

# The Impact of Private Hosting on the Integration of Ukrainian Refugees in Germany

## Method & Data Report

Version: June 16th, 2025



### **Publisher**

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

Forschungsdatenzentrum DeZIM.fdz | Mauerstraße 76 | 10117 Berlin

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

### **Citation**

Herpell, Mathis; Marbach, Moritz; Harder, Niklas; Orlova, Alexandra; Hangartner, Dominik;

Hainmueller, Jens (2025) Method & Data Report, Version: 16th June, 2025. Berlin:

Deutsches Zentrum für Integrations- und Migrationsforschung (DeZIM).



## Inhalt

<b>1 Introduction .....</b>	<b>4</b>
<b>2 Background on registration.....</b>	<b>5</b>
<b>3 Sampling.....</b>	<b>7</b>
<b>4 Questionnaire .....</b>	<b>7</b>
<b>5 Additional Registration data .....</b>	<b>9</b>
<b>6 Survey non-response weighting.....</b>	<b>10</b>
<b>7 Anonymization strategy .....</b>	<b>12</b>

## 1 Introduction

We surveyed a sample of both matched and unmatched Ukrainian refugees on the platform #UnterkunftUkraine. Our survey focused on measuring integration outcomes across six domains: social, psychological, navigational, linguistic, economic, and political. The survey was conducted in 2023 via an online questionnaire available in Ukrainian, Russian, English, and German. It targeted adult refugees from Ukraine who had registered on the platform and were reachable via email. Non-response adjustments were implemented through inverse probability weighting based on pre-survey characteristics from the platform registration data. All personally identifying information in the registration data was removed or recoded before analysis, following established anonymization procedures and in consultation with the platform. The final dataset consists of registration data and survey responses that have been anonymized accordingly.

This document serves as supplementary material for utilizing our dataset, including the replication of our study. It outlines the data sources, sampling procedures, survey instrument, integration measures, and the anonymization strategy adopted. The following sections detail each of these components. Section 2 provides background on the registration and matching process. Section 3 explains the sampling frame and recruitment procedures. Section 4 presents the questionnaire content. Section 5 summarizes the registration data used. Section 6 details the non-response weighting strategy, and Section 7 describes the anonymization approach taken to ensure respondent confidentiality.

## 2 Background on registration

In this section we summarize essential information about the registration of refugees and hosts by the non-profit organization #UnterkunftUkraine (UU) as reported in our published paper. The registration process for hosts and refugees on the platform involved providing essential information that was then used by UU to match refugees to hosts.

**Refugees Registration.** In total, more than 117,000 Ukrainian refugees registered on the online platform provided by UU. A small share of refugees (less than 2%) got registered through other organizations. To register, refugees had to provide several pieces of information, including their name, gender, date of birth, family size, the total number of beds required, languages spoken, intended date and place of arrival in Germany, and their preferred municipality of residence. As of November 2023, the registration forms are no longer online (for screenshots see the supplementary information to the published paper).

**Host Registration.** In total, more than 150,000 individuals signed up to host on the platform provided by UU. During the registration, hosts had to provide details such as the location of their accommodation, the type of accommodation (e.g., shared room, shared house), the number of available beds, the presence of other family members at the accommodation, languages spoken, and the periods during which the accommodation was available. As of November 2023, the registration forms are no longer online (for screenshots see the supplementary information to the published paper).

**Matching Hosts with Refugees.** The matching process was conducted by a professional call center contracted by UU. The number of call center agents hired by UU fluctuated between 100 in the initial months and 40 in the later months. Call center agents were responsible for two main tasks: vetting potential hosts and matching refugees with hosts.

Before a host could be matched with a refugee, they had to provide evidence that the information on their ID matched the self-registration data. Hosts could use a third-party service called “Postident”<sup>1</sup> or confirm their identity through a video call with the call center. In the latter case, call center agents followed procedures similar to the “Postident” service, requesting hosts to present their ID in a video call and hold it at different angles for the agent to verify holograms and other security features.

The matching process was partially automated. When a call center agent initiated the matching process, the UU software randomly assigned them a refugee registered on the platform and prevented other agents from working on that refugee's case. The agent then contacted the refugee to confirm if they still required accommodation. If so, the software proposed hosts who matched the refugee's requirements in terms of the number of beds, desired duration of stay, location preferences, and the availability of at least one common language. If there was a potential match, the agent contacted the host to confirm their availability as specified in the registration form. Upon confirmation, the agent shared the refugee's contact details with the host and informed the refugee that a host would soon

---

<sup>1</sup> Postident is a service offered by the German postal service that provides identity verification for various business and legal purposes. It involves checking a person's identity in person at a postal office or digitally, using an official identity document. It is commonly used for processes like opening bank accounts.

contact them. It's important to note that for safety reasons, UU had a policy of exclusively considering women or heterosexual couples as hosts for female refugees. The registration data is described in more detail in section "Registration data" below.

### 3 Sampling

Our core analysis sample is based on an original survey we conducted with the help of UU in June 2023. UU handled the survey invitations and hosted the survey through a Qualtrics account. UU invited all refugees who had registered with the UU platform, provided valid contact information, and whom UU was legally able to contact.<sup>2</sup> In total UU was able to invite 49,380 refugees.

On June 1st, 2023, a total of 41,237 refugees were contactable by UU through their MailChimp email system (contact rate: 83.5%).<sup>3</sup> The invitation said that UU was partnering with IPL and DeZIM researchers to evaluate and improve the service of UU. Each invitation contained a personalized link. The invitation also announced that 5 euros will be donated to a charity working in Ukraine for completing the survey. The median time to complete the survey was 14.8 minutes (among respondents who answered all questions).

In total 2,905 individuals participated in the survey. We define participation as someone who consented to the terms of the survey. We screen out 94 respondents who were ineligible to participate. This leaves a total of 2,811 complete and partial responses. 1,870 respondents answered all survey questions.

Assuming that the proportion of ineligible non-respondents is the same as  $\frac{94}{41,237}$

### 4 Questionnaire

The questionnaire was composed of four main sections. After an introductory message and the request to consent to the data collection and data linkage with the UU registration data, it collected basic information about the respondents' migration experience since 2022 (i.e., if and when they came to Germany and if and when they left Germany again). The respondents remaining in Germany were then asked the main outcome questions measuring integration outcomes using the Immigration Policy Lab Integration Index (Harder et al., 2018). Respondents who were not in Germany at the time of the survey were not asked these questions since the questions presuppose residence in Germany. In the next section, we asked respondents who were in Germany at some point since February 2022 about their living arrangements – whether they had been matched by UU, moved into the matched accommodation, or which other accommodation they moved into. Then we asked for details regarding the accommodation the respondents moved into. This section included questions about cohabitants and the size of the accommodation, contact with and support from the host, overall satisfaction with the host and accommodation as well as the possible termination of the accommodation arrangement. In the last section of the survey, we collected additional information about the respondents including current living situation, public

---

<sup>2</sup> To comply with the EU General Data Protection Regulation (GDPR), UU deletes all contact information that was submitted or last changed more than a year ago every month. Furthermore, UU deletes the contact information of refugees who explicitly opted out from all further communications.

<sup>3</sup> The survey was tested in two pilots in May 2023. The data from these pilots are not included in the analysis.

and non-governmental support, demographic information, as well as their childcare situation.

A key interest of the survey was to learn more about the accommodations that were brokered by UU as well as accommodations used by respondents who were not matched by UU. For this, we had to clearly define which accommodation a respondent should provide information about. To do so, we asked respondents who lived in a UU accommodation to answer questions with respect to the accommodation they got matched into by UU. If they lived in multiple UU accommodations, we asked them to consider their first UU accommodation. Respondents who did not move into a UU accommodation were asked to answer for the first accommodation they lived in two months after their registration with UU. We use two months as a reference period, since we estimated based on various data sources that this is about the time most matched respondents would have moved into a UU accommodation. This estimate proved correct, as the survey data indicates that about 77% of those that moved into a UU accommodation lived in such an accommodation two months after registration.

The survey questionnaire was developed in English, then translated into Ukrainian and Russian by a professional translation service and into German by the authors. The Ukrainian and Russian translations were reviewed by the authors and colleagues, who are native speakers, to ensure conceptual homogeneity.



## 5 Additional registration data

We utilize anonymized registration records for all 117,254 refugees registered with UU. This data is employed for non-response analysis and to control for the characteristics that influenced the matching of refugees to hosts in our effect estimation.

The following variables were made available by UU to the authors:

- **adult\_male\_count** Number of adult males in group (number)
- **adult\_female\_count** Number of adult female in group (number)
- **adult\_diverse\_count** Number of adult diverse in group (number)
- **children\_count** Number of children in group (number)
- **beds** Number of beds requested (number)
- **languages** Languages spoken by the refugee (list)
- **from\_date** Arrival date in Germany (date)
- **weeks** Estimated stay in weeks (number)
- **place\_of\_arrival** Place of arrival (free text)
- **final\_destination** Destination preference (free text)
- **message** Registration includes free-text message (binary indicator)

In addition, the authors had access to the following process-based variables:

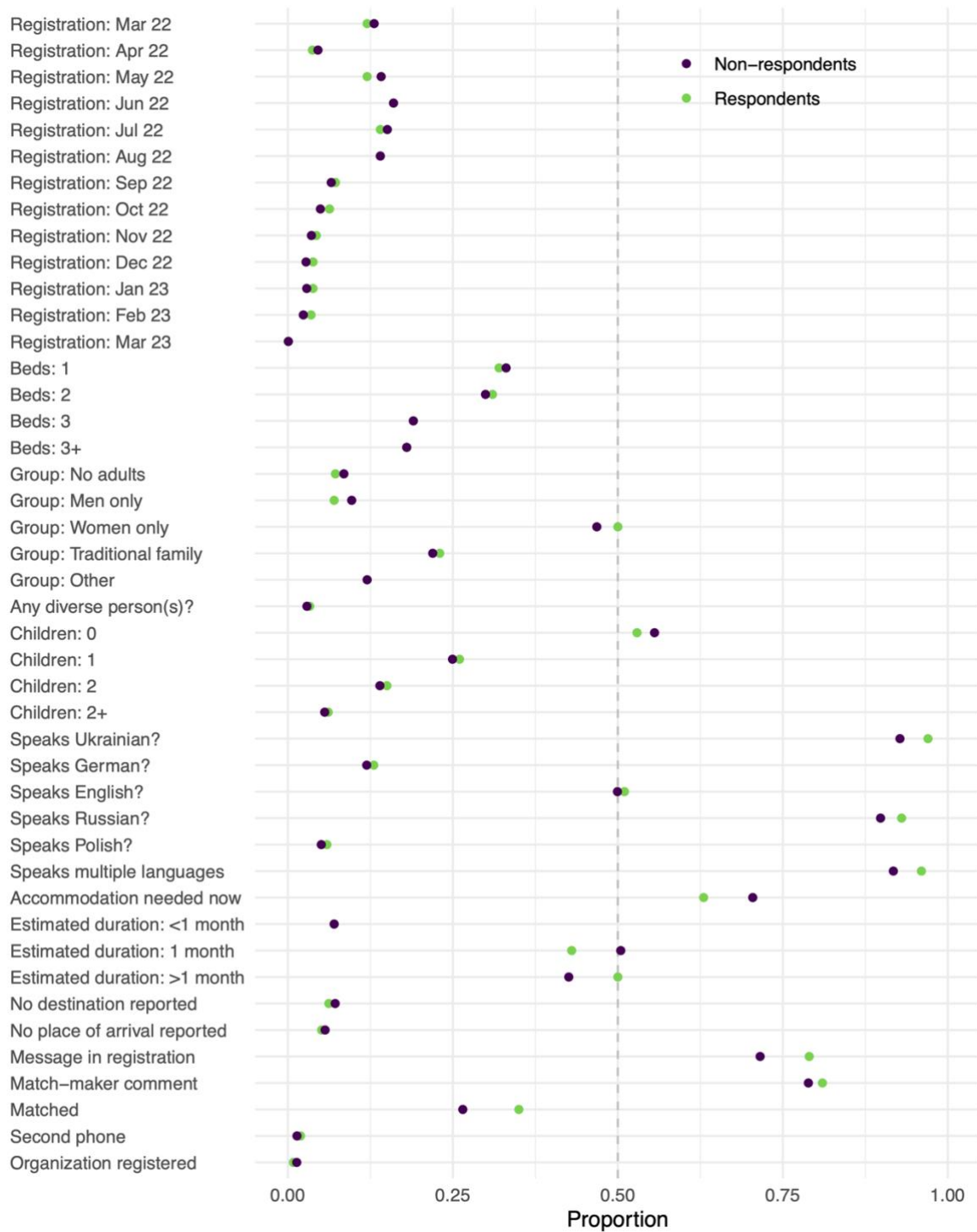
- **created** Timestamp of registration (timestamp)
- **comments\_count** Number of comments entered by match-maker (number)
- **Match.created** Timestamp of match creation (timestamp)
- **phone\_alt** Registration includes second phone number (binary indicator)
- **organization\_id** Registration was done by an organization (binary indicator)

We constructed a series of binary indicator variables based on the raw data. To limit the effect of outliers, variables counting the number of adults by gender and the number of children were winsorized at the value 10 and the variable counting the estimated stay in weeks was winsorized at 48. Figure 1 reports means for each of the binary indicator variables.

Refugees' place of arrival and destination preferences were entered into a free-text field. To encode these places, the authors created a dictionary containing names of all German municipalities, the federal states as well as their Anglicized equivalents and Ukrainian and Russian transliterations. For the 300 largest cities and all federal states, the names were translated and transliterated manually. The remaining municipalities were transliterated via the service "DeepL". The places of arrival and destination preferences in the register data were then matched with the entries in the dictionary. For places for which no match was found, partial or approximate places were searched. If the refugees had entered multiple possible arrival places or multiple destination preferences, their entries were assigned to the place with the highest population.

## 6 Survey non-response weighting

While we demonstrate in the published paper that respondents and non-respondents are broadly similar across all characteristics, we still construct non-response weights to adjust for remaining differences. To construct these weights, we use entropy balancing (Hainmueller, 2012), which adjusts the weights of the sample units so that the weighted sample moments match population target moments. These weights are constructed based on all variables shown in Figure 1 as long as the proportion is larger than 0.02. This restriction removes the indicator for arrivals in March 2023, an indicator for a second phone number, and an indicator for an organization registering the refugee. We then winsorize weights larger than the 0.99 quartile.



**Figure 1:** Non-Response Analysis: Differences Between Respondents and Non-Respondents as Observed in the Registration Data (N=52,300)

## 7 Anonymization strategy

To ensure the privacy and confidentiality of participants in our study we implemented the anonymization strategy for the shared dataset. Given the dataset combines survey responses and registration data from the UU and the sensitive nature of the data including detailed personal information such as household composition, location preferences, and dates of arrival, we developed an anonymization strategy for the dataset in alignment with GDPR compliance to exclude the risk of re-identification.

The anonymization approach followed a variable-by-variable assessment, guided by the principle of including only variables necessary for replication or meaningful secondary analysis. We classified variables according to their sensitivity, uniqueness, and potential for indirect identification. Leaving aside variables we assessed did not carry a risk of disclosing participants' anonymity, the following techniques were applied to the remaining variables.

- We completely excluded variables containing personal information (e.g. email) and sensitive variables that could directly or indirectly identify participants from the shared dataset. In particular, this concerns duration of living in accommodation, possession of citizenships other than Ukrainian, native languages other than Ukrainian, and free-text variables (e.g., “other” answer option).
- Variables for postal code and location of accommodation were merged and recoded into a new variable representing Community Identification Number (Amtlicher Gemeindeschlüssel).
- Location-specific variables from the registration data were recoded to leave only the top 45 places, putting the remaining options in “other”. We only retain month-year level values for the date of arrival and the date of making a match between host and refugee.

Thus, variables that were essential for key analyses but potentially sensitive were generalized in a way that retains their interpretative value while reducing identifiability. We ensured that the final shared dataset retains sufficient detail for replication of the main analysis and for further analyses while adhering to ethical standards for data protection.