



***Trusted Smart Statistics:
methodological developments based on new data sources***

2022-IT-TSS-METH-TOO
Project n. 101132744

Work Package 4

Proof of concept of an ad-hoc survey to improve MNO data

Deliverable 4.2

Report on a possible ad-hoc survey to improve the usage of MNO data in official statistics

June 2025

Partner in charge: Statistics Austria – *Alexander Kowarik*

Authors: Alexander Kowarik, Tiziana Tuoto, Loredana Di Consiglio, Gloria Deetjen, Maurice Brandt, Remy Kamali, Kira Gylling, Cristina Faricelli, Sara Piombo and Tiziana Pichiorri

[MNO-MINDS | Eurostat CROS \(europa.eu\)](https://eurostat.cros.europa.eu)



Co-funded by
the European Union

Deliverable 4.2

Report on a possible ad-hoc survey to improve the usage of MNO data in official statistics

Summary

Mobile Network Operator (MNO) data holds great promise for enhancing statistical outputs across various domains, including mobility, commuting, and tourism. Yet, the use of such data in official statistics is hindered by multiple sources of bias—most notably, the mismatch between users and contract holders, the use of multiple devices or SIM cards, and varying behaviour across socio-demographic groups.

This report proposes the design and implementation of an ad-hoc survey to systematically identify and correct these biases. It outlines survey objectives: understanding device usage, identifying user-contract relationships, and capturing relevant socio-demographic information.

A friendly user test was conducted to validate the proposed questionnaire and revealed useful insights for refining question wording, logic, and technical implementation. Moreover, the report presents potential complementary survey approaches, such as area-based sampling using antennas and border-point surveys, to address populations not captured in standard sampling frames—e.g., tourists or irregular residents.

Index

1. Introduction	4
2. Background and motivation.....	4
2.1. Current challenges with MNO data.....	4
2.2. The need for a dedicated survey.....	5
2.3. Segmentation of MNO User Groups	5
3. Survey design and implementation	6
3.1. Survey objectives.....	6
3.2. Questionnaire design	6
3.3. Sample design.....	6
4. Expected outcomes and implications	8
4.1. Analysis methodology	8
4.2. Enhancements in Data Quality	9
5. Report on the Friendly User Test	9
5.1. Results from the Friendly User Test	10
5.2. Summary of feedback.....	17
5.3. Feedback regarding Questionnaire design.....	17
6. Complementary Approaches	18
6.1. Design of an area sample	19
6.1.1. Sample size calculation.....	21
7. Conclusion and next steps	22
7.1. Next Steps.....	22
8. References.....	24
9. Annex: Questionnaire for area sample	25
10. Annex: Questionnaire: Use of mobile data (updated after friendly user test)	26

1. Introduction

Mobile network operators (MNO) data is recognised as a valuable source for official statistics, offering granular and high-frequency information on population mobility and behavioural patterns. However, several methodological challenges – such as the ambiguity between devices users and contract holders, the prevalence of multiple SIM cards and/or phones per person and heterogeneity in the behaviour concerning the mobile phone usage across different demographic groups – make MNO data not directly usable for proper statistical analyses in which a whole population is represented in an unbiased way (see chapter 2 for more details).

In WP3 (see, e.g., *D3.2 Report on methodologies Zhang et al. (2025)*) a methodological framework is developed to tackle the aforementioned challenges. In several of the proposed methodologies information is missing for its successful application. Here survey-based strategies to quantify, understand and correct the structural biases are mentioned. The ad-hoc survey strategy presented in this document is adding to this methodological framework as another piece of the puzzle of unbiased official statistics using MNO data as a main source.

This deliverable serves as a standalone continuation of Deliverable 4.1, providing a detailed proposal for a focused questionnaire module (see chapter 9), supplemented by insights from a friendly user test (see chapter 5), and exploring alternative and complementary sampling designs (see chapter 6 and 10). It aims to offer a solid basis for further development and piloting within the European Statistical System. The document ends with a conclusion and outlines next steps that can be conducted for this ad-hoc survey approach.

2. Background and motivation

2.1. Current challenges with MNO data

MNO data usage, while providing high frequency data that cover almost the full area of the country and most of the population, has several challenges:

- **Device-User Ambiguity:** Often, the individual paying for a service (the contract holder) is not the primary user of the device or SIM card. Family plans and business contracts create additional differences between the contract holder and the actual users. Therefore, often socio-demographic information used in current analytical project with MNO data is not related to the correct person, the (main) user of a specific device or SIM card.
- **Multiple Devices and/or SIM cards:** Individuals may use multiple devices or SIM cards, leading to duplication in records or skewed usage statistics, causing over-coverage of certain groups within MNO data. Deduplication according to movement patterns is possible, however, in several scenarios it will be error-prone, e.g. a group of tourists spending the whole time together, a work phone that is partially left at home/work, etc..
- **Different User Behaviour:** Usage patterns can vary dramatically between users. These heterogeneities could be influenced by factors like age, employment status, and lifestyle, therefore affecting the generalizability of derived statistics. The usage behaviours include that some groups might switch off their (or one of their) phone(s) during the night, during work or during vacation (in case of business phones).

The mentioned phenomena are not just based on hearsay, but they can also be observed in quasi-randomization (QR) experiments as the one conducted by Statistics Norway in WP3 (see *D3.2*

Report on methodologies Zhang et al. (2025)). Also, with the Italian QR pseudo experiment with the sample survey “Aspect of Daily Live” those phenomena are studied, and the effect of user ambiguity is studied in a simulation setting (see *D3.2 Report on methodologies Zhang et al. (2025)*).

2.2. The need for a dedicated survey

To decrease the influence of the aforementioned issues on the statistical output, it is advisable to design and implement dedicated surveys to systematically collect data on the number of devices per individual, the actual usage patterns, and the socio-demographic profiles of users. This approach is aimed at both quantifying the biases and providing corrective measures to enhance data accuracy of MNO data. The survey proposed in this document is intended to support the quasi-randomization approach proposed for the use of MNO data in the Work Package 3, deliverable D3.1, aiming to assess a general selection model for MNO data to improve representativeness for MNO-based measurements.

Three types of sample surveys are mentioned in *D3.2 Report on methodologies Zhang et al. (2025)*:

A survey to estimate a conversion factor between the target total and the MNO macro counts. This is specialized survey, e.g., on the usage of the mobile devices during travelling.

A survey that is designed to measure the target total directly, e.g., if a person is a commuter, and collects information on the mobile device usage, so that the relevant MNO macro data, e.g., an origin-destination matrix, can be used in the estimation step.

An opt-in survey in which the consent from respondent is requested to collect their MNO micro or nano data from the MNO. The target output can then be estimated based on a random sample of MNO data with necessary auxiliary variables without the issue of user ambiguity.

2.3. Segmentation of MNO User Groups

It is important to acknowledge the heterogeneity of MNO user groups when designing a single survey or multiple survey instruments to tackle the challenges mentioned in subsection 2.1. . While our traditional surveys are suitable for targeting registered residents using established population sampling frames, they do not account for several other user groups. A possible and useful segmentation divides MNO users into (1) registered residents (including their children, which can be surveyed by proxy), (2) non-registered or irregular residents such as seasonal workers, expats, or undocumented individuals, and (3) tourists and short-term visitors. Importantly, all three segments may use either foreign or national SIM cards, making simple assumptions based on the nationality of the SIM card incorrect. This segmentation has direct implications for the survey design: traditional sampling frames can definitely only reliably cover the first group. Additional strategies, such as area-based sampling at antennae level or border-point surveys are necessary to get information about the second and third groups. To get a complete picture of mobile phone usage in a given country or region, all three groups are important since to produce statistical results just on registered resident (as it is common in official statistics) we must subtract the other segments from the MNO macro counts. If we are interested in foreigner, e.g., tourist visiting a certain region, we must subtract local residents with foreign sim cards from MNO counts.

So, integrating this segmentation into a strategy for planning a survey or several surveys allows for better calibration of MNO-based indicators and supports the development of generalized adjustment models to enhance representativity.

3. Survey design and implementation

3.1. Survey objectives

The primary objectives of the survey are:

- To determine the actual number of devices and SIM cards used by individuals and explained by socio-demographic background variables.
- To identify and quantify the discrepancies between contract holders and device users.
- To analyse device usage patterns across different demographic groups.
- To collect comprehensive socio-demographic data of both contract holders and actual device users. The sampling units are individuals that correspond to the actual device users. Information on socio-demographic background might not be surveyed but used directly from an information-rich sampling frame.

In addition to determining the actual number of devices and SIM cards, the survey may provide insights on the distribution of service providers and network usage to enable a better (understanding of) extrapolation. Details on device usage facilitates further analysis of people's movements and can potentially provide a base e.g. to estimate working from home, commuting, touristic travel, and to identify mobility patterns.

3.2. Questionnaire design

The questionnaire is meant as a core module for this topic with the possibility to be extended for specific statistical domains. It includes sections on:

- Device and SIM Card Usage: Questions on the number of devices, their purposes, and details on SIM card usage.
- Contractual Relationships: Information on the contract owner, duration with current service provider, and previous providers.
- User Demographics and Lifestyle: Questions on age, gender, household composition, employment status, commuting habits, and usage during different activities and travels.

3.3. Sample design

The questionnaire is designed to survey individuals about a specific, mostly legal, age limit, e.g. in the case of Austria, 16-year-olds. A random sample of individuals registered in the country should be drawn from a sampling frame, e.g. based on a central population register. Stratification can be used to increase the representativeness across different demographics and regions. Oversampling of certain groups can also be used to compensate for differences in response behaviour.

For several applications including the application scenario developed in WP3 on commuting, the main target variable is the number of phones per person. So, this variable will be used in drafting a possible sample design.

In the commuting application scenario, NUTS2 regions are used for the adjustment factors. Therefore, we use it as estimation domain and additionally add age groups as stratification variable as this is expected to be a major influence on the number of phones as can be seen by the quasi-randomization experiment in WP3 by ISTAT on the “Aspects of Daily Life survey”. We use just 3 age groups, 16 to 24 years, 25 to 64 years and 65 and older, to reflect that the employment is probably the main driver for having a second mobile phone.

Since we sample from a big population and do not go beyond NUTS2 regions, we ignore the finite population correction. This leads to conservative sample size estimations, because variance of sample estimates is slightly overestimated.

The number of additional phones, so the number of phones minus 1 can be modelled as geometric distribution and we assume – based on the friendly user test – the parameter, the probability, equal to 80%. This means about 80% have no additional phone and about 20% have at least one additional phone.

So, this computation is based on assumptions derived from the friendly user, but since this was not a representative random sample, these assumptions might be very wrong for the population. Before running this on a larger scale, it would be advisable to run a pilot study to get a better picture, see section 7.1 for proposed next steps.

The variance of the geometric distribution is defined as $Var(X) = \frac{1-p}{p^2}$. Therefore, the standard error of the sample mean is $SE(\bar{X}) = \sqrt{\frac{1-p}{np^2}}$ and we can compute the sample size as $n = \frac{1-p}{SE^2 p^2}$ for a given probability and standard error. We would like to have a standard error of 0.02 or the confidence interval $\hat{\bar{X}} \pm 1.96 SE$ of about 0.04 by estimation domain, therefore we need about 781 persons by estimation domain. This would lead to a confidence interval of (0.21, 0.29) which seems small or even too small, but this would be used as factor in weighting the MNO data. For the confidence interval of the average number of phones used by a person is then 1.21 to 1.29. For example, if we deflate a number of 100 000 active sim cards by these factors we would end up with a confidence interval for this estimate of 77 500 to 82 600.

In case of Austria, the proposed strategy of sampling 781 by NUTS2 region stratified by age class could lead to the sample design shown in table 1. N is the approximate number of persons in the given stratum in the population. n is the sample size, allocated within a NUTS2 region proportionally to the age classes. p is the sampling probability in the stratum. Probably, it would help to oversample the “working age” age class, to oversample persons with a higher probability of having multiple phones, but given that there was no information on this in the friendly user test, this needs to be tested further in a pilot study.

Table 1: Illustration of a possible sample design in Austria

NUTS2	Age class	N	n	p
Burgenland	16-24	24000	73	0.30
Burgenland	25-64	159000	484	0.30
Burgenland	65+	74000	224	0.30
Carinthia	16-24	48000	78	0.16
Carinthia	25-64	301000	485	0.16
Carinthia	65+	135000	218	0.16
Lower Austria	16-24	150000	80	0.05
Lower Austria	25-64	925000	499	0.05
Lower Austria	65+	373000	202	0.05
Upper Austria	16-24	141000	87	0.06
Upper Austria	25-64	827000	507	0.06
Upper Austria	65+	305000	187	0.06
Salzburg	16-24	52000	85	0.16
Salzburg	25-64	311000	507	0.16
Salzburg	65+	116000	189	0.16
Styria	16-24	109000	80	0.07
Styria	25-64	690000	502	0.07
Stria	65+	274000	199	0.07
Tyrol	16-24	69000	83	0.12
Tyrol	25-64	429000	515	0.12
Tyrol	65+	152000	183	0.12
Vorarlberg	16-24	39000	89	0.23
Vorarlberg	25-64	223000	514	0.23
Vorarlberg	65+	77000	178	0.23
Vienna	16-24	203000	93	0.05
Vienna	25-64	1158000	535	0.05
Vienna	65+	329000	153	0.05

4. Expected outcomes and implications

4.1. Analysis methodology

As mentioned, the primary objective of the survey is to adjust MNO data with respect to duplication and representativity, for the resident target population. Hence, the survey is not target specific and will provide results useful in different domains. However, some specific statistical outputs might require specific adjustment, including the inclusion of dedicated questions in the questionnaire, or a different survey design. E.g. inbound tourism statistics will require the definition of a different target population compared to the country residents, and consequently different survey design and techniques, as outlined in section 6.

4.2. Enhancements in Data Quality

The survey is expected to be used in combination with methodologies from the methodological framework developed in WP3 (see *D3.2 Report on methodologies Zhang et al. (2025)*) to significantly improve the representativity of MNO data. Accurate assessments of the user behaviours connected to demographic profiles, should deliver the possibility to compute accurate models to adjust MNO data, far more precise than using mobile market shares for the expansion or correction of MNO data. This could be also to the advantage of the mobile operator as their results become more reliable. Improved data quality will bolster the reliability of policy decisions made based on mobile data analytics.

5. Report on the Friendly User Test

To refine and to improve the developed questionnaire, a *Friendly User Test* (FrUsT) was carried out to receive feedback regarding language, structure, and content. Therefore, participants were asked to focus on the following aspects:

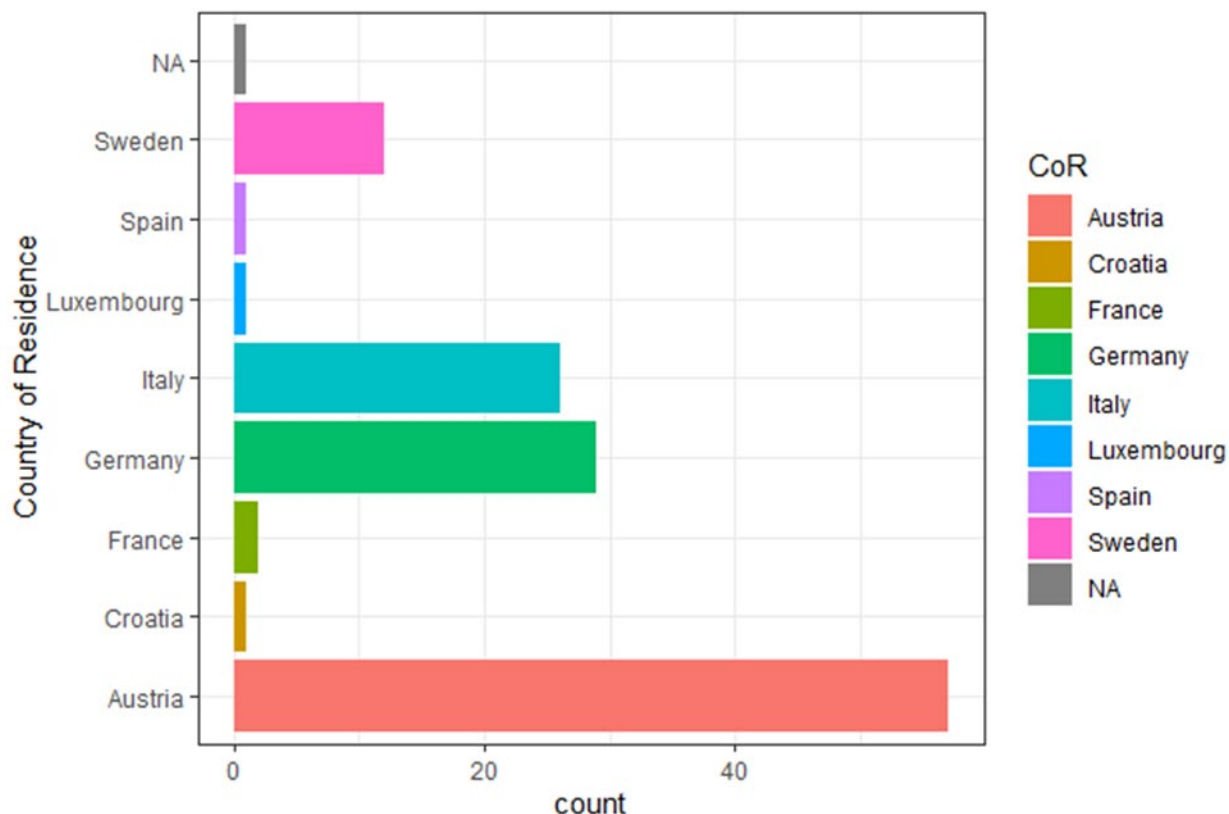
- Are the questions clear and easy to understand?
- Are there any questions you believe are too intrusive from a privacy perspective?
- Are questions easy to answer?
- Do you have any remarks regarding wording / tone of the questions?

Because of the friendly user test's aim to gain valuable feedback and insights to improve the questionnaire afterwards, the main group of invited participants were work colleagues from the project partners who were participating in this work package. This provided a couple of advantages: By inviting work colleagues, the overall organisation and bureaucratic matters were kept commensurate. Further, response rates were expected to be much higher when inviting especially direct work colleagues while at the same time providing a low-threshold for extensive feedback via e-mail and the possibility to ask questions. Then, some countries involved their survey specialists for additional and more detailed feedback. Another aspect was that some technical shortcoming in this early testing could be explained in more detail.

When looking at the results from the FrUsT, one must keep in mind that this main group of respondents is quite homogenous which reflects not only in the background information provided (e.g. employment status) but in all sections in the questionnaire (e.g. the use of business phones).

The second group of participants were members of the Task Force MNO because they could provide feedback having their knowledge and expertise on the topic in mind and may provide feedback especially regarding the questionnaire's content.

Figure 1: Number of respondents by country of residence.



The friendly user test was prepared using the EUSurvey tool¹, which comprised some benefits and contained some drawbacks. The main benefit was that it is centrally provided by the European Commission and is free to use. Additionally, it is rather easy to set up a survey and to develop the questionnaire. The main drawback is that it is not built to be used with a complex questionnaire as the one developed for this use-case and it does not provide many of the functions that are standard in tools commonly used in official statistics. The filtering of questions is possible but is limited to very basic checks and conditions. The same applies for routing within the questionnaire. However, the focus of this friendly user test was not to work on production-ready implementation, but to collect feedback on the developed questionnaire, to see if respondents are generally able to answer the questions from a knowledge and a privacy point of view.

In the next subsections some descriptive results of this very selective group of respondents are shown. Additionally, the feedback is summarized, and the resulting updated questionnaire design is presented.

5.1. Results from the Friendly User Test

In total, 130 persons participated in the friendly user test. Most of these respondents were from the partner countries of this work package: Sweden, Italy, Germany and Austria (see Figure 1). Due to the limited number of participants as well as the relatively homogenous sample, results from

¹ <https://ec.europa.eu/eusurvey/home/welcome>

the friendly user test are not representative and one cannot draw conclusions from it. However, results illustrate

Mobile devices and contracts

The question on the number of phones showed most of the respondents only have one phone (Figure 2). However, this fact is influenced a lot by the country of residence which in our case equals a specific national statistical institute (see figure 3). Therefore, the more general conclusion and somehow obvious conclusion is that most respondents will have a second phone only if this second phone is provided by their employer.

The distribution of number of sim cards (see figure 4) used by the respondents is a bit more diverse with up to four sim cards used by a respondent. However, also for this question most users only used one sim card.

How accurate is the demographic information that MNOs can provide based on the contract information? A possible hint is that almost 75% owned all the contracts themselves (see figure 5). All the contracts from respondents were in their country of residence. About 86% of the contracts from respondents are older than one year (see figure 7). Many respondents did not indicate their service provider and there is no information whether respondents did not want to share this information or whether they do not know their service provider.

Use of mobile devices in everyday life

About 90% of respondents have a mobile phone with them all the time (see figure 8). The few respondents who occasionally switch off their phone or do not always carry their phone with them, mainly named “during sleep” or “other activity” as a reason.

Figure 2: Number of respondents with 1 or 2 phones.

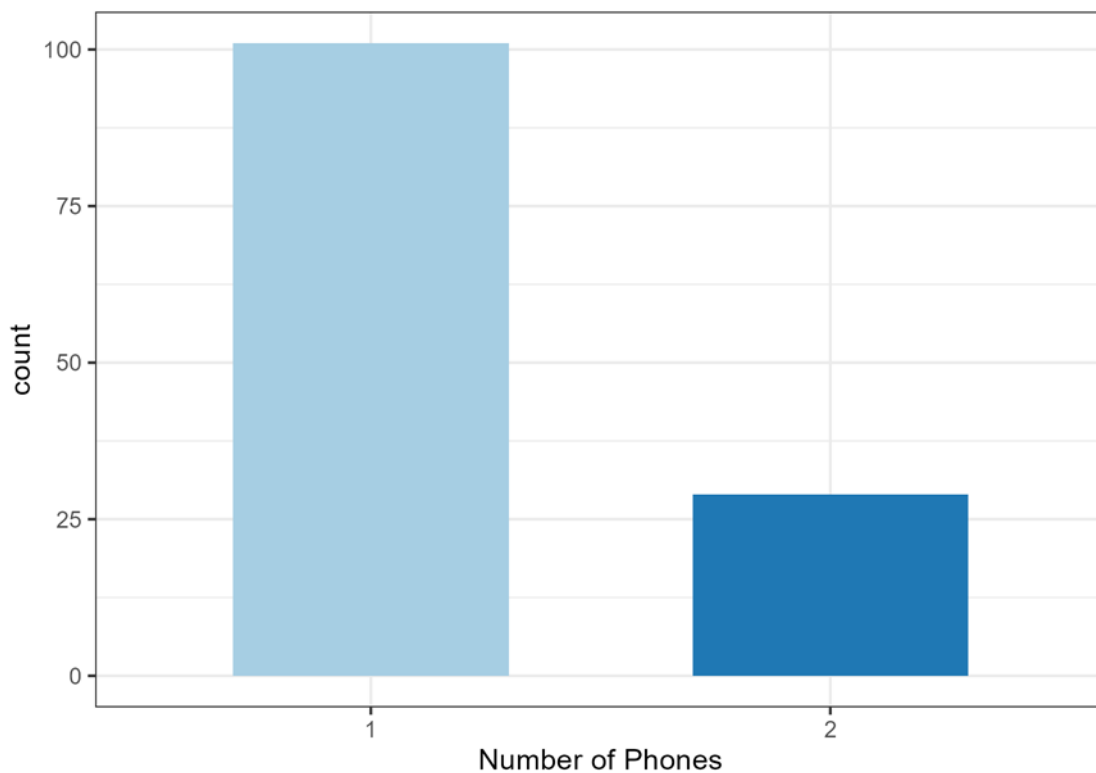


Figure 3: Number of respondents with 1 or 2 phones by country of residence.

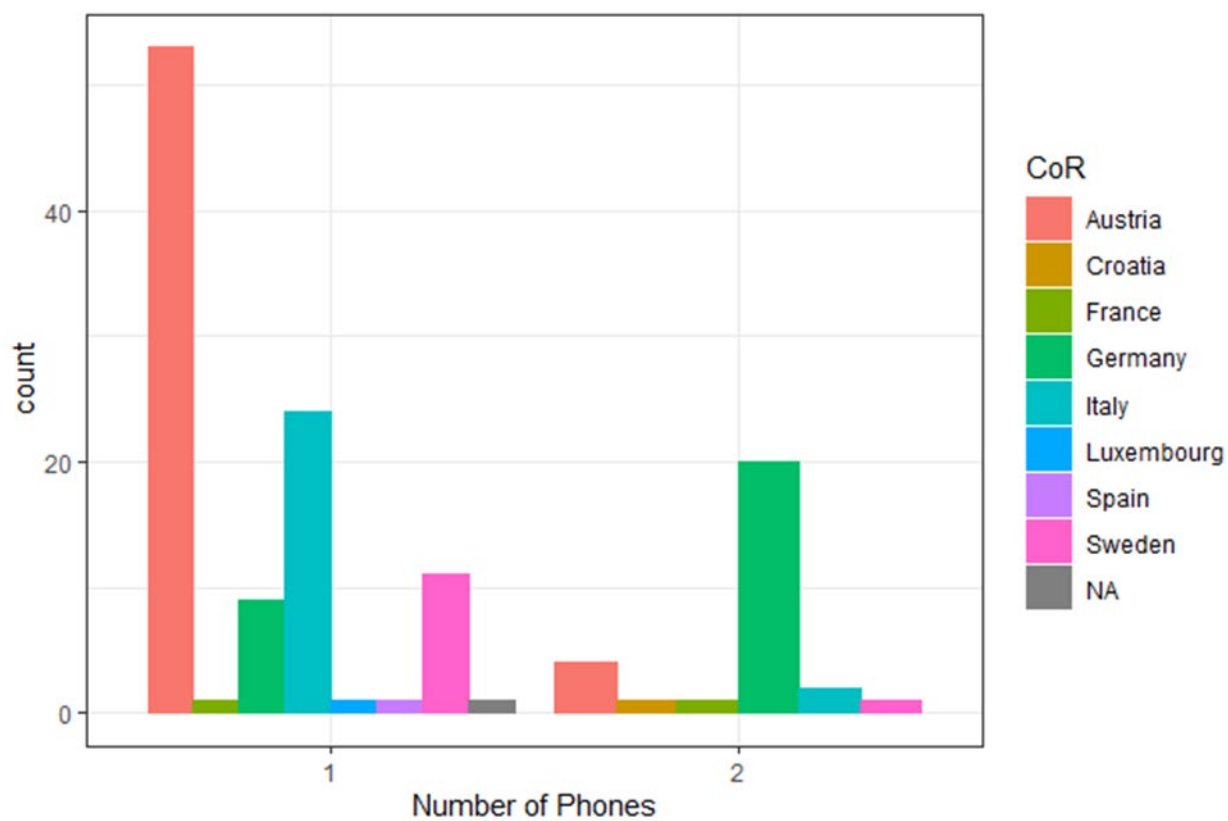


Figure 4: Number of respondents by the number of sim cards.

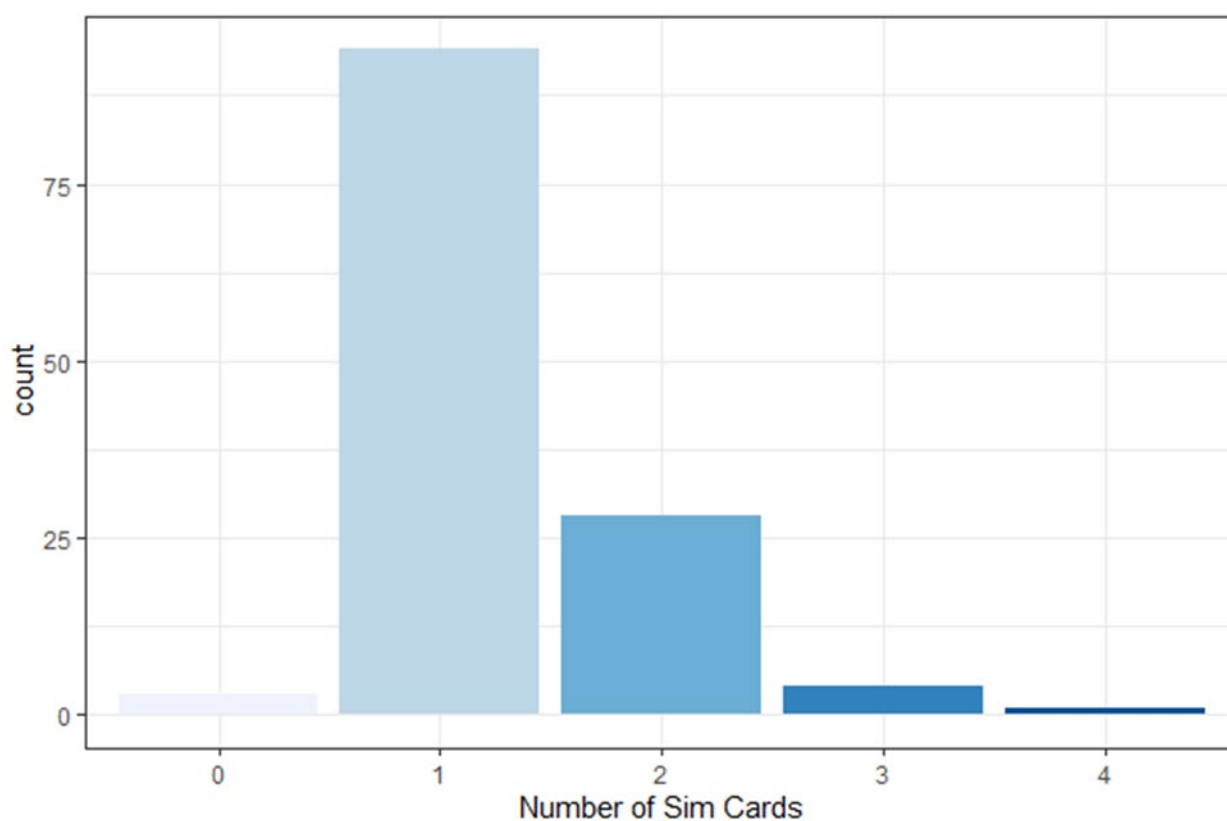


Figure 5: Number of respondents that owned all the contracts of their sim cards.

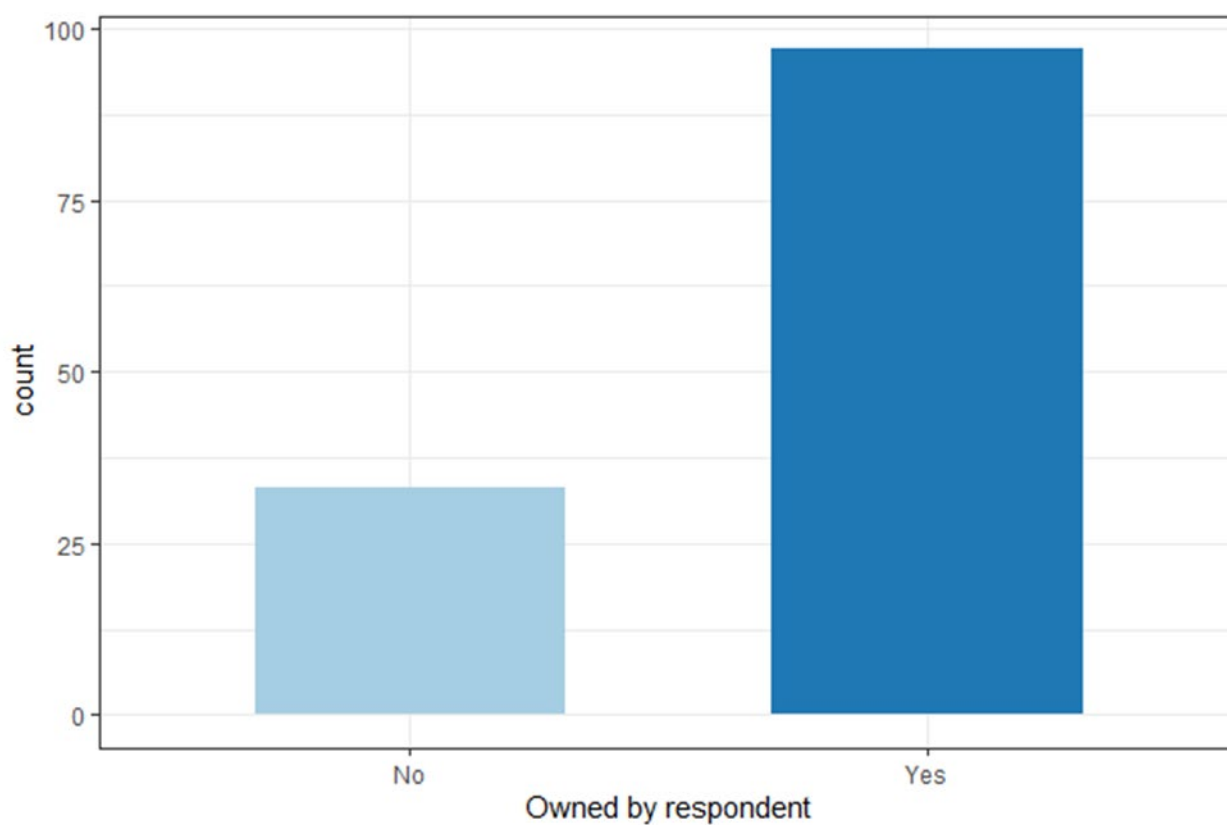


Figure 6: Number of phones by purpose of use.

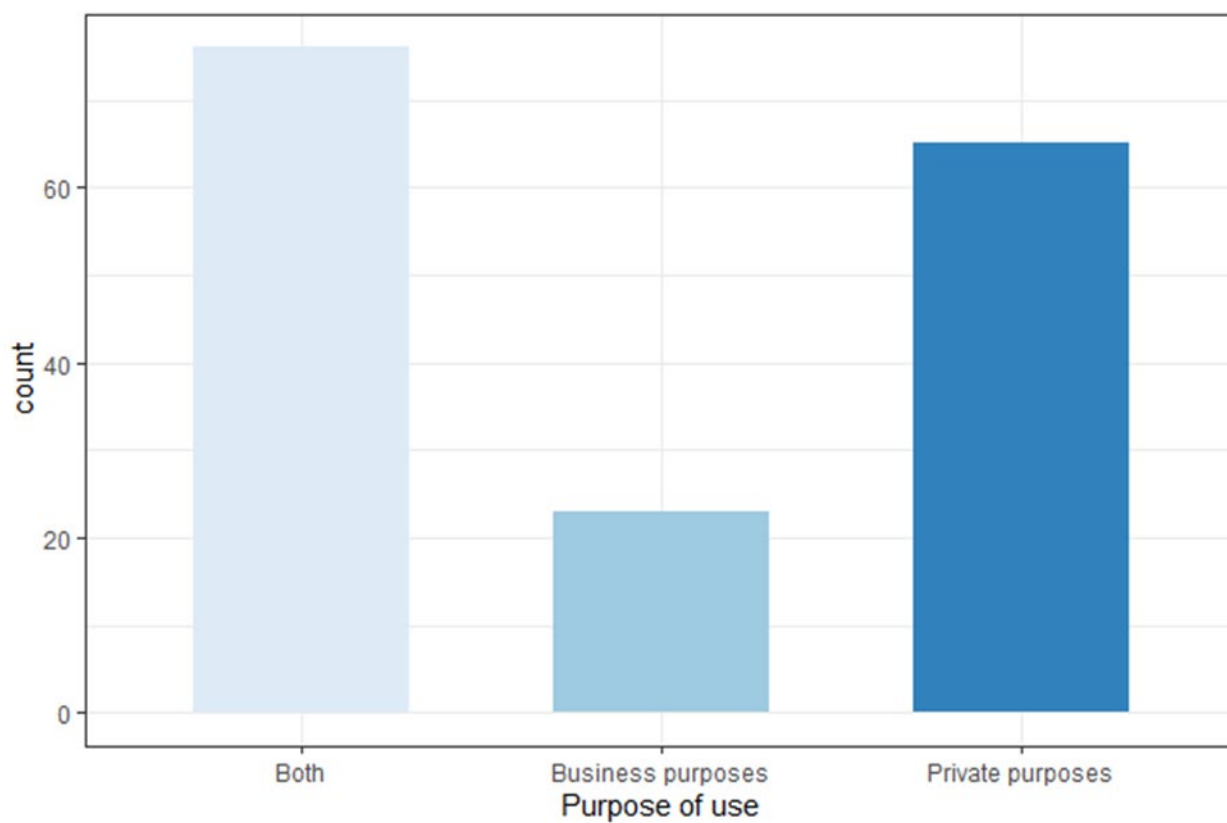
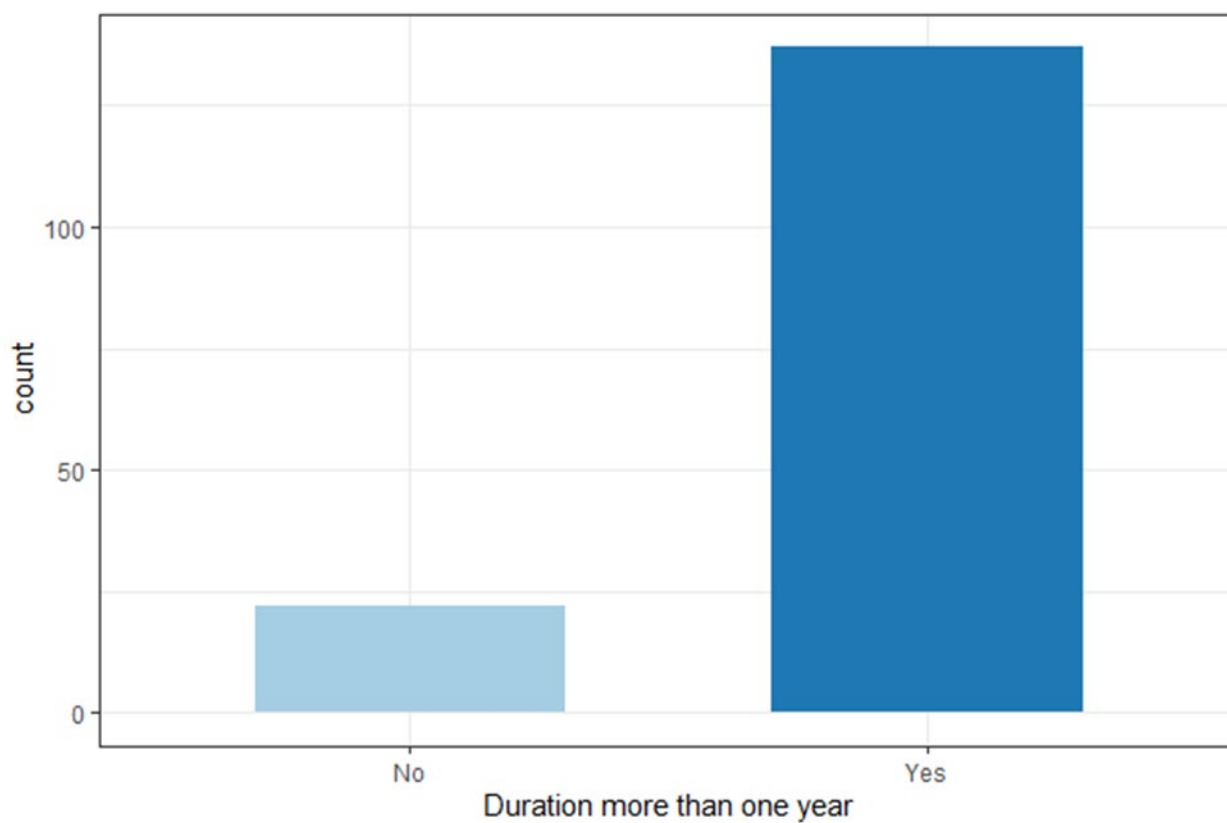


Figure 7: Number of contracts by duration more than one year.



About 90% of respondents have a mobile phone with them all the time (see figure 8). The few respondents who occasionally switch off their phone or do not always carry their phone with them, mainly named “during sleep” or “other activity” as a reason.

Most respondents indicated that they are commuting several days per week to the location of their main activity (e.g. office, university, ...) inside their country of residence (Figure 9).

Figure 8: Number of respondents that have their phone with them all the time.

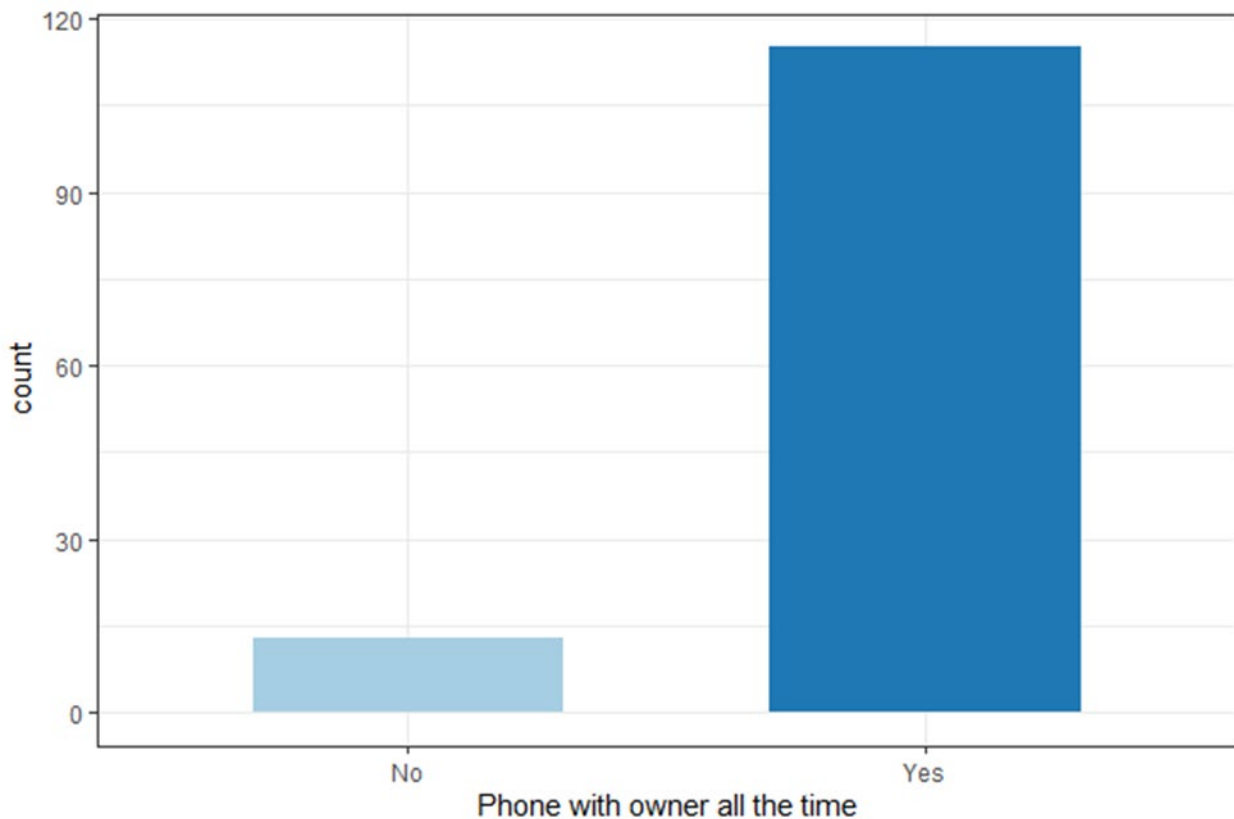
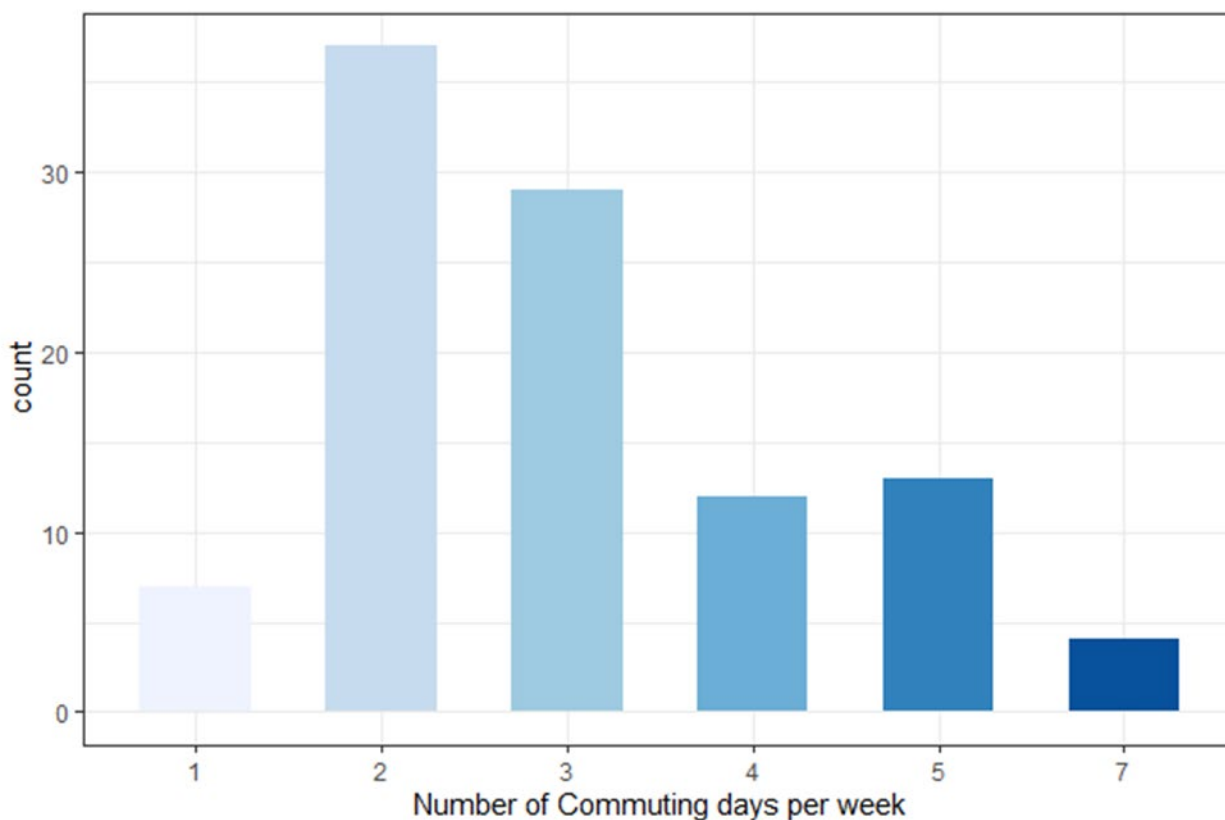


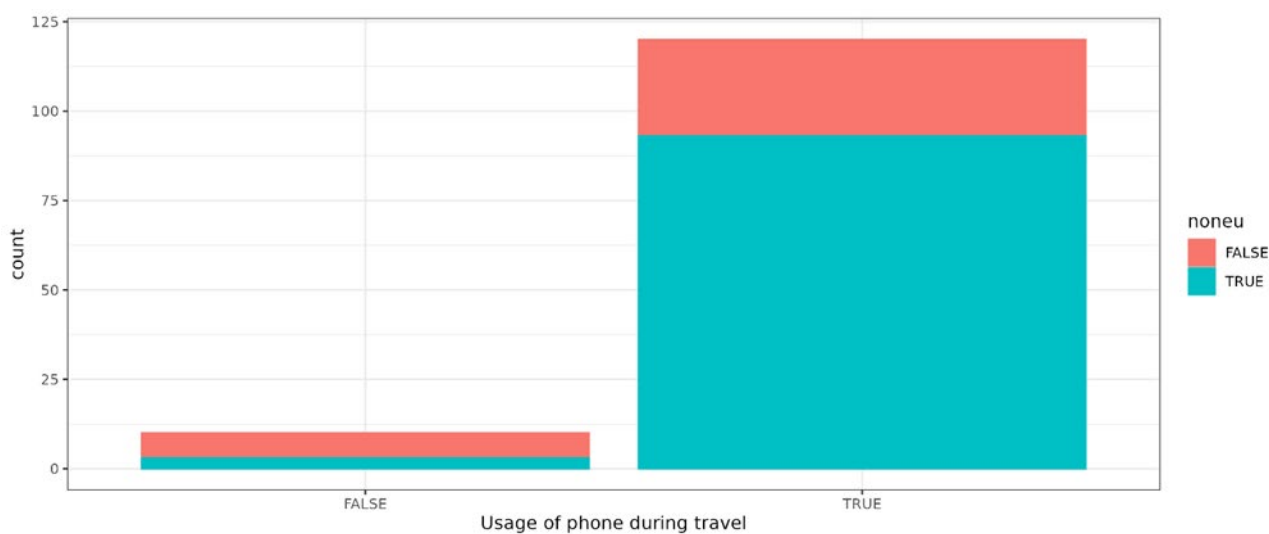
Figure 9: Number of days that commuters travel to the location of their main activity



Use of mobile devices during travels

A majority of respondents use their phones during travel in EU countries (see figure 10). The same is true for non-EU travel but to a smaller extent.

Figure 10: Number of respondents that use their phone during EU travel coloured according to usage during non-EU travel.



5.2. Summary of feedback

During the friendly user test, the freely available online tool EUSurvey by the European Commission was used. This tool provides an easy way to run multi-language surveys and convenient features like a list of all EU countries (or all countries) can be selected as possible answers for questions. However, this tool is not powerful enough to run as complicated survey questionnaires as we are used to in official statistics. This was only obvious after the decision was made to use this tool and part of the questionnaire was already implemented. Therefore, we choose to use the EUSurvey tool anyway for the friendly user test and to add a disclaimer to all invited friendly users with an explanation that the questionnaire is not implemented in an official statistics ready survey tool and issues directly related to the tool can be ignored. For respondents it seemed difficult to separate between issues of the questionnaire and issues of the tools, so the feedback is still partly on issues of the tool.

The feedback strongly emphasizes the importance of logical flows and user-friendly formats. Improved routing logic and better phrasing would enhance the survey's usability. Technical terms like "ISCED" should not be used and clear explanations and examples should be provided where complicated questions are asked. The phrasing of similar questions should be consistent and ensure that the questions are understood. Ambiguous terms, e.g., "commute" or "switched on" need a clear definition. Dropdowns could be used instead of radio buttons for easier navigation and to reduce the cognitive load. Sensitive questions, like income, should include options like "Prefer not to say."

5.3. Feedback regarding Questionnaire design

The current questionnaire that incorporates the feedback from the friendly user test can be found in the annex. Here are examples of specific user comments on questions.

2) How many SIM cards (incl. eSIM) do you use in your first mobile phones?

- The term "First mobile phones" is confusing, maybe the use of "main mobile phone" or similar would be better.
- Ensure clarity on whether the question applies to all phones or the main one.

3) For which purposes do you use your mobile phone?

- Add "Both, private and business purposes" as an option for clarity.

4) Which service provider do you use for your mobile phone?

- Include "I don't know" as an answer option for respondents unsure of their provider.

7) Which service provider did you have before your current one?

- Add an "I don't know" option for respondents who don't remember.

8) In which country is the service provider you had before your current one located?

- Specify if the question refers to the company headquarters or the local branch.

21) Is your mobile phone with you all the time?

- Clarify whether "all the time" includes sleep or other periods when it's nearby but not actively used.

22) Is your mobile phone switched on all the time?

- Specify if "switched on" includes airplane mode or low-power settings.

36) How many people living in your household are under the age of 14?

- Ensure consistency with earlier questions about children under 16.
- Include an option for "None."

37) What is your net household income per month?

- Provide an option for "Prefer not to say."
- Include income ranges or categories for easier responses.
- Clearly define "net income" and whether to include all household members.

40) Where is your place of residence? (City/Country)

- Clarify whether both city and country are required or just one.
- Use dropdown options for easier selection.

41) What is your highest level of education?

- Replace ISCED levels with national equivalents or examples like "High School" or "University."

6. Complementary Approaches

The outlined approach to make a traditional sample survey will cover well the registered resident population of a specific countries. However, as mentioned before in section 2.3, registered residents are only a subgroup of all mobile phone user on the territory of a country. One way of segmenting mobile phone users into groups are registered residents, non-registered "residents", e.g. regular commuters, expats etc., and tourists staying just for a short time span. All three groups might use a foreign or local SIM card, so a simple distinction between the groups by the nationality of the sim card is not possible. There are many reasons for this "mixture", we can only list some examples here:

Non-EU tourists might buy an EU roaming (data) SIM card at their first arrival in the EU, which could be outside of the research country, e.g., tourists visiting Vienna, could have arrived in Frankfurt before and bought a SIM card there.

Residents with a connection to another country, e.g., family, a second home, etc. could use a SIM card of this country.

Obviously, the proposed traditional sample survey can only cover part of all mobile phone users, the residents. So, an alternative sample design is to sample areas (antennae) and consider in the sample all individuals connected to the selected antennae in a given time interval. This design would have the benefit of considering the actual present persons in the country. However, it presents technical and possibly legal issues for implementation.

Ideally, an area-based survey could use a similar, but adapted, questionnaire as the traditional survey. A combination of the approaches of a traditional survey from a sampling frame and the area-based survey could be a very powerful tool to improve the quality of MNO data. The sampling design of the area-based survey should be stratified in space and time. Areas with a suspected higher density of non-resident mobile phone users could be oversampled, because

residents are already well covered by the traditional survey. The time points for the sample should be distributed along different times of the day, different weekdays and different seasons.

An alternative approach compared to the design and carrying out of a devoted survey, can be based on a register-based model for the MNO pseudo-selection mechanism. For a proposal see deliverable D.3.1 paragraph 3.2.1.

6.1. Design of an area sample

Mobile Network Operator (MNO) data, while providing high-frequency spatiotemporal data that cover nearly the entire national territory and most of the population, presents several methodological challenges. These challenges include heterogeneity in population coverage and variability in data utilization patterns. To rigorously assess the impact of these issues, it is advisable to design and implement dedicated surveys. These surveys can facilitate the identification and mitigation of coverage gaps and usage discrepancies, thereby enhancing the representativeness and reliability of the collected data.

To accurately target the coverage and usage characteristics of the resident population, a list sampling methodology was proposed in deliverable D3.1. Comprehensive details regarding this approach can be found in section 3 of this volume.

In the current section, we explore various alternative designs aimed at addressing distinct quality issues inherent in MNO data:

- a. When utilizing MNO data to estimate characteristics targeting the resident population, it may be necessary to adjust the counts to account for over-coverage due to non-residents owning a national SIM card. This adjustment is crucial to ensure the accuracy and representativeness of the population estimates. MNOs likely possess the requisite information to differentiate between resident and non-resident subscribers based on the registration details of the SIM cards. By leveraging this subscriber information, it is possible to identify only the resident population and possibly cleaning the MNO data from data generated by non-usual residents, thereby enhancing the validity of the data.
- b. If this is not the case, in order to adjust the national SIM cards for non-resident users, it is essential to estimate the proportion of out-of-scope units (non-residents) by adopting a sampling strategy that targets all the national SIM cards. In this context, a list sampling of the resident households may not be the most suitable approach. Instead, a sample of active SIM cards is more appropriate. This can be achieved through an area sampling method, focusing on SIM cards connected to specific antennas.

More specifically, a two-stage sampling approach should be applied. In this design, the first-stage units are the network antennas, and the second-stage units are the SIM cards connected to these antennas at a given time.

Alternatively, after selecting the antennas, all the SIM cards connected to them can be included in the sample (cluster sampling). However, we recommend the first option of a two-stage sample, since the average number of SIM cards connected to an antenna is approximately 1,000, that is a quite large number for the expected sample size of this kind of survey.

By implementing this two-stage sampling strategy, we can ensure a more manageable and representative sample.

To improve the sampling strategy, antennas can be stratified on the basis of the number of foreign SIM cards that are connected to. This stratification assumes a correlation between the number of foreign SIM cards and the proportion of national SIM cards owned by non-residents (or foreigners).

We suggest not including all SIMs connected to the antennas in an interval of time to avoid overlapping units and difficulties in sampling inclusion probabilities calculation.

For this reason, the sample should be taken at fixed time points, stratifying for time of the day, weekdays and other relevant seasonal characteristics.

This area sample that takes antennas of the operator networks as sample units can be weighted on the number of actual connected mobiles to reflect the actual distribution of mobile users across the operator's network, improving the reliability of any analysis based on this data.

c. To explore tourists' behaviour in the use of SIM cards, a similar ad hoc area sample might be designed, focusing exclusively on foreign SIM cards. The stratification of antennas or balancing of the antennas sample should follow the same methodology as previously described, a second stage sample should then be related only to no national SIM.

However, the main objective of assessing the MNO coverage of tourists, a parameter needed for the application of the quasi-randomization approach (see deliverable 3.2), is not attainable with the described area sample based on MNO data only, in particular for tourists' households where no member uses the mobile phone and cannot be reached by this type of survey.

As a result, the sample miss part of the tourist population, making it impossible to correctly assess the MNO coverage of tourists and apply the quasi-randomization technique to estimate the target measures accurately.

Moreover, the antenna sample survey presents technical and possibly legal issues for its implementation. First, the survey can be performed only via the MNO and the contact details for the sampled devices, such as phone number, would need to be shared with institutes running the survey. This approach has obvious privacy, legal and operational obstacles to be removed before it can be further explored.

Finally, when a survey is distributed via a link (for example, through SMS or email), the response rate is expected to be very low. Many recipients may ignore the message, find it inconvenient, or not trust it. Because of this, it is difficult to predict or control how many people will actually complete the survey.

d. For the objective of estimating the foreign MNO coverage, an alternative area sampling method could be then more effective. The alternative area sample design involves selecting "exit points", similarly to the approach used by the EU central banks for estimating the tourists' components of the balance of payment. By targeting tourists, we can estimate the actual use of devices (national or foreigner) and more accurately estimate the coverage of data generated by foreign SIM cards.

The technique adopted for collecting data on the tourism balance of payments involves interviewing a representative sample of travellers—both residents and non-residents—crossing the borders,

while simultaneously conducting qualified counts to determine the number and nationality of travellers in transit.

Sampling is carried out independently at each type of border (roads, railways, airports, and international ports).

- Sampling is usually conducted using systematic sampling (observing every n -th unit),
- Face-to-face interviews, conducted using a structured questionnaire administered to a random sample of travellers at the end of their stay (i.e., when non-residents leave the country).
- The sample is stratified according to different variables for each type of border. Stratification of the sample could be by type of carrier (road, rail, air, sea)

Selection of the border points can be done based on annual traffic, but some minor points should be also included to capture the otherwise underrepresented origin-destination flows.

For selection of road borders, additional stratification variables include:

- Day of observation
- Time band (morning, afternoon, night)
- Day type (weekday, holiday)

For rail, air, and sea borders, administrative data (e.g., train or flight schedules) allow direct sampling of the carrier or route rather than time-based sampling.

6.1.1. Sample size calculation

To determine the minimum sample size n for the survey one can consider different hypotheses e.g. on the value of the proportion of foreign tourists not using the mobile phone when visiting the country. One possibility to estimate the sample size is by applying the formula from Cochran (1977, p.76ff). Here, one would need also the expected number of foreign tourists for example from previous years or n_0 is used in case $\frac{n_0}{N}$ is negligible (no finite population correction). The formulas are

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$$

and

$$n_0 = \frac{Z^2 p (1 - p)}{e^2}$$

.

Where:

- Z-score (the number of standard deviations from the mean) based on desired confidence level
 - p : estimated proportion of the population: in this case proportion of the tourists not using the mobile phone (use 0.5 if unknown, for maximum variability)
- e : desired margin of error (in decimal form, e.g., 0.05 for 5%)

Assuming $\frac{n_0}{N}$ is negligible, which is generally true if the estimation domains are not too small, the two-sided confidence interval at 95% ($Z=1.96$) with the desired error = 4%, with the proportion $p=0.5$ the minimum sample size is 600. So, 600 (net) units need to be available per estimation domain (e.g., a certain region) to estimate the proportion with given quality. This value should then be inflated by the clustering effect expected for the survey. For the survey at the border this value can be assumed to be equal to the one estimated for the current survey by the Central Banks on similar variables.

7. Conclusion and next steps

The use of Mobile Network Operator (MNO) data offers a promising avenue for enhancing statistical analysis and policy-making, given its wide coverage, high frequency, and rich spatiotemporal detail. However, several limitations—such as device-user ambiguity, multiple SIM cards per user, and behavioural heterogeneity—pose serious challenges to its representativity and statistical usability. This report has systematically addressed these issues by proposing a dedicated survey aimed at identifying and correcting the structural biases inherent in MNO data.

The proposed survey framework focuses on understanding real-world device and SIM card usage patterns, socio-demographic disparities, and contract ownership details. Its aim is not merely descriptive but corrective—designed to support a quasi-randomization model that can improve the representativity of MNO data across different statistical domains. The pilot user test provided encouraging signals regarding the feasibility of collecting such data, while also highlighting key areas for improvement in survey logic, language clarity, and question routing.

Complementary to the traditional survey design, this report explored alternative sampling strategies—most notably, an area-based sampling design that uses antennae as primary sampling units. While this approach has considerable potential for capturing the "present population", including non-residents like tourists and commuters, it faces significant legal, operational, and technical challenges that need to be carefully navigated.

In its current form, the proposed methodology promises to drastically improve the quality and applicability of MNO data, enhancing its use in areas such as commuting analysis, tourism statistics, remote work estimation, and mobility pattern detection. These improvements will, in turn, lead to better-informed policy decisions at both the national and European levels.

7.1. Next Steps

To translate this proposal into actionable outcomes, the following steps are recommended:

1. Finalize and pilot the enhanced survey

The feedback from the friendly user test was integrated in the survey questionnaire. Now the next step is to pilot the revised survey in a pilot to test its effectiveness and gather additional feedback before full deployment.

2. Determine the sample design and size

Conduct sample size calculations based on desired precision levels and potential response rates. Explore oversampling techniques for underrepresented groups.

3. Develop the area-based and border-based survey approaches

Draft a comprehensive implementation framework for the antenna-based sample.

Identify legal and operational prerequisites, especially regarding access to and anonymisation of contact data.

Develop a parallel questionnaire tailored to the constraints and opportunities of this sampling design.

Consider coordination with MNOs to test the technical feasibility and data extraction protocols.

4. Engage stakeholders and secure collaboration

Involve national statistical institutes, legal experts, and MNOs in the next design phase to address data protection and ethical concerns.

Encourage partnerships with institutions responsible for mobility, transport, and tourism statistics to ensure cross-sectoral utility.

5. Develop adjustment models for quasi-randomization

Use survey results to build and calibrate statistical models that adjust MNO data for biases related to device ownership, SIM usage, and user behaviour.

Validate these models through comparison with traditional statistical indicators where available.

6. Explore alternative and complementary sampling frames

Evaluate the feasibility of an exit-point sample for assessing tourism-related MNO data.

Investigate the potential for register-based modelling as outlined in Deliverable D3.1, paragraph 3.2.1, as a complementary approach.

7. Develop an EU-wide repository for survey tools and results

Creating a shared platform where NSIs can contribute, access, and benchmark results from their

respective MNO-related surveys will promote knowledge transfer, reduce duplication of efforts, and support methodological innovation.

8. Invest in capacity building and knowledge transfer

Eurostat should support workshops, training sessions, and technical guidelines to help NSIs implement the new survey methodology and integrate MNO data into their statistical production processes.

8. References

- Zhang, L-C, J.K. Haug, J. Fosen, D’Orazio M., Di Consiglio L., Pichiorri T., Sanguiao-Sande L., Barragán S., Saez Calvo C., Novas Filgeira M. "D3.2 Report on methodologies" MNO.MINDS (2025)
- Kowarik A., Tuoto T., Di Consiglio L., Deetjen G., Brandt M., Kamali R., Gylling K., Faricelli C., Piombo S., Pichiorri T. "D4.1 Questionnaire design and main survey design aspects for an ad-hoc survey as a potential source to improve MNO data " MNO.MINDS (2024)
- Cochran W.G., Sampling Techniques, 3rd Edition, John Wiley & Sons (1977)

9. Annex: Questionnaire for area sample

- Gender
- Age
- Live in the country (Resident) : Yes/No
 - o If not:
 - ☐ Reason for the travel? (tourism, business)
 - ☐ Nationality (classes)
 - ☐ Own national operators on/off (for duplicates)

Questionnaire outline for area sample 2 (antenna-foreign SIM only)

- Gender M,F
- Age
- Live in the country (Resident): Yes/No
 - o If not resident:
 - ☐ Reason for the travel? (tourism, business)
 - ☐ Nationality (classes)
- Do you have a country SIM? (duplication)

Questionnaire outline for border's sample

- Gender M,F
- Age
- Reason for the travel? (tourism, business)
- Nationality (classes)
- Have you used mobile phones during the stay?
- Always ON? Few questions on time as for the resident version
- Do you have a country SIM? (duplication)

10. Annex: Questionnaire: Use of mobile data (updated after friendly user test)

I. Mobile devices and contracts

1. How many mobile phones do you use? Please include all mobile phones used for private and business purposes.

2. How many SIM cards (incl. eSIM) do you use in your (alias: first/second/third) mobile phone?

3. For which purposes is your (alias: mobile phone/first/second/third SIM card) used?

- a. Private purposes
- b. Business purposes
- c. Both, private and business purposes

4. A) Which service provider do you use for your mobile phone?

B) Which service provider do you use for your (alias: first/second/third) SIM card in your (alias: first/second/third) mobile phone?

Provide a dropdown menu with all service providers. The menu should include the options "Other service provider located in my country of residence" and "Other service provider located abroad"; if "Other service provider located abroad" -> 5, else -> 6 ; The list should also include a "Don't know" option.

5. A) In which country is the service provider of your mobile phone located?

B) In which country is the service provider of your (alias: first/second/third) SIM card in your (alias: first/second/third) mobile phone located?

Provide a dropdown menu with all countries.

6. A) Have you been with your service provider for more than a year?

B) Have you been with the service provider of your (alias: first/second/third) SIM card in your (alias: first/second/third) mobile phone for more than a year?

- a. Yes -> 7
- b. No -> 9

7. A) Which service provider did you have before your current one?

B) Which service provider did you have before your current one for your (alias: first/second/third) SIM card in your (alias: first/second/third) mobile phone?

Provide a dropdown menu with all service providers. The menu should include the options “Other service provider located in my country of residence” and “Other service provider located abroad”; if “Other service provider located abroad” -> 8, else -> 9. The list should also include a “Don’t know” option.

8. A) In which country is the service provider you had before your current one located?
B) In which country is the service provider you had before your current one for your (alias: first/second/third) SIM card in your (alias: first/second/third) mobile phone located?

Provide a dropdown menu with all countries.

9. A) Who is the contract owner of your mobile phone?
B) Who is the contract owner of your (alias: first/second/third) SIM card in your (alias: first/second/third) mobile phone?
- a. Myself -> 13
 - b. Other person -> 10
 - c. My business (for self-employed) -> 13
 - d. My employer -> 13

10. What is the relationship to the contract holder?
- a. My parent
 - b. My partner (e.g. wife/husband, girlfriend/boyfriend)
 - c. My child
 - d. Other person not listed here: _____

11. How old is the contract holder?

Don't know

12. Which gender does the contract holder identify with?
- a. Female
 - b. Male
 - c. Non-binary
 - d. A gender not listed here
 - e. Prefer not to say
 - f. Don't know

13. Do children under 16 live in your household?
- a. Yes -> 14
 - b. No -> 17

14. How many children under 16 live in your household?

15. How many of them have their own mobile phone with an active SIM card?
_____ 0 -> 17, >0 -> 16

16. Please provide for every child under 16 who has access to their own mobile phone with an active SIM card, how old they are.
-

II. Use of mobile devices in everyday life

17. What best describes your livelihood situation?
- Employed -> 18
 - Self-Employed -> 18
 - In school or university -> 18
 - Unemployed -> 21
 - Out of labor force (e.g. homemaker) / Retired-> 21
18. Do you commute to (alias: work/school or university) on at least one day per week?
- Yes, within the country -> 20
 - Yes, crossing borders to another country -> 19
 - No -> 21
19. To which country are you commuting?
Provide a dropdown menu with all countries. Put the neighboring countries first.
20. On average, how many days a week do you commute?
-
21. Is your mobile phone with you all the time?
Provide an explanation: "You have your mobile phone with you all the time if you leave the phone in another room, e.g., during sleeping or charging the phone. The phone is not with you all the time, if you leave the phone at home to do outdoor activities, go to work or travel."
- Yes
 - No
22. Is your mobile phone switched on all the time?
Provide an explanation: "Switched on means in this context that the phone is powered on and connected to a network (if a network is currently available at that location)."
- Yes -> if 21a, then 24, if 21b, then 23
 - No -> 23
23. When is your mobile phone not with you or switched off?
- During work, school or university
 - During sleep
 - Other activity not listed here: _____

III. Use of mobile devices during travels

24. Did you travel abroad in the last 12 months?
- Yes -> 25
 - No -> 35

25. Where did you travel abroad in the last 12 months?
- Inside the EU
 - Outside the EU
 - Both
26. What was the purpose of your travels abroad?
- Business purposes -> 31
 - Private purposes -> 27
 - Both -> 27
27. A) In your travels abroad for private purposes to an EU country, do you use your mobile phone?
B) In your travels abroad for private purposes to an EU country, do you use your (alias: first/second/third) SIM card in your (alias: first/second/third) mobile phone?
- Yes -> Routing
 - No -> Routing
28. A) Do you leave your mobile phone on most of the time each day during your travels for private purposes within the EU?
B) Do you leave your (alias: first/second/third) SIM card in your (alias: first/second/third) mobile phone on most of the time each day during your travels for private purposes within the EU?
- Yes -> Routing
 - No, it is switched off or in flight mode most of the times -> Routing
29. A) In your travels abroad for private purposes to a Non-EU country, do you use your mobile phone?
B) In your travels abroad for private purposes to a Non-EU country, do you use your (alias: first/second/third) SIM card in your (alias: first/second/third) mobile phone?
- Yes -> Routing
 - No -> Routing
30. A) Do you leave the mobile phone on most of the time each day during your travels for private purposes outside the EU?
B) Do you leave your (alias: first/second/third) SIM card in your (alias: first/second/third) mobile phone on most of the time each day during your travels for private purposes outside the EU?
- Yes -> Routing
 - No, it is switched off or in-flight mode most of the times -> Routing
31. A) In your travels abroad for business purposes to an EU country, do you use your mobile phone?
B) In your travels abroad for business purposes to an EU country, do you use your (alias: first/second/third) SIM card in your (alias: first/second/third) mobile phone?
- Yes -> Routing
 - No -> Routing
32. A) Do you leave the mobile phone on most of the time each day during your travels for business purposes within the EU?

B) Do you leave your (alias: first/second/third) SIM card in your (alias: first/second/third) mobile phone on most of the time each day during your travels for business purposes within the EU?

- a. Yes -> Routing
- b. No, it is switched off or in-flight mode most of the times -> Routing

33. A) In your travels abroad for business purposes to a Non-EU country, do you use your mobile phone?

B) In your travels abroad for business purposes to a Non-EU country, do you use your (alias: first/second/third) SIM card in your (alias: first/second/third) mobile phone?

- a. Yes -> 34
- b. No -> 35

34. A) Do you leave the mobile phone on most of the time each day during your travels for business purposes outside the EU?

B) Do you leave your (alias: first/second/third) SIM card in your (alias: first/second/third) mobile phone switched on most of the time each day during your travels for business purposes outside the EU?

- a. Yes
- b. No, it is switched off or in-flight mode most of the times

IV. Background information

Participating NSIs may use administrative data instead of collecting the following information directly from respondents.

35. How many people live in your household?

_____ >1 -> 36, 1 -> 37

36. How many people living in your household are under the age of 14?

37. What is your net household income?

Use categories that are suited to the national situation. This should also include a "Don't know" and "Prefer not to say" option.

38. How old are you?

39. Which gender do you identify with?

- a. Female
- b. Male
- c. Non-binary
- d. A gender not listed here
- e. Prefer not to say

40. Where is your place of residence?

41. What is your highest level of education?

Use nationally known levels that correspond to the ISCED levels.

- a. ISCED 0-2

- b. *ISCED 3-4*
- c. *ISCED 5-8*