Methodology note

Overview of Study Design

The study employed a multistage cluster sample survey design in the six settings. Cross-sectional household surveys were conducted with the female head of household and were then followed with surveys with 1-2 adolescent household members.

While it was initially planned for all studies to use stratified cluster sampling, Djibouti deviated from the planned design owing to the nature of the concentration of refugees in this setting. Instead, a purely stratified design with random or in some places a full census of eligible households was undertaken in which refugees as well as host community comparators were sampled in four locations: Djibouti City and Obock refugee urban locations, Holl Holl and Ali Addeh refugee camps. Local data collectors were recruited by the Ministry of Women and underwent human subjects research training. At the start of data collection in each zone, researchers used a random number generator to select an initial household as a starting point. This was method was applied only for the samples of refugees. For the host community sample, an initial host community household was selected using 5-6 geographically randomized start-points in the vicinity of the refugee camps (or non-camp community in Djibouti City), and 15-20 next-nearest households were interviewed in each of those areas. Data collection took place on tablets and data was uploaded onto Magpi software (Version 6.1.1, Magpi, 2019).

In Bangladesh, the study employed a cluster stratified design with strata of registered refugees as well as FDMNs. Fifty-one clusters were selected with probability proportional to size from 27 camps. In addition to including female adolescents, male adolescents were surveyed; however, this analysis only focuses on responses of adolescent FDMN females. A camp population enumeration was provided by the International Organization for Migration’s (IOM) Needs and Population Management which divides the camp population into roughly evenly sized blocks. For the FDMN strata, the first sampling unit comprising 45 blocks (considered clusters) were selected, with sampling proportional to size. Households constituted the second sampling stage and were selected using a random start point. A random direction was chosen from the center point of each block, and households were selected by identifying the next nearest doorway. For the FDMN clusters, around 20 households were selected per cluster. Data collection took place on tablets and the data was uploaded onto Magpi software (Version 6.1.1, Magpi, 2019).

In Nepal, the sample was stratified evenly between Sindhupalchowk and Dolakha districts and 30 clusters were selected in each district with probability proportional to size. In each of the 30 clusters, around 20 households were selected. The first household was selected in each cluster by choosing the center of the district as the starting point and selecting the closest house/residence. Then, households were chosen by identifying the next nearest doorway. Like Bangladesh, male adolescents were surveyed in Nepal; however, the analysis presented in this paper only focuses on responses of adolescent girls. All data collection in Nepal and Bangladesh was conducted by local enumerators who were trained in human subjects research. In both locations, data collection teams were composed of a male and female to guarantee that male and female respondents were surveyed by individuals of their own sex. Data collection took place on tablets and the data was uploaded onto Magpi software (Version 6.1.1, Magpi, 2019).

In Lebanon, the sample was stratified across three districts – Tyre, Saida, and Nabatieh – and respondents were sampled from a total of 125 clusters in the three locations. All data collectors were females who received a five-day training on the research protocol, survey design and sampling methodology, human subjects research, data quality and study tools. Because of the complexity of
Methodology note

conducting surveys among urban refugees, snowball sampling was used to initiate data collection in each cluster. The sampling started every day at charity foundations serving Syrian refugees. A randomly selected woman selected from the charity and asked to lead the female interviewers to her neighborhood. When the interviewers reached a household that agreed to participate, the first interview in the cluster was conducted. As was done in the other settings, once the first household was identified, data collectors used a recruitment script to introduce the study and ascertain study eligibility. Upon completing the survey, respondents were asked for a referral or introduction to the nearest Syrian household. To ensure representativeness and avoid geographic clustering, no more than three households in the same building were included. Data collection took place on paper and data was then entered electronically.

In Yemen, the sample was stratified between displaced and host communities in three governorates – Sana’a, Ibb and Yemen – and 13 clusters were selected by district with probability proportional to size. Within each governorate, clusters were selected by district with probability proportional to size. Sampling at the cluster level was done by selecting starting points at district central locations and proceeding in a randomly selected direction, choosing the next nearest house that met study eligibility until the targeted cluster number was achieved. Restrictions in Yemen on electronic data collection resulted in paper data collection. Data was then later entered into Magpi.

In Iraq, the sample was stratified into three governorates – Erbil, Sulaimani, Erbil, and Dohuk – and three populations - Iraqi Kurdish host communities, Iraqi IDPs and Syrian refugees. Clusters were identified in both camp and non-camp settings, utilizing UN High Commissioner for Refugees (UNHCR) and International Organization for Migration (IOM) sampling frames, and then randomly selected at the district level with probability proportional to size (PPS). Random number generators were utilized to select random starting locations within clusters to begin household selection. Households were then visited sequentially from the starting location until eligible households were interviewed per cluster. Data collection took place on tablets using Magpi.

Sample size
In all locations, sample sizes were calculated with the intent of measuring proportion of child marriage among adolescent girls with sufficient precision. Absent data on child marriage prevalence, a prevalence of 50% was presumed to maximize sample size. The margin of error was set at ±5% and the design effect for cluster sampling was assumed to be 2. Accordingly, a minimum sample size of approximately 800 households was needed to estimate proportion of adolescents married as children in each setting. This number was divided among strata in settings where a stratified sample was used. In some locations, the sample size was not fully met due to budget constraints or logistical problems.

Selection criteria
In all locations, respondents were screened for eligibility. Once the first household was identified, enumerators used a recruitment script, through which they introduced participants to the study, its aims, and objectives. They then solicited information from a household representative about household composition to ascertain household eligibility. Households were eligible to participate if they fulfilled the following criteria:

Inclusion criteria:

- The household must have had at least one female adult and at least one adolescent female aged 10-19 living in the household for at least one month in the previous 12 months.
Methodology note

- Any adolescent females aged 10-19 who resided in the household for at least one month in the past year were eligible to complete the adolescent survey. Only one married 10-19 year old female, and one unmarried 10-19 year old female (maximum 2) were interviewed. If there were more than two girls in this age group, a maximum of two were randomly selected.
- In Bangladesh, households were only eligible if they arrived before 1992 or after October 2016. In Nepal, households were only eligible if they resided in the district during the time of the earthquake.

Exclusion Criteria:

- Households with no females 10-19 living in the household at least one month in the previous 12 months were not included in the survey.
- After three visits, if an eligible female respondent for the household survey was not found, the household was excluded.
- For internally displaced persons, any individuals who reported being displaced more than 4 years ago were not included.

If the household was deemed eligible and the household representative expressed willingness to participate, the study team member either identified a time to return to resume the interview or consented the respondent into the study.