ERF COVID-19 MENA Monitor sampling, response rates, and weights
Version: March 30, 2021

FOR WAVE 2 RELEASE

Summary
This document describes the sampling, response rates, and creation of the ERF COVID-19 MENA Monitor sample weights. Weights should be used in all analyses to ensure the basic characteristics of samples reflect the underlying universe. Weights cannot overcome unobservable dimensions of non-response bias.

Sample
The sample universe for the household survey was mobile phone users aged 18-64. Random digit dialing, within the range of valid numbers, was used, with up to three attempts if a phone number was not picked up/answered, was disconnected or busy, or picked up but could not complete the interview at that time. Samples were stratified by country-specific market shares of mobile operators.

For follow-up waves, previous wave respondents were recontacted if they consented to follow-up in the previous wave. Up to three attempts were used, including contacting second and family/friend numbers, if provided in wave one, on the third call. If the individual could not be reached or refused, a refresher individual was added to the sample in their place, randomly selected as with base wave respondents.

Quota in Jordan
The sample in Jordan collected responses from only Jordanians, Syrians, and Palestinians. It over-sampled Syrians (quota of 500; resulting sample 516 Syrians). The weights include nationality and are based on an in-person survey with nationality-specific weights to account for this quota (Krafft & Assaad, 2018; OAMDI, 2018).

Cross sectional data

The cross sectional (one point in time) or repeated cross sectional (multiple points in time) data structure should be used for all point-in time-analyses. The data are structured so that a unique observation is an individual-wave combination (long data). Individuals may have multiple observations for each wave; the variable wave identifies the wave and indid the individual identifier in that wave.

Cross-sectional weight variable names:
- Individual weight: ind_wt
- Household weight: hh_wt
- Household member weight: hh_mem_wt

See details below on weight creation.

Panel data

The panel data are structured so that an observation is a unique individual (wide data), and time-varying variables have a _w# suffix where # denotes the wave. This data structure should be used for any analysis of transitions, whether from February 2020 (#=0) or across specific wave point in times (#), or any combination of these times.

Panel weight variable names [where first # is base wave and second # is subsequent wave]:
- Individual weight: panel_ind_wt_#_#
- Household weight: panel_hh_wt_#_#
- Household member weight: panel_hh_mem_wt_#_#

Current combinations include:
*_0_1 (Feb. 2020 to Nov. 2020)
*_1_2 (Nov. 2020 to Feb. 2021)
*_0_2 (Feb. 2020 to Feb. 2021)

See details below on weight creation.

<u>**Response rates**</u>
Details of unsuccessful calls were not available for Egypt in wave 1, but Table 1 includes responses and response rates for all other waves, distinguishing between panel (unweighted) and refresher respondents as applicable. Phones that were not in service, disconnected/busy (after multiple calls) and individuals who were not eligible are excluded from the response rate calculations. The responses are based on the final result, which may have been on the first, second, or third attempt.

Table 1. Responses and response rates, by country

Response	Morocco (Nov. 2020)	Tunisia (Nov. 2020)	Egypt (Feb. 2021)	Jordan (Feb. 2021)	Morocco Panel (Feb. 2021)	Morocco Refresher (Feb. 2021)	Tunisia Panel (Feb. 2021)	Tunisia Refresher (Feb. 2021)
Phone disconnected/busy	9	26	9	1	8	19	5	21
Not in service	67	35	38	76	25	57	4	42
Did not answer	4	11	4	2	4	9	10	26
Picked up and refused	4	8	35	6	11	6	8	3
Incomplete, and refused	3	3	2	1	4	2	5	1
Incomplete, return call	0	1	0	0	0	0	0	0
Complete	12	14	11	11	48	7	68	6
Not Eligible	1	3	1	1	0	1	0	1
Total	100	100	100	100	100	100	100	100
Response rate	50	38	21	53	71	29	74	18

Note: Responses are for individuals who consented to follow-up in the previous wave. Not shown are 85 Tunisia and 190 Morocco Nov. 2020 respondents who did not consent to follow-up and 16 panel Feb. 2021 respondents in Tunisia excluded due to quality control issues.

Initial Weights
This section discusses the initial, base wave weights, and a subsequent section discusses panel weights and then refresher and combined weights.

Inverse probability weighting was undertaken to reduce bias along a number of observable dimensions. Weights were created on three levels: Individual, household, and household member. Weights had the following inputs:
- Telephone operators and their market shares, provided by the data collection firm
- Number of phones by operator for individuals (individual weight) and household members (household weight and household member weight)
- Representative data with comparable demographic and household characteristics to weight for non-response

Denote individuals as i (ranging from 1 to N) and households as h (ranging from 1 to N). Denote the number of phones from a particular operator, o, as t_o (operators ranging from 1 to M). Denote as the total number of phones there should have been in the sample from o, given the total number of phones observed and market shares, as T_o.

We then generated initial market-share individual weights as:
$$w_i = 1/\sum_1^M[(\{\sum_1^N t_{o,i}\}/T_o) * t_{o,i}]$$

With these individual weights, we then pooled the phone surveys with representative in-person surveys and used a probit model weighted with survey weights (for the representative survey) and w_i (for the COVID-19 monitor data) to estimate the probability an individual with particular characteristics was in the phone survey data. The predicted probability from that model, p_i, was used to generate individual weights for the COVID-19 monitor data as:
$$w'_i = w_i * (1 - p_i)/p_i$$

We likewise generated initial market-share household weights as:
$$w_h = 1/\sum_1^M[(\{\sum_1^N t_{o,h}\}/T_o) * t_{o,h}]$$
Which accounts for the number of phones in the household, across all members, and thus for a higher probability of sampling a household with more members or more phones. The predicted probability from the individual level model was combined with the market-share household weight to generate a household weight as:
$$w'_h = w_h * (1 - p_i)/p_i$$
Household member weights were calculated by multiplying household weights by household size. Household and individual weights (but not member weights, for internal consistency) were all winsorized at the 99th percentile to ensure that no outlier weight drove statistics. Weights were then normalized by dividing by the mean weight.

Individual weights should be used for all analyses where the outcome is at the individual level. If outcomes are at the household level (e.g. household income, household food security) then the

weight will depend on whether you are generalizing to households (e.g. X% of households are food insecure) or household members (e.g. X% of individuals live in a food insecure household).

Panel weights

All respondents who consented to follow up in the prior wave were contacted in an attempt to include them in the subsequent wave. Varying degrees of follow-up occurred; 64.7% (1,294 of 2,000) Nov. 2020 respondents in Tunisia were successfully tracked to Feb. 2021 and 43.0% (863 of 2,007) in Morocco.

We compute a response adjustment factor, r, to weight the households and individuals retained in the panel, based on the predicted probably of attrition, $\Pr(A)$, from a probit model with attrition as the binary outcome, as follows:

$$r = \frac{1}{1 - \Pr(A)}$$

This response adjustment factor multiplies the base wave household, household member, and individual weights for panel households that were retained, in order that they can represent the base wave universe.

The panel attrition models use a few base wave variables in addition to those used for initial weighting (Table 2). Specific additional variable are:
- Telephone operator
- Household income (categorically) in Feb. 2020
- Base wave labor market status (employed, unemployed (search required), out of labor force)
- Interactions with sex for categorical income and base wave labor market status

Refresher weights

The refresher weights are created in an identical fashion to the base wave, initial weights, but for the refresher samples within the subsequent waves of the panel.

Combined weights

For subsequent waves (waves after the base wave), cross-sectional weights combine the panel and refresher data. Weights are normalized to one within each of the panel and refresher samples and then combined into a single, representative cross-sectional weight.

Representative in-person national surveys

The representative in-person national survey samples used to generate weights were as follows:
- Egypt: Egypt Labor Market Panel Survey 2018 (Krafft, Assaad, & Rahman, 2019; OAMDI, 2019).

- Jordan: Jordan Labor Market Panel Survey 2016 (Krafft & Assaad, 2018; OAMDI, 2018).
- Tunisia: Tunisia Labor Market Panel Survey 2014 (Assaad, Ghazouani, Krafft, & Rolando, 2016; OAMDI, 2016).
- Morocco: Morocco Household and Youth Survey (MHYS) 2009-10 (World Bank, 2010).

These were selected as the most recent publicly available data with individual phone ownership and relevant demographic and labor market characteristics.

Table 2 displays the covariates included in the weighting models by country.

Table 2. Covariates included in weighting models, by country

Covariate	Egypt (June 2020 Poll)	Egypt (2021)	Jordan	Tunisia	Morocco
Sex	X	X	X	X	X
Age group	X	X	X	X	X
Education level	X	X	X	X	X
Household size (categorically)	X	X	X	X	X
Labor mkt. status in Feb. 2020	X	X	X	X	X
Administrative geography	Governorate	Governorate	Governorate	Region	Region
Urban v. rural	X	X	X	X	X
Int. b/w admin. geo. and urban	X	X	X	X	X
Marital status		X	X	X	X
Presence of kids 0-5		X	X	X	X
Presence of kids in school		X	X	X	X
Nationality			X		
Int. of covariates and sex	X	X	X	X	X

Notes: Marital status was not available in Egypt (June 2020 Poll) data. Labor market status in Feb. 2020 was mapped onto labor market status at the time of the representative national survey for Egypt (June 2020 Poll) data. See do files and questionnaires for additional information on how equivalent variables were created.

References

Assaad, R., Ghazouani, S., Krafft, C., & Rolando, D. J. (2016). Introducing the Tunisia Labor Market Panel Survey 2014. *IZA Journal of Labor and Development*, *5*(1).

Krafft, C., & Assaad, R. (2018). *Introducing the Jordan Labor Market Panel Survey 2016. Economic Research Forum Working Paper Series No. 1186*. Cairo, Egypt.

Krafft, C., Assaad, R., & Rahman, K. W. (2019). *Introducing the Egypt Labor Market Panel*

Survey 2018. Economic Research Forum Working Paper Series No. 1360. Cairo, Egypt.

OAMDI. (2016). Labor Market Panel Surveys (LMPS). Version 2.0 of Licensed Data Files; TLMPS 2014. Egypt: Economic Research Forum (ERF).

OAMDI. (2018). Labor Market Panel Surveys (LMPS). Version 1.1 of Licensed Data Files; JLMPS 2016. Cairo, Egypt: Economic Research Forum.

OAMDI. (2019). Labor Market Panel Surveys (LMPS). Version 2.0 of Licensed Data Files; ELMPS 2018.

World Bank. (2010). Morocco Household and Youth Survey (MHYS) 2009-2010. *Ref. MAR_2009_MHYS_v01_M_PUF*. Retrieved from http://microdata.worldbank.org/index.php/catalog/1546