Approach to population estimation in Ukraine using mobile operators' data

Volodymyr Sarioglo[1], Maryna Ogay[2]

ABSTRACT

Estimating the size and places of residence of the population of Ukraine has been this country's problem for the past decade, and is related to the lack of census data for the 2010 round, large-scale processes of external and internal labour migration, and Russia's armed aggression against Ukraine that started in 2014. This already disadvantageous situation has been significantly exacerbated by Russia's full-scale war against Ukraine which began on 24th February 2022. Conducting statistical surveys, especially surveys regarding the population, turned out to impossible under war circumstances. Therefore, the task of developing effective approaches to estimating the population size using data from existing sources, in particular the data of mobile operators regarding the number, location and mobility of subscribers, has become even more pressing. The article highlights the results of a study on the use of data from mobile operators, data from administrative registers, and the results of a special population sample survey on the use of mobile communication for the purpose of estimating the population. It also provides the results of experimental calculations of the population size in Ukraine as a whole and in particular regions. The study moreover showed that the size of Ukraine's population in November 2019, unlike the official estimate of 41,940.7 thousand people, was probably about 37,289.4 thousand people.

The developed approaches can be used to estimate the number and location of the population of Ukraine during the intercensal period or significant population movements due to environmental disasters or military conflicts.

Key words: population estimation, mobile operators, mobile subscribers, administrative data, sample survey.

JEL Classification: C82, C83, J10

---

[1] Ptoukha Institute for Demography and Social Studies of the National Academy of Science of Ukraine, Ukraine. E-mail: sarioglo@ukr.net.net. ORCID: https://orcid.org/0000-0003-4381-9633.

[2] Ptoukha Institute for Demography and Social Studies of the National Academy of Science of Ukraine, Ukraine. E-mail: ogay@ukr.net. ORCID: https://orcid.org/0000-0001-9292-5710.

Problem statement

In independent Ukraine, the census was conducted only once, in 2001. Accordingly, in recent years, in the implementation of the census of large-scale processes of virtually uncontrolled labour migration, lack of reliable data on migration within the country, official estimates of the population as a whole and by individual territories and settlements have long been criticized by specialists. The situation was significantly complicated by the impact of Russian aggression in the East of Ukraine in 2014 and became one of the most relevant for the country in the full-scale war of 2022, when a significant number of people were forced to seek new places to live in other regions of Ukraine or abroad, or they were forcibly moved to the occupying country or found themselves in the temporarily occupied territories. Human losses due to the death of military and civilians should also be taken into account. All this has exacerbated the issue of finding and applying approaches to estimating and monitoring the population based on existing data sources in Ukraine without conducting a traditional census.

In our opinion, for the development of approaches to monitoring the population of Ukraine in the current conditions, the national experience of 2019-2020 is very useful to clarify the population of Ukraine based on available information from various sources, including data from mobile operators (hereinafter – DMO). Relevant surveys were conducted by the State Statistics Service of Ukraine (hereinafter – the State Statistics), Institute of Demography and Social Research named after M. Ptukha of the National Academy of Sciences of Ukraine, members of a working group from other organizations, at the initiative of Minister of Ukraine D. Dubilet.

The aim of the study is to highlight approaches to the operational assessment of the current population of Ukraine and its regions based on data from administrative registers and mobile operators.

In preparing the article, methods of statistical data processing from different sources, methods of data aggregation, methods of evaluation of indicators based on the results of sample surveys, etc., were used**.**

Overview of existing research and experience. According to the international experience, traditional data collection methods, such as censuses, cannot meet the needs of users for prompt and timely information when it is necessary to estimate the population living in certain areas in the event of natural disasters, armed conflicts or epidemics. In fact, censuses conducted once every ten or five years have not been considered satisfactory to users lately, even under normal conditions, without cataclysms. This is explained by the fact that the availability of reliable and up-to-date

information on the population is now an increasingly important prerequisite for the development and implementation of effective and efficient measures of socio-economic policy at all levels of administrative-territorial division. Under such conditions, more and more attention of researchers, primarily statisticians and data scientists, is paid to the development of methodological approaches and applied tools for estimating population, demographic phenomena and processes based on available data from sources other than traditional censuses. An important result of these studies is also the possibility of a radical modernization of the methodology and tools of censuses. A promising process in this regard in many countries is the use of data from administrative registers, which contain information about the population as a whole, its individual groups, households, etc., as well as DMO on the location and/or movement of subscribers.

The article focuses on approaches to estimating the number and location of the population based on data on mobile network subscribers and data from some other sources. Mobile network or DMO data is the data collected by mobile network operators as a product of the operation of a mobile network. DMOs consist of two types of data: communication data – mainly calls and text messages; and positioning data – determination and fixation by mobile operators of the approximate location of the device based on the identification of the antenna used for communication (see, for example, Handbook on the use of Mobile Phone data for Official Statistics (2017)).

Data of mobile operators are increasingly used in statistics to estimate population movements (labour migration, pendulum migration, tourism, number of internally displaced persons, etc.), population density, population of certain territories, economic activity, etc. The main advantages of DMO are their efficiency (data are recorded on a regular basis and can be accessed with a slight delay, approximately a week, due to the need to prepare and verify data by a mobile operator) and completeness (almost all data are recorded and stored for a long time).

As with the problems of estimating the number and density of the population on the basis of DMO, their main disadvantages are: the presence of several mobile operators in the country, which are commercial organizations, have different databases, heterogeneous coverage areas and with whom it is necessary to agree on data format and content; inaccuracy in determining the positioning of mobile devices for small areas, in particular for small settlements; the need for special control and processing of data by mobile operators for the purposes of use in population estimates to exclude cases of double counting, understatement or other errors; shifting population estimates when using DMO due to the fact that one person may own several phones or SIM-cards, which leads to overestimation, while a person may not use a mobile phone at all, which will lead to underestimation for certain social groups; one or more operators may

refuse to provide mobile data, then the coverage of the population under study may be incomplete; the possibility of atypical behaviour of mobile phone owners (phone owners may be temporarily in a certain area in connection with a business trip, on vacation, in the hospital, etc.), which leads to systematic errors in assessments.

With reference to these shortcomings of the DMO as a source of data in estimating the population, their use implies a mandatory "link" to the existing estimates of population size or density based on special statistical models at the initial stage of implementation of the relevant tools. Such models should be developed or improved, first of all, based on the results of population censuses. Under this approach, the model is developed according to the census, and then the estimates are carried out with the necessary frequency according to the current DMO and the model. Under other conditions, models for estimating the population can be developed on the basis of data from relevant administrative registers, which contain up-to-date, reliable and complete data on individual populations, such as children of a certain age, persons of retirement age, etc. It is advisable to have information on the use of mobile devices by the population.

A wide range of studies demonstrate the potential of using DMO as a source of data in the following main areas: estimating the spatio-temporal distribution of the population within certain areas, given, for example, in Deville P., Linarde C., Martine S., and others (2014); estimates of the extent of movement of people within individual settlements at different times of the day or different days of the week and holidays covered in particular in the publications Järv O., Ahas R., Saluveer E., and others (2012), Ratti C., Pulselli R. M., Williams S., and Frenchman D. (2006), Reades J., Calabrese F., Ratti C. (2009); measuring the extent of labour migration (see, for example, De Meersman F., Seynaeve G., Debusschere M., and others (2016); real-time estimates of the number of people during large-scale events (such as a concert) or at the scene of an emergency covered, in particular, in the publication Lu H., Frauendorfer D., Rabbi M., and Mast M. (2012), etc.

The study Deville P., Linarde C., Martine S., and others (2014) based on datasets of MO call records from Portugal and France shows how spatially and temporarily estimations of population densities can be produced at national scales. Therefore, the parameters of the created models, which represent the ratio between mobile phone user density and population density, are defined by using the census-derived national population.

In the Handbook on the use of Mobile Phone data for Official Statistics (2017) results of some research on the use of mobile positioning data in population statistics are mentioned. In addition to the study discussed above, references are given to "The Study on Urban Mobility and Dynamic by Using Aggregate Mobile Phone

Sources" (Japan), "Overview of the Sources and Challengers of Mobile Positioning Data for Statistics" (Estonia), "Estimating Population Density Distribution from Network-Based Mobile Phone Data" (Pan-European level), "Use of Mobile Phone Data to Estimate Mobility Flows. Measuring Urban Population and Intercity Mobility Using Big Data in an Integrated Approach" (Italy), and a research on dynamic population monitoring platform based on mobile phone data and methods for estimating population numbers in China. Unfortunately, some study materials were not available through the links provided. In general, given the lack of up-to-date population census data in Ukraine, as well as data of population registers, the use of approaches presented in publications is impossible. Although the presented study unquestionably is based on the ideas underlying existing research.

The study data sources. In Ukraine, at the end of the second decade of the 21st century, the population counting situation was as follows: as previously noted, the last census was conducted only in 2001, the demographic register and other population-related registers were not introduced, although regulations on some of them were adopted. Experts and scientists, individual politicians, and international organizations have increasingly raised the question of the need to conduct a census based on traditional statistical methods or using modern technologies, in particular, by filling out census materials online.

An attempt to quickly estimate the number and location of the population of Ukraine was made using data from mobile operators and some existing data from administrative registers, which were partially covered in such publications as the Population of Ukraine. Demographic dimension of living standards: joint monograph (2019), Sarioglo V., Dubilet D., Werner I. (2020). To carry out this operational assessment of the population of Ukraine, data were used from the following main sources:

- data from a special survey of the population on the use of mobile devices, the number of devices, mobile operators, etc.;
- data of three main Ukrainian mobile operators – Kyivstar, Vodafone Ukraine, lifecell on the number of subscribers by administrative-territorial units;
- data of the Pension Fund of Ukraine on the number of persons aged 60 and older by administrative-territorial units;
- administrative data on the number of children under 14 inclusive by administrative-territorial units.

The main purpose of the survey was to estimate the use of mobile devices by the population of Ukraine, taking into account the mobile operator that provides network

services. The survey was conducted by the State Statistics in October-November 2019 on the basis of a representative (probabilistic) stratified multi-stage sample of households. About 32,000 households were selected for the survey, and the participation rate was about 84%. When developing the survey tools, it was taken into account that in order to estimate the population by the number of subscribers, it should be considered that the activity of mobile communication is different among different population groups (by age, region, type, etc.) together with the duplication of data of mobile operators in the part of persons who have several SIM-cards of different mobile operators (including several SIM-cards of one operator). Therefore, DMOs should be adjusted to accept these effects. In this context, mobile operators, as a rule, do not have their own reliable data for such an adjustment and it should be implemented based on the results of the survey.

The results of the survey were extended to the entire population of Ukraine, in view of the following basic assumptions. First, it was assumed that the vast majority of the institutional population that is not covered by state statistical sample surveys (individuals in prisons, on compulsory military service, in care houses, etc.) use mobile communications similarly to individuals, relevant sex and age groups living in private households. Second, the population (individuals living in settlements located in the territory affected by radioactive contamination as a result of the Chernobyl disaster, homeless individuals, etc.) is a small population, which is statistically insignificant.

The survey results provided an opportunity to reliably estimate the main characteristics of mobile users in Ukraine by gender and age groups, type of area – large city (with a population of 100 thousand people and more), small town, rural area – and by consolidated geographical regions. In total, five geographical regions were formed: Northern region (Kyiv, Kyiv region, Zhytomyr region, Sumy region, Chernihiv region); Central region (Cherkasy region, Poltava region, Kirovohrad region, Vinnytsia region); Eastern region (Dnipropetrovsk region, Donetsk region, Zaporizhia region, Luhansk region, Kharkiv region); Southern region (Odesa region, Mykolaiv region, Kherson region); Western region (Ivano-Frankivsk region, Khmelnytsky region, Chernivtsi region, Lviv region, Rivne region, Ternopil region, Volyn region, Zakarpattia region).

The survey found, in particular, that men and women use mobile communication approximately with the same intensity, and, accordingly, it can be assumed that the use of mobile communication by the population does not depend on gender. Mobile usage varies slightly by location and age group. As can be expected, the number of respondents using mobile communication decreases slightly in the older age groups for all three types of residence and is most noticeable for rural areas (Figure 1).
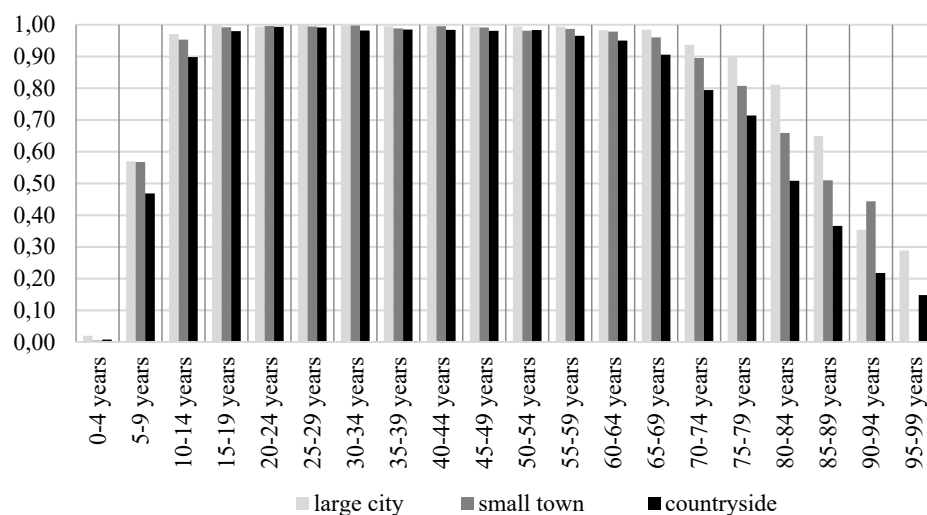
**Figure 1**. Proportion of mobile users by age group

Source: built by authors using data of the State Statistics survey.

In this regard, the use of mobile communications by type of residence differs markedly only for the population in the older age groups, in fact for people aged 65 years and older. It should be noted that for individuals aged 85 years and older, the survey results are not statistically reliable. As for children, at the age of 10–14 in rural areas the share of children who use mobile communication is about 6 pp (percentage points) smaller than in cities, and this difference increases for younger age groups.

Characterizing the peculiarities of the use of mobile communications by region, it should be noted that here the differences are observed only for older people. At the same time, the level of mobile communication use is the lowest in the western regions of Ukraine, and the highest – in its eastern regions. For example, for the age group 75–79 years the difference in the share of the population that uses mobile communication reaches 20 pp, in the western regions in this age group mobile communication is used by about 70% of the population, and in the eastern – about 90% of the population. The differences of the respective shares of the population are similar for individuals aged 80 and older. This is primarily due to the fact that the share of urban population in Ukraine is significantly higher for the eastern regions than for the western ones. Other regions of Ukraine – northern, central and southern – have an intermediate position on the use of mobile communications. It should be noted that

the results of the survey of the use of mobile communication by the population of Ukraine provided an opportunity to reliably quantify the relevant indicators.

The research has shown that the degree of coverage of mobile services in the regions for the three main operators was also assessed (Figure 2). As can be seen from the above data, Kyivstar had the largest number of subscribers in the network in 2019, which was the most significant ahead of other companies in the western, central and northern regions of Ukraine. Vodafone Ukraine had the main share of subscribers in the eastern and southern regions of Ukraine. The mobile operator lifecell had the largest share of subscribers in the northern regions, but, in general, was significantly behind the number of subscribers to two other companies. Information on the coverage of the territory of Ukraine by mobile communication services provided by various operators is useful in estimating the population at the level of regions and administrative-territorial units.
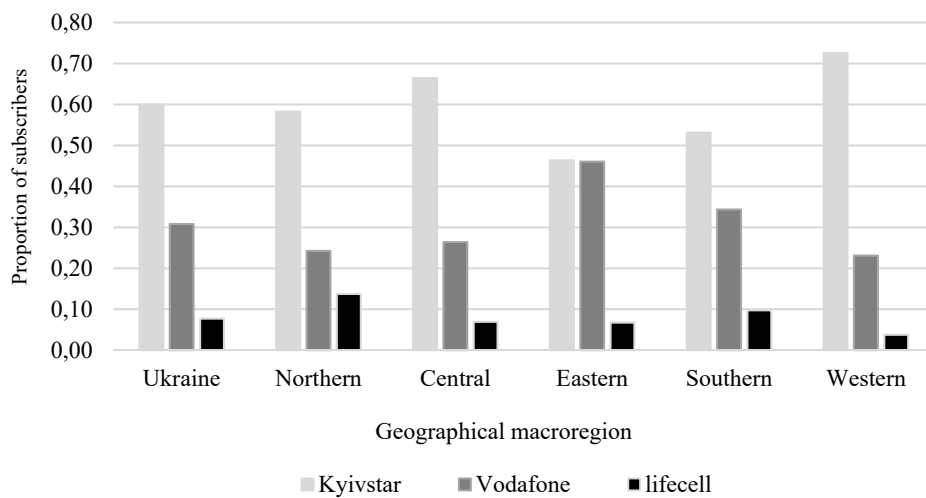
Figure 2. Proportion of subscribers of the main mobile operators of Ukraine by geographical macroregions in 2019

Source: built by authors using data of the State Statistics survey.

The information received from mobile operators had the form of a table with aggregated data by administrative-territorial units, the form of which is presented in Table 1. In the formation of the data set, operators have carried out significant work on their preparation. First of all, telephone numbers that belong to people but not to devices were set. Then the place of predominant stay of individuals (user telephone numbers) was established during a certain period and time (determination of home

location). The basic number of the subscriber was also determined provided that many subscribers can use the services of different operators, as well as have several SIM-cards of one operator. It should be noted that mobile operators in the preparation of data tried to determine the sex of subscribers, but per their attribution, such estimates were very approximate and therefore were not used in estimating the population.

Table 1. Example of aggregated data for Zhytomyr region, obtained from major mobile operators, units

Territory	Mobile operator			
	number of phone numbers (total)	number of phone numbers		
		basic phone number	nonbasic phone number	phone number type not defined
Zhytomyr Region	960533	710757	125792	123984
Zhytomyr	244265	164923	29693	49649
Berdychiv	68642	49681	8441	10520
Korosten	52018	37358	6324	8336
Malyn	19720	14620	3088	2012
Novograd-Volynsky	33805	22124	6038	5643
Andrushiv District	24854	19324	3195	2335
Andrushivka	9895	7695	1134	1066
Baraniv District	31780	24842	4256	2682
…				

Source: data of mobile operators.

Estimation of population

The results of the population survey and the data of mobile operators make it possible to make estimates of the population. For example, based on the data of three mobile operators, the total number of subscribers (numbers) for the study period was 41,916 thousand. According to the results of a sample survey, the average number of SIM-cards per mobile user was 1.2133, and the average share of the population using mobile connection was 0.8908. Thus, the estimate of the total number of mobile users is about 41916 / 1.2133 = 34547.1 thousand people. The estimate of the total population of Ukraine, made on the basis of the above data, is: 34547.1/0.8908 = 38782.1 thousand people. It should be noted that according to the State Statistics, the current population

of Ukraine as of November 1, 2019 was 41,940.7 thousand people. The calculations at the national level took into account PFC data on the number of individuals aged 60 and over and 65 years and older, and estimates of the proportion of these individuals according to demographic statistics and the results of state sample surveys of the population (households). It was taken into account that some retirees live in the temporarily occupied territories of Donetsk and Luhansk regions, the relevant data were obtained from the PFC. The calculation resulted in an average estimate of the current population of Ukraine, which amounted to 37289.4 thousand people.

The calculations of population estimates by region also took into account administrative data on the number of children under 14 and the number of individuals aged 60 and older, and estimates of the number of people aged 15-59 years obtained from mobile operators. It should be noted that the level of mobile use by the population of this age group is the highest and, accordingly, estimates of the number of subscribers are the most accurate. Regarding the estimate of the number of SIM-cards per person for this age group and the proportion of people who use mobile communications, the relevant characteristics were calculated following the survey by geographical regions, taking into account the type of place of residence. For mobile use, the corresponding estimates are 0.995 for all urban regions and 0.982 for rural areas. The number of SIM-cards for people aged 15-59 by geographical regions is presented in Table 2.

Table 2. Estimation of the number of SIM-cards per mobile user aged 15–59 by geographical macroregion and type of settlement

Geographical macroregion	Large city	Small town	Countryside
Northern	1.2781	1.2722	1.2680
Central	1.4376	1.2888	1.2057
Eastern	1.3462	1.2674	1.2855
Southern	1.3775	1.3733	1.3110
Western	1.2737	1.3016	1.2617

Source: built by authors using data of the State Statistics survey.

At the regional level as a whole, the estimated population was calculated as the sum of the number of children under 15 according to the administrative register, the number of individuals aged 60 and older as shown on the PFC records and the estimated number of individuals aged 15-59 determined by mobile operators, based on the proportion of such users among all users, the number of SIM-cards per user and the

proportion of mobile users in the entire population of the region. After the calculations for all regions, the estimates were adjusted so that the total number of available population by region corresponds to the accepted estimate of the population of Ukraine. The results of population estimates by region are presented in Table 3.

Table 3. Estimation of the current population by regions of Ukraine

Region	The number of the available population according to the state statistics dated 1.11.2019	Estimation of the available population
Ukraine	**41940726**	**37289395**
Vinnytsia Region	1547575	1331403
Volyn Region	1032178	903669
Dnipropetrovsk Region	3181481	3229959
Donetsk Region	4137246	1981154
Zhytomyr Region	1210242	1062657
Transcarpathian Region	1254327	924677
Zaporizhzhia Region	1690423	1656736
Ivano-Frankivsk Region	1368761	1125696
Kyiv Region	1778132	2286396
Kirovohrad Region	935158	826817
Luhansk Region	2138533	1127460
Lviv Region	2513820	2290052
Mykolaiv Region	1121778	1053218
Odesa Region	2376607	2347885
Poltava	1389289	1337002
Rivne Region	1153702	943623
Sumy Region	1069941	932520
Ternopil Region	1039702	763582
Kharkiv Region	2661509	2794956
Kherson Region	1029615	913743
Khmelnytsky Region	1256356	1024697
Cherkasy Region	1194584	1088149
Chernivtsi Region	901908	727676
Chernihiv Region	993758	912547
Kyiv	2964101	3703121

Source: built by authors.

The population of the regions was distributed by administrative-territorial units (cities and districts) in proportion to the number of mobile subscribers. The admissibility of such a procedure was confirmed by the very high correlation between the distributions of mobile subscribers, individuals of retirement age according to the PFC and the population according to the State Statistics by administrative-territorial units for all oblasts. At the same time, the reliability of population estimates by administrative-territorial units, especially for small settlements, was expertly assessed as insufficient for use.

It should be noted that the obtained population estimates using the data of mobile operators fully confirm the existence of more noticeable effects of urbanization of Ukraine than shown by state statistics. For example, as can be seen from Table 3, according to estimates, the current population of Kyiv is almost 25% larger than according to the State Statistics. This is also confirmed by estimates from other studies conducted over the last decade, which estimated the population of Kyiv at even 4 million people. Similar trends are typical for some other large cities in Ukraine, although the effects are not so significant.

Conclusions

The development and application of modern approaches to population estimation and distribution based on data from such sources as administrative population records, data of mobile operators, results of special sample surveys, etc., are very relevant for Ukraine, especially in conditions of significant population movements as a result of the military aggression by Russia in 2014 and 2022, large-scale processes of external and internal labour migration, as well as the lack of opportunities for traditional censuses.

The results of the research showed that the data on number of subscribers by the main mobile operators – Kyivstar, Vodafone Ukraine, lifecell – are quite suitable for use in estimating and monitoring the population of Ukraine in case of availability of reliable and detailed data on the use of mobile devices by population. The potential efficiency of the use of data from mobile operators can increase significantly if census data are available, as this would provide an opportunity to build adequate statistical models to estimate the population at different levels of administrative-territorial division and control the reliability of such estimates.

The main advantages of these mobile operators are their efficiency and completeness, which is due to the fact that the results of mobile use are recorded in real time, stored for a long time period and mobile devices are used by the vast majority of the population. At the same time, it should be borne in mind that mobile data belong to commercial organizations, and this limits their availability, requires special processing and unification, which is due in particular to the fact that mobile operators

have their own database architectures, record and store slightly different characteristics of subscribers and transactions, use different classifiers of territories, etc.

The experimental estimates of the population in Ukraine as a whole, by regions and administrative-territorial units, were significantly different from those calculated by the State Statistics on the basis of the 2001 census and administrative data on births, deaths and migration. Thus, the total population of Ukraine was estimated at about 37,289.4 thousand people, in contrast to the State Statistics estimate, which is 41,940.7 thousand people. In general, the population estimates obtained have not raised serious remarks from experts in Ukraine.

The developed approaches can be used to assess and monitor the number and location of the population of Ukraine, provided the availability and proper preparation of data of mobile operators, the availability of administrative records containing information about the population, the availability of sample surveys, in particular on the peculiar use of mobile communications by the population. At the same time, the presence of operational estimates of the population does not eliminate the need to conduct, if possible, a census of the population of Ukraine, including with the use of modern technologies for collecting information about the population.

References

Handbook on the use of Mobile Phone data for Official Statistics, (2019). UN Global Working Group on Big Data for Official Statistics, September 2019. Retrieved from:
https://unstats.un.org/bigdata/taskteams/mobilephone/MPD%20Handbook%202 0191004.pdf.

Lyudsʹkyy rozvytok v Ukrayini. Otsinka ta prohnoz rivnya zhyttya naselennya: kol. monohr/NAN Ukrayiny, In-t demohrafiyi ta sotsialʹnykh doslidzhenʹ imeni M. V., (2019). Ptukhy — Kyiv, 270 p [in Ukrainian].

Sarioglo, V., Dubilet, D., Verner, I., (2020). Otsinka chyselʹnosti nayavnoho naselennya Ukrayiny. Prezentatsiya, Kyiv,. Retrieved from:
https://buhgalter.com.ua/upload/news/2020/Оцінка_наявного_населення_Укр аїни.pdf [in Ukrainian].

Deville P., Linarde, C., Martine, S., Gilbert, M., Stevens, F. R., & Gaughan, A. E., (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences,* 111(45), pp. 15888–15893. DOI: 10.1073/pnas.1408439111.

Järv, O., Ahas, R., Saluveer, E., Derudder, B. and Witlox, F., (2012). Mobile phones in a traffic flow: A geographical perspective to evening rush hour traffic analysis using call detail records. *PLoS One,* 7 (11), e49171. Retrieved from: https://doi.org/10.1371/journal.pone.0049171.

Ratti, C, Pulselli, R. M., Williams, S. and Frenchman, D., (2006). Mobile landscapes: Using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design,* 33(5), pp. 727–748. DOI:10.1068/b32047.

Reades, J., Calabrese, F. and Ratti, C., (2009). Eigenplaces: Analysing cities using the space-time structure of the mobile phone network. *Environment and Planning B: Planning and Design*, 36(5), pp. 824–836. Retrieved from: https://doi.org/10.1068/b34133t.

De Meersman, F., Seynaeve, G., Debusschere, M., Lusyne, P., Dewitte, P., Baeyens, Y., (2016). Assessing the Quality of Mobile Phone Data as a Source of Statistics. *European Conference on Quality in Official Statistics (Q2016).* Madrid, 31 May – 3 June 2016. Retrieved from: https://ec.europa.eu/eurostat/cros/system/files/assessing_the_quality_of_mobile_ phone_data_as_a_source_of_statistics_q2016.pdf.

Lu, H., Frauendorfer, D., Rabbi, M., Mast, M., (2012). StressSense: Detecting stress in unconstrained acoustic environments using smartphones. *UbiComp'12.ACM Conference on Ubiquitous Computing,* Pittsburgh, 5–8 September, pp. 351–360. DOI:10.1145/2370216.2370270.