COLLECTING DATA ON MIGRANTS' HEALTH STATUS AND ACCESS TO HEALTH SERVICES: THE EXPERIENCE OF THE MOBILE APP "COMESTAI"

Chiara Allegri, Alessandro Avellone, Elisa Barbiano di Belgiojoso, Sara Khatab, Dario Pescini, Stefania M. L. Rimoldi, Antonella Zambon

Abstract. Despite the process of migrant settlement in Italy, there is still a lack of data on migrant health and access to health services. We have developed the mobile app "ComeStai" to collect new data on this population.
ComeStai collects new and original information on migrants' self-assessed health status and access to health care through short surveys. The main questionnaire includes questions about the respondent's immigrant status, health status and health habits, access to and use of health care, and socio-demographic characteristics. Every month, the app displays a follow-up survey to track key variables over time, including respondents' health status and experiences of barriers to accessing health care.
After submitting the answers to the main questionnaire, the user can access an in-app platform that provides an overview of the Italian public health system and regulations, a brief description of the main public health services, and a list of NGOs providing health services or support in the Metropolitan area of Milan, as well as their location on a map.
Our study aims to test the efficacy of a mobile application as a data collection tool. We present the fieldwork and discuss the advantages and challenges of such a tool. Our evaluation considers the sample size achieved, the heterogeneity and representativeness of the sample to the real population. Based on the results, we will evaluate whether to extend the research to a larger geographical scale.

1. The use of mobile apps in social research

The increase in the number of activities reliant on mobile phones and mobile apps has improved users' familiarity with their devices and has paved the way to use mobile devices (tablets, smartphones, mobile Web) for conducting surveys and collecting data (Couper, 2013; Marcano Belisario *et al.*, 2015) even in migration studies (Rocheva *et al.*, 2022). Social studies have therefore analysed big data from social networking sites, or online services targeting migrants for their studies, moreover, surveys online have become more frequent. Mobile apps can represent a useful and additional data-collecting method as they can both passively collect data (e.g. geolocalisation) or be used to conduct surveys (Couper, 2013).

Similarly to online surveys, mobile apps offer several advantages for studying migrants: they enable respondents to complete surveys anywhere and at any time

(Marcano Belisario *et al.*, 2015), reach populations geographically dispersed on a territory, highly mobile and not included in a Population Register, eliminate the need for interviewers and their knowledge of migrants' native languages, since they can employ multilingual interfaces easing the self-administration of the questionnaire (Rocheva *et al.*, 2022; Jacobsen and Kühne, 2021). In addition, the cost of a survey using mobile apps is considerably lower than a traditional survey requiring interviewers. Finally, apps allow researchers to solicit the collection of longitudinal data through mobile notifications rather than through recontact calls or emails. However, many challenges remain: the target population needs to have a mobile device; the mobile app should be designed to run on different operating systems; respondents' reluctance to download a mobile app for privacy or data security concerns; a demanding task, such as daily activities or a long period of the study (Jacobsen and Kühne, 2021; Jäckle *et al.*, 2022, 2023).

The evidence regarding the quality of data collected with mobile apps presents mixed results: some studies found similar quality compared to traditional surveys, while others found the opposite (Jacobsen and Kühne, 2021; Marcano Belisario *et al.*, 2015; Oakeley-Girvan *et al.*, 2018). Previous studies unanimously agreed that the response rate in mobile app surveys is lower compared to other methods (Zhang *et al.*, 2018; Jäckle *et al.*, 2022, 2023), unless there is an interviewer persuading survey respondents to participate or offering a browser-based alternative to those reluctant to download the app (Jäckle *et al.*, 2022). Little evidence on non-respondent rates highlights that personal characteristics may affect participation: younger and frequent users of other apps are more likely to participate (Jäckle *et al.*, 2022; Jacobsen and Kühne, 2021). The characteristics of the survey are determinant too: the length of the questionnaire, the way the questions are formulated (a compromise between the available space and the need for clarity), the compatibility of the app with the device, a demanding task, finding the app on the store and downloading the app, setting relevant permissions in the app (e.g. notifications, location tracking) and invitation message (Jäckle *et al.*, 2023; Wenz and Keusch, 2023). All these aspects are found to affect the participation rate and the completeness and quality of the responses. According to Fan and Yan (2010), the design of the invitation to participate in a survey indicating the organisation's name, the purpose of the study, as well as the contact delivery modes, also affect the response rate.

2. Selection bias: to weight, or not to weight?

Literature unanimously recognises that the auto-selection in surveys conducted by mobile apps is evident, therefore, the use of data collected by mobile apps deserves caution. Usually, young people and women are more likely to complete the

surveys (Haddad *et al.*, 2022). Scholars have used different solutions: poststratification or weighting, ranking, propensity score adjustment or a combination of these methods. However, there is little consensus on whether or not to use weighting to correct for selection bias. According to Mercer *et al.* (2018), the choice of the variable for weighting is more important than the statistical methods used, and the basic weighting method performs as well as more elaborate ones. Moreover, a larger non-probability sample reduces the variability of the estimates but does not significantly increase the accuracy (Mercer *et al.*, 2018; Haddad *et al.*, 2022).

3. Questionnaire design

The population under study consists of individuals with foreign citizenship, naturalised or Italians born abroad, present in Lombardy during the data collection period.

The research design employs two questionnaires, with the main one being presented upon the first download of the app, while a second set of questions is proposed (conditional upon acceptance) at three subsequent deadlines, every month from the first compilation.

The questionnaires were designed to investigate the interaction between health status and socio-demographic factors. Therefore, in the main questionnaire, the questions were grouped into two sets: the health set and the socio-demographic set.

As for the former, we used the European Health Interview Survey (EHIS) and the Istat survey "Condizione e Integrazione Sociale dei Cittadini Stranieri" as references for formulating the questions. This approach ensures that questions have already been validated and allows us to compare our results with the national and European averages. The total number of questions for the health set is 20.

The socio-demographic set is made of 22 questions. A crucial point of this set of questions was to identify immigrants. We operationalised this by asking about the respondents' birthplace, their first and second nationality, parents' birthplace and their type of residence permit. The socio-demographic set was presented after the health status set to reduce the non-response rate[1] (Lor *et al.*, 2017) as well as social desirability bias linked to identity priming. However, as we wanted to ask some questions only to foreign citizens, we moved the question about the place of birth ("Were you born in Italy or abroad?") at the beginning of the questionnaire and filtered the following questions accordingly.

[1] Some demographic questions are largely recognized as sensitive and may be left unanswered or respondents may even choose to skip the entire survey due to confidentiality concerns.

As the questionnaire had to be self-completed, it was crucial to have a limited number of questions in order to limit the burden on respondents and increase the response rate. The total number of questions was 42. Of these, 4 were not proposed to people born in Italy and 3 were conditional on a previous affirmative answer, thus not always appearing[2]. The estimated average time of compilation was between 5 and 7 minutes.

The follow-up questionnaires are designed with two objectives. Firstly, to collect longitudinal data on three key aspects: current health status, health services used and barriers experienced in accessing healthcare. Secondly, to further investigate specific respondents' health behaviours, e.g. regarding alcohol and drug use in case they declared a frequent use of these substances in the main questionnaire. The estimated average time of compilation was 1 minute.

The questionnaires were designed in Italian and translated into 7 additional languages: English, French, Spanish, Arab, Chinese, Russian and Albanese. Languages were chosen according to the diffusion of the largest immigrant communities in the Lombardy region.

4. App design

As suggested by Jäckle *et al.* (2023) to increase the number of potential respondents, the ComeStai app was made available for both Android and iOS operating systems. As the app size is only 11 MB, it is fairly simple to download it both on WIFI and data, and it is unlikely that someone won't be able to install it because of a lack of storage space on their phone. When first opening the app, it presents to the users the following steps:

1. Choice of language
2. Informed consent regarding privacy issues
3. Informed consent regarding app notifications
4. Main questionnaire
5. Surfable information about healthcare in Italy and Milan area

At further openings of the app, the user is redirected directly to step 5.
In step 1, people using Android operating systems are forced to choose between the 8 available languages. Unfortunately, it was not possible to implement this feature on the iOS operating system. For Apple devices, by default, the app is set on the language used by the mobile phone, if this is among the 8 languages available, and

[2] Therefore, the total number of questions for a foreign-born individual spanned from 39 to 42, while the number of questions for a person born in Italy spanned from 35 to 38.

on Italian otherwise. Users can change the app language from the Settings area of their smartphone.

In step 2, people are informed about the objectives of the research and about the fact that the app does not collect personal data that allow the identification of the respondent such as phone numbers, emails, names, and geolocation of the respondents. Previous studies showed that asking permission for geolocalisation and collecting emails or addresses considerably reduces the consent to install apps (Jäckle *et al.* 2023; Jacobsen and Kühne, 2021; Wenz and Keush, 2023). All data collected through the surveys is therefore anonymous. Upon completion of the surveys, the responses are assigned to a progressive ID number and directly stored in a database of the University of Milano-Bicocca. As no data is stored on the respondents' phones, third parties cannot access the data even in case of loss of the device.

In step 3, people are informed that in the following 3 months, the app will send 3 notifications to check again on their health status. They can accept or refuse to receive notifications. If they accept, after 30, 60 and 90 days from the first compilation, a notification is sent to the mobile phones redirecting to the 3 follow-up questionnaires. Over the 3 months, no other notifications are sent apart from these. This is consistent with the best practices identified in previous studies (Jäckle et al. 2023), which suggest that users are more likely to allow notifications when the reason is given and the task is not demanding, such as a long period of analysis or frequent requests (daily or weekly).

The in-app platform provides information about healthcare for foreign citizens in Italy and Milan, organised in 9 tabs. The users can navigate through brief and simple descriptions about how to access public health services in Italy and some information on prevention services (free screening campaigns) and on the correct behaviours to maintain a healthy life (nutrition, active lifestyle, smoking, sun exposure); a list of all basic public health services available in the Milan Metropolitan area (emergency rooms, centres for addictions, centres for family planning, centres for mental health); as well as a list of the third-sector associations offering free healthcare to migrants in Milan. Links redirecting to the relevant external web pages are provided, including Google Maps links to geolocate the listed services. We also provide a summary map displaying all the points of interest, which was constructed using the "MyMaps" function of Google Maps. Users can also contact the research team through a dedicated email account, as shown within the app. All the contents can be updated over time by the research team through the University server, with no need for action on the respondents' side, i.e. no need to re-download or update the app. This can be considered an incentive to participate although frequently the incentive is monetary. However, Jäckle *et al.* (2023) highlighted that monetary incentives could not compensate for the burden. Jacobsen and Kühne (2021) provided the users with news

on the situation of refugees in Europe without testing the effect of this incentive, however, the consent to use the mobile app was higher among those who arrived in the last 3 years compared to long-term refugees.

5. Invitation design

To recruit the participants, we used different channels. Firstly, we organised, with CSV Milano, several meetings with the selected NGOs that have participated in the design of the app, to help us disseminate ComeStai among their services' users. We provided them with some full-colour flyers in Italian and two posters (one in Italian and one multilingual, Spanish, Chinese, Russian and Arab) showing the purpose of the survey, the name and the logo of the app, the QR code to download the app from both Apple's and Android's app stores, and the logo of the University Bicocca. These elements help to find the app reducing the risk of non-recognition of the app in the stores. According to Wenz and Keush (2023), the identification of the organisation sponsoring the study is crucial to increasing the response rate with apps sponsored by universities showing a higher response rate than those sponsored by market research companies. Secondly, we emailed migrant communities or other associations involved in helping migrants based in the Milan Metropolitan area, introducing the project and the survey, and providing the flyers. Thirdly, the university's press office informed the newspaper. Two of them agreed to publish a short article about the app. Fourth, we promoted the survey through Facebook. Finally, we solicited each staff members' personal contacts to participate in the survey. As shown in Table 1, we sent 192 messages, and we received feedback from 17 of these contacts who decided to promote the app among their users or associates. In addition to overcoming the reluctance of migrants to download the app, the research team or the volunteers invited in-person the users of some services to download the app, helped them with the download and answered their doubts about the questionnaire. As suggested by previous studies (Jäckle et al. 2023), in fact, an in-person invitation increases the participation rate.

Table 1 - *Recruitment of the participant*

Contact	Total
Health services	28
Cultural or community association	64
Place of worship	5
Social networks - newspaper	29
Other associations (Italian language schools, legal services, housing…)	66
Total	192

6. Sample

In the first 4 months since the app was published on the stores' web pages (March-June 2024), we collected 224 answers, of which 92 were from natives. We excluded the latter cases and we focused only on foreign-born and naturalised persons (132 observations). Our sample is unbalanced in terms of gender composition with a higher percentage of women (64.4%) which confirms the higher participation rate of women. The age distribution is quite homogeneous across the age groups, and we interviewed people from 38 different countries of origin. The distribution by country of origin reflects the distribution of the resident foreign population in the metropolitan area of Milan with a higher percentage of Egyptians, Peruvians, Filipinos and Senegalese. The sample is biased with respect to the length of stay: a large proportion of respondents (40%) had recently arrived in Italy (last 2 years). Despite our efforts to disseminate the app through different channels, it was easier to reach foreign-born people who had recently arrived. The availability of information about health services probably acted as an incentive especially for those who needed it, confirming previous studies (Jacobsen and Kühne 2021).

7. Discussion

While using an app to interview migrant populations presents many advantages, it also comes with several limitations and challenges, most of them anticipated in Section 1. Among these limitations, we found that language barriers can impact sensibly on results. Although the app supports multiple languages, ensuring accurate and culturally appropriate translations is challenging, misunderstandings or mistranslations can affect the quality of the data collected. For example, in our main questionnaire, the question "Was one of your parents born abroad?" was sometimes misunderstood, i.e. some immigrant respondents took their country of origin as country of reference and answered that their parents were *not* born abroad (meaning not born in a third country), while the phrasing intent was to consider any country other than Italy as "abroad". This generated a few incorrect answers. The issue could be solved by asking "Was one of your parents born outside Italy?". Another important issue refers to privacy concerns. Migrants might have concerns about the confidentiality and security of their data, particularly if they are undocumented or fear repercussions. These concerns can lead to lower participation rates or less honest responses. This can indeed be the case of the app here presented, as it deals with health and sanitary issues, which are sensitive data. One way to mitigate this limitation is to present the app to the target population with the help of trusted testimonials, such as NGO volunteers, especially if they belong to the same ethnic

background. Indeed, planning a close collaboration with the main NGOs in the area is one of the key operational strategies needed, both to ensure higher participation in the survey and to improve the overall quality of the responses. The collaboration with the NGOs is a crucial factor especially for longitudinal studies, as in the case of the app presented here. Establishing a close collaboration with NGOs is far from simple. Ultimately, it entails creating a tool, the app, that must be useful to three stakeholders with different needs: researchers, users, and NGOs. In this sense, the app's area dedicated to "content feedback" must be functional to the needs of users and NGOs, and therefore it must be carefully designed and maintained, possibly also with the collaboration of the NGOs themselves. Finally, we also suggest that combining the app-based surveys with other data collection methods, such as face-to-face interviews or paper surveys, can also help reach a broader and more representative sample.

We sent 192 messages through different channels, but most of these messages did not receive a response. The sample size is in line with other studies using apps, as, according to Zhang *et al.* (2018), the median sample size is 95. As mentioned above, our sample is quite heterogeneous in terms of age and country of origin, while it shows some specificity in terms of gender and length of stay.

As for the timeline, each step takes time, which is difficult to predict. The dissemination of the app took longer than expected because, to ensure user participation, we had to contact NGOs to present the app with ad hoc meetings, which always take time to be organised, and we had to allocate additional time for the volunteers to answer the questionnaire and navigate the in-app information on health services in order to receive their evaluation of the potential usefulness for their users. Volunteers disseminated the app to their users through social media posting, distributing the flyers, sending emails and WhatsApp messages to their contacts, or they encouraged their users to download the app during classes or other in-person meetings.

In conclusion, we can take stock of the experiment described in the preceding pages. The use of the ComeStai app as a data collection tool proved to be quite costly in terms of the time dedicated to promoting the tool among members of a "rare and elusive" population. It should also be noted that in this specific case, the topic (health) and the target population (migrants) presented numerous complexities regarding the sensitive points of administering a questionnaire via app (privacy, language, etc.). Numerically, the results were limited, although in line with the experiences cited in the literature. The low number of responses does not allow for comparisons between the different components of the reference population (e.g., legally and illegally present migrants); however, we can discern the positive value of this experience in two particular aspects:

- Enhanced Engagement: Despite the challenges, the process of engaging with the migrant population through the ComeStai app demonstrated the potential for building trust and fostering participation, especially when facilitated by trusted intermediaries such as NGO volunteers.
- Data Integration Potential: The information collected, although limited in quantity, holds significant potential for integration with other data sources. This can enrich the overall understanding of health issues within the migrant population and support the development of more targeted and effective health interventions.

Acknowledgements

This research is funded by Cariplo Foundation, project grant 2022-1689 Multiple Ethnic Inequalities.

References

COUPER M.P. 2013. Is the Sky Falling? New Technology, Changing Media, and the Future of Surveys, *Survey Research Methods* Vol. 7, No. 3, pp. 145-156.

FAN W., YAN Z. 2010. Factor affecting response rate of the web survey: A systematic review, *Computers in Human Behavior,* Vol. 26, pp. 132-139.

HADDAD C., SACRE H., ZEENY R.M., HAJI A., AKEL M., ISKANDAR K., SALAMEH P. 2022. Should sample be weighted to decrease selection bias in online surveys during the COVID-19 pandemic? Data from seven datasets, *BMC Medical Research Methodology*, Vol. 22 No. 63.

JÄCKLE A., BURTON J., COUPER M. P. 2023. Understanding Society: minimising selection biases in data collection using mobile apps, *Fiscal Studies*, Vol. 44, pp. 361–376. DOI: 10.1111/1475-5890.12351.

JÄCKLE A., WENZ A., BURTON J., COUPER M.P. 2022. Increasing Participation In A Mobile App Study: The Effects Of A Sequential Mixed mode Design And In-Interview Invitation, *Journal of Survey Statistics and Methodology,* Vol. 10, pp. 898–922.

JACOBSEN J., KÜHNE S. 2021. Using Mobile App When Surveying Highly Mobile Populations: Panel Attrition, Consent, and Interviewer Effect in a Survey of Refugees, *Social Science Computer Review*, Vol. 39, No. 4, pp. 721-743.

LOR M., BOWERS B. J., KRUPP A., JACOBSON N. 2017. Tailored explanation: A strategy to minimize nonresponse in demographic items among low-income racial and ethnic minorities, *Survey Practice*, Vol. 10, No. 3, pp. 1-11.

MARCANO BELISARIO J.S., JAMSEK J., HUCKVALE K., O'DONOGHUE J., MORRISON C.P., CAR J. 2015. Comparison of self-administered survey questionnaire responses collected using mobile apps versus other methods (Review), *Cochrane Database of Systematic Reviews*, Vol. 7, No. MR000042.

MERCER A., LAU A., KENNEDY C. 2018. For Weighting Online Opt-In Samples, What Matter Most? Pew Research Center https://www.pewresearch.org/methods/2018/01/26/for-weighting-online-opt-in-samples-what-matters-most/

OAKLEY-GIRVAN I., LAVISTA J. M., MILLER Y., DAVIS S., ACLE C., HANCOCK J., NELSON L. M., 2018. Evaluation of a Mobile Device Survey System for Behavioral Risk Factors (SHAPE): App Development and Usability Study, *JMIR Formativ Res*, vol. 3, No. 1 e10246.

ROCHEVA A., VARSHAVER E., IVANOVA N. 2022.Targeting on Social Networking Sites as Sampling Strategy for Online Migrant Surveys: The Challenge of Biases and Search for Possible in S. PÖTZSCHKE, S. RINKEN (Eds.), *Migration Research in a Digitized World*, IMISCOE Research Series, https://doi.org/10.1007/978-3-031-01319-5_3, pp. 35- 58.

WENZ A., KEUSCH F. (2023) Increasing the Acceptance of Smartphone-Based Data Collection, *Public Opinion Quarterly*, Vol. 87, No. 2, pp. 357–388.

ZHANG J., CALABRESE C., DING J., LIU M., ZHANG B. 2018. Advantages and challenges in using mobile apps for field experiments: A systematic review and a case study, *Mobile Media & Communication*, Vol. 6, No. 2, pp. 179–196.

Chiara ALLEGRI, Università degli Studi di Milano-Bicocca, chiara.allegri@unimib.it

Alessandro AVELLONE, Università degli Studi di Milano-Bicocca, alessandro.avellone@unimib.it

Elisa BARBIANO DI BELGIOJOSO, Università degli Studi di Milano-Bicocca, elisa.barbiano@unimib.it

Sara KHATAB, Università degli Studi di Milano-Bicocca, sara.khatab@unimib.it

Dario PESCINI, Università degli Studi di Milano-Bicocca, dario.pescini@unimib.it

Stefania M. L. RIMOLDI, Università degli Studi di Milano-Bicocca, stefania.rimoldi@unimib.it

Antonella ZAMBON, Università degli Studi di Milano-Bicocca, antonella.zambon@unimib.it