**Supplementary Material**

## TRANSMIT Study Design and Sampling.

### Target population and sampling strategy

The target population for the TRANSMIT Surveys included individuals with a minimum age of 15 years who were in either Lebanon or Turkey at the time of the surveys who were living in private housing (e.g. formal camps , detention centers or barracks,). The survey was stratified by country of birth of the head of household, such that 50 percent of respondents were from Syrian households and 50 percent were from non-Syrian households. The sample of the present study consists of the second wave of data collections in Lebanon an the first wave of data collection in Turkey, both collected between September 2020 and February 2021 as part of the TRANSMIT research project.

### Baseline Sampling and Follow-up Lebanon

The initial survey in Lebanon took place in 2019. The number of respondents was 1252. As there are no publicly accessible population registers in Lebanon, respondents were selected using area sampling combined with the random walk method. For this purpose, five hierarchical lists were created on five geographical levels (governorates, zones within the governorates, sub-areas within the zones, clusters within the sub-areas and neighbourhoods within the clusters). In order to create a sample that is as representative as possible, information on the demographic and socio-demographic composition of the population (religious affiliation, socio-economic status), as well as the degree of urbanisation and geographical characteristics, is taken into account when selecting the sub-areas. From the entirety of the neighbourhoods, 200 neighbourhoods are then randomly selected as so-called "Primary Sampling Units" (PSU). In order to achieve the target size of the Syrian sample while remaining economically viable, only neighbourhoods with an above-average proportion of Syrians are included in this selection. For this purpose, data from previous surveys and the field expertise of our survey institute are used Within the selected neighbourhoods, households are randomly selected by random walk and asked to participate.

The participants from the baseline survey were contacted again on the basis of the addresses where the previous interview took place. If the interviewees could not be found, two further visits were made a few days apart. If it was still not possible to make contact, an attempt was made by telephone to arrange an interview appointment in those cases where telephone data had previously been provided. If this was also not possible or the respondent refused to participate in the study again, the contact attempts were cancelled and the respondent's contact details were permanently deleted.

If participants could not be interviewed again, new people were recruited for the survey using a random walk in the same neighbourhood ("refreshment sample"). The nationality of the head of household was again used as the selection criterion for a household, so that half of all respondents in the overall survey again came from Syrian households. As before, the selection within a household took place by means of random selection among all household members present who were 15 years or older.

### Basline Sampling Turkey

The number of respondents in the baseline survey was 3045. As in Lebanon, this survey was also stratified according to the country of birth of the head of household, so that 50% of the respondents came from Syrian households and 50% from non-Syrian households. Equivalent to Lebanon, the population here also includes all non-institutionalised persons aged 15 years or older in the country at the time of the survey. In Turkey, too, no register data on the Syrian population in the country is publicly available, so the respondents were also selected via area sampling. For this purpose, based on publicly available, aggregated data on the distribution of the Syrian population, the districts with the highest and second-highest proportion of Syrians in each province (cf. federal states) of the country were selected. Within these districts, neighbourhoods were again randomly selected as PSUs. As in Lebanon, respondents were selected by random walk and by randomly selecting the household member to be interviewed.

Since no reliable registry data for the Syrian population in Turkey was available, the selection of respondents took place through multilevel area sampling combined with the random walk method. Based on publicly available aggregate data on the distribution of the Syrian population, the districts (cf. counties) with the highest and second highest Syrian population shares were selected in each region (cf. states) of Turkey. This increases cost efficiency of the survey and is commonly done for hard-to-reach population samples which require random walk sampling due to a lack of registry data. Within these districts, neighborhoods were again randomly selected as Primary Sampling Units (PSU). The selection of respondents took place via random walk. For this, interviewers walked the streets in the selected neighborhood according to predetermined rules to select households for participation. Within the households that agreed to participate the choice of the interviewee (s) is also randomized (within age requirements). If the random walk in the respective neighborhoods did not yield sufficient observations, the survey continued in previously defined alternative neighborhoods until the respective quota of Syrian and Turkish respondents was met. For the refreshment sample, the same sampling procedure was applied. After fraudulent behaviour by a few enumerators was detected by means of rigorous statistical benchmarks (see Gundacker et al., forthcoming for details), 313 observations were removed from the sample, yielding the final sample size of 2732.

Gundacker, L., Hertner, L., Ruhnke, S. (forthcoming). Six years after the EU-Turkey Agreement: A quantitative assessment of the living conditions of Syrians in Turkey. IAB-Forschungsbericht.

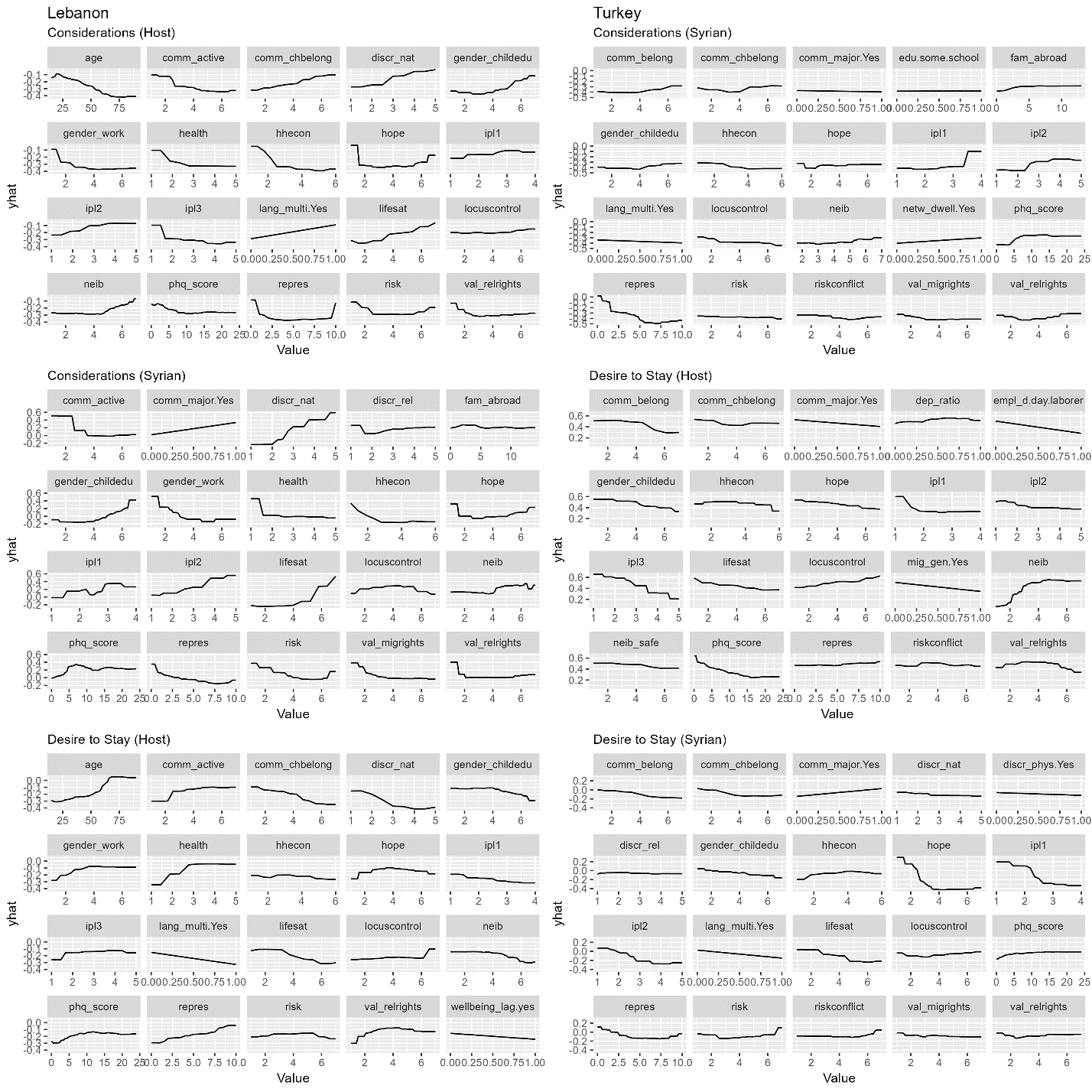


Figure S1: Partial Dependence Plots after model estimation using Random Forest

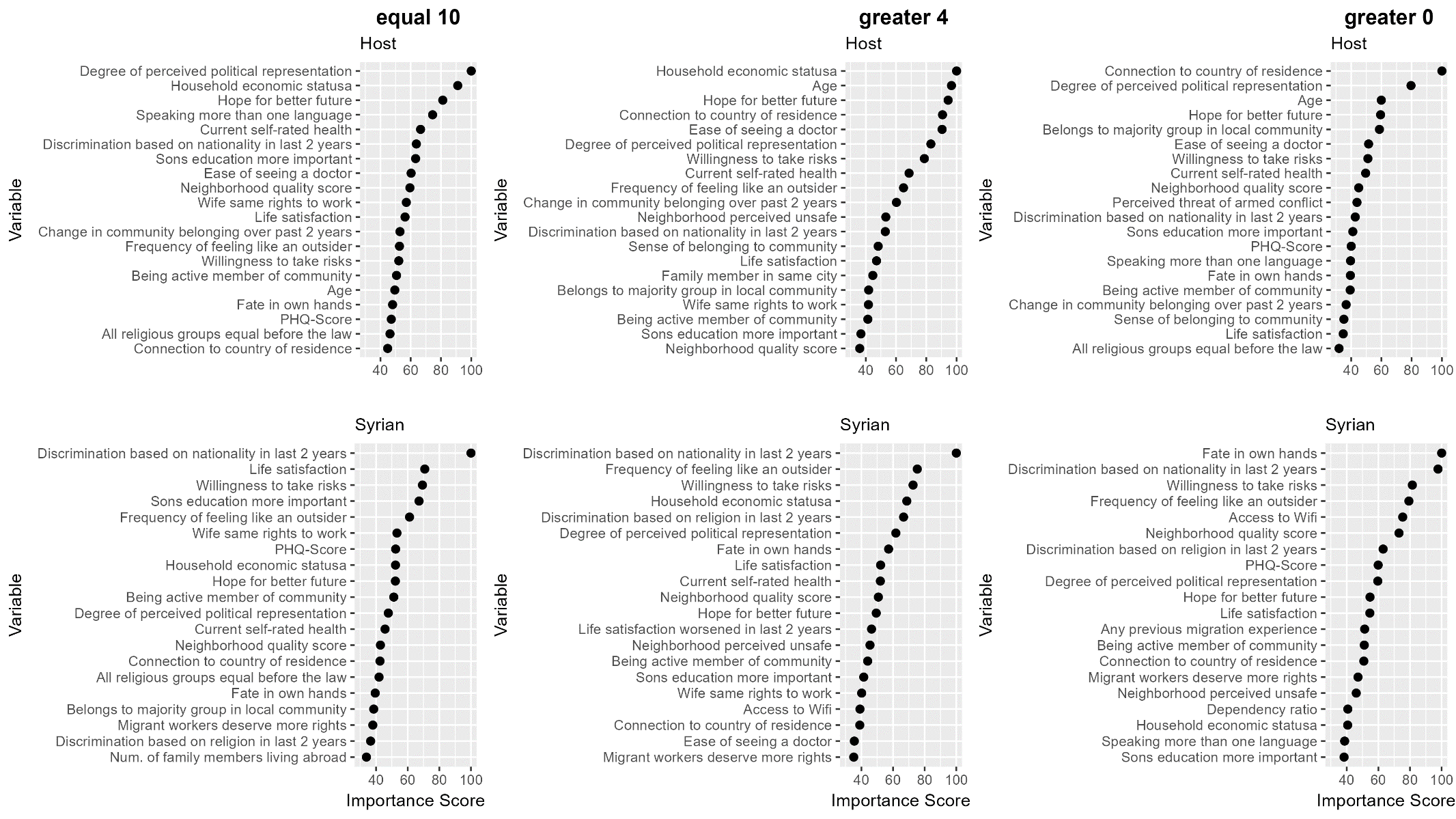
**LEBANON** 

Figure S2: Permutation based importance scores derived from Random Forest for Syrian and Host Population for consideration to move in Lebanon at different cut-off points for the binary coding of the 0-10 Scale in Lebanon

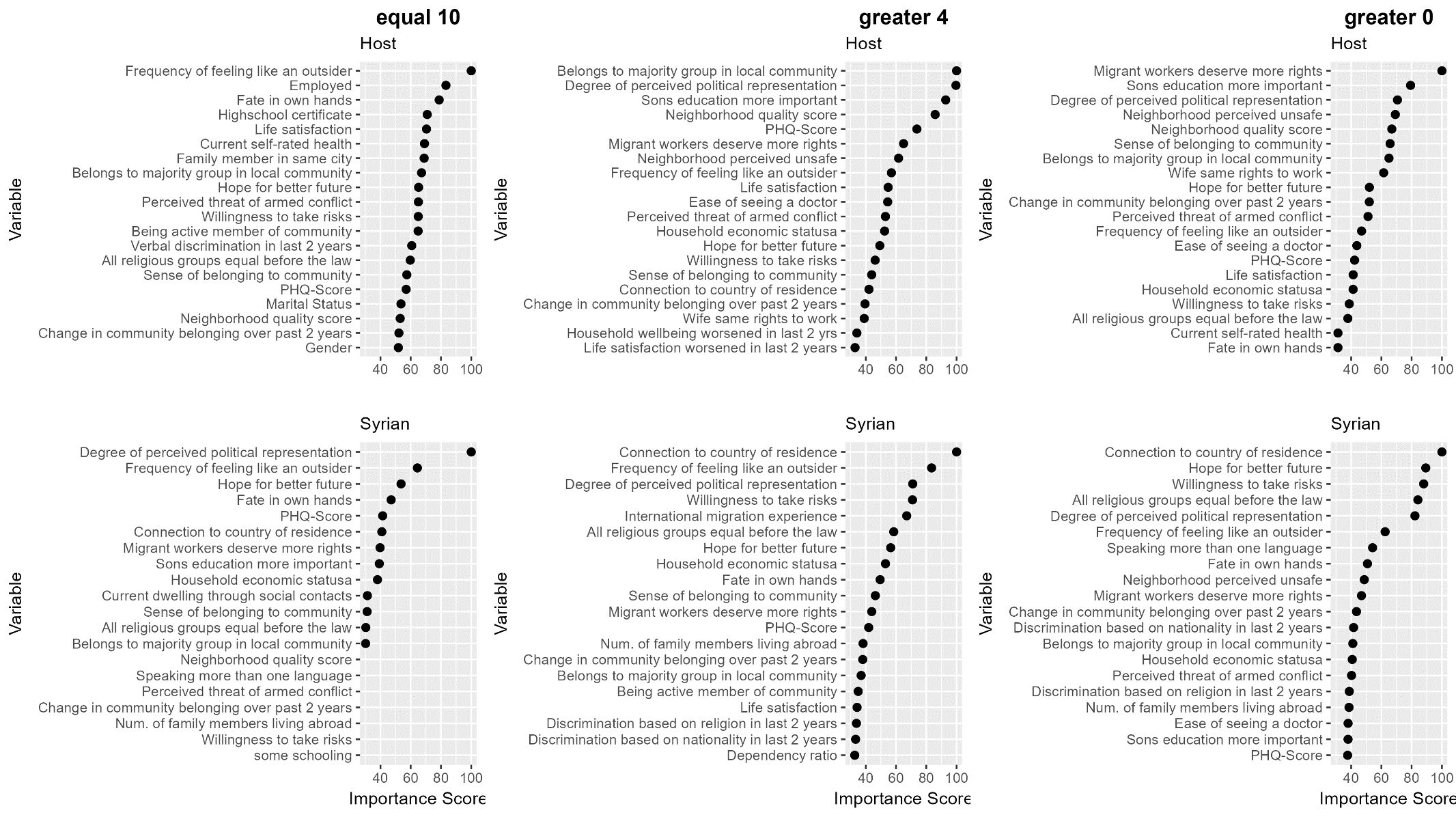
**TURKEY**

Figure S3: Permutation based importance scores derived from Random Forest for Syrian and Host Population for consideration to move in Lebanon at different cut-off points for the binary coding of the 0-10 Scale in Turkey

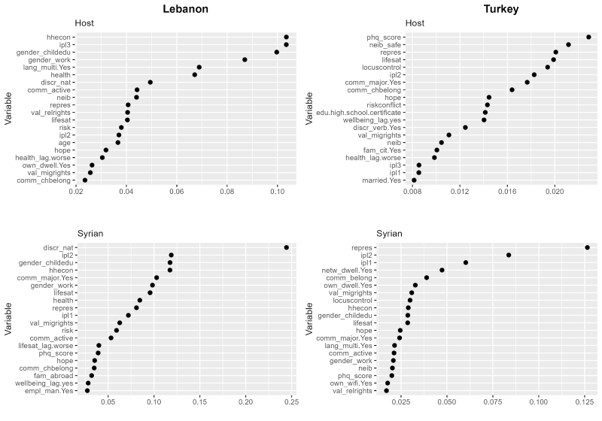


Figure S4: Conditional Importance Scores derived from a Conditional Random Forest Model of Individual Considerations to Move

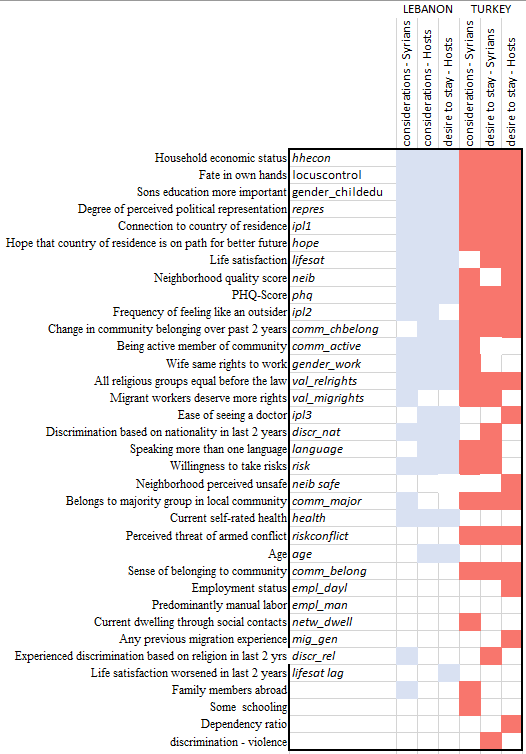


Figure S4: Overview of top 20 most important factors across outcome indicators and samples (Figure 3)