

TRANSMIT Surveys of Syrian nationals and their neighbors in Lebanon and Turkey (2019-2023)

Data Manual

Version 1.0.0 | May 2025



Editor

Deutsches Zentrum für Integrations- und Migrationsforschung (DeZIM) e.V.

Forschungsdatenzentrum DeZIM.fdz | Mauerstraße 76 | 10117 Berlin

+49 (0)30 804 928 93 | fdz@dezim-institut.de

Authors

Lidwina Gundacker (Institute for Employment Research, IAB, Nürnberg, Germany, and University of Bamberg, Germany, Lidwina.gundacker@iab.de, corresponding author)

Laura Hertner (Berliner Institute for Empirical Integration and Migration Research, Humboldt University, Berlin, Germany)

Nora Kühnert (Berliner Institute for Empirical Integration and Migration Research (Humboldt University, Berlin, Germany)

Ramona Rischke (Deutsches Zentrum für Integrations- und Migrationsforschung, DeZIM, Berlin, Germany)

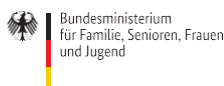
Simon Ruhnke (Berliner Institute for Empirical Integration and Migration Research, Humboldt University, Berlin, Germany)

Nader Talebi (Berliner Institute for Empirical Integration and Migration Research (Humboldt University, Berlin, Germany)

Caroline Trocka (Institute for Employment Research, IAB, Nürnberg, Germany)

Herbert Brücker, (Institute for Employment Research, IAB, Nürnberg, Germany, and Berliner Institute for Empirical Integration and Migration Research, Humboldt University, Berlin, Germany)

Funding through



1 Inhalt

2	Introduction	2
3	Target population and sampling strategy	3
3.1	Baseline Sampling Lebanon	3
3.2	Baseline Sampling Turkey	4
3.3	Follow-up Strategy	4
3.4	Deviations	5
4	Survey design	5
4.1	Interview process.....	5
4.2	Questionnaire Design	6
4.3	Ethical review.....	7
5	Data quality controls.....	7
6	Panel composition and attrition	11
	Data access	12
	Citation of data	13
	Publications based on data	14
	References	14
	Appendix.....	17

List of Tables

Table 1. Fieldwork timing and number of observations by country and wave	4
Table 2. Mean and median interview durations by country and wave	6
Table 3. Interviews affected by irregular patterns, by country, wave, and type of irregularity	11
Table 4. Number of observations by wave, strata and panelist status (excluding potential fraud cases)	12
Table 5. Provision of household geo information by data access type and country.....	12

Executive summary

Migration scholars commonly study international migration and socio-economic integration separately, typically focusing on a single receiving country in the Global North. The TRANSMIT project adopts a transnational perspective, studying key migration routes to Europe via large-scale surveys among Syrians and their host communities in two major countries of mixed-migration contexts, namely Lebanon and Turkey. Both countries, in addition to their own history of emigration, host large refugee populations. The surveys cover both migration biographies and aspects of the socioeconomic participation of Syrian nationals. Surveying the host populations in equal terms allows researchers to examine local contexts and group differences.

2 Introduction

Knowledge about migration and integration dynamics is growing steadily. Thereby, patterns of international migration and socio-economic integration are usually studied separately from each other and typically focus on a single receiving nation state (often in the Global North). However, a transnational perspective is necessary in order to understand the diverse links between countries of origin, transit and destination that arise from the cross-border phenomenon of migration. This is particularly true for forced migration which typically does not unfold in a linear and predictable manner, given the uncertainties and constraints involved (Collins, 2021; Jung, 2023). The TRANSMIT Project aims to provide the empirical foundation for such a transnational perspective, collecting data across origin and transit countries along key migration routes to Europe. Between 2019 and 2023, the project has administered annual large-scale panel surveys among Syrians and their host neighbors in Lebanon and Turkey, covering both migration biographies and major aspects of (post-migration) socioeconomic participation and living conditions. Aligning the questionnaires with the IAB-BAMF-SOEP Survey of Refugees in Germany (Brücker et al., 2017) wherever feasible allows comparisons with a major European receiving country. Surveying the host population in equal terms enables researchers to account for local context factors and examine differences between groups.

Both Lebanon and Turkey have emerged as important mixed-migration contexts as in addition to their own long-standing history of emigration, they host some of the largest refugee populations in the world. Located at Syria's

Western border, Lebanon was among the first countries to receive significant numbers of Syrian refugees when the Syrian civil war broke out in 2011 (Gundacker et al., 2024a). In 2022, the UNHCR counted over 800.000 officially registered refugees in Lebanon (UNHCR, 2024, likely a considerable undercount given the lack of reliable registration data). This rendered the small Mediterranean country the state with the highest refugee-to-population ratio in the world. With an estimated 3.2 million in 2023, Turkey is host to the largest population of displaced Syrians and in absolute terms host to one of the largest refugee population in the world (UNHCR, 2024). Both Turkey and Lebanon thus represent important mixed-migration contexts that allow for unique insights into the dynamics of human mobility, displacement and social cohesion.

3 Target population and sampling strategy

The target population for the TRANSMIT Surveys included individuals aged 15 years or older who were residing in either Lebanon or Turkey at the time of the surveys and who were residing in private housing (excluding individuals residing in formal refugee camps, detention centers and barracks). The survey was stratified according to the country of birth of the head of household, such that 50 percent of respondents were from Syrian households and 50 percent were from non-Syrian households. In addition to the main questionnaire, basic demographic information on household members and family members outside the household was collected. To reduce the length of the survey and to address the concerns noted in some respondents about providing information on individual family or household members, this was reduced to collecting basic aggregate information on children, spouses, and parents starting in wave 4 of the Lebanon survey and wave 3 of the Turkey survey. Note that these additional data do not follow a panel structure, i.e., household and family members of panelists were not explicitly followed up on. In other words, these family and household rosters reflect the constellations at the time of the interview and should rather be treated as cross-sectional data. The number of observations are provided in Table A1 in the appendix.

3.1 Baseline Sampling Lebanon

The initial survey in Lebanon took place in 2019. The number of respondents was 1,252 (see Table 1). As there are no publicly accessible population registers in Lebanon, respondents were selected using area sampling combined with the random walk method. This is a common procedure for hard-to-reach population samples (Bauer, 2014; Thomson et al., 2020). For this purpose, five hierarchical lists were created on five geographical levels (governorates, zones within the governorates, sub-areas within the zones, clusters within the sub-areas and neighborhoods within the clusters). In order to create a sample that is as representative as possible, information on the demographic and socio-demographic composition of the population (religious affiliation, socio-economic status) as well as the degree of urbanization and geographical characteristics were considered when selecting the sub-areas. From the entirety of the neighborhoods, 200 neighborhoods were then randomly selected as Primary Sampling Units (PSU). In order to achieve the target size of the Syrian sample while remaining

economically viable, only neighborhoods with an above-average proportion of Syrians were included in this selection. For this purpose, data from previous surveys and the field expertise of our implementation partners were used. Within the selected neighborhoods, Syrian and host households were recruited in equal numbers via random walk, i.e., interviewers walked the streets in the selected neighborhood according to predetermined walking rules to select households for participation. Within the households that agreed to participate the choice of the interviewee was also randomized (within age requirements). If the random walk in the respective neighborhoods did not yield sufficient observations, the survey continued in previously defined alternative neighborhoods until the respective quota of Syrian and host respondents was met.

3.2 Baseline Sampling Turkey

The baseline survey in Turkey took place in 2020. The number of respondents was 3,045. As in Lebanon, this survey was stratified according to the country of birth of the head of household, so that 50 percent of the respondents came from Syrian households and 50 percent from non-Syrian households. In Turkey, too, no registry data on the Syrian population in the country is publicly available, so the respondents were also selected via area sampling combined with random walk. Based on publicly available aggregate data on the distribution of the Syrian population, the districts (cf. counties) with the highest and second highest Syrian population shares were selected in each region (cf. states) of Turkey. Within these districts, neighborhoods were again randomly selected as PSUs, and respondents were recruited via random walk (see 2.1).

3.3 Follow-up Strategy

Following the baseline (wave 1) survey, follow-ups were conducted on an annual basis (see Table 1). In up to three attempts, the participants from the baseline surveys were contacted again on the basis of the addresses where the previous interview took place. If the respondent could not be located, the agents tried to arrange an in-person interview appointment via phone given the respondent had provided a telephone number in the previous wave. If this also proved unsuccessful or if the respondent refused to participate in the study again, the contact attempts were cancelled and the respondent's contact details were permanently deleted. To compensate for the dropout of participants, a refreshment sample was recruited, following the target group requirements and baseline sampling strategies described in 2, 2.1 and 2.2.

Table 1. Fieldwork timing and number of observations by country and wave

Country	Wave	Start of Fieldwork	End of Fieldwork	N ¹
Lebanon	1	Sep. 25, 2019	Dez. 03, 2019	1252
Lebanon	2	Nov. 30, 2020	Apr. 13, 2021	2403
Lebanon	3	Dez. 21, 2021	Jan. 24, 2022	1010
Lebanon	4	Oct. 20, 2022	Dez. 07, 2022	2498

Lebanon	5	Oct. 05, 2023	Dez. 08, 2023	2512
Turkey	1	Jan. 12, 2021	Apr. 02, 2021	2925
Turkey	2	Dez. 05, 2021	Jan. 25, 2022	2216
Turkey	3	Nov. 02, 2022	Dez. 20, 2022	2472

Sources: TRANSMIT Surveys Turkey (2020/21, 2021/22, 2022/23) and Lebanon (2019/20, 2020/21, 2021/22, 2022/23, 2023) ¹excluding potential fraud cases, see Section 4.

If the quota of Syrian and non-Syrian households could not be met through interviews within the neighborhoods of the dropped-out participants, additional neighborhoods were selected for the recruitment of participants. The selection of these additional neighborhoods in Turkey and Lebanon was equivalent to the area sampling strategy of the baseline surveys.

3.4 Deviations

Due to unforeseen events, there were instances where the survey deviated from the outlined sampling and follow-up strategies. First, data collections in 2020 and 2021 took place when restrictions due to the COVID-19 pandemic were in place. These included nightly curfews and lockdowns on the weekends. With appropriate hygiene and social distancing protocols fieldwork was still possible, but the restrictions caused major disruptions and fieldwork delays. These delays in turn lead to a funding shortfall in 2021, so that the third wave data collection in Lebanon was limited to a follow-up of 1,000 panelists. Panelists were only re-contacted up to the point that this quota was realized. No comprehensive follow-up nor refreshment-sampling took place.

Second, a series of earthquakes in February 2023 caused wide-spread devastation in the border region between the Syria and Turkey; a region from which roughly half of the TRANSMIT sample was drawn. Consequently, the Turkey fieldwork planned for 2023 was deemed unfeasible under ethical and logistical standpoints. Instead, a limited phone-follow up was conducted to inquire about the well-being of the survey participants. The phone-follow up is not described in this article, see Ruhnke et al. (2024) for details.

4 Survey design

4.1 Interview process

Interviews were conducted using Computer Assisted Personal Interviews (CAPI), based on questionnaires designed by the project team. Wherever possible, the items were aligned with the IAB-BAMF-SOEP Survey of Refugees in Germany to facilitate country comparisons. The interviews were conducted by private, for-profit research institutes, namely by the Research & Consulting House (REACH) in Lebanon and Yöntem Research Consultancy Ltd. (Yöntem) in Turkey. Both institutes are members of the European Society for Opinion and Market Research (ESOMAR) and had previously conducted surveys for international research projects (e.g., the United Nations and the International Organization for Migration).

The interviews in Lebanon were conducted in Arabic, while in Turkey, they were held in either Arabic or Turkish according to the participant's preference. All interviewers were native speakers of the respective language. The questionnaires and participant information were prepared in English by the research team and translated by the contracted survey companies. The companies also checked the questionnaires for coherence and piloted the baseline questionnaire among a pre-selected population.

Potential participants were approached by employees of the implementation partners at their place of residence and, upon explanation of the study's purpose and content, were asked to participate. Upon recruitment, interviewers provided participants with an information sheet containing key information about the study, data protection measures and contact information and confirmed the participant's informed consent. Interviews were conducted in a preferably private location within the respondent's premises, chosen by the respondent. The aim was to create a safe environment that would allow respondents to share private and potentially sensitive information without interference from third persons.

Interviews on average took 55 minutes in Lebanon and 47.6 minutes in Turkey (Table 2). In some cases (e.g., for respondents with a more complex migration or family history), this time was exceeded. To reduce the burden on participants, the questionnaire was shortened starting in 2021/22 (wave 3 for Lebanon and wave 2 for Turkey), lowering median interview lengths. Participants were free to pause the interview at any time, skip individual questions, or terminate the interview. Splitting the interview over several days was not permitted. No financial compensation was offered for participation in the study to avoid creating any coercion to participate.

Table 2. Mean and median interview durations by country and wave

	1	2	3	4	5	Total
Lebanon						
Mean	59,7	63,6	51,6	50,0	50,3	55,0
Median	59,0	63,0	49,0	48,0	48,0	54,0
Turkey						
Mean	59,0	44,5	36,9			47,6
Median	53,0	41,0	34,5			42,5

Sources: TRANSMIT Surveys Turkey (waves 1, 2, and 3) and Lebanon (waves 1, 2, 3, 4, and 5), own calculations.

4.2 Questionnaire Design

The surveys cover a wide range of demographic (age, gender, family structure) and socioeconomic characteristics (education, income, employment, housing). Furthermore, questions about individual migration biographies and intentions, societal participation (networks, discrimination), and (mental) health and wellbeing were included. Additional modules on events such as the Covid-19 pandemic or the Beirut Port explosion were

added on short notice. In addition to the main questionnaire collecting individual information on the respondent, basic demographic information was collected on selected household members and family members outside the household (for details, see 2 and Table A1 in the appendix). If items (including response options) differ over waves, they enter the data set as wave-specific variables with the respective affix in the variable name (i.e., varname_W3 for a wave 3 variable deviating from previous waves). The wave-specific availability and continuity of items is furthermore summarized in the country-specific item overview that is provided upon data access.

4.3 Ethical review

Ethical considerations were an integral part of the survey design process, including ensuring the informed consent of all survey participants. Starting in 2022, all survey waves (Lebanon waves 4 and 5, Turkey wave 3), were approved by the Ethics Commission of the Faculty of Humanities and Social Sciences at Humboldt Universität zu Berlin. Earlier waves which employed near-identical procedures, data protection measures and similar questionnaire designs as the subsequent waves, did not undergo independent ethical review prior to fieldwork as no responsible review body was established at the onset of the project.

5 Data quality controls

The survey data underwent multiple stages of data quality controls before, during and after the data collection. Before going to the field, the scripted questionnaire and the automated data export were comprehensively tested to ensure that all items were correctly programmed and logically structured. Field teams underwent thorough briefing and training to ensure a clear understanding of survey objectives, methodologies, and protocols. During data collection, field supervisors accompanied at least 10 percent of each interviewer's work. To monitor compliance with the sampling plan and to track the progress of the survey, supervisors and interviewers were required to maintain detailed field logs. The locations of all interviews were tracked via GPS signal of the devices (tablets)¹. Date, start and end times of interviews were automatically recorded. These logs were used for subsequent quality checks carried out by both survey companies and TRANSMIT team.

Callbacks were carried out on 20 percent of the collected surveys. This involved re-contacting respondents to verify their existence and the accuracy of responses. Simultaneously, data was systematically checked for consistency and plausibility upon reception by the TRANSMIT Team.

During the standard quality controls in wave 2 of the Turkey survey, the TRANSMIT team observed some irregular patterns in the data. Given the difficulties involved with recruiting respondents (following random route instructions, finding respondents of a hard-to-reach population, and convincing respondents to

¹ In Lebanon, only the GPS location of the initial starting point of the random work was confirmed and recorded.

participate) and the incentives to save time and effort, fraudulent behavior among interviewers or survey companies is not uncommon (Schwanhäuser et al., 2022). Incidents of data fabrication trouble even internationally well-known surveys such as the World Value Survey and the European Social Survey (Blasius and Thiessen, 2012, 2015). The uncommonly high similarity of some interviews for selected interviewers in the Turkey survey, wave 2, led us to systematically check the particular wave and, subsequently, the entire survey data for statistical irregularities. We firstly ran the statistical tests suggested by Schwanhäuser et al. (2022). Their indicators target the identification of fully fabricated interviews. Applied to our survey data, the indicators did not deliver a coherent picture of fraud. Most indicators did not confirm systematically different response patterns for the suspicious interviewers. In addition, the callbacks to respondents did not detect non-existing interviewees. We thus ruled out the assumption that our survey data systematically suffers from full fabrication of interviews and instead tested the option of partial data manipulation by single interviewers.

To quantify the high similarity we occasionally found between interviews, we calculated and examined the maximum matching rates of interviews in duplicate analyses by wave and country both within and between interviewers. The maximum matching rate for each interview denotes the maximum proportion of identical responses with another interview. For seven interviewers in Turkey, wave 2, the analyses confirmed average matching rates of their interviews with their own interviews of 80 percent or higher, responsible for 302 interviews² in total. In other words, their interviews were identical to another interview conducted by the same interviewer on 8 out of 10 items, which is a statistically highly unlikely event (Kuriakose and Robbins, 2016). These interviewers furthermore completed an uncommonly high number of interviews per day: while the average number of completed interviews per day was 5 in the respective wave, this number ranged between 8 and 12 for the identified interviewers. Finally, three out of the seven interviewers had uncommonly straight-lined answers on selected survey instruments. In the affected interviews, answers within item batteries with standardized 5- or 7-scaled answer options (e.g., “strongly agree” to “strongly disagree” in seven steps) showed very low or no variation (referred to as straight-lining or nondifferentiation, Schwanhäuser et al., 2022). In other words, in at least 6 out of 7 Likert-scale item batteries, responses were identical (e.g., always middle response option “neither agree nor disagree”) within each battery. This unlikely scenario supports the assumption of partial fraud with some interviewers skipping through items during the interview and randomly entering answers without reading out the question to save time. Notably, the Syrian sample was overrepresented among the affected cases (see Table A2 in the appendix), underlining the recruitment difficulties for this hard-to-reach population.

We checked two more indicators of fraudulent time-saving behavior by interviewers: firstly, the manipulation of answers such that filters are activated which reduce the amount of follow-up questions (Kosyakova et al.,

² The difference to the 294 reported interviews in Gundacker et al. (2024b) stems from the sample selection in the latter.

2015). Relatedly, when entering information on household and family members, non-reporting some persons saves time. We therefore, secondly, consulted the average number of reported network members per interviewer. For these indicators, no systematic irregularities by interviewer could be observed in wave 2 of the Turkey survey. 96 interviews randomly distributed across interviewers, however, both showed above-average shares of triggered filters and were conducted on a day when the interviewer completed an above-average number of interviews (see Table A3 in the appendix for the applied threshold).

In sum, the duplicate analysis (i.e., the maximum matching rate) remained our strongest predictor of potentially fraudulent behavior and was supported by some, but not all other tested indicators. Based on the analyses and respective literature, we designed a set of indicators for potential fraud and calculated them for the entire panel. With 16.2 and 10.2 percent of the observations characterized by irregularities in wave 1 and 2, respectively, the early data collections in Turkey were disproportionally affected. The expanded and adapted data quality controls for subsequent data collection resulted in an improved data quality in wave 3. Lebanon data collections proved largely unproblematic (Table 3).

To be precise, we flagged all interviews from interviewers who had an average matching rate of 80 percent or higher. In addition, we flagged the interviews of interviewers who showed irregular patterns for two or more additional indicators (see Box 1 for an overview). These cases were deleted from the data set for the Turkish panel, affecting 119 cases in wave 1, 302 cases in wave 2, and 25 cases in wave 3 (Table 3, columns 2 and 3). In the Lebanon data collection, only one wave out of five was affected, and the irregularities were much smaller in size. In wave 2, one interviewer with a total of 55 interviews had an average matching rate above 80 percent. These cases were deleted from the data. Another interviewer running 42 interviews had an average matching rate above 80 percent and additionally showed an uncommonly high deviation from the sample mean of the number of average interviews per day at the same time. Other than these two, Lebanese interviewers did not show widespread irregular behavior. The fieldwork of Lebanon, wave 2, was heavily affected by COVID restrictions (see Table 1 and Section 2.4) including curfews and temporary lockdowns. Consequently, time-windows for completing the fieldwork continued to change, and a wider variance in the interviews per day is to be expected. We thus do not consider this as a stringent indicator of potential fraud for Lebanon, wave 2. For these 42 cases, we instead provide a respective indicator with the data (*dctrl*). We recommend checking analyses for robustness with regard to this indicator but do not expect the cases to bias the results.

We furthermore tested the deviation of single interviews from the ‘statistical normal’ independent of who ran them. We flagged interviews which showed irregular patterns for two or more indicators out of four. Box 1 lays out the identification framework. The respective thresholds for irregular patterns depend on the questionnaire design and the respective sample averages, see Table A3 in the appendix for a detailed overview. As displayed in Table 3, an additional 6.3 percent, 4.2 percent and 6.3 percent of the interviews in waves 1, 2, and 3 were flagged as irregular in Turkey. This was mostly driven by an uncommonly high share of triggered filters and

because they were conducted on a day when the interviewer ran an uncommonly high numbers of interviews per day (see Table A4 in the Appendix). The same pattern was found in Lebanon. An additional 2.6 percent, 7.1

Box 1 Fraud identification framework

Interviewer level indicators

1. *matching rate*: the median matching rate of an interviewer's interviews is 80 percent or higher, i.e., his/her interviews are on average identical on at least 80 percent of interview questions
2. *number of interviews per day*: the average number of an interviewer's completed interviews per day is higher than the sample median plus one median absolute deviation (MAD)
3. *straightlining*: the interviews of an interviewer show straightlining in an uncommonly high share of item batteries. We refer to a straightlined item battery if the answers are identical for all items (e.g. always taking value 1)
4. *network size*: the median reported network size (for Turkey, W1 and W2, and Lebanon W1, W2 and W3) or the median number of reported children (for all subsequent waves) is smaller than the respective sample median minus one MAD

Interview level indicators

1. *number of interviews per day*: interview was completed on a day when an interviewer completed more than the sample median plus one MAD
2. *matching rate*: matching rate with another interview of the same interviewer is 85 percent or higher (Kuriakose/Robinson, 2016)
3. *triggered filters*: the share of triggered filters (relative to the number of possible filters) is more than the sample's median share plus one MAD
4. *straightlining*: in the interview, the responses in an uncommonly high share of batteries take the same value

percent, 7.1 percent, 5.4 percent, and 5.9 percent of the interviews in waves 1, 2, 3, 4, and 5 hit at least two criteria, mostly regarding triggered filters and the interviewer's number of interviews on the interview day (see Table A4 in the Appendix). Again, particularly for the waves affected by the COVID-19 restrictions, is it questionable whether these indicators actually indicate fraudulent behavior, because curfews and lockdowns might have forced interviewers to run a high number of interviews on particular days when fieldwork was possible. We do not drop the cases from the data set but flag them in the respective indicator (*dctrl*) provided with the data. An overview of basic demographics by type of irregularity identified in our data quality controls is provided in the Appendix (Table A5).

Table 3. Interviews affected by irregular patterns, by country, wave, and type of irregularity

Type of irregularity		(1)	(2)	(3)	(4)
		No irregularities	Interviewer median matching rate at least 80%	Interviewer hit at least 2 fraud criteria	Interview hit at least 2 fraud criteria
Wave					
Lebanon					
1	%	97.4			2.6
	N	1219			33
2	%	88.9	2.2	1.7	7.2
	N	2222	55	42	180
3	%	92.9			7.1
	N	938			72
4	%	94.6			5.4
	N	2370			136
5	%	94.1			5.9
	N	2368			149
Turkey					
1	%	89.8	0.7	3.2	6.3
	N	2733	22	97	192
2	%	83.8		12	4.2
	N	2110		302	106
3	%	92.8		1	6.3
	N	2316		25	156

Sources: TRANSMIT Surveys Turkey (waves 1, 2, and 3) and Lebanon (waves 1, 2, 3, 4, and 5), own calculations. For details on affected interviews by strata and panelist status see Table A2 in the Appendix. Thresholds for irregular patterns are specified in Table A3 in the Appendix. For details on affected interviews by indicator see Table A4 in the Appendix.

6 Panel composition and attrition

In times where both Lebanon and Turkey faced major disruption through the COVID-19 pandemic, economic crises and anti-refugee campaigns, among others, panel follow-up presented a major challenge. As can be seen in Table 4, field teams in Lebanon were considerably more successful in re-contacting survey participants than those in Turkey. Basic demographics of panelist and refreshment samples are provided in Table 5. While minor differences in demographic composition exist, their magnitude and changing directionality over waves suggest no systematic bias in demographic characteristics resulting from attrition. Given low follow-up rates particularly in Turkey, we suggest to account for attrition in longitudinal analyses of the presented data (i.e., to apply attrition weights). Note furthermore that the deletion of presumably fraudulent observations as part of the data quality controls described in Section 4 resulted in deviations from the parity of host and Syrian respondents in some waves.

Table 4. Number of observations by wave, strata and panelist status (excluding potential fraud cases)

	Host			Syrian			All		
	Panelist	Total	Follow-up Rate (%)	Panelist	Total	Follow-up Rate (%)	Panelist	Total	Follow-up Rate (%)
Lebanon									
Wave 1	0	550	-	0	702	-	0	1252	-
Wave 2	314	1209	0.57	442	1194	0.63	756	2403	0.60
Wave 3	495	510	0.41	483	500	0.40	978	1010	0.41
Wave 4	613	1249	0.51	664	1249	0.56	1277	2498	0.53
Wave 5	725	1279	0.58	739	1233	0.59	1464	2512	0.59
Turkey									
Wave 1	0	1393	-	0	1532	-	0	2925	-
Wave 2	351	1237	0.25	285	979	0.19	636	2216	0.22
Wave 3	137	1214	0.11	156	1258	0.16	293	2472	0.13

Sources: TRANSMIT Surveys Turkey (waves 1, 2, and 3) and Lebanon (waves 1, 2, 3, 4, and 5), own calculations.

Data access

The TRANSMIT surveys are available to the scientific research community as Scientific Use Files (SUF). To receive access, researchers have to submit a request form to the DeZIM.fdz (Forschungsdatenzentrum des Deutschen Zentrums für Integrations- und Migrationsforschung, DeZIM, Berlin, Germany, URL: <https://fdz.dezim-institut.de/>) and sign a data usage contract. Upon approval, the anonymized data can be accessed via secure download or remote access. To access geo-referenced data, researchers can apply for onsite access at the DeZIM.fdz (see Table 5). For an overview of the anonymization of variables by access type, please refer to Tables A7 and A8 in the appendix. The Lebanon panel (2019-2023) and the Turkey panel (2020-2022) are provided as two separate data files. The household rosters, family rosters, child rosters and household information packages are provided as separate files for each wave and country and can be linked to the main respondent panels via unique household identifiers. Additional information on the data sets is provided in Gundacker et al. (2025). More information on the TRANSMIT project can be found on the project homepage (<https://www.projekte.hu-berlin.de/en/transmit>).

Table 5. Provision of household geo information by data access type and country

	Turkey	Lebanon
Download	none	none
Remote	Anonymised PSU identifier, province	Anonymised PSU identifier, governorate
Onsite	Province, District, PSU (neighborhood), spatial midpoint (latitude, longitude) of households in each neighborhood, relocated by a random distance between 1 and 2 km in a random direction if number of households in the same neighborhood is less than 5 over all waves. Missing values in latitude and longitude indicate missing or implausible data.	Governorate, PSU (neighborhood), latitude and longitude of starting point of random walk per neighborhood. Missing values in latitude and longitude indicate missing or implausible data.

Citation of data

A short version of the data manual was published as Gundacker, Lidwina, Laura Hertner, Nora Kühnert, Ramona Rischke, Simon Ruhnke, Nader Talebi, Caroline Trocka & Herbert Brücker (2025), Transnational Perspectives on Migration and Integration - TRANSMIT Surveys of Syrian nationals and their neighbors in Lebanon and Turkey (2019-2023), Journal of Economics and Statistics (2025), DOI: 10.1515/jbnst-2024-0091. It serves as the main publication to accompany the data. Please additionally cite the respective data file:

Access Type Download: Gundacker, Lidwina, Laura Hertner, Nora Kühnert, Ramona Rischke, Simon Ruhnke, Nader Talebi, Caroline Trocka & Herbert Brücker (2025). TRANSMIT Surveys of Syrian nationals and their neighbors in Lebanon and Turkey (2019-2023). Dataset. Version: 1.0.0. Berlin: Deutsches Zentrum für Integrations- und Migrationsforschung (DeZIM). <https://doi.org/10.34882/dezim.transmit1.download.1.0.0>

Access Type Remote: Gundacker, Lidwina, Laura Hertner, Nora Kühnert, Ramona Rischke, Simon Ruhnke, Nader Talebi, Caroline Trocka & Herbert Brücker (2025). TRANSMIT Surveys of Syrian nationals and their neighbors in Lebanon and Turkey (2019-2023). Dataset. Version: 1.0.0. Berlin: Deutsches Zentrum für Integrations- und Migrationsforschung (DeZIM). <https://doi.org/10.34882/dezim.transmit1.remote.1.0.0>

Access Type Onsite: Gundacker, Lidwina, Laura Hertner, Nora Kühnert, Ramona Rischke, Simon Ruhnke, Nader Talebi, Caroline Trocka & Herbert Brücker (2025). TRANSMIT Surveys of Syrian nationals and their neighbors in Lebanon and Turkey (2019-2023). Dataset. Version: 1.0.0. Berlin: Deutsches Zentrum für Integrations- und Migrationsforschung (DeZIM). <https://doi.org/10.34882/dezim.transmit1.onsite.1.0.0>

When referring to details laid out in this Data Manual, please cite:

Gundacker, Lidwina, Laura Hertner, Nora Kühnert, Ramona Rischke, Simon Ruhnke, Nader Talebi, Caroline Trocka & Herbert Brücker (2025). TRANSMIT Surveys of Syrian nationals and their neighbors in Lebanon and Turkey (2019-2023) Data Manual, Version 1.0.0. Deutsches Zentrum für Integrations- und Migrationsforschung (DeZIM). Berlin.

Funding

Funding for development and implementation of the panel study was provided by the German Federal Ministry for Family Affairs, Senior Citizens, Women and Youth (grant no. 3920405WZB).

Acknowledgments

We gratefully acknowledge the willingness of all survey participants to share their information with us. We furthermore thank our local cooperation partners for their efforts. Our thanks also goes to Lukas Olbrich, Silvia Schwanhäuser, and Simon Wagner who generously shared their expertise on survey quality controls with us. Finally, the support of our student assistants Felix Rahberger, Paula Niemöller, Fadi Wahbi, Vico Kutz, Büşra Lütfüoğlu, and Lukas Suttner has been tremendously helpful.

Publications based on data

Gundacker, Lidwina, Sekou Keita, & Simon Ruhnke (2024a). Unequal access to protection? Selection patterns over arrival cohorts of Syrians seeking refuge in Lebanon, Turkey, and Germany. *Frontiers in Human Dynamics*, 5, DOI: <https://dx.doi.org/10.3389/fhumd.2023.1171885>.

Gundacker, Lidwina, Laura Hertner, & Simon Ruhnke (2024b). Six years after the EU-Turkey Agreement: A quantitative assessment of the living conditions of Syrians in Turkey. *IAB-Forschungsbericht* 18/2024, DOI:10.48720/IAB.FB.2418.

Rischke, Ramona & Nader Talebi (2021). Lebanon at a critical juncture—Perspectives of Syrians and Lebanese in Lebanon 2019-2021 (SSRN Scholarly Paper 3848223). Social Science Research Network. <https://doi.org/10.2139/ssrn.3848223>

Ruhnke, Simon (2021), The EU-Turkey Deal as a successful Blueprint? New Data on the Wellbeing of Syrian Refugees in Turkey, MERGE Data Brief Series #1, <https://www.projekte.hu-berlin.de/en/merge/publications/data-brief-the-eu-turkey-deal-as-a-successful-blueprint>

Ruhnke, Simon, Laura Hertner, Lidwina Gundacker & Simon Wagner (2024). Going from bad to worse? Well-being of Syrian refugees in Turkey in the aftermath of the February 2023 earthquakes (BIM News). Humboldt-Universität zu Berlin. <https://doi.org/10.18452/28152>

Ruhnke, Simon, Laura Hertner, Judith Köhler & Ulrike Kluge (2024). Social ecological determinants of the mental distress among Syrian refugees in Lebanon and Turkey: A transnational perspective. *Social Science & Medicine*, 116700. <https://doi.org/10.1016/j.socscimed.2024.116700>

Ruhnke, Simon & Ramona Rischke (2024). Predicting mobility aspirations in Lebanon and Turkey: A data-driven exploration using machine learning. *Data & Policy*, 6, e47. <https://doi.org/10.1017/dap.2024.32>

Ruhnke, Simon, Michele Scala, Fadi Wahbi & Nader Talebi (2023). Healthcare Access Crisis in Lebanon, MERGE Data Brief Series #3, https://www.projekte.hu-berlin.de/en/merge/publications/data-brief_healthcare_access_crisis_in_lebanon.

Ruhnke, Simon & Nader Talebi (2022), Food insecurity in Lebanon and the potential ripple effects of the war in Ukraine, MERGE Data Brief Series #2, <https://www.projekte.hu-berlin.de/en/merge/publications/food-insecurity-in-lebanon-and-the-potential-ripple-effects-of-the-war-in-ukraine>

References

Bauer, Johannes J. (2014). Selection Errors of Random Route Samples, *Sociological Methods and Research* 43(3), 519–544. DOI: <https://doi.org/10.1177/0049124114521150>.

- Blasius, Jörg & Lukas Sausen (2023). Detecting Fabricated Interviews Using the Hamming Distance. *Survey Research Methods*, 17(2), 131-145. DOI: <https://doi.org/10.18148/srm/2023.v17i2.7961>.
- Blasius, Jörg & Victor Thiessen (2012). *Assessing the quality of survey data*. London: Sage.
- Blasius, Jörg & Victor Thiessen (2015). Should we trust survey data? Assessing response simplification and data fabrication. *Social Science Research* 52, 479-493. DOI: <https://doi.org/10.1016/j.ssresearch.2015.03.006>.
- Brücker Herbert, Nina Rother & Jürgen Schupp (2017). IAB-BAMF-SOEP-Befragung von Geflüchteten 2016: Studiendesign, Feldergebnisse sowie Analysen zu schulischer wie beruflicher Qualifikation, Sprachkenntnissen sowie kognitiven Potenzialen. IAB-Forschungsbericht 14/2016.
- Collins, Francis L. (2021). 'Give me my pathway!': multinational migration, transnational skills regimes and migrant subjectification. *Global Networks* 21(1), 18–39. DOI: <https://doi.org/10.1111/glob.12294>.
- Gundacker, Lidwina, Sekou Keita & Simon A. Ruhnke (2024a). Unequal access to protection? Selection patterns over arrival cohorts of Syrians seeking refuge in Lebanon, Turkey, and Germany. *Frontiers in Human Dynamics*, 5, DOI: <https://dx.doi.org/10.3389/fhumd.2023.1171885>.
- Gundacker, Lidwina, Laura Hertner & Simon A. Ruhnke (2024b). Six years after the EU-Turkey Agreement: A quantitative assessment of the living conditions of Syrians in Turkey. IAB-Forschungsbericht 18/2024, DOI:10.48720/IAB.FB.2418.
- Jung, Philipp R. (2023). Multinational Migration in the Global South: Complex and Non-linear Trajectories of Senegalese Migrants in Brazil. In Jill Ahrens & Russell King (eds.), *Onward Migration and Multi-Sited Transnationalism: Complex Trajectories, Practices and Ties*, 159–178. Cham: Springer.
- Kosyakova, Yuliya, Jan Skopek & Stephanie Eckman (2015). Do Interviewers Manipulate Responses to Filter Questions? Evidence from a Multilevel Approach. *International Journal of Public Opinion Research* 27(3), 417–431. DOI: <https://dx.doi.org/10.1093/ijpor/edu027>.
- Kuriakose, Noble & Michael Robbins (2016). Don't get duped: Fraud through duplication in public opinion surveys. *Statistical Journal of the IAOS* 32(3), 283-291. DOI: <https://doi.org/10.3233/sji-160978>.
- Ruhnke, Simon A., Laura Hertner, Lidwina Gundacker & Simon Wagner (2024). Going from bad to worse? Well-being of Syrian refugees in Turkey in the aftermath of the February 2023 earthquakes. Humboldt-Universität zu Berlin, BIM News 1. DOI: <https://doi.org/10.18452/28152>.
- Robbins, Michael (2019). New Frontiers in Detecting Data Fabrication. In Timothy P. Johnson, Beth-Ellen Pennell, Ineke A.L. Stoop & Brita Dorer (eds.), *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC)*, 771–805. Hoboken: John Wiley & Sons, Inc.

Schwanhäuser, Silvia, Joe W. Sakshaug & Yuliya Kosyakova (2022). How to catch a falsifier: Comparison of statistical detection methods for interviewer falsification. *Public opinion Quarterly* 86(1), 51-81. <https://doi.org/10.1093/poq/nfab066>.

Thomson, Dana R., Dale A. Rhoda, Andrew J. Tatem & Marcia C. Castro (2020): Gridded Population Survey Sampling: A Systematic Scoping Review of The Field and Strategic Research Agenda. *International Journal of Health Geographics* 19(34), 1-16. DOI: <https://doi.org/10.1186/s12942-020-00230-4>.

UNHCR. (2024). United Nations High Commissioner for Refugees (UNHCR) Refugee Data Finder. URL: <https://www.unhcr.org/refugee-statistics/download/?v2url=dbd9e3>, last accessed 19 Nov 2024.

Appendix

Table A1. Rosters on family and household members, number of observations by country, wave, and strata

Lebanon	Wave		Syrian	Host
	1	Household roster	2667	1479
	1	Family roster	3316	2130
	2	Household roster	4611	3353
	2	Family roster	5162	3678
	3	Household roster	2010	1439
	3	Family roster	2453	1699
	4	Child roster	1249	1026
	4	Household info	1249	1249
	5	Child roster	1330	960
	5	Household info	1233	1279
Turkey	Wave		Syrian	Host
	1	Household roster	4474	3214
	1	Family roster	2158	3083
	2	Household roster	2917	2557
	2	Family roster	759	1421
	3	Child roster	2708	1348
	3	Household info	1258	1214

Sources: TRANSMIT Surveys Turkey (waves 1, 2, and 3) and Lebanon (waves 1, 2, 3, 4, and 5), household rosters, family rosters, child rosters and household information rosters, own calculations.

Table A2. Share of interviews affected by irregular patterns, by country, wave, strata, and panelist status

	Wave	Strata	Panelist status	% of sample affected
Lebanon	1	Syrian	baseline	2.3
		Host	baseline	3.1
	2	Syrian	panelist	10.8
		Syrian	refreshment	11.2
		Host	panelist	10.9
		Host	refreshment	12.2
	3	Syrian	panelist	8.4
		Host	panelist	5.9
	4	Syrian	panelist	3.4
		Syrian	refreshment	5.2
		Host	panelist	8.3
		Host	refreshment	3.1
	5	Syrian	panelist	6.2
		Syrian	refreshment	6.1
		Host	panelist	5.3
		Host	refreshment	6.2
Turkey	1	Syrian	baseline	7.1
		Host	baseline	13.4
	2	Syrian	panelist	24.1
		Syrian	refreshment	29.1
		Host	panelist	4.7
		Host	refreshment	5.1
	3	Syrian	panelist	3.7
		Syrian	refreshment	5.1
		Host	panelist	3.7
		Host	refreshment	10.4

Sources: TRANSMIT Surveys Turkey (waves 1, 2, and 3) and Lebanon (waves 1, 2, 3, 4, and 5), own calculations. *Notes:* See Box 1 and Section 4 for deletion rules. See Table A4 for the affected interviews by country, wave, and indicator.

Table A3. Thresholds for irregular patterns by country, wave, indicator, and sample (if applicable). Indicator hits if...

Lebanon	Wave	1	2	3	4	5
Average matching rate of interviewers at or above... (in %)		80	80	80	80	80
Matching rate of single interview at or above... (in %)		85	85	85	85	85
Number of interviews per day at or above...		7	6	7	8	7
Straight-lining: number of straight-lined item batteries at or above...		5 out of 5	6 out of 7	3 out of 3	3 out of 3	3 out of 3
Network size (number of recorded family members) below...		4	2	3		
Network size (number of reported children) below... ¹					2	2
Share of triggered filters is at or above...						
<i>Host sample, baseline sample or refreshment</i>		76.5	74.4		80.2	80.7
<i>Syrian sample, baseline sample or refreshment</i>		73.6	68.2		75.3	76.3
<i>Host sample, panelist</i>			74.2	74.7	76.5	73.1
<i>Syrian sample, panelist</i>			67.7	62.7	73.4	67.5
Turkey	Wave	1	2	3		
Average matching rate of interviewers at or above... (in %)		80	80	80		
Matching rate of single interview at or above... (in %)		85	85	85		
Number of interviews per day at or above...		8	8	9		
Straight-lining: number of straight-lined item batteries at or above...		6 out of 7	6 out of 7	4 out of 4		
Network size (number of recorded family members) below...		1	1			
Network size (number of reported children) below... ¹				2		
Share of triggered filters is at or above...						
<i>Host sample, baseline sample or refreshment</i>		80.4	80.5	83.3		
<i>Syrian sample, baseline sample or refreshment</i>		66.1	75.6	70.3		
<i>Host sample, panelist</i>			81.3	68.8		
<i>Syrian sample, panelist</i>			76.2	68.8		

Sources: TRANSMIT Surveys Turkey (waves 1, 2, and 3) and Lebanon (waves 1, 2, 3, 4, and 5), own calculations. *Notes:*¹ Given a change in the questionnaire design in 2021 (see 3.2), we consult the number of reported children from Lebanon, wave 4, and Turkey, wave 3, onwards. See Box 1 and Section 4 for deletion rules.

Table A4. Share of interviews affected by irregular patterns, by country, wave, and indicator

Lebanon										
wave	1		2		3		4		5	
	%	N	%	N	%	N	%	N	%	N
<i>Interviews identified on interviewer level:</i>										
Matching rate	0	0	3.9	97	9	0.9	0	0	0.6	14
Number of interviews per day	1.1	14	5.7	143	3.3	33	2.9	72	1.4	34
Straight-lining	0	0	0	0	0	0	0	0	0	0
Network size	0	0	0.2	4	0.3	3	0	0	0	0
<i>Interviews additionally identified on interview level:</i>										
Number of interviews per day	1.5	19	3.2	79	3.5	35	2.3	58	3.5	88
Matching rate	0	0	0.3	7	0.9	9	0	0	0.6	14
Triggered filters	2.6	33	6.9	173	6.9	70	5.4	134	5.2	132
Straight-lining	0	0	0.4	9	0.2	2	0.4	9	1.5	38
Turkey										
wave	1		2		3		4		5	
	%	N	%	N	%	N	%	N	%	N
<i>Interviews identified on interviewer level:</i>										
Matching rate	3.9	119	12	302	0	0				
Number of interviews per day	2.1	64	15.3	386	3.5	88				
Straight-lining	0.7	22	6.1	154	0	0				
Network size	0.3	9	0	0	1.1	27				
<i>Interviews additionally identified on interview level:</i>										
Number of interviews per day	4.7	144	20	0.8	3.3	82				
Matching rate	0.3	9	0.4	9	0	0				
Triggered filters	6.1	185	3.9	97	6.1	153				
Straight-lining	0.4	11	0.8	19	0.1	14				

Sources: TRANSMIT Surveys Turkey (waves 1, 2, and 3) and Lebanon (waves 1, 2, 3, 4, and 5), own calculations. *Notes:* see Box 1 and Section 4 for deletion rules.

Table A5. Demographic Characteristics by country, wave and type of irregularity

(1) no irregularity, (2) interviewer hit 2+ fraud criteria, (3) interviewer median matching rate >= 80% (4) interview hit 2+ fraud criteria

Lebanon

Characteristic	Wave 1				Wave 2				Wave 3			
	(1) N = 1,219 ¹	(2) N = 0 ¹	(3) N = 0 ¹	(4) N = 33 ¹	(1) N = 2,222 ¹	(2) N = 42 ¹	(3) N = 55 ¹	(4) N = 180 ¹	(1) N = 938 ¹	(2) N = 0 ¹	(3) N = 0 ¹	(4) N = 72 ¹
Age	33 (25, 44)			43 (21, 50)	33 (26, 44)	34 (26, 37)	33 (25, 41)	33 (20, 47)	34 (26, 44)			35 (29, 46)
Gender												
Male	620 (51%)			6 (18%)	1,153 (52%)	26 (62%)	20 (36%)	54 (30%)	468 (50%)			20 (28%)
Female	599 (49%)			27 (82%)	1,070 (48%)	16 (38%)	35 (64%)	126 (70%)	470 (50%)			52 (72%)
Educational Attainment												
Never attended	132 (11%)			12 (36%)	197 (8.9%)	12 (29%)	4 (7.3%)	20 (11%)	59 (6.3%)			11 (15%)
Less than high school	765 (63%)			13 (39%)	1,310 (59%)	24 (57%)	30 (55%)	110 (61%)	544 (58%)			44 (61%)
High school certificate	167 (14%)			4 (12%)	351 (16%)	2 (4.8%)	9 (16%)	33 (18%)	161 (17%)			13 (18%)
Attended university	33 (2.7%)			3 (9.1%)	92 (4.1%)	0 (0%)	1 (1.8%)	6 (3.3%)	44 (4.7%)			1 (1.4%)
Bachelors or higher	121 (9.9%)			1 (3.0%)	273 (12%)	4 (9.5%)	11 (20%)	11 (6.1%)	130 (14%)			3 (4.2%)
Strata												
Syrian	686 (56%)			16 (48%)	1,115 (50%)	32 (76%)	25 (45%)	79 (44%)	458 (49%)			42 (58%)
Host	533 (44%)			17 (52%)	1,108 (50%)	10 (24%)	30 (55%)	101 (56%)	480 (51%)			30 (42%)

Sources: TRANSMIT Surveys Turkey (waves 1, 2, and 3) and Lebanon (waves 1, 2, 3, 4, and 5), own calculations. Notes: ¹Median (IQR); n (%), “other” responses removed. See Box 1 and Section 4 for deletion rules.

Lebanon (Table A5 continued)

Characteristic	Wave 4				Wave 5			
	(1) N = 2,370 ¹	(2) N = 0 ¹	(3) N = 0 ¹	(4) N = 136 ¹	(1) N = 2,368 ¹	(2) N = 0 ¹	(3) N = 0 ¹	(4) N = 149 ¹
Age	35 (26, 45)				36 (27, 46)			
Gender								
Male	1,177 (50%)				1,097 (46%)			
Female	1,193 (50%)				1,270 (54%)			
Educational Attainment								
Never attended	208 (8.8%)				211 (8.9%)			
Less than high school	1,418 (60%)				1,437 (61%)			
High school certificate	363 (15%)				340 (14%)			
Attended university	101 (4.3%)				102 (4.3%)			
Bachelors or higher	280 (12%)				278 (12%)			
Strata								
Syrian	1,206 (51%)				1,159 (49%)			
Host	1,164 (49%)				1,209 (51%)			

Sources: TRANSMIT Surveys Turkey (waves 1, 2, and 3) and Lebanon (waves 1, 2, 3, 4, and 5), own calculations. *Notes:* ¹Median (IQR); n (%), “other” responses removed. See Box 1 and Section 4 for deletion rules.

Turkey (Table A5 continued)

Characteristic	Wave 1				Wave 2				Wave 3			
	(1) N = 2,733 ¹	(2) N = 22 ¹	(3) N = 97 ¹	(4) N = 192 ¹	(1) N = 2,110 ¹	(2) N = 302 ¹	(3) N = 0 ¹	(4) N = 106 ¹	(1) N = 2,316 ¹	(2) N = 0 ¹	(3) N = 25 ¹	(4) N = 156 ¹
Age	33 (25, 41)	35 (29, 42)	24 (19, 36)	38 (29, 48)	31 (25, 40)	31 (25, 37)		45 (28, 56)	35 (28, 42)		27 (24, 30)	42 (32, 52)
Gender												
Male	1,407 (51%)	15 (68%)	42 (43%)	79 (41%)	1,067 (51%)	184 (61%)		33 (31%)	1,122 (48%)		15 (60%)	22 (14%)
Female	1,324 (48%)	7 (32%)	55 (57%)	113 (59%)	1,042 (49%)	118 (39%)		73 (69%)	1,194 (52%)		10 (40%)	134 (86%)
Educational Attainment												
Never attended	295 (11%)	0 (0%)	0 (0%)	107 (56%)	197 (9.3%)	35 (12%)		30 (28%)	179 (7.7%)		0 (0%)	77 (49%)
Less than high school	1,270 (46%)	3 (14%)	23 (24%)	36 (19%)	1,171 (55%)	196 (65%)		55 (52%)	1,281 (55%)		7 (28%)	73 (47%)
High school certificate	751 (27%)	19 (86%)	56 (58%)	40 (21%)	512 (24%)	39 (13%)		13 (12%)	563 (24%)		14 (56%)	3 (1.9%)
Attended university	112 (4.1%)	0 (0%)	4 (4.1%)	3 (1.6%)	57 (2.7%)	3 (1.0%)		3 (2.8%)	65 (2.8%)		0 (0%)	1 (0.6%)
Bachelors or higher	304 (11%)	0 (0%)	14 (14%)	6 (3.1%)	173 (8.2%)	29 (9.6%)		5 (4.7%)	228 (9.8%)		4 (16%)	2 (1.3%)
Strata												
Syrian	1,423 (52%)	0 (0%)	0 (0%)	109 (57%)	934 (44%)	302 (100%)		45 (42%)	1,197 (52%)		0 (0%)	61 (39%)
Host	1,310 (48%)	22 (100%)	97 (100%)	83 (43%)	1,176 (56%)	0 (0%)		61 (58%)	1,119 (48%)		25 (100%)	95 (61%)

Sources: TRANSMIT Surveys Turkey (waves 1, 2, and 3) and Lebanon (waves 1, 2, 3, 4, and 5), own calculations. Notes: ¹Median (IQR); n (%), “other” responses removed. See Box 1 and Section 4 for deletion rules.

Table A6. Age, gender and educational attainment by wave and panelist status (excluding potential fraud cases²)

Characteristic	Wave 1	Wave 2	Wave 3		Wave 4		Wave 5		
	Refreshment, N = 1,252 ¹	Refreshment, N = 1,647 ¹	Panelist, N = 756 ¹	Refreshment, N = 0	Panelist, N = 1,010 ¹	Refreshment, N = 1,221 ¹	Panelist, N = 1,277 ¹	Refreshment, N = 1,048 ¹	Panelist, N = 1,464 ¹
Lebanon									
Age	33 (25, 44)	33 (25, 44)	34 (27, 44)		34 (26, 45)	35 (25, 46)	35 (27, 45)	36 (26, 47)	36 (27, 46)
Gender									
Male	626 (50%)	836 (51%)	371 (49%)		488 (48%)	562 (46%)	644 (50%)	452 (43%)	682 (47%)
Female	626 (50%)	811 (49%)	385 (51%)		522 (52%)	659 (54%)	633 (50%)	596 (57%)	781 (53%)
Educational Attainment									
Never attended	144 (12%)	129 (7.8%)	88 (12%)		69 (6.9%)	131 (11%)	101 (7.9%)	114 (11%)	125 (8.5%)
Less than high school	778 (62%)	938 (57%)	482 (64%)		588 (58%)	726 (59%)	762 (60%)	645 (62%)	877 (60%)
High school certificate	171 (14%)	285 (17%)	99 (13%)		174 (17%)	199 (16%)	188 (15%)	133 (13%)	227 (16%)
Attended university	36 (2.9%)	77 (4.7%)	21 (2.8%)		45 (4.6%)	34 (2.8%)	69 (5.4%)	37 (3.5%)	71 (4.8%)
Bachelors or higher	122 (9.8%)	218 (13%)	66 (8.7%)		133 (13%)	131 (11%)	157 (12%)	119 (11%)	164 (11%)
Turkey									
	Refreshment, N = 2,925 ¹	Refreshment, N = 1,580 ¹	Panelist, N = 636 ¹	Refreshment, N = 2,179 ¹	Panelist, N = 293 ¹				
Age	33 (25, 41)	31 (25, 40)	33 (26, 42)	35 (28, 43)	33 (27, 40)				
Gender									
Male	1,486 (51%)	753 (48%)	347 (55%)	990 (45%)	154 (53%)				
Female	1,437 (49%)	826 (52%)	289 (45%)	1,189 (55%)	139 (47%)				
Educational Attainment									
Never attended	402 (14%)	157 (9.9%)	70 (11%)	240 (11%)	16 (5.5%)				
Less than high school	1,306 (45%)	916 (58%)	310 (49%)	1,204 (55%)	150 (51%)				
High school certificate	791 (27%)	358 (23%)	167 (26%)	485 (22%)	81 (28%)				
Attended university	115 (3.9%)	36 (2.3%)	24 (3.8%)	57 (2.6%)	9 (3.1%)				
Bachelors or higher	310 (11%)	113 (7.2%)	65 (10%)	193 (8.9%)	37 (13%)				

Notes: ¹Median (IQR); n (%), “other” responses removed. ²See Box 1 and Section 4 for deletion rules. Sources: TRANSMIT Surveys Turkey (waves 1, 2, and 3) and Lebanon (waves 1, 2, 3, 4, and 5), own calculations.

Table A7. TRANSMIT Surveys Lebanon: Anonymization of variables by data access type

	variables	Download	Remote	Onsite
Locational information (province)	origin_p_LEB origin_p_Other displ_from_x1_p_LEB displ_from_x2_p_LEB prov_prev prov_prev_Other	no release	no release	release
Open response categories	mig_dest_Other mig_dest_Other_1 mig_dest_Other_2 mig_dest_Other_3 mig_dest_panel_Other own_dwell_Other hhinc_Other whynschool_Other edu_school_Other edu_ctype_Other outmig_ctype_Other1 outmig_ctype_Other2 prof_cert_country_Other ethnicity_Other national_idnt_Other lang_Other_1_LEB lang_Other_2_LEB lang_Other_3_LEB lang_Other_Other displ_why_x1_Other_LEB displ_why_x2_Other_LEB migleb_why_Other migleb_why_other mess_friend_Other mess_migr_Other migcontact_dest_Other_1_LEB migcontact_dest_Other_2_LEB migcontact_dest_Other_3_LEB retcontact_dest_Other1_LEB retcontact_dest_Other2_LEB plans_ctype_Other plans_reas_Other stay_reas_Other mig_desire_Other_1 mig_desire_Other_2	no release	no release	release
Variables that contain country information	citizenship_2 mig_dest_x1_LEB mig_dest_x1_W4W5_LEB mig_dest_x2_LEB mig_dest_x2_W4W5_LEB mig_dest_x3_LEB mig_dest_x3_W4W5_LEB mig_dest_x4_LEB mig_dest_x4_W4W5_LEB mig_dest_panel_1_LEB mig_dest_panel_2_LEB mig_dest_panel_3_LEB mig_dest_panel_4_LEB own_dwell_c1 own_land_c2_LEB own_land_c1_LEB own_land_c2_LEB outmig_ctype_1 outmig_ctype_2 outmig_ctype_3 outmig_ctype_4 outmig_ctype_5 prof_cert_country_LEB intro_1 intro_2 intro_3 displ_from_x1_c displ_from_x1_c_W1 displ_from_x2_c_W1 migcontact_dest_1_LEB migcontact_dest_2_LEB migcontact_dest_3_LEB migcontact_dest_4_LEB migcontact_dest_5_LEB retcontact_dest_1_LEB retcontact_dest_2_LEB retcontact_dest_3_LEB retcontact_dest_4_LEB retcontact_dest_5_LEB	no release	release	release
Ethnicity/Natioanality	ethnicity_1_LEB ethnicity_2_LEB ethnicdiv national_idnt_LEB nationaldiv	no release	release	release
Year information	mig_time_1 mig_time_2 mig_time_3 mig_time_4 displ_when_x2_y	no release	release	release
Additional survey information	iid datestart timestart timeend duration	no release	release	release
Other sensitive information	occ occ_last occ_last_host occ_last_syr occ_last_syria occ_loc occ_syr profession married_y2 married_y3 married_y4 married_y5 married_lag_1-married_lag_11 mig_return_y2 mig_return_y3 mig_return_y4 label_Other_2_LEB label_Other_3_LEB	no release	release	release
Aggregations	hhsz cowives_f sons cowives_m_LEB_W1W2 daughters years_univ selfempl workexp_prof_LEB workexp_gen_LEB age origin_c citizenship_1 entry_y ysm edu_year edu_ctype_LEB years_school married_y1 workstart religion displ_when_x1_y mig_return_y1 resid_time resid_time_prov	Release with aggregations	Release without aggregations	Release without aggregations

Table A8. TRANSMIT Surveys Turkey: Anonymization of variables by data access type

	variables	Download	Remote	Onsite
Locational information (province)	province origin_p_Other origin_p_W1_TK origin_p_W2_TK origin_p_W3_TK displ_from_x1_p_Other displ_from_x1_p_TK displ_from_x2_p_Other displ_from_x2_p_TK prov_prev prov_prev_Other	no release	no release	release
Open response categories	hhinc_Other whynoschool_Other edu_univ_cert_Other crises_Other_TK ethnicity_Other religion_Other lang_Other_Other displ_why_x1_Other displ_why_x2_Other migleb_why_Other migleb_mode_Other discrleb_hier_Other discrleb_hier_Other_nat discr_govwhy_Other discr_jobwhy_Other discr_policewhy_Other plans_ctry_Other plans_reas_Other stay_reas_Other course_priv_Other chroncon_Other chroncon_hh_Other	no release	no release	release
Variables that contain country information	citizenship_2 citizenship_3 mig_dest_x1_TK mig_dest_x2_TK mig_dest_panel_TK own_land_c1_TK own_land_c2_TK edu_ctry_TK prof_cert_country_TK lang_Other_1_TK lang_Other_2_TK lang_Other_3_TK displ_from_x1_c displ_from_x2_c migcontact_dest_1_W1_TK migcontact_dest_1_W2_TK migcontact_dest_1_W3_TK migcontact_dest_2_W1_TK migcontact_dest_2_W2_TK migcontact_dest_2_W3_TK migcontact_dest_3_W1_TK migcontact_dest_3_W2_TK migcontact_dest_3_W3_TK migcontact_dest_4_W1_TK migcontact_dest_4_W2_TK migcontact_dest_4_W3_TK migcontact_dest_5_W1_TK migcontact_dest_5_W3_TK retcontact_dest_1_TK retcontact_dest_2_TK retcontact_dest_3_TK retcontact_dest_4_TK retcontact_dest_5_TK mig_desire_TK mig_labor_rand_ctry	no release	release	release
Ethnicity/Natioanality	ethnicity_1_TK ethnicity_2_TK ethnicity_3_TK ethnicity_4_TK ethnicity_5_TK ethnicity_6_TK national_ident_TK nationaldiv	no release	release	release
Year information	mig_time_1 mig_time_2 mig_return_y2 migleb_lastentry_y	no release	release	release
Additional survey information	iid datestart timestart timeend	no release	release	release
Other sensitive information	occ occ_last occ_last_host occ_last_syr occ_last_syria occ_loc occ_syr profession married_lag_0-married_lag_10	no release	release	release
Aggregations	hhsize cowives_f sons cowives_m_TK daughters years_univ selfempl workexp_prof_TK workexp_gen_TK age origin_c citizenship_1 entry_y ysm edu_year years_school married_y workstart religion	Release with aggregations	Release without aggregations	Release without aggregations