multistage design is

.

The NR adjustment for a listed sample frame design is

for the sample selected for stratum h, or

for a sample selected for PSU i for stratum h.

The probability of inclusion for a business is the product of its probability of selection and its response rate adjustment factor,

Several choices are made to simplify this expression:

* Fix the number of completes to be same for each SSU2, and
* Use comparable measures of size at each stage of the sampling.

If this is done, the probability of inclusion simplifies to

,

and an equal probability of selection with a stratum for all businesses such as 1000/ Mosh, if the numbers of selections are constant through each stage. The probability of inclusion for a business from a listed sample follows this, and it reduces to

*Weighting Effect #3 – Eligibility Adjustments to Weights*

Weighting Effect #3 adjusts for inaccuracies in the database used in listed sample frame designs, and businesses are contacted believing they qualify for the survey but are found to be too large.

Sample designs use listed sample frames for medium-sized business stratifies medium-sized businesses in the sample frame by region. The sample design then selects businesses with equal probabilities from the frame within each stratum. The sample frame provides a population count for each stratum and is the measure of size for the probabilities of selection. Selected businesses are then contacted and screened. If a selected business was classified as medium-sized but is determined to be a large-sized business, the survey is terminated. Since the probabilities of selection are based on a measure of size from the sample frame, using the number of medium-sized businesses from the sample frame biases the probability of selection calculation. In this case, the measure of size requires adjustment.

The eligibility adjustment factor for stratum h is:

where is the number of screened eligible businesses in stratum h, and is the number of screened businesses in stratum h. This is an estimate of the eligibility rate of businesses in the sample frame for stratum h. The adjusted measure of size is

.

These revised are substituted into the weight formulae for the listed sample frame.

*Weighting Effect #4 – Oversampling to meet Quotas for Special Population Segments*

Segments of the population have been targeted for oversampling. This section provides adjustments to the sample weights account for this. This adjustment is being incorporated into the sample weighting plan.

In some countries, segments of businesses are oversampled to provide sufficient samples for analytical or strategic purposes. One population segment are women-owned businesses, and in another country, IT related business were asked to be oversampled. Oversampling a population segment requires screening to find additional population members, and it also requires interviewers to terminate interviews with otherwise eligible and cooperative sampled businesses. Specific sample size targets are created for the targeted population subgroups. The result is that the oversampled population subgroups will have a disproportionately large number of businesses in the sample, while the remainder of the sample would be underrepresented. An adjustment for the oversample comes directly from the screening process and information collected on the electronic contact form during the screening process.

The over-sample quota adjustment factor for businesses in the oversample population subgroup is:

The adjustment factor for the businesses not part of the oversampled subpopulation is

The adjustments are applied to the weights, where the weights after the adjustment are:

, if business is in the oversample subpopulation, and

, if business is not in the oversample subpopulation.

*Weighting Effect #5 – Mode Adjustments*

Weighting Effect #5 deals with adjustments due mode of interviewing. The dual-frame studies will have some businesses contacted, recruited, and interviewed in-person by an interviewer and other businesses contacted, recruited, and interviewed by telephone. This section deals with prospective impact by mode of interviewing on response rates and responses to the survey question. It is the recommendation to not adjust the weights for mode effects, and this section provides a discussion behind this recommendation.

Four countries have dual-frame designs where micro and small business are sampled using the stratified multistage area probability clustered sample design, and medium-sized businesses are sampled from listed sample frame with contact and interview to take place by telephone using CATI. Since the contact and interview takes place by telephone, there is a concern that the mode of interviewing may impact response rates and responses. As Table W2 shows, most countries have their interviews conducted using the same mode, so there is no need to consider an adjustment for a mode effect in the single mode countries. Four countries use a dual-frame mixed mode design where mode is incorporated explicitly into the sample design itself.

In the dual-frame designs, the sample design purposively stratifies businesses by size of business and sample frame. This effectively removes the possibility of sampling of medium-sized business from the listed sample frame and the area sample frame. This also eliminates the overlap of the mode of contact and recruiting for the same businesses. In these instances, any possible issues with two different modes of contact and response rates are removed, and any effect on response rates by mode of contact are compensated in the response rate weighting adjustments.

**Table W2: Methodology by Country**

| **Country** | **Interviews done by F2F** | **Interviews done by CATI** |
| --- | --- | --- |
| **Argentina** | 1,000 (100%) |  |
| **Bangladesh** | 1,000 (100%) |  |
| **Brazil** |  | 1,000 (100%) |
| **Cambodia** | 1,000 (100%) |  |
| **Colombia** | 900 (90%) mostly micro, small, and informal businesses | ~ 100 CATI interviews (mostly formal and medium sized) |
| **India** | 1,000 (100%) |  |
| **Indonesia** | 800 (80%) representing mostly micro, small, and informal businesses | 200 (20%)3 representing mostly formal and medium sized businesses |
| **Peru** | 800 (80%) representing mostly micro, small, and informal businesses | 200 (20%)3 representing mostly formal and medium sized businesses |
| **Philippines** | 1,000 (100%) |  |
| **Sri Lanka** | 700 (70%) representing mostly micro, small, and informal businesses | 300 (30%) representing mostly formal and medium sized businesses |
| **Thailand** | 1,000 (100%) |  |
| **Vietnam** | 1,000 (100%) |  |

Literature Review on Mode Effects on Survey Responses

The survey research literature has extensively investigated differences in survey results due to the mode of interviewing, i.e., mode effects. Most research focuses on the differences between self-administered and interviewer-administered surveys. Past research studied these questions when surveys began migrating from face-to-face interviewer-administered surveys to self-administered surveys, from mail surveys to telephone surveys, and more recently, from telephone surveys to online surveys. Much of the research focuses on social psychological constructs such as the impact of social desirability on interviewer-administered surveys versus self-administered surveys. Social desirability is the unintended desire to paint oneself or one’s business in the best light for an interviewer.[[1]](#footnote-1) Another aspect is, when answering sensitive questions, the desire for privacy or anonymity afforded by a self-administered versus an interviewer-administered survey.

Another area of focus has been in the differences in the ability of respondents to process and respond in self-administered and interviewer-administered surveys. For example, interviewer-administered surveys do not require respondents to have reading skills. However, any survey question requires a respondent to comprehend the question, evaluate and recall information to answer the question, combine the information with the question and make a response.[[2]](#footnote-2) An interviewer-administered survey question relies solely on the respondent to listen, process and recall by memory alone to answer the question, while respondents in a self-administered survey has the question written out in front them along with all of the scale points to help answer the question. Past research and common survey research experience show that interviewer-administered respondents will migrate their answers toward the endpoints of the scale while self-administered survey respondents use of the middle of the scale more often.

The study also uses Ipsos’s iField tablet-based computer assisted interviewing system for both the in-person and telephone interviewing in countries where face-to-face or a combination of face-to-face and telephone interviewing are planned. Computer assisted interviewing controls questionnaire logic flow and prompts for verification or correction to out-of-range responses.

These concerns do not exist in the MSMe study since all surveys are interviewer-administered. This reduces mode effects for the survey responses. The survey does not ask sensitive questions about a business also reducing concerns about the need anonymity. Effects from social desirability will exist similarly for both modes, but the lack of sensitive questions mitigates this concern. Finally, the use of computer-assisted interviewer interviewing – CATI and CAPI, eliminates errors that could occur from interviewers asking and recording surveys into paper forms.

The conclusion is mode effect is not a substantive issue for the MSMe study since both modes use a live interviewer and computer-assisted interviewing to administer the survey.

***Mitigating and Investigating Mode Effects on Survey Responses***

Although, the conclusion is that mode effect is not a substantive issue, there are aspects of sample design that mitigate mode effects. Statistical sampling has addressed using multiple frames within a single survey since the 1960s. Past uses of multiple frames have assumed the same mode of interviewing respondents is used in both frames. In the dual-frame countries, the impact of using two frames and modes are mitigated by the sample design. This section suggests analyses to explore if any impact from mode exists. They are:

* Addressing overlap in sample frames -- An advantage to using a business registry is that it increases the reach of the sample for medium-sized from the PSUs and SSUs to most of the country. To address sample frame overlap for medium-sized businesses, businesses located in the sampling plans’ PSUs, are excluded or the area probability sample is restricted to micro and small businesses. Additionally, business names will be recorded to ensure no business is interviewed twice.
* Identifying mode effects and impact - Stratification by size of business use different sample frames and modes of contact and interviewing. This would allow for analyzing the impact of mode on survey measures’ responses. Survey methods research and practice suggest that the impact on the measures should be limited, and its impact on response rates may be greater. There may be confounding effects by geography and business with size of business and other factors. A suggestion is to conduct a post-survey analyses that examines the mix of businesses and its impact on a couple of key survey variables. Note that there may be legitimate differences by size of business regardless of the mode of interviewing. Suggestions for an analysis include:
  + Within a country, conduct either a regression or logistic regression for 2 or 3 key survey variables.
  + The key survey variables will serve as dependent variables and dictate the regression methodology
    - A binary dependent variable (or recoded top-box) will require a logistic regression analysis
    - A scalar dependent variable with require a multiple regression analysis
  + Include size of business, mode of interviewing, urban/rural status, business sector, connected/ non-connected business status among a few possible explanatory variables in the regression analyses. If the mode of interviewing is significant after all other possible differentiators, then we will know whether a mode effect exists, and we will have a gauge on the size of its impact. If it is not significant, then mode effect is not substantial enough to be registered by the analysis.

*Weighting Effect #6 – National Adjustments to Calibrate and Control Weights*

Most surveys use national post-stratification to calibrate weights and survey results. At this time, reliable national survey statistics of businesses have not been identified and available for each country. Current available statistics are up to 10 years-old and more recent statistics are from before the pandemic. The conclusion is to not incorporate national statistics and calibration of the weights until after a review of the weights and survey results can be conducted.

The sample weighting plan creates design weights and adjusts them for differential response rates, oversampling of specific population segments, and using potentially obsolete measures of size as part of the weighting plan. These can possibly create large weights and unstable results. Survey sampling typically uses known population statistics to calibrate and mitigate the impact of large weights. Although, many of the countries in the MSMe study have economic censuses and recent sophisticated surveys of businesses conducted within the past 8 years, the Covid-19 pandemic has rendered most economic and business statistics obsolete or, at the least, suspect.

The problem is most surveys use the government and reliable survey statistics as targets for post-survey adjustment of the weights. Methods such as raking-ratio adjustments calibrate design weights by adjusting the weights to known marginal totals such as the number of businesses by size, by economic sector, and other characteristics. Many of the measures predate the pandemic. Anecdotally, we have heard that people in many countries survived economically by setting up informal cottage industry businesses to endure the pandemic. These business types are often not represented in government statistics and using the pre-pandemic statistics might suppress the impact of the informal cottage, micro and small businesses.

The MSMe team recommends not using existing official government and non-official survey results as part of the sample weighting plans to directly calibrate the weights. On the other hand, the distribution of the weights and their results population characteristics from the survey data will need to be thoroughly reviewed. Post-survey calibration of the weights may be needed based on best estimates of characteristics of businesses, but this may include a combination of older surveys, economic and country experts, and subjective assessments to provide rational targets for calibration.

Our recommendation is to not yet adjust weights to national statistics, but analyze the weights, confer with the Facebook team, and decide on strategies if the weights produce uncertain or unreliable results.

*Weighting Effect #7 - Cross-National (across countries*) *Weighting*

The last possible effect is to create weights across the 12 countries for reporting and comparisons. Combined reporting and direct statistical comparisons will not be done, so no effort will be made to consolidate the weighting for the 12 countries.

1. Kreuter, Frauke, Stanley Presser and Roger Tourangeau. 2008. “Social Desirability Bias in CATI, IVR, and Web Surveys the Effects of Mode and Question Sensitivity.” Public Opinion Quarterly. [↑](#footnote-ref-1)
2. Tourangeau R. Cognitive sciences and survey methods. In: Jabine T, Straf M, Tanur J, Tourangeau R, eds. *Cognitive aspects of survey methodology: building a bridge between disciplines*. Washington DC: National Academy Press, 1984. [↑](#footnote-ref-2)