

Main analysis from Survey (and estimation)

Our main analysis will be a within-subject design, where the treatment is being asked about scenarios if you were to move to the U.S., compared to the same scenario of you were to stay in Mexico. We randomize at the individual level the order in which we ask these scenarios, and will control for order in our analysis.

Our main estimating equation will then be for individual i facing scenario s

$$Y_{i,s} = \alpha_i + \beta US_s + \gamma USaskedfirst_{i,s} + \varepsilon_{i,s}$$

Where here $Y_{i,s}$ is the response for individual i under scenario s , and then we are interested in the effect of the scenario being the US vs Mexico. This will be the coefficient β , on a dummy variable US which takes value 1 if the scenario being asked about is the U.S., and 0 if the scenario is with respect to Mexico. We control for the randomized order $USaskedfirst_{i,s}$, a binary taking one if the questions in the US scenario are asked first, i.e., before the questions in the Mexico scenario. We control for individual fixed effects, and then cluster the standard errors at the individual level (since we have 2 responses per person).

Primary Outcomes

1. *Subjective Probability of getting married in 10 years* (B_US1 and B_Mex1). This is an ordinal variable with values 1 through 5. We will look at the distribution of this variable under the Mexico scenario, and create a binary outcome that is either “Extremely likely”, or if the proportion saying extremely likely is less than 25%, “Extremely likely” or “likely”. We will also look at the impact on the ordinal variable (coded 1 through 5).
2. *Unable to say what characteristics of future spouse will be*. This is the sum of the unable to say responses for B_Mex2, B_Mex3 and B_Mex4 for the in Mexico case, and of B_Us2, B_Us3, and B_Us4 for the U.S. case – measuring unsureness about the age, place of birth, and education of the spouse. We will also look at the unable to say response separately for each of the three spouse characteristics: age, education and place of birth.
3. *Hard to picture other characteristics of future spouse*. This is an ordinal variable ranging from 1 = very uncertain to 5 = certain, asked in B_Mex5 and B_US5. We will use this as an ordinal variable, but also look at the impact on the binary variable “very uncertain”. If the proportion of the latter is less than 25%, we will also look at the binary variable taking one if “very uncertain” or “uncertain”.

Descriptive direct responses:

1. *Direct response on whether it is harder to picture the spouse if they migrate*: proportion answering B9 = 3 (no regression).
2. *Variance in the responses about spouse characteristics (secondary outcome)*
In each scenario (Mexico and US), we will compute the variance of the responses given for each spouse characteristics (age, education, birthplace), excluding the “unable to say” response. We will then compare the variance of the responses between the US and the Mexico scenario, separately for each of these three characteristics. We will also construct an index of spouse characteristics as the first component of a PCA using responses about spouse age,

education and birthplace, and examine the variance of this index in the Mexico and US scenario.

3. *Direct responses on whether marriage is important in the migration decision:*
 - a. Absolute terms: Based on Outcome C4_4 (how important is marriage)
 - b. Relative terms: Comparing the proportion who say Outcome C4_4 is very important or important (3 or 4), to those saying this for wages, legal status, and career prospects.

Secondary analysis: We also conduct a between-individuals experiment to look at whether asking questions about marriage and migration change the expressed desire to migrate to the U.S.

Here, we will look at how migration intentions differ between respondents who

- answered questionnaire Block C before Block B (control)
- answered questionnaire Block B before Block C (treated)

to assess whether getting asked about migration and family first change migration intentions. We randomize at the individual level the order in which we ask Block C and Block B.

We estimate the following regression (between-subject design):

$$Y_i = \alpha + \beta \text{ Block B before C} + \varepsilon_i$$

Migration intentions are measured with three outcomes.

Outcome 1: ideal desire to migrate (Question C1)

Outcome 2: Concrete steps (Question C2)

Outcome 3: likelihood of migrating in next couple of years (Question C3)