Appendix

The Dynamics of Refugee Return:

Syrian Refugees and Their Migration Intentions*

Ala Alrababah[1,2], Daniel Masterson[2,3], Marine Casalis[2],
Dominik Hangartner[1,2,4], Jeremy Weinstein[2,5]

[1]Center for International and Comparative Studies, ETH Zurich
[2]Immigration Policy Lab, Stanford University and ETH Zurich
[3]Department of Political Science, University of California, Santa Barbara
[4]Department of Government, London School of Economics and Political Science
[5]Department of Political Science, Stanford University

October 31, 2022

Contents

1 Sampling Strategy **1**

2 Survey questions **3**

2.1 Dependent variables 3

2.2 Independent variables 4

2.2.1 Push factors from Lebanon 4

2.2.2 Pull factors in Syria 5

2.2.3 Confidence in information and mobility 7

2.3 Control variables 8

3 Index construction and descriptives **9**

3.1 Push factors from Lebanon 10

3.2 Pull factors in Syria 11

*All analyses, unless otherwise noted, were pre-specified in EGAP registration 20190914AB. This project was reviewed and approved by the Stanford Human Subjects Committee under IRB protocol 49387 and by the Ethics Committee of ETH Zurich and the Human Subjects Committee of the University of California, Santa Barbara.

3.3 Confidence in information . . . 13
3.4 Mobility . . . 13
3.5 Preparation to return . . . 13
3.6 Covariates . . . 13

4 Scree plots for principal component analysis 14
4.1 Push factors from Lebanon . . . 14
4.2 Pull factors in Syria . . . 17
4.3 Confidence in information . . . 19
4.4 Preparation for return . . . 20

5 Deviations from PAP 20

6 Robustness and Additional Tests 21
6.1 Return in two years and ever . . . 22
6.2 Safety, services, and additional covariates . . . 22
6.3 Excluding locality fixed effects . . . 23
6.4 Return plans for household members and uncertainty about return . . . 23
6.5 Additive indices . . . 24
6.6 Information and return intentions . . . 25
6.7 Testing the models' predictive power . . . 27
6.8 Map of conditions in Lebanon . . . 29
6.9 Conjoint results . . . 30

7 Jordan survey 30

8 Ethical Considerations: Survey 34

9 Qualitative Interviews: Details and Ethical Considerations 37

1 Sampling Strategy

We conducted stratified multistage sampling. In the first stage, we selected localities based on two dimensions: the prevalence of Syrians and the majority sect. We drew Syrian population data from UNHCR registration records and Lebanese population data from voter registration records, since there is no current census available. In the second stage, we used a random walk procedure to select households within each locality. In the third stage, we selected a head of household—defined as a person regardless of gender who plays a large role in household decision-making.

For this study, we wanted to draw inferences about all individuals from Syria in Lebanon as of the study's start date irrespective of their legal status and access to services and support. Thus, this study includes Syrian citizens regardless of whether they registered with UNHCR, and also Palestinian refugee from Syria (PRS), regardless of whether they registered with UNRWA.

First sampling stage: Locality selection

The sampling frame for the first stage is the list of localities published by the Lebanese Council for Development and Reconstruction (CDR) in 2018 and the 2018 UNHCR data on the number of registered Syrians by locality. Each locality is identified by way of its administrative affiliation—Kaza and Mohafza.

The localities were sorted into nine strata depending on their prevalence of Syrian population and the majority Lebanese sectarian group, as follows:

- Dimension 1: Prevalence of Syrian refugees
  - Low prevalence: where the Syrian population accounted for less than 20% of the total population.
  - Medium prevalence: where the Syrian population is between 20% and 50% of the total population.
  - High prevalence: where the Syrian population accounted for over 50% of the total population.
- Dimension 2: Sectarian component of the non-Syrian population
  - Sunni non-Syria majority: More than 50% of the documented non-Syrian population of the village is Sunni.
  - Non-Sunni non-Syrian majority: More than 50% of the documented non-Syrian population of the village is of a single non-Sunni sectarian group.

– Mixed: No single sectarian group makes up more than 50% of the village population.

In the first stage sample we selected 150 Lebanese localities with replacement, leading to 93 unique localities. We then randomly sampled people in each locality. Our goal is to say something about the drivers of return in the overall population. Hence, in our main analysis, we cluster standard errors by locality, since there are localities in the population of interest beyond those captured in the sample (Abadie et al., 2017).

The distribution of the sample of 150 localities into strata faced the classical dilemma of whether doing it in proportion to the population of the strata or rather selecting the same number of localities in each stratum. Since both are important considerations for our study, and we want to study subnational variation while also making nationwide claims, we followed a first-stage sampling distribution in accordance to Markward's rule (also known as the '50/50 equal/proportional allocation' rule), which is generally considered a good compromise between the two extremes. Given the small number of mixed localities in the pure PPS sample, the 50/50 equal/proportional allocation has an additional desirable feature of avoiding any bins with a very small number of localities.

Second-stage sampling

Enumerators interviewed 20 households for each sample. In order to capture all Syrians and not only registered refugees, we opted for a random walk strategy rather than sampling from UNHCR's registration database. To do so, team leaders initially met with local key informants (such as the head of the municipal government) and had a map print out of the locality. The local key informant was asked to draw boundaries around the neighborhoods that include Syrians and the rough proportion of Syrians in each part of the town. The team leader then assigned enumerators in proportion to Syrians across the towns and provided them with a starting direction (N, NE, E, SE, S, SW, W, NW) for the day. Given that streets limit the direction of movement, the enumerator were asked to proceed along the street that is closest to the starting direction indicated in the enumerator schedule. The enumerators were told to interview a household in every third building.

Third-stage sampling: Respondent Selection

Enumerators interviewed a head of household—a person with primary decision-making responsibility in the household. An eligible 'head of household' should know the details of the household, its economic situation, and important household decisions. If a household had multiple adults who share decision-making responsibility, this classifies as there being multiple heads of household, in which case enumerators would interview whichever head of household was at home.

Response Rate

The team leaders recorded the total number of doors they knocked on each day and kept track of refusal to participate as well as non-answers. Out of 3,882 interview attempts, enumerators conducted 3,003 interviews. This resulted in a response rate of 77%.

Replacement

We needed to drop one research site in the Baalbek area from our sample because a shawish refused us entry to the informal settlement. We returned to our original sampling methodology and drew another town in order to replace this site.

2 Survey questions

2.1 Dependent variables

Our outcome of interest is intention to return. The survey included multiple questions about stated short-term and long-term return intentions as well as return preparations. Table A1 shows the list of questions used to measure these concepts.

Short term return intentions	**Long term return intentions**	**Return preparations**
Do you (the respondent) plan to return to Syria in the next 12 months?	Two years from now, where do you expect to actually be living?	Have you or your immediate family been saving resources in order to prepare for your return to Syria?
Are other members of your household planning to return in the next 12 months?	Do you hope to move back to Syria and live there one day?	Have you or your immediate family prepared any legal paperwork, such as marriage documents, birth certificates, or proofs of property to prepare for your return to Syria?
	Would you say it is correct that you don't want to return no matter what happens?	Have you or your immediate family reached out to Lebanese authorities to discuss returning to Syria?
		Have you or your immediate family reached out to UNHCR to discuss returning to Syria?
		Have you or anyone in your immediate family made a scoping trip back to Syria to learn about the situation there?
		Have there been times in the past 12 months when you were planning to return and aborted those plans?

Table A1: Questions about return intentions and preparations

2.2 Independent variables

2.2.1 Push factors from Lebanon

We use several measures of push factors in Lebanon. Table A2 shows questions used to build the economic well-being and access to services in Lebanon indices. Table A3 shows the indices that measure social well-being, legal conditions, and networks in Lebanon.

Economic well-being in Lebanon	**Services in Lebanon**
Do you possess the status that allows you to work legally in Lebanon (do you currently possess a work permit or Lebanese residency)?	Have you been sick in the past 6 months to the point of requiring medical treatment? (To enumerator: this does not mean that they actually saw a doctor, just that they needed to be treated. This includes chronic illnesses but not common illnesses like cold)
During the past 4 weeks, how many days did you work outside home to make money? (If they did not work, enter 0.)	(If yes) Were you able to see a doctor? (To Enumerator: Going to a pharmacy does not count as seeing a doctor)
(If worked) On the days that you worked during the past 4 weeks, how many hours did you usually work per day?	(For each household member) Has (household member) has been sick in the past 6 months to the point of needing medical treatment? (To enumerator: this does not mean that they actually saw a doctor, just that they needed to be treated. This includes chronic illnesses but not common illnesses like cold)
(If worked) During the past 4 weeks, how much money did you make in total?	(If yes) Was (household member) able to see a doctor? (To Enumerator: Going to a pharmacy does not count as seeing a doctor)
Can you please tell me which sources your household normally receives income from? By household we mean people who are family members or close relatives and who live under the same roof and share meals with you: Aid or assistance from other organizations (such as the United Nations, other international NGOs, local Lebanese charities, etc.)	In Lebanon, how difficult or easy would it be for you to do each of the following? See a doctor (aside from the cost)
In the last month, how much money did your household withdraw in total using cards from humanitarian organizations? To clarify, I do not mean using a card to buy from certain shops. I mean using a card to go to and ATM and withdraw cash. (This includes the 260 and other cash aid.) Enumerator: enter the sum of all cash aid used by household members in the last month.	Do you think it is hard for your or your family to access healthcare in Lebanon because you are Syrian? (To enumerator: if hard but not because they are Syrians, select "Not hard because we are Syrians")
In the last month, how much money did your household spend in total using WFP (World Food Program) support, that is, using a card to buy food only from certain shops? (We are referring to the taghziye program.) Enumerator: enter the sum of all food card aid used by household members in the last month.	Need school: Whether any children between ages of 6 and 18 never studied OR (did not finish primary school and are above 10 years old) OR (are not currently attending school)
To enumerator: This question is written in colloquial Arabic. Read it loud as it is and do not try to summarize it. If you had to live solely on your remaining savings and assets, without any income or debt, approximately how many months of expenses and spending would your savings and assets support you? Note that we're talking about the assets and savings of your household in Lebanon.	Now we would like you to think about all the areas you have lived since moving to Lebanon since you arrived here in (insert year). How many different towns have you lived in Lebanon (including this place) since you came here in (insert year)? (To measure stability in Lebanon, if always in the same town, the variable gets a value of 3 (most stable). If lived in two towns, the variable gets a value of 2. If lived in more than two towns, the variable gets a value of 1 (least stable))
Does this dwelling have the following items that you are able to use? Refrigerator, Washing Machine, Oven, Desktop or Laptop computer at home, Car, Microwave oven, Television, Internet connection at home (other than through a smartphone, not through a neighbor but owned at home), Indoor toilet, Central heating	Which year did you start living in in this area (neighborhood or town)? (To code stability in current town, we subtracted answer from 2019 to find number of years in this town then we cut the answers into quantiles)
How does the aid that you're currently receiving compare to the amount of aid that you were receiving a year ago? (Enumerator: Here we are referring to cash and all other aid.)	(If age of a child <18 AND child not currently attending school) Why is (child) not attending school? (Do not read options. Let respondent answer and select all that apply).
To enumerator: This question is written in colloquial Arabic. Read it loud as it is and do not try to summarize it. Then read all the answer options. In a typical month, what share of your household's monthly expenses and spending needs you are you able to satisfy from household members' income?	In Lebanon, how difficult or easy would it be for you to do each of the following? Get help with legal problems
What is the approximate total value of assets and cash that you possessed when you first came to Lebanon? Note that we're talking about the assets and savings of your household in Lebanon and not anything you left in Syria.	Does this dwelling have the following items that you are able to use? Running water
Does your household income vary from month to month?	
Did you or anyone in your household work in Lebanon before 2011?	
What was your total household income in the past month? By household we mean people who are family members or close relatives and who lived under the same roof and share meals with you. Enumerator: This does not include income from aid.	

Table A2: Push factors: Economic well-being and access to services in Lebanon

Social well-being in Lebanon	Legal conditions in Lebanon	Networks in Lebanon
How would you describe your relationship with Lebanese people?	Do you possess the status that allows you to work legally in Lebanon (does (respondent) currently possess a work permit or Lebanese residency))?	How many of (original household members from Syria) (excluding yourself) are living in Lebanon now?
Have you been detained by Lebanese authorities since arriving to Lebanon? Please note that we will not share this information.	What is your status with UNHCR? Please note that we will not share this information with anyone (Enumerator: If the respondent says registered or recorded, please ask to see the UNHCR registration file with names of registered individuals).	Please think about the Lebanese people in your phone contacts. With how many of them did you have a conversation–either by phone, messenger chat, face-to-face, or text exchange –in the last week? Note that this does not include service provides such as the United Nations or NGOs
Do you think it is hard for you or your family to get housing in Lebanon because you are Syrian? To enumerator: if hard but not because they are Syrians, select "Not hard because we are Syrians"	(If this person is a Palestinian from Syria) What is your status with UNRWA? Please note that we will not share this information with anyone.	Please think about the Syrians in Lebanon in your phone contacts. With how many of them did you have a conversation–either by phone, messenger chat, face-to-face, or text exchange –in the last week?
In your personal experiences, How friendly or hostile would you describe your personal experiences with Lebanese authorities in this area (town/neighborhood)?		Outside of your household, do you have any close Lebanese relatives in this area or elsewhere in the country?
In your personal experiences, how friendly or hostile would you describe your personal experiences with the Lebanese public in this area (town/neighborhood)? We are asking about the Lebanese general public, not authorities.		In the last 12 months, how often did you share a meal with Lebanese people who are not part of your family? (To enumerator: Those do not need to be friends. They can be people at work or other people.)
How well do you understand the important issues facing Lebanon?		
In the last 12 months, how often did you typically discuss major issues facing Lebanon with others?		
In Lebanon, how difficult or easy would it be for you to do each of the following? Search for a job		
Which year did you move to Lebanon to stay here until now?		
Does this town currently have curfews for Syrians?		
(If no) Has this town had curfews for Syrians in the last two years?		
How often do you feel like an outsider in Lebanon?		
(To enumerator) Did the respondent speak Arabic well?		
What is the highest level of education you have completed?		
How well can you read and write?		
How often are you personally able to travel freely and safely around this area of Lebanon?		
Are all your household members able to move freely in this town?		

Table A3: Push factors: Social well-being, legal conditions, and networks in Lebanon

2.2.2 Pull factors in Syria

For pull factors in Syria, Table A4 shows questions used to build the safety, control, and economic well-being in Syria indices. Table A5 shows the indices that measure services and networks in Syria.

Safety in Syria	**Control in Syria**	**Economic well-being in Syria**
How would you describe the current risk to civilians' physical safety in (place of origin)?	Who mainly controlled (place of origin) in the month before you left? Syrian army; Opposition forces such as the FSA; Jabhat al-Nusra; ISIS; Kurdish forces; Russian forces; Turkish forces; It was contested	How would you describe the current availability of jobs at present in (place of origin) currently?
To measure sympathy with opposition, we examine difference between trusting two anti-regime media (Al-Jazeera and Al-Arabiya) and two pro-regime media (Al-Mayadeen and Al-Manar) using the question: How trustworthy would you say each of the following news sources is?	Who mainly controls (place of origin) currently? [Same options]	What is the total amount of outstanding debts you currently have in Syria? This includes any debts on unpaid electricity, water, or other bills while you were away
Were there anti-regime protests in (place of origin) in 2011 and 2012?	Did ISIS control (place of origin) at all during the conflict?	Did you or your immediate family own (not rent) the following items in Syria? (ask for each): House (not an apartment); Apartment (other than their house, if they owned a house); Land
Have you suffered physical or psychological harm because of violence in Syria?		(If they stated that they own land in previous question) Do you think you would be able to continue as owner and operator of this land if you returned to Syria?
How do you expect the safety situation to be in (place of origin) one year from now?		(If owned house/apartment/land) Do you or your immediate family have property documents that prove you are the owner?
Conscription: To examine if someone in household is of/near conscription age, we saw if household includes any males born between 1977 and 2003.		

Table A4: Pull factors: Safety, control, and economic well-being in Syria

Services in Syria	**Networks in Syria**
As far as you know, how many hours per day is there electricity in (place of origin) currently?	How many of (household members from Syria before leaving) are living in Syria now?
As far as you know, how many hours per day is there running water in (place of origin) currently?	Approximately how many Syrian relatives or friends who have lived in Lebanon have gone back to Syria with the goal of staying there?
As far as you know, are schools operating in (place of origin) during the school year?	Next, think about your Syrian friends and relatives from (place of origin) who have lived in Lebanon. Approximately how many of them have gone back to (place of origin)?
As far as you know, are health centers operating in (place of origin) currently?	How many people who were in your household in Lebanon at some point since 2011 have gone back to Syria, regardless of where they are now?
How good do you think that public service provision in (place of origin) (such as health centers, schools, infrastructure) will be one year from now?	

Table A5: Pull factors: Services and networks in Syria

2.2.3 Confidence in information and mobility

Table A6 describes our measures of confidence in information and mobility costs.

Confidence in information	Mobility cost
To enumerator: This question is written in colloquial Arabic. Read it loud as it is and do not try to summarize. Then read all the answer options. When thinking of your knowledge about the safety situation in (place of origin), would you say that... : I know enough to be confident I understand the situation (1); I don't know enough, and I want to know more (0); I don't know much but do not feel the need to know (0)	We calculate travel distance from each survey respondent's town of residence in Lebanon to their hometown in Syria, via the Beirut–Damascus highway and border crossing. Travel routes were calculated using the Google Maps API. We used the R package mapsapi and commands mp_directions() and mp_get_routes().
To enumerator: This question is written in colloquial Arabic. Read it loud as it is and do not try to summarize. Then read all the answer options. When thinking of your knowledge about employment opportunities in (place of origin), would you say that... : (same options as above)	The log of household size
To enumerator: This question is written in colloquial Arabic. Read it loud as it is and do not try to summarize. Then read all the answer options. When thinking of your knowledge about the availability of public services (such as health centers, schools, or water) in (place of origin), would you say that... : (same options as above)	
To enumerator: This question is written in colloquial Arabic. Read it loud as it is and do not try to summarize. Then read all the answer options. When thinking of your knowledge about the conscription requirements by the Syrian military, would you say that... : (same options as above)	
Now, please think about the person you communicate with the most who is currently living in Syria. (If respondent answers: I don't communicate with anyone in Syria, this is coded as 0, otherwise 1)	
In general, how often do you communicate with people in (place of origin)?	

Table A6: Confidence in information and mobility costs

2.3 Control variables

We also adjust for a range of control variables in Equation 1 in the paper, including household-level covariates and locality-level fixed effects. As defined in the pre-analysis plan, the control set includes the following: indicators for being from an urban area in Syria, living in an informal (tented) settlement, a household member being seriously ill, the head of household having finished secondary school or higher, a toddler household member, an elderly household member, female-headed household, whether an area is a Hizbullah-controlled area, and Syrian governorate of origin and Lebanese district of residence. All variables come from the survey data except the final covariate. Data for Hizbullah control is coded based on whether a Hizbullah-aligned candidate won a seat in the 2018 parliamentary elections, using Lebanese Ministry of Interior data released by the Data Liberation Project. In regressions including travel distance on the right-hand side, we drop controls for location in Lebanon and hometown in Syria, since travel distance is a deterministic function of these variables.

3 Index construction and descriptives

We measured four key concepts with multiple independent variables using the first principal component.

1. Well-being in Lebanon
2. Expected well-being in Syria
3. Confidence in information
4. Preparation to return

The component variables are mentioned in Section 2. For the first three independent variables, we have many input variables. Note that all results for the survey in Lebanon impute missing values using multivariate imputation by chained equations. We specify 10 imputations and use random forest to predict missing values using the `mice()` package in R. We use Rubin's rules (Rubin, 1987) when pooling estimates across imputations.

Our primary analysis was based on regression models with indices constructed using polychoric PCA unless any variable in the index had too many categories (above 8 categories, in which case we used Pearson correlations) (Hainmueller, Hangartner and Pietrantuono, 2017). Scales for individual questions were reversed as necessary to simplify interpretation (to make sure they were all in the same positive direction). In addition to the indices constructed using the first principal components, we ran mean effects indices (aka z-scores) as a robustness check (Kling, Liebman and Katz, 2007). We constructed those indices by standardizing each variable (demeaning and dividing it by the standard deviation). We then summed the standardized variables and then standardized the sum again.

This section shows the component variables of each of the indices as well as descriptive statistics for these variables. We also show descriptive statistics for the control variables we included in the main regressions.

3.1 Push factors from Lebanon

Economic well-being in Lebanon

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Aid: atm card	0.63	0.58	1.39	0.00	0.00	7.00	0.30%
Aid change from last year	1.49	1.48	0.56	1.00	1.00	3.00	9.16%
Aid: wfp card	1.12	1.02	1.68	0.00	0.00	7.00	0.13%
Assets: months left	0.13	0.15	0.62	0.00	0.00	12.00	0.00%
Assets: value upon arrival	0.62	0.70	2.26	0.00	0.00	16.00	0.27%
Ability to cover expenses	2.78	2.88	1.26	1.00	3.00	5.00	0.17%
Income	1.75	1.90	3.07	0.00	0.00	13.00	0.10%
Work days past 4 weeks	5.82	6.25	9.91	0.00	0.00	28.00	0.00%
Work hours past 4 weeks	2.92	3.13	4.74	0.00	0.00	24.00	0.00%
Able to work legally	0.10	0.11	0.31	0.00	0.00	1.00	0.03%
Income source: aid	0.36	0.33	0.47	0.00	0.00	1.00	0.03%
Stable household income	0.81	0.79	0.71	0.00	1.00	2.00	0.20%
Household income	4.36	4.72	3.50	0.00	5.00	17.00	0.93%
HH worked in Leb. before 2011	0.23	0.24	0.43	0.00	0.00	1.00	0.00%
Own refrigerator	0.80	0.81	0.39	0.00	1.00	1.00	0.00%
Have indoor toilet	0.83	0.83	0.37	0.00	1.00	1.00	0.00%
Have central heating	0.05	0.05	0.22	0.00	0.00	1.00	0.13%
Own washing machine	0.68	0.67	0.47	0.00	1.00	1.00	0.00%
Own oven/stove	0.79	0.79	0.41	0.00	1.00	1.00	0.00%
Own computer	0.02	0.02	0.13	0.00	0.00	1.00	0.00%
Own car	0.02	0.02	0.14	0.00	0.00	1.00	0.00%
Own microwave oven	0.04	0.05	0.21	0.00	0.00	1.00	0.00%
Own television	0.78	0.79	0.41	0.00	1.00	1.00	0.00%
Have internet	0.24	0.25	0.44	0.00	0.00	1.00	0.07%

Table A7: Summary statistics of variables included in constructing the economic well-being in Lebanon index

Social well-being in Lebanon

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Never had curfews	0.65	0.63	0.48	0.00	1.00	1.00	0.93%
Authorities discrimination (higher is less)	2.71	2.73	0.74	1.00	3.00	4.00	15.38%
Ease of mobility	3.16	3.16	1.09	1.00	4.00	4.00	0.13%
Ease of mobility for household	2.38	2.39	0.78	1.00	3.00	3.00	0.53%
Public discrimination (higher is less)	2.97	2.98	0.68	1.00	3.00	4.00	1.60%
Arabic speaking ability	2.87	2.87	0.34	1.00	3.00	3.00	1.27%
Relation with Lebanese	3.66	3.68	0.95	1.00	4.00	5.00	0.33%
Discuss Lebanese politics	1.35	1.38	0.94	1.00	1.00	5.00	0.47%
Ease job search	1.59	1.61	1.02	1.00	1.00	5.00	0.83%
Literacy level	2.10	2.09	0.72	1.00	2.00	3.00	0.00%
Feeling outsider (higher is less)	2.95	2.96	1.29	1.00	3.00	5.00	0.03%
Know Lebanese politics	2.33	2.33	1.24	1.00	2.00	5.00	0.83%
Time in Lebanon	5.52	5.46	2.05	0.00	6.00	9.00	0.07%
No curfew now	0.75	0.73	0.45	0.00	1.00	1.00	0.50%
Housing discrimination (higher is less)	2.48	2.49	0.69	1.00	3.00	3.00	0.33%
Never detained	0.94	0.94	0.23	0.00	1.00	1.00	0.20%

Table A8: Summary statistics of variables included in constructing the social well-being in Lebanon index

Services in Lebanon

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Can access legal services	1.66	1.66	1.07	1.00	1.00	5.00	2.73%
No healthcare discrimination	2.37	2.37	0.73	1.00	3.00	3.00	0.47%
Not sick	0.83	0.83	0.38	0.00	1.00	1.00	0.07%
Received treatment (if sick)	0.94	0.95	0.23	0.00	1.00	1.00	0.00%
Can access doctor	2.80	2.87	1.66	1.00	3.00	5.00	0.70%
No Kids need school	0.59	0.59	0.49	0.00	1.00	1.00	0.03%
Have running water	0.80	0.81	0.39	0.00	1.00	1.00	0.00%
No HH member sick	0.78	0.79	0.41	0.00	1.00	1.00	0.17%
HH members treated if sick	0.91	0.91	0.28	0.00	1.00	1.00	0.03%
School not preventive	0.96	0.96	0.20	0.00	1.00	1.00	0.07%
Period in current town	2.39	2.37	1.16	1.00	2.00	4.00	0.10%
Towns lived in Lebanon	2.64	2.65	0.63	1.00	3.00	3.00	0.10%

Table A9: Summary statistics of variables included in constructing the services in Lebanon index

Legal situation in Lebanon

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Registered with UNHCR/UNRWA (or resident)	0.81	0.80	0.40	0.00	1.00	1.00	0.00%
Legal resident in Lebanon	0.04	0.05	0.21	0.00	0.00	1.00	0.00%

Table A10: Summary statistics of variables included in constructing the legal situation in Lebanon index

Networks in Lebanon

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Syria HH members living in Leb. now	4.58	4.45	3.51	0.00	4.00	15.00	0.00%
Lebanese phone contacts	1.71	1.71	1.11	1.00	1.00	5.00	0.53%
Share meals with Lebanese	1.51	1.53	1.08	1.00	1.00	5.00	0.17%
Syrian phone contacts	3.22	3.20	1.37	1.00	3.00	5.00	0.30%
Lebanese relatives	0.22	0.22	0.55	0.00	0.00	2.00	0.10%

Table A11: Summary statistics of variables included in constructing the networks in Lebanon index

3.2 Pull factors in Syria

Safety in Syria

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
HH male at conscription age	0.80	0.81	0.39	0.00	1.00	1.00	0.00%
Exposed to violence	0.28	0.27	0.45	0.00	0.00	1.00	0.03%
Follow anti-regime media more than pro-regime media	0.20	0.21	0.41	0.00	0.00	1.00	0.00%
Hometown had protests	0.67	0.67	0.47	0.00	1.00	1.00	2.40%
Expect hometown to be safe	2.54	2.53	0.86	1.00	3.00	4.00	18.81%
Current safety in hometown	2.05	2.06	0.88	1.00	2.00	4.00	7.13%

Table A12: Summary statistics of variables included in constructing the safety in Syria index. Respondents were coded as following anti-regime media more than pro-regime media if they reported following Al-Jazeera or Al-Arabiya (anti-regime) more than Manar/Mayadeen (pro-regime).

Control in Syria

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Contested Now	0.05	0.05	0.22	0.00	0.00	1.00	5.16%
Contested before leaving	0.15	0.15	0.36	0.00	0.00	1.00	3.06%
Controlled by Kurds now	0.13	0.14	0.35	0.00	0.00	1.00	5.16%
Controlled by oppsn/FSA now	0.09	0.09	0.29	0.00	0.00	1.00	5.16%
Controlled by regime now	0.69	0.67	0.47	0.00	1.00	1.00	5.16%
Controlled by Russia now	0.00	0.01	0.08	0.00	0.00	1.00	5.16%
Controlled by Turkey now	0.01	0.02	0.12	0.00	0.00	1.00	5.16%
Controlled by Kurds before leaving	0.02	0.03	0.17	0.00	0.00	1.00	3.06%
Controlled by oppsn/FSA before leaving	0.28	0.29	0.45	0.00	0.00	1.00	3.06%
Controlled by regime before leaving	0.38	0.36	0.48	0.00	0.00	1.00	3.06%
Controlled by Russia before leaving	0.00	0.00	0.06	0.00	0.00	1.00	3.06%
Controlled by Turkey before leaving	0.00	0.00	0.05	0.00	0.00	1.00	3.06%
Controlled by ISIS at some point	0.37	0.39	0.49	0.00	0.00	1.00	5.13%

Table A13: Summary statistics of variables included in constructing the control in Syria index

Economic well-being in Syria

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Debt in Syria	0.31	0.32	0.66	0.00	0.00	2.00	4.96%
Job situation in origin	1.55	1.54	0.70	1.00	1.00	4.00	9.62%
Home ownership docs (1 for some, 2 for everything)	0.58	0.62	0.91	0.00	0.00	2.00	1.90%
Can operate land in future	0.12	0.13	0.33	0.00	0.00	1.00	2.33%
Own house in Syria	0.62	0.64	0.48	0.00	1.00	1.00	0.30%
Own apt in Syria	0.07	0.07	0.25	0.00	0.00	1.00	0.37%
Own land in Syria	0.21	0.21	0.41	0.00	0.00	1.00	0.63%

Table A14: Summary statistics of variables included in constructing the economic well-being in Syria index

Services in Syria

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Electricity in origin	2.51	2.51	1.27	1.00	3.00	5.00	15.25%
Health services in origin	0.36	0.36	0.48	0.00	0.00	1.00	19.21%
Expect services to improve in 1 year	2.43	2.42	0.84	1.00	3.00	4.00	17.82%
Schools in origin	0.34	0.34	0.47	0.00	0.00	1.00	19.28%
Running water in origin	2.47	2.46	1.23	1.00	3.00	5.00	14.72%

Table A15: Summary statistics of variables included in constructing the services in Syria index

Services in Syria

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Electricity in origin	2.51	2.51	1.27	1.00	3.00	5.00	15.25%
Health services in origin	0.36	0.36	0.48	0.00	0.00	1.00	19.21%
Expect services to improve in 1 year	2.43	2.42	0.84	1.00	3.00	4.00	17.82%
Schools in origin	0.34	0.34	0.47	0.00	0.00	1.00	19.28%
Running water in origin	2.47	2.46	1.23	1.00	3.00	5.00	14.72%

Table A16: Summary statistics of variables included in constructing the services in Syria index

Networks in Syria

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
No. HH members returned to Syria	0.10	0.11	0.65	0.00	0.00	5.00	0.27%
Relatives permanently return to Syria	0.97	1.05	2.99	0.00	0.00	15.00	1.40%
Relatives return to origin	0.53	0.60	2.44	0.00	0.00	15.00	1.17%
Syria HH members living in Syria now	1.86	2.00	3.36	0.00	0.00	15.00	0.10%

Table A17: Summary statistics of variables included in constructing the networks in Syria index

3.3 Confidence in information

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Know Syr. conscription policy	0.28	0.30	0.46	0.00	0.00	1.00	0.23%
Know employment in origin	0.27	0.29	0.46	0.00	0.00	1.00	0.13%
Know safety in origin	0.26	0.27	0.45	0.00	0.00	1.00	0.10%
Know services in origin	0.25	0.27	0.44	0.00	0.00	1.00	0.20%
Communication freq. with origin	2.39	2.45	1.69	1.00	2.00	6.00	0.03%
Communication with someone in Syria	0.61	0.62	0.49	0.00	1.00	1.00	0.03%

Table A18: Summary statistics of variables included in constructing the confidence in information index

3.4 Mobility

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
household size (logged)	1.50	1.48	0.56	0.00	1.61	2.89	0.00%
travel distance (logged)	12.83	12.85	0.52	11.13	12.98	15.63	0.37%

Table A19: Summary statistics of variables to measure mobility

3.5 Preparation to return

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Planned to return but aborted	0.03	0.03	0.17	0.00	0.00	1.00	0.27%
Reached to Leb. authorities about return	0.01	0.01	0.08	0.00	0.00	1.00	0.13%
Prepared docs for return	0.02	0.03	0.17	0.00	0.00	1.00	0.10%
Saved resources for return	0.02	0.02	0.15	0.00	0.00	1.00	0.10%
Conducting scoping trip to Syria	0.03	0.03	0.17	0.00	0.00	1.00	0.10%
Reached to UNHCR about return	0.01	0.01	0.09	0.00	0.00	1.00	0.17%

Table A20: Summary statistics of variables included in constructing the preparation to return index

3.6 Covariates

The following covariates are included in the regressions. When including fixed effects, we also add locality fixed effects in Lebanon and Syria.

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Household includes elderly	0.04	0.04	0.19	0.00	0.00	1.00	0.00%
Female headed single-parent household	0.11	0.10	0.30	0.00	0.00	1.00	0.03%
High school graduate	0.11	0.12	0.32	0.00	0.00	1.00	0.03%
Hezbollah controlled area	0.16	0.17	0.38	0.00	0.00	1.00	0.00%
Location: Tental settlement	0.35	0.33	0.47	0.00	0.00	1.00	0.00%
Sick required medical treatment	0.30	0.30	0.46	0.00	0.00	1.00	0.23%
Syria origin: urban	0.23	0.23	0.42	0.00	0.00	1.00	0.20%
Household includes toddler	0.45	0.44	0.50	0.00	0.00	1.00	0.00%

Table A21: Summary statistics of variables included as controls in the regressions

4 Scree plots for principal component analysis

The following figures display how eigenvalues change with each additional component for the indices we created using PCA. Throughout the analysis, we used the first principal component.

4.1 Push factors from Lebanon

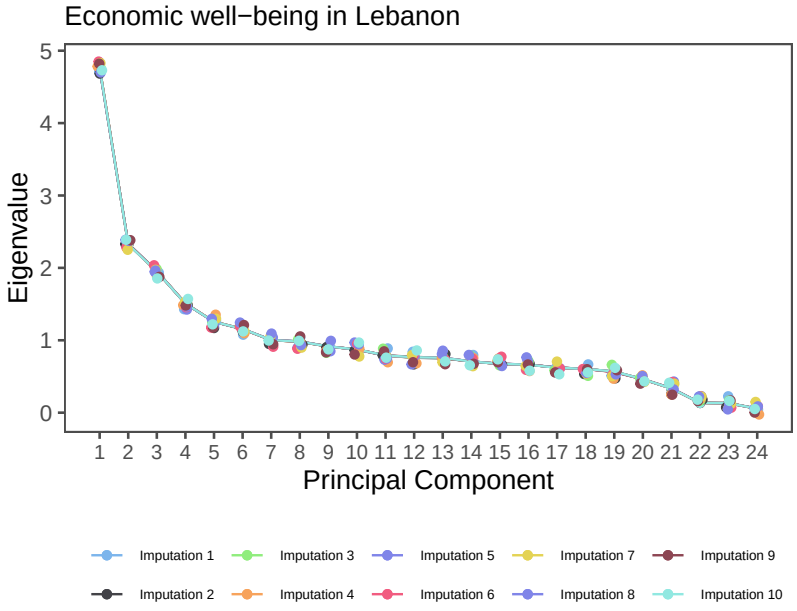

Figure A1: Screeplot for the economic well-being in Lebanon index

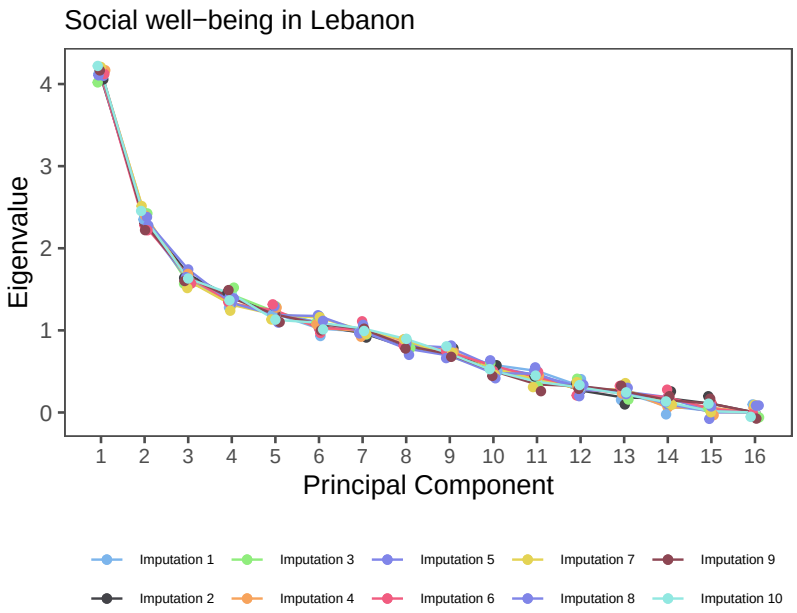

Figure A2: Screeplot for the social well-being in Lebanon index

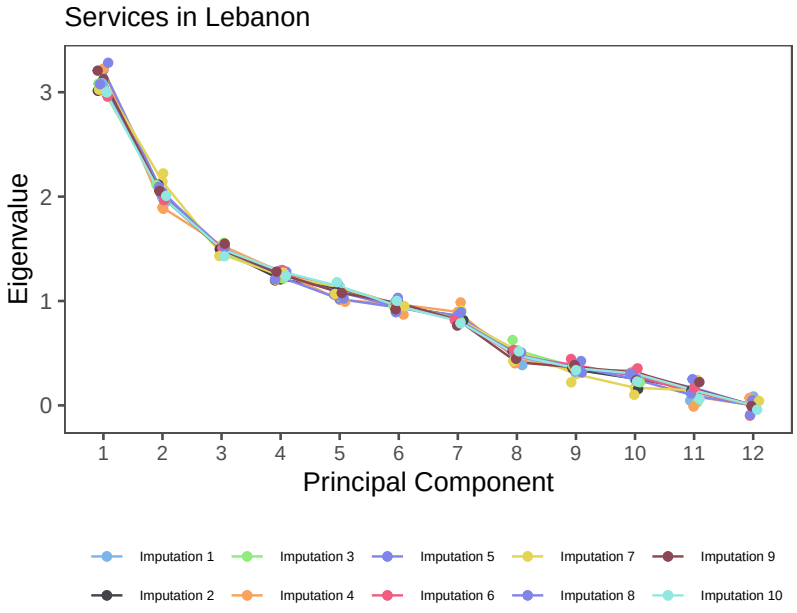

Figure A3: Screeplot for the services in Lebanon index

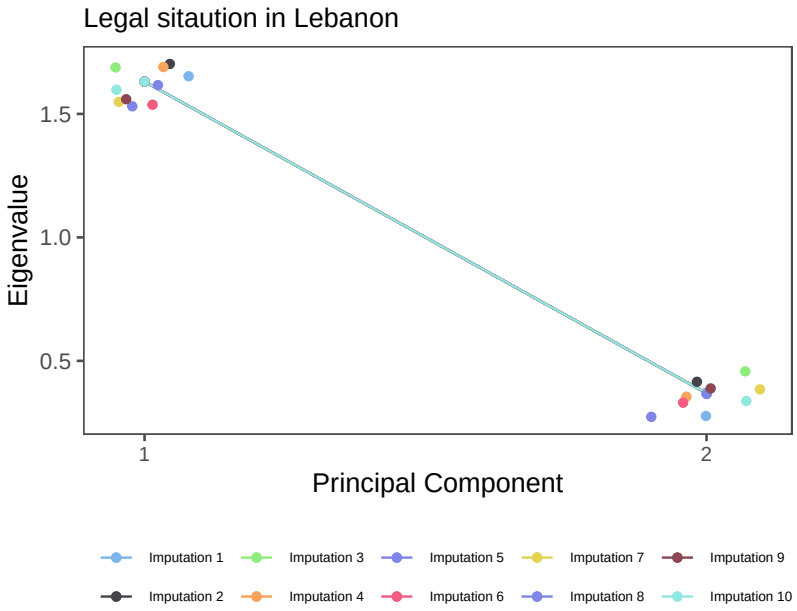

Figure A4: Screeplot for the legal situation in Lebanon index

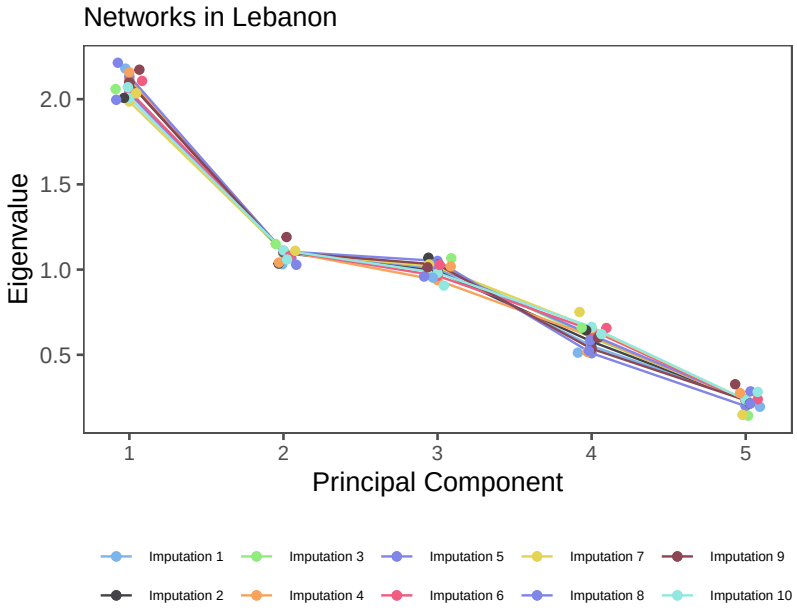

Figure A5: Screeplot for the networks in Lebanon index

4.2 Pull factors in Syria

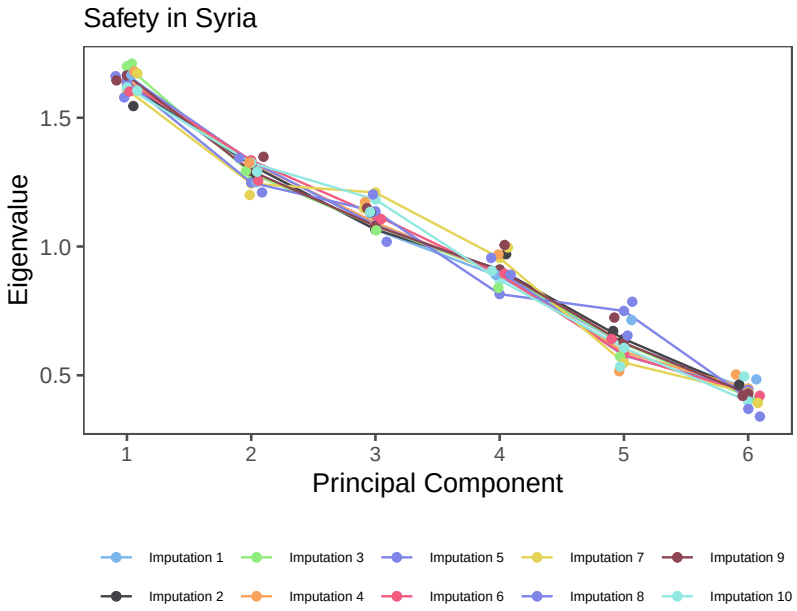

Figure A6: Screeplot for the safety in Syria index

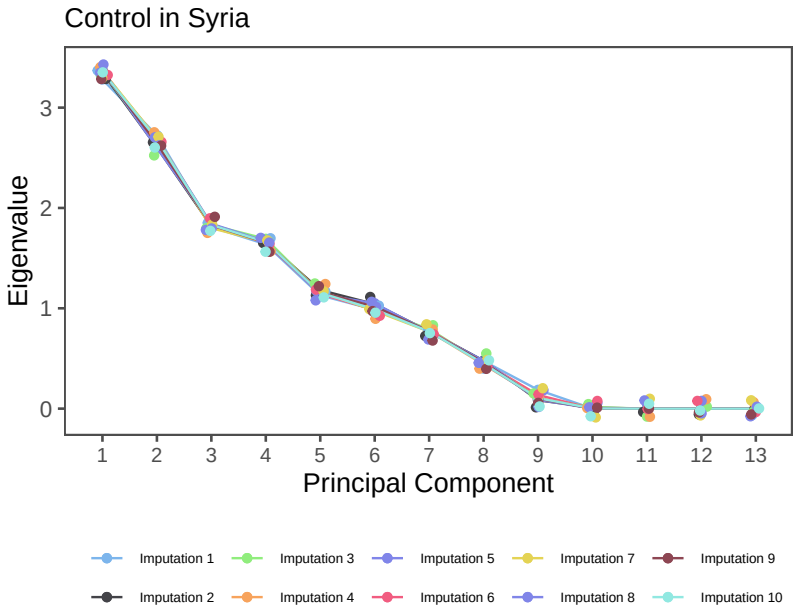

Figure A7: Screeplot for the control in Syria index

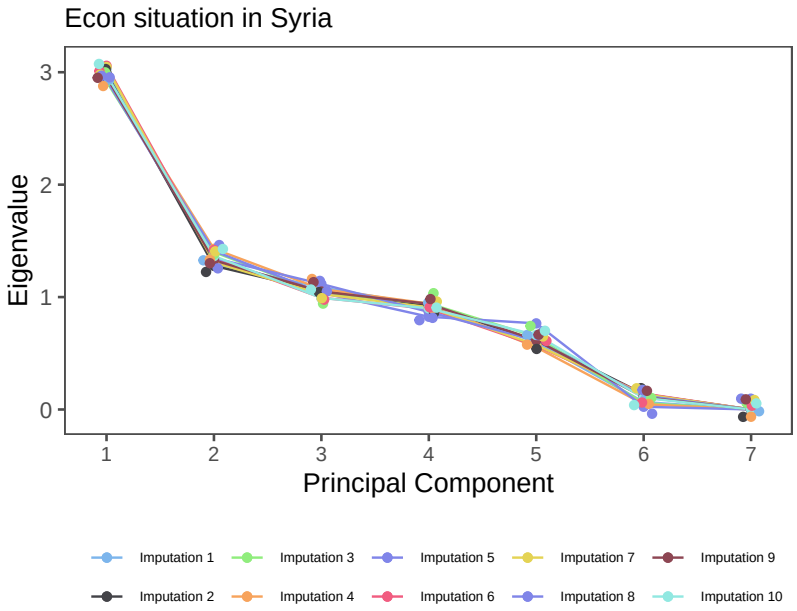

Figure A8: Screeplot for the economic well-being in Syria index

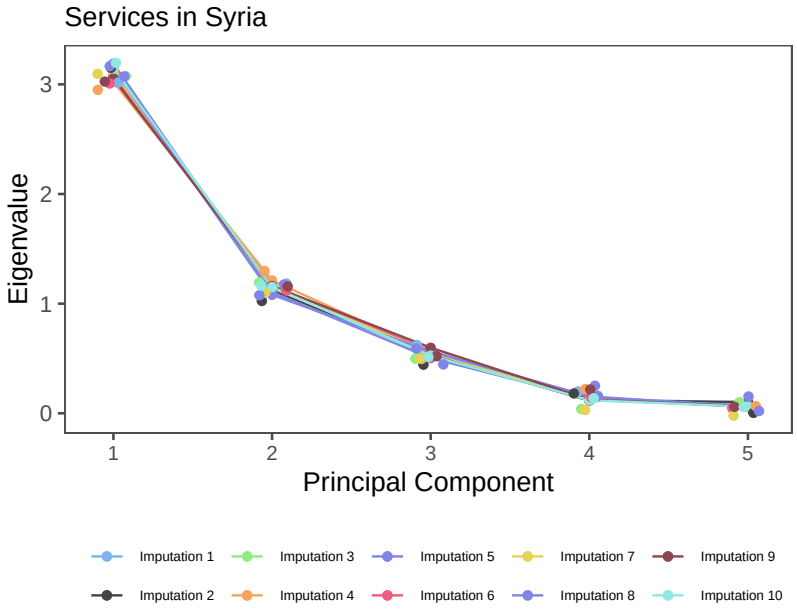

Figure A9: Screeplot for the services in Syria index

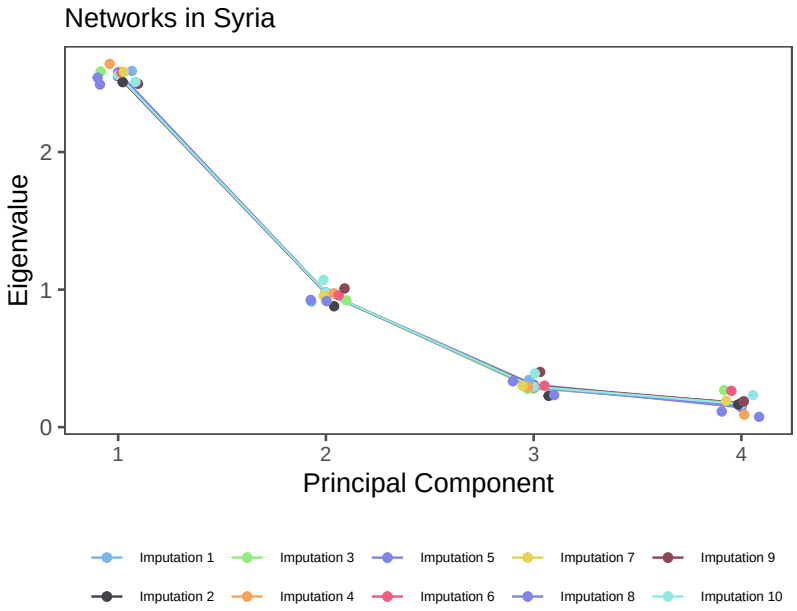

Figure A10: Screeplot for the networks in Syria index

4.3 Confidence in information

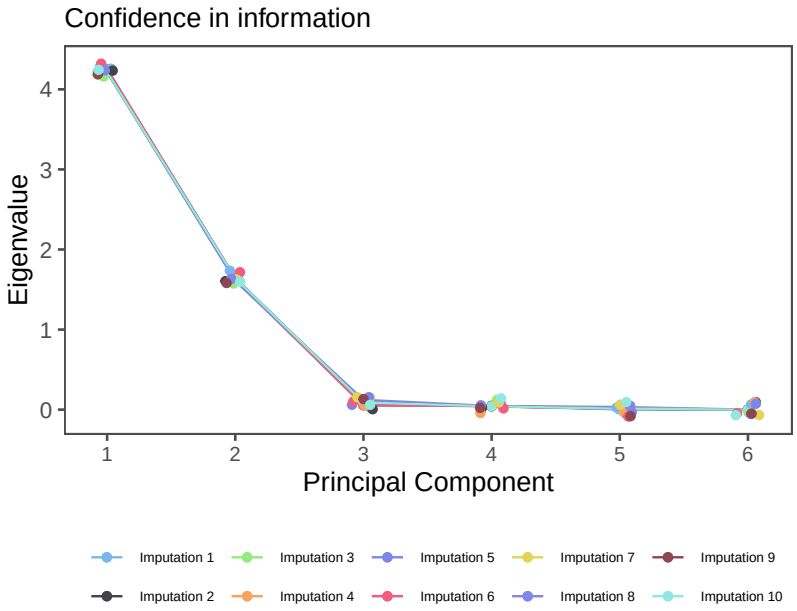

Figure A11: Screeplot for the confidence in information index

4.4 Preparation for return

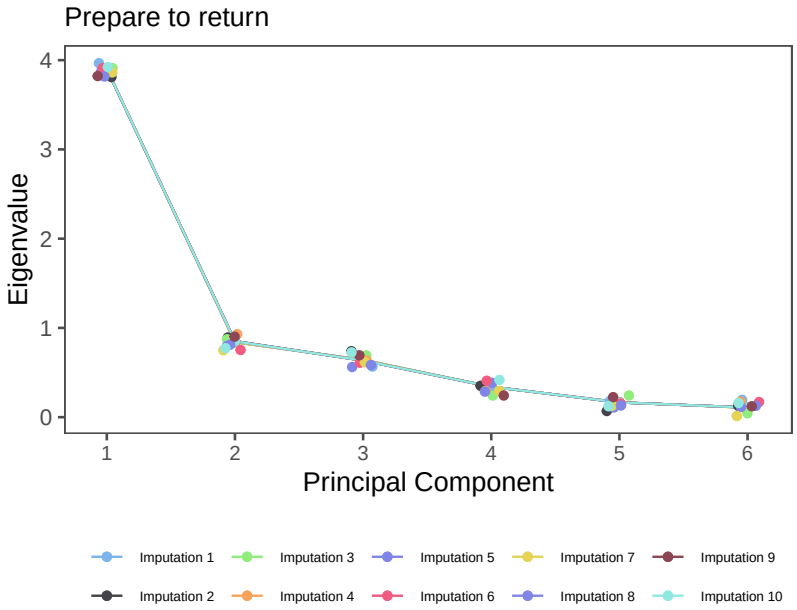

Figure A12: Screeplot for the preparation to return index

5 Deviations from PAP

Multicollinearity: In the PAP, we indicated that we would run one regression with all the indices. Because of multicollinearity, we ran separate regressions for each of the indices and also ran the pre-specified regression with a minor adjustment: we did not include the indices measuring safety and services in Syria in the same regression because of their high correlation. In the paper, we present the results results for all the indices where we control for safety. In Section 6.2, we include the results for all the indices where we control for services in Syria.

We made the choice to present an alternative regression after tests of model performance identified multicollinearity in our regression predictors. The first simple test of model performance involved examining the simple pairwise correlations between our predictors. We find that the pairwise correlation between the security index and services index in Syria is high at 0.57.

Of course, simple correlation is not multicollinearity. Therefore, second, we test for an association between predictors conditional on the other variables in the model. The variance inflation factor is a measure to analyze the magnitude of multicollinearity of model terms. Using the `performance()` package in R, we find that in the pre-specified models we identify very high variance inflation factors for a number of indices, most notably services in Syria and regime control. Evidence of multicollinearity is consistent across multiple tests in the `performance()` package.

Analysis without locality fixed effects: We re-ran our main analysis without Lebanese locality fixed effects as a robustness check. This regression can be seen in Figure A24 in Section 6 of the appendix. Furthermore, after submitting the PAP, we realized that it would be misguided to control for Lebanese locality and Syrian locality when analyzing the role of travel distance from Lebanese locality to Syrian locality. Therefore, in regressions with travel distance on the right-hand side, we do not include locality fixed effects.

Predictive analysis: The predictive analyses included in appendix section 6.7 were not pre-specified.

PCA inputs: PCA inputs were pre-specified but required a number of ex post modifications for reasons explained below.

Index 1—Economic well being in Lebanon: The PAP mistakenly indicated that a question about someone's former job in Syria would be included in the economic well-being *in Lebanon* index. This was a typo, and it was removed since it is not a dimension of economic well-being *in Lebanon*.

Index 2—Social well being in Lebanon: The PAP specifies that we intended to calculate the IPL-12 integration score to use it as an input for PCA. In later revisions to the questionnaire, some questions were modified and no longer matched IPL-12 inputs. We modified this index slightly and now use component questions of IPL-12 as inputs rather than calculating the IPL-12 score. Furthermore, household income in Lebanon was mistakenly included in both index 1 and index 2. We decided this index fits better in index 1 and removed it from index 2.

Index 3—services in Lebanon: A question included in the PAP was subsequently cut from the survey: "Have you been forced to move in the last two years, for instance because you were kicked out of your home or your home was deconstructed/demolished?" So we did not include it in the analysis.

Index 5.1—Safety: First, we separated regime control from safety conditions in Syria. These two concepts were not closely related and we decided to examine territorial control and safety separately. Furthermore, Family deaths in Syria was excluded due to flaw in measurement strategy.

Index 10—Information quality: After submitting the PAP we removed the following question: "How confident are you in your knowledge about conditions in [Piped place of origin]?" so we did not include it in the analysis.

6 Robustness and Additional Tests

6.1 Return in two years and ever

In addition to the return in 12 months and preparation to return outcomes, we have also examined the predictors of expectation to return in two years and intentions to ever return to Syria. In Table A22 the first two models present the regression results for the return ever outcome using one index per regression (Model 1) and using all indices in the same regression (Model 2). Models 3 and 4 show the same analyses for the return in two years outcome.

	Ever (Ind. Indices)	Ever (All Indices)	2 years (Ind. Indices)	2 years (All Indices)
Safety (Syr.)	0.038**	0.039**	0.049***	0.076***
	(0.016)	(0.014)	(0.017)	(0.017)
Regime control (Syr.)	0.004	−0.014	0.006	−0.021
	(0.015)	(0.014)	(0.014)	(0.019)
Economic well-being (Syr.)	0.033**	0.061***	0.026	0.062***
	(0.013)	(0.013)	(0.021)	(0.016)
Services (Syr.)	0.030*	0.018	0.023	0.052***
	(0.017)	(0.015)	(0.015)	(0.013)
Networks (Syr.)	0.031***	0.020*	−0.002	−0.004
	(0.009)	(0.010)	(0.009)	(0.009)
Economic well-being (Leb.)	−0.040*	−0.079***	−0.019	−0.055
	(0.020)	(0.020)	(0.015)	(0.035)
Services (Leb.)	−0.034	−0.005	0.044***	0.030
	(0.021)	(0.024)	(0.015)	(0.019)
Networks (Leb.)	−0.017	−0.051***	−0.015	−0.007
	(0.012)	(0.014)	(0.018)	(0.012)
Social well-being (Leb.)	0.063***	0.112***	0.004	0.029
	(0.022)	(0.022)	(0.020)	(0.018)
Legal conditions (Leb.)	0.001	0.011	−0.019*	−0.012
	(0.018)	(0.016)	(0.010)	(0.011)
Log travel distance	−0.078***	−0.056**	0.002	0.020
	(0.024)	(0.022)	(0.038)	(0.029)
Log household size	0.015	0.011	0.009	0.015
	(0.020)	(0.020)	(0.023)	(0.029)
Confidence in information	0.026	0.037**	0.020	−0.015
	(0.016)	(0.017)	(0.015)	(0.016)

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A22: Regression results using alternative outcomes. The first two models present the regression results for the return ever outcome using one index per regression (model 1) and using all indices in the same regression (model 2). The last two models present the regression results for the return in 2 years outcome using one index per regression (model 3) and using all indices in the same regression (model 4).

6.2 Safety, services, and additional covariates

Given the high correlation between safety and services in Syria, we ran a regression that included all the predictors except services in Syria, and then ran a separate regression that included all the predictors except safety in Syria. We reported the results of all the coefficients from the regression that included safety (but not services) in Syria. We then added the single coefficient for services from the second regression. In Models 1 and 2 of Table A23, we do the opposite. We include all the coefficients from the regression that included services (but not safety) in Syria and then add the single coefficient for safety in Syria from the first regression.

In models 3 and 4 of Table A23, we present the results using all the indices in the regression (including safety and services in Syria) for the return in 12 months outcome (Model 3) and the preparation to return outcome (Model 4). As can be seen here, the two indices (safety and services) get smaller point estimates and are no longer statistically significant. This difference between the individual-index models and the all-indices model aligns with the evidence of correlation and multicollinearity shown in section 5.

Finally, the main results using individual indices control for a limited set of household covariates because many of the other potential covariates are included in the appendix. Of course, when we include the full list of indices in the same regression (all indices), we control for indices that include a variety of other covariates. In addition to this approach, we also run the main model with the individual indices using a larger set of controls. In addition to the main variables included in the paper, model 5 in Table A23 controls for household income, whether household receives aid, the number of years in Lebanon, the presence of a curfew targeting refugees in the area, whether any household members are school age, and having Lebanese relatives. Note that when one of these variables is included in the index, we do not control for this variable. The full set of variables in each index can be seen in section 3.

6.3 Excluding locality fixed effects

We also test for the robustness of our results by re-running our models, but without controlling for locality fixed effects.[1] Model 1 of Table A24 shows the results for the 12 months outcome while Model 2 shows the results for the preparation to return outcome.

6.4 Return plans for household members and uncertainty about return

The main paper coded intention to return in 12 months as 1 if heads of households indicated their intentions to return. The last two models in Table A24 present alternative codings of this outcome. The first two models re-run the main analysis without fixed effects for the return in 12 months outcome (Model 1) and for the preparation to return outcome (Model 2). In Model 3, the outcome is whether anyone in the household plans to return in 12 months. In Model 4, the outcome at the head of household level is coded as 1 if the heads of household said they planned to return or were uncertain about return and 0 only if they said that they do not plan to return.

[1] This robustness check was not pre-specified.

	12 months (Services)	Prepare (Services)	12 months (Safety + Services)	Prepare (Safety + Services)	Return in 12 months
Economic well-being (Leb.)	−0.007	0.154***	−0.007	0.154***	0.000
	(0.008)	(0.055)	(0.008)	(0.055)	(0.010)
Social well-being (Leb.)	0.021***	0.020	0.020***	0.018	0.017***
	(0.006)	(0.025)	(0.006)	(0.026)	(0.006)
Services (Leb.)	−0.005	−0.010	−0.006	−0.011	−0.003
	(0.007)	(0.027)	(0.007)	(0.028)	(0.007)
Legal conditions (Leb.)	−0.003	0.000	−0.002	0.001	−0.004
	(0.005)	(0.025)	(0.005)	(0.025)	(0.006)
Regime control (Syr.)	−0.007	−0.000	−0.009	−0.002	−0.011
	(0.006)	(0.017)	(0.006)	(0.018)	(0.007)
Economic well-being (Syr.)	0.012**	0.013	0.013**	0.014	0.016**
	(0.005)	(0.022)	(0.005)	(0.022)	(0.006)
Services (Syr.)	0.012**	0.073***	0.006	0.066**	0.018***
	(0.005)	(0.025)	(0.006)	(0.026)	(0.006)
Networks (Syr.)	0.012*	0.077*	0.012*	0.077	0.017***
	(0.006)	(0.045)	(0.006)	(0.045)	(0.006)
Networks (Leb.)	−0.002	0.017			
	(0.006)	(0.024)			
Confidence in information	0.020**	0.003	0.020**	0.003	0.019**
	(0.008)	(0.022)	(0.007)	(0.022)	(0.009)
Log travel distance	0.006	−0.036	0.006	−0.036	0.002
	(0.009)	(0.044)	(0.009)	(0.044)	(0.008)
Log household size	0.013	0.001	0.014	0.001	0.010
	(0.010)	(0.049)	(0.010)	(0.049)	(0.011)
Safety (Syr.)	0.014**	0.048*	0.011	0.015	0.017**
	(0.006)	(0.026)	(0.007)	(0.026)	(0.006)
Networks (Leb.)			−0.003	0.016	0.010*
			(0.006)	(0.024)	(0.006)

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A23: Robustness tests. The first two models include the services in Syria index in the all-indices regression (instead of the safety index) for the return in 12 months outcome (model 1) and the preparation to return outcome (model 2). The next two models present the analysis that includes running both the safety and services in Syria indices in the same regression. Note that there is a high correlation between the safety and services in Syria indices. The third model uses the return in 12 months outcome and the fourth model uses the preparation to return outcome. The final model uses the main return in 12 months outcome with individual indices but uses additional covariates. Besides the ones mentioned in the paper, it controls for household income, an indicator for receiving aid, years in Lebanon, an indicator for a curfew in the locality, an indicator for age-school children, and an indicator for Lebanese relatives. Note that these additional controls are included only when the individual index in the regression does not contain any of these variables. The list of variables in each index is included appendix section 3.

6.5 Additive indices

To build indices in the paper, we use the first principal component for the predictors under push factors, pull factors, and information as well as for the preparation to return outcome. In Table A25, we present results using mean effects indices. The first two models in Table A25 present the results using one index per regression for the 12 months return outcome (Model 1) and the preparation to return outcome (Model 2). The last two models present the same analyses but using all the indices in the same regression.

	12 months (No FEs)	Prepare (No FEs)	Household member return	Head of HH return (counting uncertain)
Safety (Syr.)	0.021***	0.084***	0.018**	0.031*
	(0.006)	(0.024)	(0.008)	(0.018)
Regime control (Syr.)	−0.013*	−0.002	−0.010	−0.042*
	(0.007)	(0.021)	(0.007)	(0.021)
Economic well-being (Syr.)	0.018***	0.063**	0.020***	−0.000
	(0.005)	(0.030)	(0.007)	(0.011)
Services (Syr.)	0.021***	0.115***	0.019***	0.046**
	(0.005)	(0.028)	(0.006)	(0.017)
Networks (Syr.)	0.019***	0.110**	0.015**	0.011
	(0.006)	(0.040)	(0.006)	(0.009)
Economic well-being (Leb.)	0.007	0.183***	−0.001	0.043
	(0.007)	(0.054)	(0.008)	(0.026)
Services (Leb.)	−0.001	0.011	−0.009	0.000
	(0.007)	(0.025)	(0.008)	(0.012)
Networks (Leb.)	0.013***	0.088***		
	(0.004)	(0.021)		
Social well-being (Leb.)	0.021***	0.068***	0.020***	−0.004
	(0.005)	(0.020)	(0.007)	(0.013)
Legal conditions (Leb.)	−0.003	0.002	−0.003	−0.012
	(0.006)	(0.020)	(0.005)	(0.011)
Log travel distance	0.002	−0.079*	0.007	0.019
	(0.008)	(0.043)	(0.012)	(0.015)
Log household size	0.001	−0.103***	0.022	0.012
	(0.009)	(0.034)	(0.013)	(0.029)
Confidence in information	0.022***	0.050*	0.016**	−0.015
	(0.008)	(0.027)	(0.007)	(0.015)
Networks (Leb.)			−0.001	0.024**
			(0.006)	(0.010)

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A24: Additional tests. The first two models re-run the main analysis without fixed effects for the return in 12 months outcome (model 1) and for the preparation to return outcome (model 2). The third models uses an alternative dependent variable: household member return intentions. This outcome is coded as 1 if any household member plans to return and 0 otherwise. The final model uses another outcome. The main return intentions in 12 month is coded as 1 if the respondent indicates their plan to return in 12 months and 0 otherwise. The last model uses the return in 12 months outcome but it is coded as 1 if the respondent indicates either that they plan to return in 12 months or they are uncertain about returning in 12 months and 0 otherwise.

6.6 Information and return intentions

To further explore the relationship between information and return, we examine whether information moderates the role of perceived conditions in Syria. Specifically, we examine whether conditions in Syria have a larger effect on people's intentions when they have high levels of confidence in their information about the situation in Syria.

$$Y_i = \alpha + \beta_1 T_i + \beta_2 \left(T_i \times \mathbb{1}(I_i > 0)\right) + \gamma X_i + \epsilon_i \tag{1}$$

Equation 1 is similar to the "individual indices" specification of Equation 1 in the main paper, but includes a multiplicative interaction term between each index T and confidence in information. The indicator function, $\mathbb{1}(I_i > 0)$, denotes whether a respondent i had an index value for infor-

	12 months (Ind. Indices)	Prepare (Ind. Indices)	12 months (All Indices)	Prepare (All Indices)
Safety (Syr.)	0.012**	0.073*	0.008	0.039
	(0.006)	(0.039)	(0.006)	(0.036)
Economic well-being (Syr.)	0.021***	0.075*	0.015**	0.021
	(0.006)	(0.037)	(0.006)	(0.024)
Services (Syr.)	0.018***	0.087***	0.010*	0.053*
	(0.006)	(0.030)	(0.005)	(0.026)
Networks (Syr.)	0.016**	0.094**	0.012**	0.084**
	(0.006)	(0.039)	(0.005)	(0.040)
Economic well-being (Leb.)	0.008	0.186**	0.001	0.204**
	(0.008)	(0.074)	(0.008)	(0.082)
Services (Leb.)	−0.010*	0.026	−0.012**	−0.021
	(0.005)	(0.020)	(0.006)	(0.018)
Networks (Leb.)	0.008	0.020	0.002	−0.014
	(0.005)	(0.019)	(0.006)	(0.027)
Social well-being (Leb.)	0.004	0.033	0.011*	−0.004
	(0.006)	(0.022)	(0.006)	(0.029)
Legal conditions (Leb.)	−0.003	0.004	−0.003	−0.013
	(0.006)	(0.023)	(0.005)	(0.025)
Log travel distance	0.002	−0.079*	0.007	−0.022
	(0.008)	(0.043)	(0.008)	(0.037)
Log household size	0.006	−0.066*	0.011	−0.099**
	(0.009)	(0.033)	(0.012)	(0.046)
Confidence in information	0.022**	0.055*	0.018**	0.016
	(0.008)	(0.027)	(0.008)	(0.026)

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A25: Additive indices. While the results in the paper use the first principal component of independent variables, here we construct the independent variables using additive indices. First, we standardize each variable to have zero mean and unit standard deviation. Then we summed these variables and standardized the sum again using the same method. The first two models present the results using one index in each regression for the return in 12 months outcome (model 1) and the preparation to return outcome (model 2). The last two models present the results using all the additive indices in one regression for the return in 12 months outcome (model 3) and the preparation to return outcome (model 4).

mation confidence above the mean. X denotes the same vector of covariates as the main paper's Equation 1. Figure A13 presents regression results, displaying the estimated marginal effect of a one standard deviation change in each index for people with low (below-average) confidence in information compared to high (above-average) confidence in information. To obtain standard errors, we used bootstrapping. The confidence intervals here represent the 97.5th (95th) and the 2.5th (5th) percentiles of coefficients across all bootstraps and imputations.

The results in Figure A13 suggest that the relationship between conditions in Syria and return intentions and preparations is shaped by respondents' confidence in their information sources for some key factors. Specifically, we find evidence that information is a significant moderator for the role of regime control and economic prospects in shaping return intentions. Next, we see a differential relationship between the availability of services in people's hometowns and both return intentions and return preparations. Last, we see a differential relationship between networks in Syria and return preparations depending on information confidence.

An alternative way of obtaining confidence intervals is by finding the empirical variance of

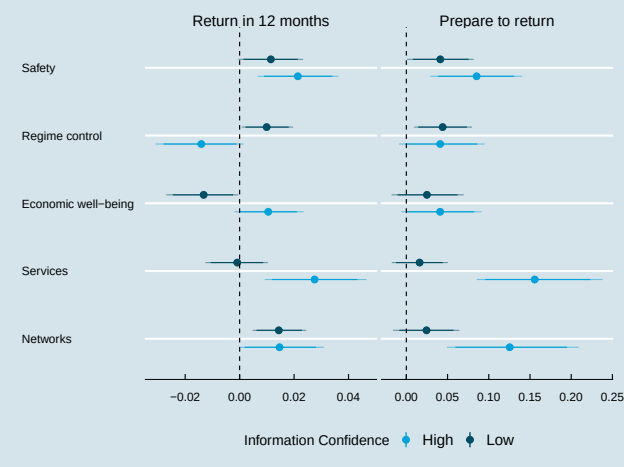

Figure A13: Interactive Effects of Information Confidence on Intentions and Preparations. Estimated marginal effects are presented with the corresponding 90% and 95% confidence intervals, with standard errors clustered by locality in Lebanon.

coefficients across bootstraps within each imputation then finding the pooled variance across all imputations using the rules of Rubin (1987). Figure A14 presents results from this approach.

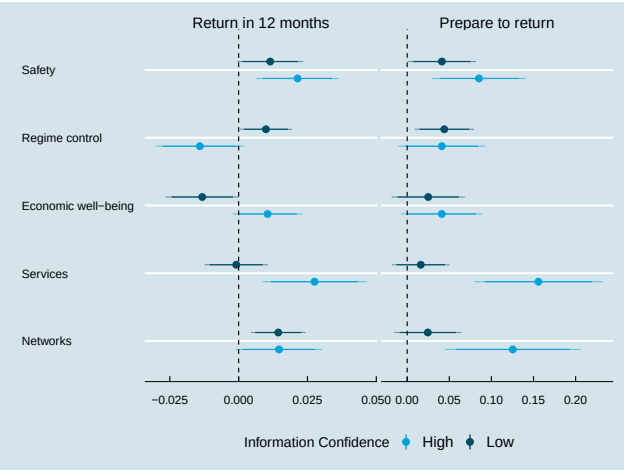

Figure A14: Interactive Effects of Information Confidence on Intentions and Preparations. Estimated marginal effects are presented with the corresponding 90% and 95% confidence intervals, with standard errors clustered by locality in Lebanon.

6.7 Testing the models' predictive power

As an additional test of the findings in Figure 2 of the main paper, we fit predictive models based on home-country factors and host-country factors using 10-fold cross validation. Below we present prediction plots, including OLS (same models as in the PAP) and lasso, with AUC results for ROC

and PR.[2] PR is often as a better performance metric than ROC for predicting rare outcomes. This is due to the fact that the ROC allows for relatively "good" performance by predicting all zeros, which is not the case with PR plots.

Looking at the results in Figures A15–A16, we see that the trends are consistent with our main findings but the differences across models are small. Looking at the ROC plot in the OLS model, we witness a ∼2.5 percentage point increase (∼4%) comparing the push model to the pull model, and a ∼4 percentage point increase (∼6%) moving from push to full (i.e., push + pull).[3] Looking at the PR curve, we see a ∼2 percentage point increase (∼25%) comparing the push model to the pull model, which is similar to the increase moving from the push model to the full model (i.e., push + pull). The gains in terms of PR AUC are large in percentage terms, although not absolute terms. The results from the lasso model are similar.

The results suggest a few key takeaways about the predictive power of the models presented in the paper. First, the Syria model is a better predictor than the Lebanon model, aligning with our main results. Second, the gains across models are large in percentage terms, but not in absolute terms. Third, overall predictive power is low and prediction is hard in our case, possibly due to studying a rare outcome, but also highlighting that understanding the aggregate drivers of return will not necessarily allow policymakers to make reliable predictions about whether an individual household will return or not.

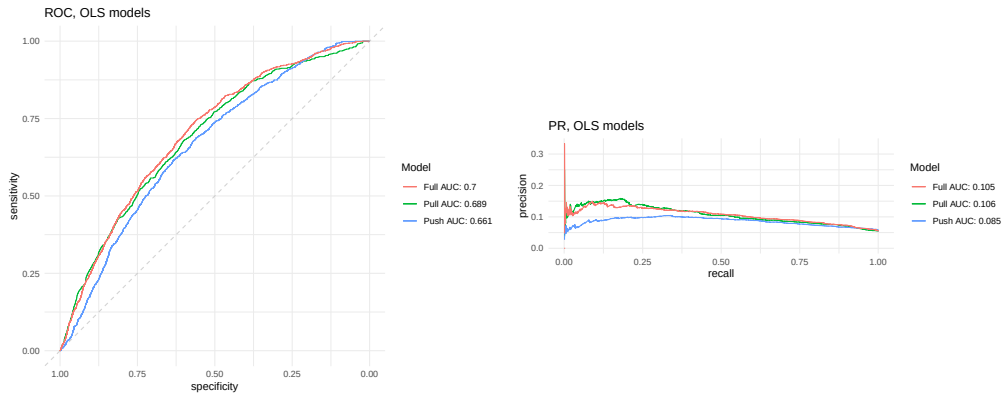

Figure A15: ROC—OLS models with pre-specified inputs of indices as predictors (left) and Principal-response curve—OLS models with pre-specified inputs of indices as predictors (right)

[2] We use mean imputation for this analysis.

[3] We indicate that these differences are approximate since their precise magnitudes will vary across different simulations.

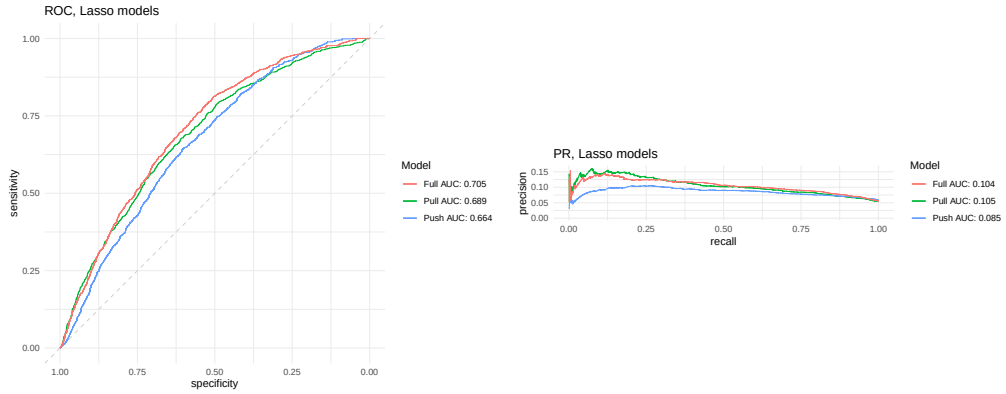

Figure A16: ROC—Lasso models with pre-specified inputs of indices as predictors (left) and Principal-response curve—Lasso models with pre-specified inputs of indices as predictors (right)

6.8 Map of conditions in Lebanon

This section further demonstrates the variation in push factors within Lebanon. Using all variables use to measure push factors in Lebanon, we construct a single index for conditions in Lebanon (extracting the first principal component from PCA).[4] Figure A17 shows the weighted average of this index by district. It should be noted that this index was constructed to have zero mean and unit standard deviations, so the variation can be measured in terms of standard deviations. The map in Figure A17 shows that push factors in Lebanon varied substantially across districts. The difference between the district with the worst conditions for respondents (El Minieh-Dennie) and the district with the best conditions for respondents (El-Nabatieh) is about 2.4 standard deviations. This map suggests that variation across districts was associated with a significant change in push factors in our survey. It should be noted that disaggregating results from a nationally representative survey does not produce estimates that are representative of the subnational units (here, districts). Nonetheless, Figure A17 clearly demonstrates that respondents in our sample exhibit significant variation in push factors across Lebanon's different districts.

[4] We use mean imputation for this analysis.

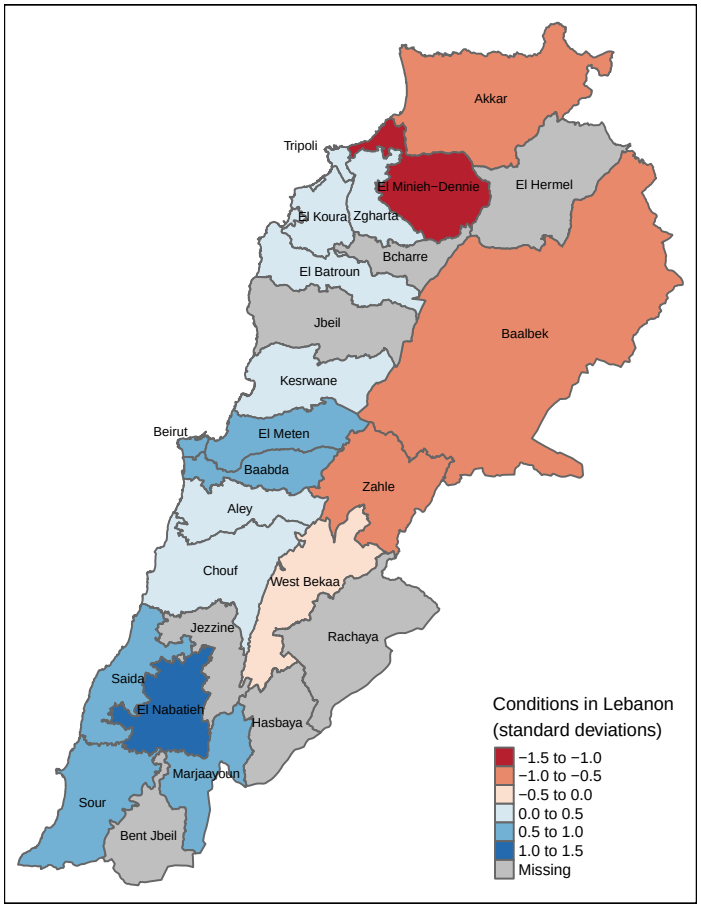

Figure A17: Map of respondents' conditions in Lebanon.

6.9 Conjoint results

As many respondents had relatively low literacy levels, the enumerators read them a vignette with the conjoint experiment. To make sure that the vignette was manageable and could be remembered by the respondent, each vignette randomly included either the condition about jobs in Lebanon (Syria) or the condition about public service provision in Lebanon (Syria). We report the regression results for the conjoint in Table A26.

7 Jordan survey

In this section, we present the set of questions we used to construct each index from the Jordan survey data. As with the Lebanon indices, we constructed these indices by extracting the first components from PCA of the input variables. Some of the questions differ from the wording used in Lebanon in order to fit the Jordan context. Due to space constraints in the Jordan survey, the

Table A26: Model 1: Main conjoint results. Model 2: Conjoint results for vignettes that included hometown is not safe AND military conscription remains. Model 3: Conjoint results for vignettes that included (hometown is safe OR all Syria is safe) AND military conscription ended.

	Model 1	Model 2	Model 3
Intercept	-0.046 $(0.012)^{***}$	0.032 $(0.012)^{**}$	0.526 $(0.024)^{***}$
Hometown Safe	0.352 $(0.010)^{***}$		
All Syria Safe	0.417 $(0.011)^{***}$		
Many Jobs in Syria	0.076 $(0.011)^{***}$	0.006 (0.014)	0.111 $(0.022)^{***}$
Public Services Available in Syria	0.077 $(0.011)^{***}$	0.006 (0.011)	0.126 $(0.022)^{***}$
Conscription Ended	0.177 $(0.009)^{***}$		
Good job in Lebanon	-0.022 $(0.011)^{*}$	-0.008 (0.012)	-0.042 $(0.021)^{*}$
Public Services Available in Lebanon	-0.028 $(0.011)^{**}$	-0.012 (0.013)	-0.044 $(0.022)^{*}$
Friends in Lebanon	-0.001 (0.009)	0.021 $(0.009)^{*}$	0.010 (0.019)
Friends in Syria	0.050 $(0.010)^{***}$	0.046 $(0.011)^{***}$	0.058 $(0.019)^{**}$
Num. obs.	14728	2487	4840
RMSE	7.344	3.608	8.430
N Clusters	2998	1778	2549

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

survey did not contain the full set of questions used in Lebanon.

Safety in Syria	**Economic well-being in Syria**	**Services in Syria**	Networks in Syria
How would you describe the risk to civilians physical safety (such as fighting, kidnapping, IEDs, crimes) in your place of origin?	How would you describe the availability of jobs at present in your place of origin?	As far as you know, how many hours per day is there electricity in your place of origin?	Approximately how many of your relatives or friends in Jordan have gone back to Syria? (capped at 15 max)
What were the main reasons for you to leave your home country Syria? (Violence/bombardment selected)		As far as you know, how many hours per day is there running water in your place of origin?	Approximately how many of your relatives or friends in Jordan have gone back to your place of origin in Syria? (capped at 15 max)
Conscription eligible (< 42 male)		As far as you know, are schools operating in your place of origin?	
	As far as you know, are health centers operating in your place of origin?		

Table A27: Jordan questionnaire: Pull factors in Syria

Economic well-being in Jordan	Services in Jordan	Networks in Jordan	Social well-being in Jordan	Legal situation in Jordan
If you think back about the situation one year ago in terms of access to assistance, has the situation improved, stayed the same, or deteriorated?	On 24 January [2019] the Prime Minister's office signed a decree informing that Syrian refugees in MOH hospitals and health centers will be requested to pay directly 80% of the applicable 'foreigners rate,' whereas from November 2014, they were treated like Jordanians who did not have health insurance and were able to access health services at subsidized rates. Has it impacted your ability to access health services?	In the last 12 months, how often did you share a meal with Jordanians who are not part of your family?	Does anyone in your family face verbal or physical harassment, meaning verbal or other actions meant to annoy, threaten, intimidate, or make someone feel scared for their safety, in the area around your house?	Do all your family members have a valid registration with UNHCR?
Have you received assistance from [government organizations, local organizations, NGOs, UN in the last two months]?	Number of school aged children out of school	Please think about the Jordanians in your phone contacts. With how many of them did you have a conversation–either by phone, messenger chat, face-to-face, or text exchange–in the last week?	Would you describe the relations with your neighbors as mostly positive/neither positive nor negative/mostly negative?	Do you have a government service card, currently called "MOI Card"?
Are your currently working?		Please think about the Syrians in Jordan in your phone contacts. With how many of them did you have a conversation–either by phone, messenger chat, face-to-face, or text exchange–in the last week?	What is your level of education?	
Is any member of your household currently working?			When did the first member of your family arrive to Jordan?	
Do you or any of your household have a work permit?			How connected do you feel with Jordanian society?	
Have you received food vouchers during the last month?			How often do you feel like an outsider in Jordan?	

Table A28: Jordan questionnaire: Push factors from Jordan

Outcome	Confidence in information
Do you think it will ever be possible to return to your place of origin in Syria?	What do you have information about [in Syria]: Respondent selects all that applies. 1 if selected: (a) Safety/security news (b) Status of infrastructure in a particular location (c) News about friends/family in Syria
	How confident are you in your knowledge about conditions in your place of origin?

Table A29: Jordan questionnaire: Outcome and confidence in information

Variable	Mean	Std.Dev	Min	Median	Max	Pct.Missing
Age (Years)	38.69	11.92	18.00	37.00	85.00	0.00%
Female	0.58	0.49	0.00	1.00	1.00	0.00%
Female head of household	0.34	0.47	0.00	0.00	1.00	0.00%
Household size	5.06	2.32	1.00	5.00	15.00	0.00%
Plan to ever return to Syria	0.04	0.19	0.00	0.00	1.00	0.00%

Table A30: Summary statistics of the outcome and demographic variables used in the Jordan study

Variable	Mean	Std.Dev	Min	Median	Max	Pct.Missing
No. relatives/friends returning to origin (censored at 15)	1.05	3.10	0.00	0.00	15.00	0.00%
No. relatives/friends returning to Syria (censored at 15)	2.24	4.65	0.00	0.00	15.00	0.00%
Eligible for conscription (male 18-42)	0.28	0.45	0.00	0.00	1.00	0.00%
Know about friends and family in Syria	0.21	0.40	0.00	0.00	1.00	0.00%
Know about status of infrastructure in Syria	0.12	0.33	0.00	0.00	1.00	0.00%
Know about safety and security in Syria	0.50	0.50	0.00	1.00	1.00	0.00%
Left Syria because of violence	0.92	0.27	0.00	1.00	1.00	0.00%
Hours of electricity in origin (0: none, 4: 24 hours a day)	1.69	1.24	0.00	1.00	4.00	29.47%
Are health centers operating in origin?	0.55	0.50	0.00	1.00	1.00	48.83%
Availability of jobs in origin (0: very bad, 3: very good)	0.44	0.65	0.00	0.00	3.00	35.38%
Confidence in info about origin (0: not at all, 4: very)	2.21	1.19	0.00	2.00	4.00	0.00%
Safety in origin (0: very dangerous, 3: very safe)	0.44	0.70	0.00	0.00	3.00	24.26%
Are schools operating in origin?	0.69	0.46	0.00	1.00	1.00	41.99%
Hours of water in origin (0: none, 4: 24 hours a day)	1.61	1.18	0.00	1.00	4.00	30.40%

Table A31: Summary statistics of the conditions in Syria used in the Jordan study

Variable	Mean	Std.Dev	Min	Median	Max	Pct.Missing
Aid change in last year (0: deteriorated, 1:same, 2: improved)	0.41	0.58	0.00	0.00	2.00	0.00%
Received aid in last 2 months	0.17	0.37	0.00	0.00	1.00	0.00%
Education Level (1: no educ, 2: primary, 3: secondary/vocational, 4: higher)	2.33	0.74	1.00	2.00	4.00	0.00%
Household member employed	0.62	0.49	0.00	1.00	1.00	0.00%
Received food voucher in last month	0.84	0.37	0.00	1.00	1.00	0.00%
Household member reported harassment in local area	0.08	0.27	0.00	0.00	1.00	0.00%
Lack access to healthcare	0.56	0.50	0.00	1.00	1.00	0.00%
Received help from neighbors	0.17	0.38	0.00	0.00	1.00	0.00%
Possess Ministry of Interior ID	0.98	0.12	0.00	1.00	1.00	0.00%
Share meal with Jordanians (1: Never, 5: Almost daily)	2.19	1.16	1.00	2.00	5.00	0.00%
Conversations with Jordanians in past week (1: 0, 5: 15 or more)	2.53	1.16	1.00	2.00	5.00	0.00%
Conversations with Syrians in Jordan in past week (1: 0, 5: 15 or more)	3.36	1.16	1.00	3.00	5.00	0.00%
Lack access to school	0.14	0.35	0.00	0.00	1.00	0.00%
Relation with neighbords (0: negative, 1: neither, 2: positive)	1.81	0.41	0.00	2.00	2.00	0.00%
Connection to Jordanian society (1: not at all, 5: extremely close)	3.90	1.02	1.00	4.00	5.00	0.00%
Feel like an outsider (1: always, 5: never)	4.12	1.05	1.00	5.00	5.00	0.00%
All household members registered with UNHCR	0.98	0.14	0.00	1.00	1.00	0.00%
Any household member has work permit	0.25	0.44	0.00	0.00	1.00	0.00%
No. years since first household member arrived in Jordan	1.69	0.94	0.00	1.00	3.00	0.00%

Table A32: Summary statistics of the conditions in Jordan used in the Jordan study

8 Ethical Considerations: Survey

The ethical imperative to do no harm is especially pressing in research with refugees, given their extreme vulnerability (Masterson and Mourad, 2019). We designed this project to reduce potential harm, maximize policy relevance, and increase opportunities for direct benefits to research participants. The achieve the first two goals, the authors drew on exploratory fieldwork, interviews with international and local humanitarian actors, and a research planning workshop with the humanitarian community in Beirut during which we discussed our research plan and questionnaire to minimize potential harm and ensure that the design can provide the humanitarian community with required evidence to fulfill their needs. To address the first and the third goals, the research team partnered with NGOs to provide protection training to enumerators and established a referral mechanism through which research participants in need of humanitarian services were connected to available resources. Below we discuss some of the measures we have taken.

Compensation

Survey interviews took about 30-40 minutes to complete and we provided survey respondents with $10 cash compensation for their time. We decided to compensate respondents following extensive interviews with humanitarian actors, as many humanitarian actors suggested that it is only fair to compensate participants for their time following a somewhat long survey.

Engagement with humanitarian actors

To strengthen the research contribution, we closely consulted with humanitarian actors at all stages of the project. At the outset, we worked with humanitarian actors to conduct focus groups and meetings with refugees in Lebanon and Jordan. We consulted with humanitarian actors closely on the development of the questionnaire. Prior to data collection, we conducted a workshop with humanitarian actors in Beirut, Lebanon, in March 2019 to present the whole project and get feedback, including on the questionnaire. We then revised the questionnaire based on comments (for instance, we removed questions that directly measured political attitudes, which they suggested were too sensitive for the context). In June 2020, we conducted several online workshops with humanitarian actors in Lebanon to share results and answer questions from humanitarian actors.

Referrals

When we presented the research design and questionnaire to humanitarian organizations in March 2019, they suggested developing a referral strategy for refugees who require or ask for help. The

goal of referrals is to help facilitate refugees' access to services by either (1) putting individuals in need of services directly in contact with the service providers or (2) enabling people to seek assistance and support them in receiving assistance. Enumerators and team leaders who participated in data collection received special training from a humanitarian organization in Lebanon on the goals of referrals, when to refer respondents, and how to refer respondents. In addition to the training, enumerators received guidelines that were built using material that is used by humanitarian organizations. Humanitarian organizations in Lebanon collaborate on a centralized referral platform called Referral Information Management System (RIMS), in addition to traditional referral methods (by reaching out directly to the responsible organization). As RIMS was not available for use by researchers (only humanitarian actors could use it), the research team collaborated with a local humanitarian organization to conduct the referrals and compensated them for the time they spent carrying out needed referrals for this project's research participants.

Figure A18 summarizes the referral process and was included in the enumerator guide:

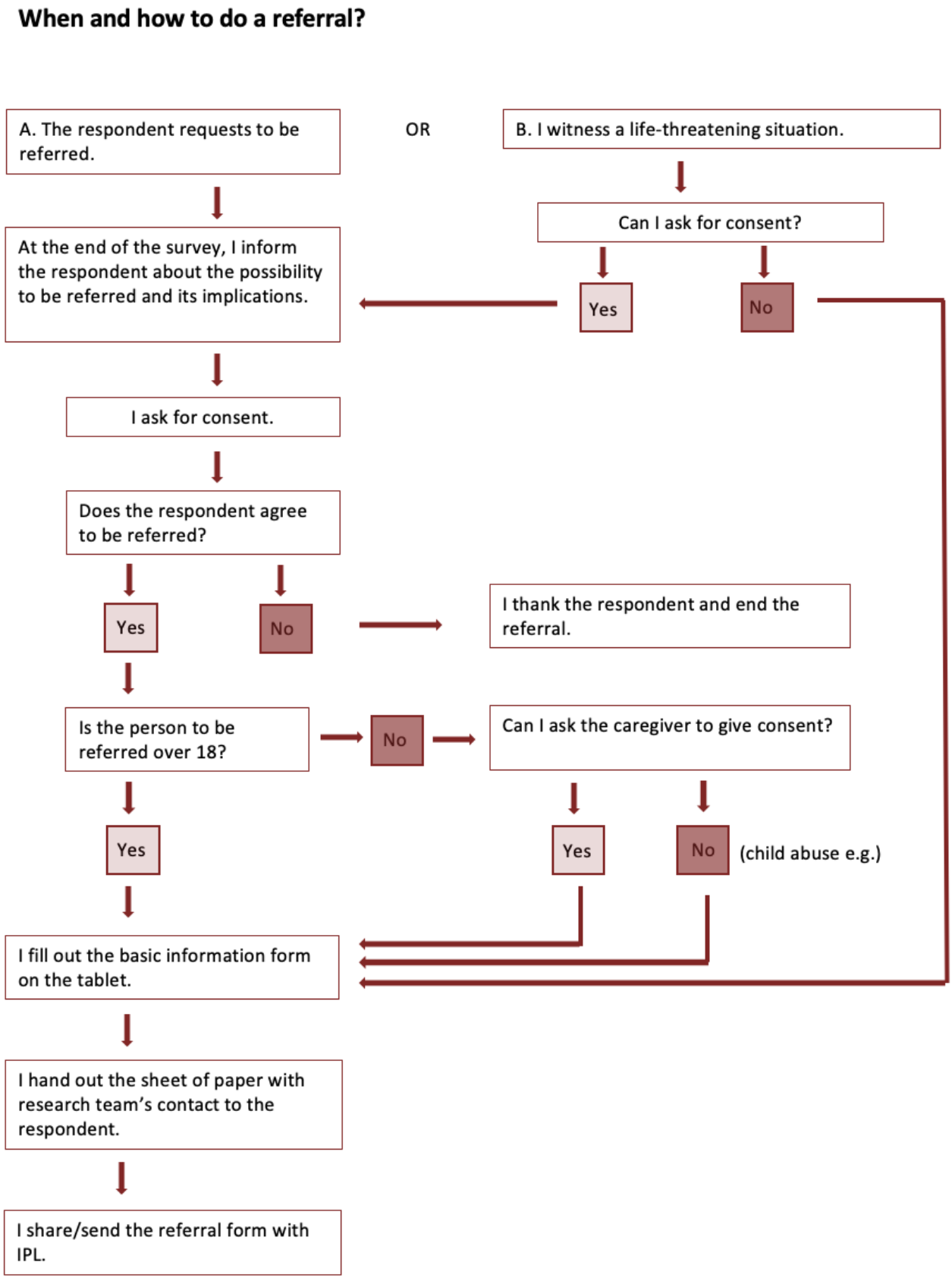

Figure A18: Referral guide summary

9 Qualitative Interviews: Details and Ethical Considerations

In order to complement the quantitative survey, we also conducted repeated in-depth interviews with 20 Syrian families who were in Lebanon as of January 2020. The ethical concerns of these qualitative interviews are distinct from, although we believe no greater than, the quantitative survey. One important aspect of the modification is that the in-depth interviews with 20 participants were audio-recorded and transcribed.

Due to the timing of the beginning of qualitative data collection and the onset of the global Covid-19 pandemic, all qualitative data collection took place through WhatsApp.

Table A33 describes some of the characteristics of the respondents in the qualitative interviews as well as when these interviews took place.

ID	Interview Number	Month/year	Residence	Origin	Gender	Age	Comment
1	1	April 2020	Beirut	Homs	Male	55-59	
1	2	February 2021	–	–	Male & Female	–	Wife took over in middle of interview
1	3	May 2021	–	–	Female	–	Wife answered this interview
2	1	May 2020	Greater Beirut	Aleppo	Female	30-34	
2	2	June 2020	–	–	–	–	
2	3	February 2021	–	–	–	–	
2	4	May 2021	–	–	–	–	
2	5	August 2021	–	–	–	–	
3	1	March 2020	Beirut	Deir Ez-Zor	Female	35-39	
3	2	December 2020	–	–	–	–	
4	1	April 2020	Akkar	Hama	Male	40-44	
4	2	December 2020	–	–	–	–	
4	3	June 2021	–	–	–	–	
5	1	March 2020	Akkar	Hama	Male	30-34	
6	1	May 2020	Beirut	Hassakeh	Female	25-29	
6	2	June 2020	–	–	–	–	
6	3	June 2021	–	–	–	–	
7	1	March 2020	Beirut	Deraa	Male	40-44	
8	1	March & April 2020	Beirut	Homs	Female	25-29	
8	2	June 2020	–	–	–	–	
8	3	February 2021	–	–	–	–	
8	4	June 2021	–	–	–	–	
9	1	April 2020	Mount Lebanon	Hama	Male	20-24	
9	2	November 2020	–	–	–	–	
9	3	January 2021	–	–	–	–	
9	4	May 2021	–	–	–	–	
10	1	February 2020	Beirut	Aleppo	Female	55-59	
10	2	January 2021	–	–	–	–	
10	3	June 2021	–	–	–	–	
11	1	February 2020	Beirut	Deir Ez-Zor	Female	40-44	
11	2	April 2020	–	–	Female & Male	–	Husband took over in middle of interview
11	3	November 2020	–	–	Male	–	Husband answered this interview
11	4	June 2021	–	–	Male	–	Husband answered this interview
12	1	February 2020	Mount Lebanon	Hama	Male	30-34	
12	2	April 2020	–	–	–	–	
12	3	February 2021	–	–	–	–	
13	1	May 2020	Mount Lebanon	Hama	Male	30-34	
14	1	March 2020	Greater Beirut	Aleppo	Male	30-34	
14	2	January 2021	–	–	–	–	
14	3	May 2021	–	–	–	–	
15	1	March 2020	Akkar	Hama	Male	30-34	

Table A33: Qualitative Interview Details

Risks

As discussed above regarding survey interviews, we do not believe that interview participation is likely to exacerbate risk for Syrians in Lebanon.

Risks from our interviews over WhatsApp calls are minimal. We developed our data collection

approach through WhatsApp calls through consultation with digital rights groups, and phone-based in-depth interviews avoided sensitive topics.

One potential risk would be that phone calls were monitored or overheard. Monitoring of WhatsApp communication would require that someone's phone be confiscated or they be overheard speaking with us. If someone's phone were confiscated, a phone call record (but not the content of the calls) would be visible. Text message records would be visible. To avoid risk from these situations we did not ask any sensitive questions over text.

If someone were overheard speaking with us by a bystander, that person would hear whatever the interview participant says, but not our questions. Although we deemed these circumstances unlikely, we wanted to take special precaution to reduce this potential threat, especially for anyone in Syria. Therefore, before any interview began, we asked participants a series of questions that gave them the opportunity to not take the call if they were not in a safe position to do so, without needing to say anything that would be suspicious. For instance, "Is now a good time for you to talk?" and "Would it be better if I called back later?" allow people to say simple "yes" and "no" responses if speaking is not safe. Four of the qualitative interview participants moved to Syria during the research time period and two others traveled elsewhere. As a result, those six people were not interviewed.

We designed the study to minimize the risk of a data breach or confidentiality breach. All data is stored on a secure, encrypted, GDPR-compliant server. All contact information and identifying information was stored separately from interview data and is not part of the audio recording but rather filed in a separate document that was password-protected and stored on a secure server. The crosswalk code is encrypted and only available to the research team.

Psychological risks

Research can create a risk of psychological harm if researchers do not anticipate the risk of re-traumatizing people. Such questions could include asking about people's wartime experiences in Syria, their journey to Lebanon and/or things they lost because of the war in Syria. To address these concerns, interview content continually aimed to avoid re-traumatizing participants, which could arise from seeking out information about violence or loss.

The semi-structured nature of baseline interviews and open-ended nature of follow-up interviews allowed participants to bring up their own stories as they choose, telling us only what they feel comfortable discussing. One important goal of the first qualitative interview with each participant was learning about potential sensitivities that they face. We did not elicit discussion of sensitive topics in general, and made specific efforts with each participant to avoid issues that are sensitive

for them personally.

Additionally, we pilot tested the interview guide to gauge and minimize risks. In the pilot, following each potentially sensitive section and at the end of the interview, the interviewer debriefed with participants about their experience and perception of the interview questions. This was intended to identify questions or sections that provoke negative reactions or anxiety.

The study author carrying out the interviews is well-positioned and trained to address potential protection concerns that may arise during the research. The author is an experienced interviewer who has conducted research work with similarly vulnerable populations in multiple countries and fluently speaks Levantine Arabic. The author has attended several protection trainings, the most recent one delivered by Oxfam in Beirut in July 2019 for enumerators of this study.

If participants needed to be referred to specific NGOs for case management, the team used the referral system described in Appendix Section 8 In the end, no participants in the qualitative interviews were referred to NGOs for support through our referral system.

Although we did not collect data that would put participants at psychological risk if the responses were revealed, we designed the study to minimize the risk of a data breach or confidentiality breach, as discussed directly above.

Privacy and confidentiality

WhatsApp calls and voice messages facilitate privacy because participants do not need to travel or even leave their homes for data collection, and are always free to refuse to answer a call. As discussed above, we also asked screening questions at the beginning of each qualitative interview that gave people the opportunity to safely decline the interview, if they were not in a good situation to speak on the phone with us.

Data was uploaded from a password-protected audio recording device to MedBox shortly after interviews, and then deleted from the audio recording device.

Informed consent

The research team obtained verbal consent rather than signed consent. There are two reasons for not asking participants to sign anything. First, many Syrian refugees are semi-literate or illiterate. Asking them to sign a form that they cannot fully understand would likely cause anxiety and mistrust. Second, even among Syrians who can read, asking refugees with precarious legal status to sign formal documents will create anxiety and fear about what exactly the document states and how it will be used.

Informed consent during the first interview explained the repeated nature of participation in the research study. Semi-structured interviews were only conducted with selected participants if and after they gave consent.

Verbal consent statement

1. DESCRIPTION: My name is Marine Casalis, I am working on a study with researchers from the universities ETH Zurich, Stanford University, and the University of California, Santa Barbara. We interviewed you .... months ago in ... and you agreed to be re-contacted. As you may remember, our goal is to understand the situation of Syrians living in Lebanon and how they see their future, which will inform the work of humanitarian agencies and the academic community. Your answers will be confidential and will not be shared with anyone outside of the research team. I would like to ask you some questions with the aim of having a better understanding of your living conditions. This interview will be different from the first one you did with us because you will have the opportunity to speak much longer and share your experience with me.

I would like to audio record this interview, only your voice, not taking pictures of you or filming you, so that I can fully focus on what you are telling me and make sure I have an accurate account of what you are sharing with me. This will also allow me when I listen to it again in the coming weeks/months to transcribe truthfully what you told me. I will not use this audio recording beyond the research team without asking you again for consent.

In addition to interviewing you today, we would like to learn about how your situation changes in the future by corresponding with you further in the future, including via WhatsApp and possibly more in-person interviews like today. We will reach in touch with you every 1-3 months, and you are also always free to contact us if you would like to follow-up on anything from the interview.

2. TIME INVOLVEMENT: This is very valuable to us so we thank you for your time. The interview will take roughly 60 minutes but you can stop it at any given time.

3. RISKS AND BENEFITS: Risks include possible loss of confidentiality, and the survey includes questions that you may find upsetting. We cannot and do not guarantee or promise that you will receive any benefits from this study.

We will provide you with approximately $10 in phone credit for each long interview.

Your decision whether or not to participate in this study will not affect your eligibility for any services the UN Refugee Agency or any other humanitarian aid provider.

4. PARTICIPANT'S RIGHTS: If you agree to participate in this project, please understand that your participation is voluntary and you have the right to withdraw your consent or discontinue participation at any time without penalty or loss of benefits to which you are otherwise entitled. The

alternative is not to participate. You have the right to refuse to answer particular questions. The results of this research study may be presented at scientific or professional meetings or published in scientific journals. Your individual privacy will be maintained in all published and written data resulting from the study.

5. Identifiers might be removed from identifiable private information and, after such removal, the information could be used for future research studies or distributed to another investigator for future research studies without additional informed consent from you.

6. CONTACT INFORMATION: a. Questions: If you have any questions, concerns or complaints about this research, its procedures, risks and benefits, contact the Protocol Director at [XXX]. This number is only reachable via WhatsApp and Viber and you can message us or call us this way for free.

b. Independent Contact: If you are not satisfied with how this study is being conducted, or if you have any concerns, complaints, or general questions about the research or your rights as a participant, please contact the Stanford Institutional Review Board (IRB) to speak to someone independent of the research team at (650)-723-2480 or toll free at 1-866-680-2906, or email at IRB2-Manager@lists.stanford.edu. You can also write to the Stanford IRB, Stanford University, 3000 El Camino Real, Five Palo Alto Square, 4th Floor, Palo Alto, CA 94306.

Do you have any questions?

Do you agree?

References

Abadie, Alberto, Susan Athey, Guido W Imbens and Jeffrey Wooldridge. 2017. "When should you adjust standard errors for clustering?".

Hainmueller, Jens, Dominik Hangartner and Giuseppe Pietrantuono. 2017. "Catalyst or crown: Does naturalization promote the long-term social integration of immigrants?" *American Political Science Review* 111(2):256–276.

Kling, Jeffrey, Jeffrey Liebman and Lawrence Katz. 2007. "Experimental analysis of neighborhood effects." *Econometrica* 75(1):83–119.

Masterson, Daniel and Lama Mourad. 2019. "The Ethical Challenges of Field Research in the Syrian Refugee Crisis." *APSA MENA Politics* 2(1).

Rubin, Donald B. 1987. *Multiple imputation for nonresponse in surveys*. John Wiley & Sons.