Cena 15,00 zł
(VAT 8%)

Indeks 381306
e-ISSN 2543-8476
PL ISSN 0043-518X

# WIADOMOŚCI STATYSTYCZNE

# THE POLISH STATISTICIAN

CZERWIEC / JUNE
ROCZNIK / VOLUME 68

2023 | 6

GŁÓWNY URZĄD STATYSTYCZNY
STATISTICS POLAND

POLSKIE TOWARZYSTWO STATYSTYCZNE
POLISH STATISTICAL ASSOCIATION

GUS

# WIADOMOŚCI STATYSTYCZNE

## THE POLISH STATISTICIAN

CZERWIEC / JUNE
ROCZNIK / VOLUME 68

2023 | 6 (745)

**ADRES REDAKCJI / EDITORIAL OFFICE ADDRESS**

Główny Urząd Statystyczny / Statistics Poland, al. Niepodległości 208, 00-925 Warszawa
tel./phone +48 22 608 32 25, e-mail: redakcja.ws@stat.gov.pl

Redakcja językowa: Wydział Czasopism Naukowych, Główny Urząd Statystyczny
Language editing: Scientific Journals Division, Statistics Poland

Redakcja techniczna, skład i łamanie, opracowanie materiałów graficznych, korekta: Zakład Wydawnictw Statystycznych – zespół pod kierunkiem Macieja Adamowicza
Technical editing, typesetting, preparation of graphic materials, proofreading: Statistical Publishing Establishment – team supervised by Maciej Adamowicz

Zakład Wydawnictw Statystycznych

Druk i oprawa / Printed and bound by:
Zakład Wydawnictw Statystycznych / Statistical Publishing Establishment
al. Niepodległości 208, 00-925 Warszawa, zws.stat.gov.pl

**Wersja elektroniczna, stanowiąca wersję pierwotną czasopisma, jest dostępna na ws.stat.gov.pl**
**The primary version of the journal, issued in electronic form, is available at ws.stat.gov.pl**



**Indeks 381306**

Informacje w sprawie sprzedaży czasopisma / Sales of the journal:
Zakład Wydawnictw Statystycznych / Statistical Publishing Establishment
tel./phone +48 22 608 32 10, +48 22 608 38 10
Prenumerata jest prowadzona przez / Subscription is available at RUCH S.A.
Zamówienia na prenumeratę można składać na stronie / Subscriptions can be ordered at
www.prenumerata.ruch.com.pl

Zam. 186/2023 – nakład 305 egz.

SPIS TREŚCI
CONTENTS

Wiadomości Statystyczne. The Polish Statistician, 2023, *68*(6), IV–V
Od redakcji / From the editorial team

# OD REDAKCJI

W czerwcowym wydaniu „Wiadomości Statystycznych. The Polish Statistician" proponujemy Państwu lekturę artykułów z zakresu metodologii badań statystycznych, analiz społeczno-ekonomicznych i prowadzenia spisów powszechnych.

Praca dr. Piotra Sulewskiego i dr. hab. Marcina Szymkowiaka, prof. UEP, pt. *Modelling income distributions based on theoretical distributions derived from normal distributions* dotyczy modelowania rozkładu dochodów z wykorzystaniem różnych typów rozkładów teoretycznych. Autorzy sprawdzają możliwości posługiwania się innymi typami rozkładów niż powszechnie stosowane, a mianowicie rozkładami asymetrycznymi wywodzącymi się z rozkładu normalnego. Przeprowadzają badanie, w którym określają charakterystyki rozkładów na podstawie danych o miesięcznym dochodzie brutto na mieszkańca w Polsce pochodzących z badania EU-SILC 2011. Dowodzą, że asymetryczny rozkład SU Johnsona może być – podobnie jak rozkład Daguma – z powodzeniem wykorzystywany do modelowania rozkładu dochodów.

Dr Olga Komorowska i dr Arkadiusz Kozłowski, autorzy artykułu *Impact of a child's disability on the probability of the mother taking up paid employment*, badają wpływ niepełnosprawności dziecka na prawdopodobieństwo podjęcia pracy zawodowej przez matkę. W analizie empirycznej wykorzystują metodę dekompozycji wywodzącą się z podejścia Blindera i Oaxaki, a także regresję logistyczną. Posługują się jednostkowymi danymi z lat 2005–2020 dotyczącymi gospodarstw domowych, zaczerpniętymi z reprezentacyjnego badania budżetów gospodarstw domowych. Z analizy wynika, że niższy wskaźnik zatrudnienia wśród matek dzieci z niepełnosprawnościami jest spowodowany przede wszystkim niepełnosprawnością dziecka, a także innymi czynnikami będącymi tego skutkiem. Zmienną, która ma największy wpływ na aktywność zawodową matek, jest ich wykształcenie.

Edgar Vielma Orozco, MSc, w pracy pt. *Use of new technologies and evidence-based decisions: key factors in the strategy for the 2020 Population and Housing Census in Mexico, in the context of the COVID-19 pandemic* przedstawia doświadczenia meksykańskiego Narodowego Instytutu Statystyki i Geografii (Instituto Nacional de Estadística y Geografía – INEGI) z realizacji Powszechnego Spisu Ludności i Mieszkań 2020, przeprowadzonego w warunkach pandemii, oraz wdrożonych strategii, z których mogą skorzystać urzędy statystyczne innych krajów. Autor podkreśla znaczenie ciągłości opracowywania statystyk w okresach kryzysu, konieczność posiadania systemu skutecznego zarządzania ryzykiem przez krajowe organy statystyczne oraz potrzebę wdrożenia innowacyjnych metod gromadzenia danych i szerszego wykorzystania technologii informacyjno-komunikacyjnych.

Ponadto zachęcamy do zapoznania się z zestawieniem nowości wydawniczych GUS przygotowanym przez Joannę Sadowy.

Zapraszamy do lektury.

# FROM THE EDITORIAL TEAM

The June issue of *Wiadomości Statystyczne. The Polish Statistician* features papers pertaining to the methodology of statistical research, socio-economic analyses and conducting censuses.

In their paper, Piotr Sulewski, PhD, and Marcin Szymkowiak, PhD, DSc, professor at the Poznań University of Economics and Business, discuss the issue of *Modelling income distributions based on theoretical distributions derived from normal distributions*. The authors examine the possibility of utilising distributions other than those commonly used, namely asymmetric distributions derived from normal distributions. They conduct a research determining the characteristics of the distributions based on data from the 2011 EU-SILC survey on gross monthly income *per capita* in Poland. They find that, as in the case of the Dagum distribution, the SU Johnson asymmetric distribution can be successfully used to model income distribution.

Olga Komorowska, PhD, and Arkadiusz Kozłowski, PhD, are authors of the article entitled *Impact of a child's disability on the probability of the mother taking up paid employment*. In their empirical analysis, they use a decomposition method derived from the Blinder and Oaxaca approach, and logistic regression. The research is based on individual household data from the Representative Household Budget Survey, covering the years 2005–2020. The analysis shows that the lower employment rate observed among mothers of children with disabilities is primarily caused by the child's disability itself, as well as by other factors resulting from it. The variable that has the greatest impact on the mothers' economic activity is their education.

The article entitled *Use of new technologies and evidence-based decisions: key factors in the strategy for the 2020 Population and Housing Census in Mexico, in the context of the COVID-19 pandemic* by Edgar Vielma Orozco, MSc, describes the experiences of Mexico's National Institute of Statistics and Geography (Instituto Nacional de Estadística y Geografía – INEGI) relating to the 2020 Population and Housing Census conducted under pandemic conditions, and the adopted strategies which may prove useful to statistical offices in other countries. The author emphasises the importance of maintaining the production of statistics during crisis periods and calls for national statistical authorities to have an effective risk management system. The author also opts for the implementation of innovative data collection methods and making greater use of information and communication technologies.

We would also like to recommend the latest publications of Statistics Poland whose compilation was prepared by Joanna Sadowy.

We wish you pleasant reading.

Wiadomości Statystyczne. The Polish Statistician, 2023, *68*(6), 1–23 https://doi.org/10.59139/ws.2023.06.1
Studia metodologiczne / Methodological studies
Zgłoszony/submitted: 12.10.2022, poprawiony/revised: 21.04.2023, zaakceptowany/accepted: 21.05.2023

# Modelling income distributions based on theoretical distributions derived from normal distributions

Piotr Sulewski,[a] Marcin Szymkowiak[b]

**Abstract.** In income modelling studies, such well-known distributions as the Dagum, the lognormal or the Zenga distributions are often used as approximations of the observed distributions. The objective of the research described in the article is to verify the possibility of using other type of distributions, i.e. asymmetric distributions derived from normal distribution (ND) in the context of income modelling. Data from the 2011 EU-SILC survey on the monthly gross income *per capita* in Poland were used to assess the most important characteristics of the discussed distributions. The probability distributions were divided into two groups: I – distributions commonly used for income modelling (e.g. the Dagum distribution) and II – distributions derived from ND (e.g. the SU Johnson distribution). In addition to the visual evaluation of the usefulness of the analysed probability distributions, various numerical criteria were applied: information criteria for econometric models (such as the Akaike Information Criterion, Schwarz's Bayesian Information Criterion and the Hannan-Quinn Information Criterion), measures of agreement, as well as empirical and theoretical characteristics, including a measure based on quantiles, specifically defined by the authors for the purposes of this article. The research found that the SU Johnson distribution (Group II), similarly to the Dagum distribution (Group I), can be successfully used for income modelling.

**Keywords:** income modelling, EU-SILC, normal distribution, SU Johnson distribution, Dagum distribution

**JEL:** C13, C15, C55, D31

# Modelowanie rozkładu dochodów z wykorzystaniem rozkładów teoretycznych wywodzących się z rozkładu normalnego

**Streszczenie.** W badaniach nad modelowaniem dochodów do aproksymacji ich rozkładów bardzo często wykorzystuje się takie znane rozkłady, jak Daguma, log-normalny czy Zengi. Celem badania omawianego w artykule jest sprawdzenie możliwości posłużenia się innymi

[a] Akademia Pomorska w Słupsku, Instytut Nauk Ścisłych i Technicznych, Polska / Pomeranian University in Słupsk, Institute of Exact and Technical Sciences, Poland. ORCID: https://orcid.org/0000-0002-0788-6567. Autor korespondencyjny / Corresponding author, e-mail: piotr.sulewski@apsl.edu.pl.

[b] Uniwersytet Ekonomiczny w Poznaniu, Instytut Informatyki i Ekonomii Ilościowej; Urząd Statystyczny w Poznaniu, Polska / Poznań University of Economics and Business, Institute of Informatics and Quantitative Economics; Statistical Office in Poznań, Poland.
ORCID: https://orcid.org/0000-0003-3432-4364. E-mail: marcin.szymkowiak@ue.poznan.pl.

typami rozkładów, tj. rozkładami asymetrycznymi wywodzącymi się z rozkładu normalnego (ND), w kontekście modelowania dochodów. Najważniejsze charakterystyki rozpatrywanych rozkładów określono na podstawie danych z badania EU-SILC 2011 dotyczących miesięcznego dochodu brutto na mieszkańca w Polsce. Rozkłady prawdopodobieństwa podzielono na dwie grupy: I – rozkłady powszechnie stosowane do modelowania dochodów (np. rozkład Daguma) i II – rozkłady wywodzące się z ND (np. rozkład SU Johnsona). Oprócz wizualnej oceny przydatności analizowanych rozkładów prawdopodobieństwa zastosowano kryteria liczbowe, takie jak: kryteria informacyjne dla modeli ekonometrycznych (Akaike Information Criterion, Schwarz's Bayesian Information Criterion oraz Hannan-Quinn Information Criterion), miary zgodności oraz charakterystyki empiryczne i teoretyczne, w tym specjalnie zdefiniowana na potrzeby artykułu autorska miara wykorzystująca kwantyle. Jak wynika z badania, rozkład SU Johnsona (II grupa), może być – tak jak rozkład Daguma (I grupa) – z powodzeniem wykorzystany do modelowania dochodów.

**Słowa kluczowe:** modelowanie dochodów, EU-SILC, rozkład normalny, rozkład SU Johnsona, rozkład Daguma

## 1. Introduction

Finding a proper model of income distribution and, consequently, examining and explaining income inequality, has been a task undertaken by economists since the times of Pareto. One of the main directions of research in the area of income distribution is the search for a theoretical model describing empirical income distributions.

Firstly, a theoretical model simplifies the analysis. When a small number of parameters is used, the estimation of different characteristics of the distribution and studying the properties of these characteristics expressed as the functions of certain parameters of the theoretical distribution become possible. Secondly, a well-fitted theoretical model allows the prediction of income distributions across different domains, both in time and space. It can be used e.g. in small area estimation as part of the model-based approach (Pratesi, 2016). Thirdly, approximations of empirical income distributions based on appropriately chosen theoretical distributions can compensate for irregularities resulting from the data collection method.

Many authors who study income distribution propose a set of economic, econometric, stochastic and mathematical properties considered as criteria used to select a particular mathematical model of income distribution. The final choice of the model depends on the degree to which it is capable of satisfying these criteria (Jędrzejczak & Trzcińska, 2018).

Aitchison and Brown (1957) defined the fundamental properties that form the most representative model of any stochastic process that generates an income distribution. The same issue was analysed by Dagum (1977) and Metcalf (1972). The suggested properties characterise a desirable income distribution model, including its foundation, interpretation, flexibility and inferential properties (Jędrzejczak, 2006). Among the most important features is the convergence to the Pareto principle for high income groups, a small number of finite moments of a distribution (heavy tails), goodness-of-fit (GoF) for the whole range of a distribution, a simple interpretation of parameters, and simplicity (or a small number of parameters).

Normal distribution (ND) is certainly the best known representative of the family of distributions used in statistical theory and practice to model distributions of phenomena defined as positive or non-negative, characterised as symmetric (e.g. people's weight, height, etc.). Obviously, ND is not used to model income distributions, which are asymmetric, i.e. most values are clustered around the left tail, whereas the right tail is considerably longer (a positively skewed distribution).

The objective of the study is to verify the possibility of using asymmetric distributions derived from ND in the context of income distribution modelling.

## 2. Literature review

As mentioned in the Introduction, there are many probability distributions describing non-negative or positive variables that can be used to approximate income distributions. This family includes the following distributions: lognormal (LOG), defined by Gaddum (1945), Birnbaum-Saunders (BS), defined by Birnbaum and Saunders (1969), Dagum (DA), defined by Dagum (1977), beta prime (BPr) and inverse gamma (IG), defined by Johnson et al. (1995), Singh-Maddala (SM), defined by Singh and Maddala (1976), Zenga (Z), defind by Zenga (2010), Pareto type IV (PIV), defined by Pareto (1895), generalised gamma (GG), defined by Stacy (1962) and generalised beta of the second kind (GB2), defined by McDonald (1984). For more information on the PIV, GG and GB2 distribution families, see the Appendix.

Distributions such as Pareto, lognormal, gamma or Dagum were used to approximate income distributions in the Polish population (see e.g. Kordos, 1968, 1973; Kot, 1999, 2000; Lange, 1967; Vielrose, 1960; Wiśniewski, 1934). Research on income and wage distribution in Poland has confirmed that the Dagum, Singh-Maddala and Zenga distributions are particularly well-fitted to empirical data (see e.g. Brzeziński, 2013; Jędrzejczak, 1993, 2006; Jędrzejczak & Trzcińska, 2018; Łukasiewicz & Orłowski, 2004; Ostasiewicz, 2013; Salamaga, 2016; Trzcińska, 2020, 2022; Wałęga & Wałęga, 2021).

To achieve the aim of this article, a competitive family of distributions based on ND is needed. For this purpose, ND can be 'plasticised' by adding a 'plasticising' parameter to the cumulative distribution function (CDF) of the ND, by 'plasticising' the formula located in the exponential function or through a combination of distributions.

The first way was introduced by Azzalini (1985), who added a skewness parameter to the CDF of ND and initiated a very interesting family of ND-derived distributions. This family includes distributions such as: skew-normal (SN), defined by Azzalini (1985), skew-generalised normal (SGN), defined by Arellano-Valle et al. (2004), flexible skew-normal (FSN), defined by Ma and Genton (2004), two-piece skew-normal (TPSN), defined by Kim (2005), power-normal (PN), defined by Gupta and Gupta (2008), generalised Balakrishnan skew-normal (GBSN), defined by Yadegari et al. (2008), Balakrishnan skew-normal (BSN), defined by Sharafi and Behboodian (2008), extended skew-generalised normal (ESGN1), defined by Choudhury and Abdul (2011), extended skew-generalised normal (ESGN2), defined by Venegas et al. (2011), skew-flexible normal (SFN), defined by Gómez et al. (2011), Kumaraswamy-normal (KN), defined by Cordeiro and de Castro (2011), flexible skew-generalised normal (FSGN1), defined by Nekoukhou et al. (2013), extended skew-generalised normal (ESGN3), defined by Kumar and Anusree (2015), flexible skew-generalised normal (FSGN2), defined by Bahrami and Qasemi (2015) and shape skew-generalised normal (SSGN), defined by Rasekhi et al. (2017).

The second approach is to 'plasticise' the formula in the exponential function of the ND. A family derived in this way includes the following distributions defined in a real domain: symmetrical sinh-normal (S-N), defined by Rieck and Nedelman (1991), SU defined by Johnson (1949), SC and expnormal (EN), defined by Sulewski (in press).

The third way involves a combination of at least two normal distributions which can fit more characteristics that the sample data might contain. Behboodian (1970) describes the conditions under which a combination of two normal distributions, called the compound normal (CN) distribution, is unimodal.

## 3. Income modelling – theoretical probability distributions

In the study, theoretical probability distributions used for income modelling are divided into two groups: Group I – consisting of distributions strictly dedicated to income modelling and Group II – mainly including ND-derived distributions, with ND being their special case. This group also consists of three distributions from the Johnson family, namely: SU, SC and EN. Some parameter values of these distributions are very similar to ND, but ND is not a special case of these distributions. It is possible to calculate the measure of the similarity of ND with the $\varphi(x;\mu,\sigma)$ PDF (probability density function) to the distribution with the $g(x;\boldsymbol{\Theta})$ PDF, where $\boldsymbol{\Theta}$ is the vector of parameters. This similarity measure ($S_M$) is provided by Sulewski (2019):

$$S_M(\mu,\sigma,\boldsymbol{\Theta}) = \int_{-\infty}^{\infty} \min[\varphi(x;\mu,\sigma), \quad g(x;\boldsymbol{\Theta})]dx. \tag{1}$$

As mentioned above, the main objective of the study is to verify the possibility of using asymmetric distributions derived from ND in the context of income distribution modelling (Group II) and to compare it with the properties of well-known distributions strictly intended for this purpose (Group I).

When analysing the structure and properties of distributions, one can group them into systems. The systems for income distributions include the Dagum, Pearson, D'Addairo, Burr and Johnson distribution. However, these systems are not always separate. For example, the lognormal distribution belongs to the D'Addairo and Johnson systems, while the gamma distribution to the Pearson and D'Addairo systems, the Dagum distribution belongs to the Dagum, Pearson and Burr systems and the Pareto distribution to the Dagum and D'Addairo systems;, whereas the Singh-Maddala distribution to the Dagum and Burr systems (Jędrzejczak & Pekasiewicz, 2020).

There is no need to select all distribution systems for a Monte Carlo simulation to assess their properties; only certain members need to be chosen. The use of distributions which happen to be special cases of other distributions is also convenient. PIV, GG and GB2 are examples of such distribution groups, with the exception of SM, which is a special case of GB2 and, as mentioned earlier, recommended for income modelling.

Group I distributions includes the LOG as a member of the D'Addairo and Johnson systems, BS as a member of the Johnson system, BPr and IG as members of the Pearson system, DA as a member of the Dagum system, SM as a member of the Dagum and Burr systems, Z as the youngest distribution in this group and the PIV, GG and GB2 families.

Group I distributions intentionally includes not only the most popular distributions for income modelling (e.g. Dagum, Singh-Maddala, Zenga), but also distributions never or very rarely used for income modelling (e.g. BS, BPr, IG).

Distributions derived from ND (Group II) include SN, SGN, ESGN1, ESGN2, ESGN3, SSGN, PN, SFN, FSN, FSGN1, FSGN2, KN, BSN, and GBSN. The Johnson system distributions, such as SU, SC and EN also belong to ND-derived distributions, because their measure of similarity to ND calculated with (1), as shown in the Appendix, exceeds 96%.

Let $\varphi(x)$ and $\Phi(x)$ be PDF and CDF of $N(0,\ 1)$, respectively. Let $\mu \in R$ be the position parameter, $\sigma > 0$ the scale parameter and $\alpha, \beta, \gamma$ the shape parameters. Probability distributions are described by PDF and CDF. The appropriate formulas for the probability distributions (either PDF or CDF were presented depending on the distribution and their simplicity) included in the empirical part of the article are presented in the Appendix.

Group II distributions, after eliminating distributions that are special cases of other ones, includes: ESGN1, ESGN2, ESGN3, SSGN, PN, FSGN1, FSGN2, KN, GBSN, TPSN, SU, and EN.

In summary, 22 distributions (10 from Group I and 12 from Group II) were used in the empirical study to model income.

## 4. Goodness-of-fit measures

In order to select the best theoretical models described in Section 3 and in the Appendix for income modelling, the authors assessed them in a two-step procedure. In the first step, the GoF measures were applied to assess the consistency of the estimated models with the empirical data from the EU-SILC. In the second step, the very popular theoretical characteristics of models selected in the first step were compared with the empirical characteristics computed for the EU-SILC data (see Section 5 for more details and Table 5 for comparison).

It is a well-known problem in studies of income distribution that in the case of large samples, GoF statistical tests lead to the rejection of the null hypothesis, even if the studied model describes the empirical distributions very well (Kunte & Gore, 1992). Therefore, to assess the properties of the distributions used for income modelling, the authors decided not to use statistical tests, but rather to apply other numerical measures: the information criteria, GoF measures, as well as empirical and theoretical characteristics, including a specifically defined measure using quantiles.

Let $M(\boldsymbol{\Theta})$ be a model with a $\boldsymbol{\Theta}$ vector of parameters used for describing the distribution of income. Let $f_M(x;\boldsymbol{\Theta})$ and $F_M(x;\boldsymbol{\Theta})$ be the respective PDF and CDF of this model. Let $x^*_{(1)},\ x^*_{(2)},\dots,x^*_{(n)}$ be a sample of size $n$. Our goal is to estimate the unknown values of parameters $\boldsymbol{\Theta}$ by using the maximum likelihood estimation (MLE) method. The likelihood function is given by

$$L(\boldsymbol{\Theta}) = \prod_{i=1}^{n} f_M(x^*_i;\boldsymbol{\Theta}) \tag{2}$$

Then the log-likelihood function is defined as

$$l(\boldsymbol{\Theta}) = \ln\big(L(\boldsymbol{\Theta})\big) = \sum_{i=1}^{n} \ln[f_M(x^*_i;\boldsymbol{\Theta})]. \tag{3}$$

Formulas of derivatives $dl/d\boldsymbol{\Theta}$ have complex forms. In practice, it is not necessary to calculate them. It is better to maximise the log-likelihood function using mathematical software instead of struggling with a system of complicated non-linear equations that may have extraneous roots.

To avoid any local maxima of the log-likelihood function, the optimisation routine is run repeatedly each time from different starting values that are widely scattered in the parameter space. The maximum likelihood estimates of parameters $\boldsymbol{\Theta}$ can be easily calculated, e.g. in the R software (R Core Team, 2021) using the *fitdistr* function (package MASS), or in Mathcad. Information criteria, such as the Akaike Information Criterion ($AIC$), the Bayesian Information Criterion ($BIC$) and the Hannan-Quinn Information Criterion ($HQIC$) are used for the model comparisons. Let us recall that:

$$AIC = -2l + 2p, \quad BIC = -2l + p\ln(n), \quad HQIC = -2l + 2p\ln(\ln(n)), \quad (4)$$

where $l$ is the log-likelihood function (3), $n$ is the sample size and $p$ is the number of model parameters.

Let $k$ be the number of intervals in which $n$ individual values are grouped. Let $n_i$ and $\hat{n}_i$ $(i = 1, 2, \ldots, k)$ be the observed and the estimated counts of the $i$-th interval, respectively, then $w_i = n_i/n$ and $\hat{w}_i = \hat{n}_i/n$ $(i = 1, 2, \ldots, k)$ represent empirical and theoretical frequencies. Estimated counts $\hat{n}_i$ are given by

$$\hat{n}_i = \begin{cases} nF_M(x_i; \boldsymbol{\Theta}) & i = 1 \\ n[F_M(x_i; \boldsymbol{\Theta}) - F_M(x_{i-1}; \boldsymbol{\Theta})] & i = 2, 3, \ldots, k, \end{cases} \quad (5)$$

where $x_i$ $(i = 1, 2, \ldots, k - 1)$, $x_k = \infty$ are the upper bounds of the $k$ intervals.

The first GoF measure, proposed by Egon Vielrose, a Polish economist, demographer and statistician, is calculated using the following simple formula (Vielrose, 1960):

$$W_p = \sum_{i=1}^{k} \min(w_i, \hat{w}_i). \quad (6)$$

The higher the value of $W_p$, the better the consistency of the compared distributions.

The second and the third GoF measures are Mortara's $A_1$ index and quadratic Pearson's $A_2$ index defined as (Zenga et al., 2012)

$$A_1 = \frac{1}{n}\sum_{i=1}^{k}|n_i - \hat{n}_i|, A_2 = \sqrt{\frac{1}{n}\sum_{i=1}^{k}\frac{(n_i - \hat{n}_i)^2}{\hat{n}_i}}, \tag{7}$$

respectively.

The smaller the value of $A_i$ $(i = 1, 2)$, the better the consistency of the compared distributions.

The fourth GoF measure takes into account a coefficient based on the relative difference between the mean value of empirical distribution $\bar{X}$ and the expected value of theoretical distribution $E(X)$ (Jędrzejczak & Pekasiewicz, 2020):

$$A_3 = \frac{|\bar{X} - E(X)|}{E(X)} \cdot 100\%. \tag{8}$$

The smaller the $A_3$ value, the better the consistency of the compared distributions.

The fifth GoF measure, relative error $RE$, calculated by comparing the theoretical and empirical characteristics, $TCH$ and $ECH$ is expressed as follows:

$$RE = \frac{|TCH - ECH|}{TCH} \cdot 100\%. \tag{9}$$

The smaller the $RE$ value, the better the consistency of the compared distributions.

A number of characteristics can be selected: the mean $(A)$, the lower quartile $(Q_1)$, the median $(Q_2)$, the upper quartile $(Q_3)$, the standard deviation $(SD)$ or the coefficient of variation $(CoV)$. A model is considered to be well-fitted if the differences between empirical and theoretical characteristics are less than 5%.

Let $x_p^*$ and $x_p$ $(0 < p < 1)$ be empirical and theoretical $p$-th quantiles, respectively. The last GoF measure, specifically defined for the purpose of this study, is the $QM$ given by

$$QM = \sum_{i=1}^{19}|x_{0.05i} - x_{0.05i}^*|. \tag{10}$$

The smaller the $QM$ value, the better the consistency of the compared distributions.

The GoF measures listed above can be divided into three classes. The first class includes information criteria (IC) (4), the second one the GoF coefficients (GOFC) (6)–(8), and the third one the characteristics (CH) represented by classic measures from $A$ to $CoV$, plus the new measure proposed by the authors (10).

## 5. Data

The empirical analysis is based on data for gross monthly income *per capita* in Poland from the 2011 edition of the EU Statistics on Income and Living Conditions (EU-SILC) survey. EU-SILC is a non-obligatory, representative questionnaire of individual households, where the data are collected in face-to-face interviews. The main objective of the survey is to supply the European Union with comparable data on the living conditions of the population.

EU-SILC is the basic source of information used for the calculation of indicators, including those related to income, poverty and social exclusion for the EU member states. Data from the survey are used to produce various income statistics, such as the average yearly equivalised disposable income *per capita*, selected measures of the diversification of the average disposable income (e.g. the Gini coefficient, S80/S20), at-risk-of-poverty thresholds, the relative at-risk-of-poverty rate or selected measures of average disposable income distribution in the Polish regions (Główny Urząd Statystyczny [GUS], 2021).

## 6. Results

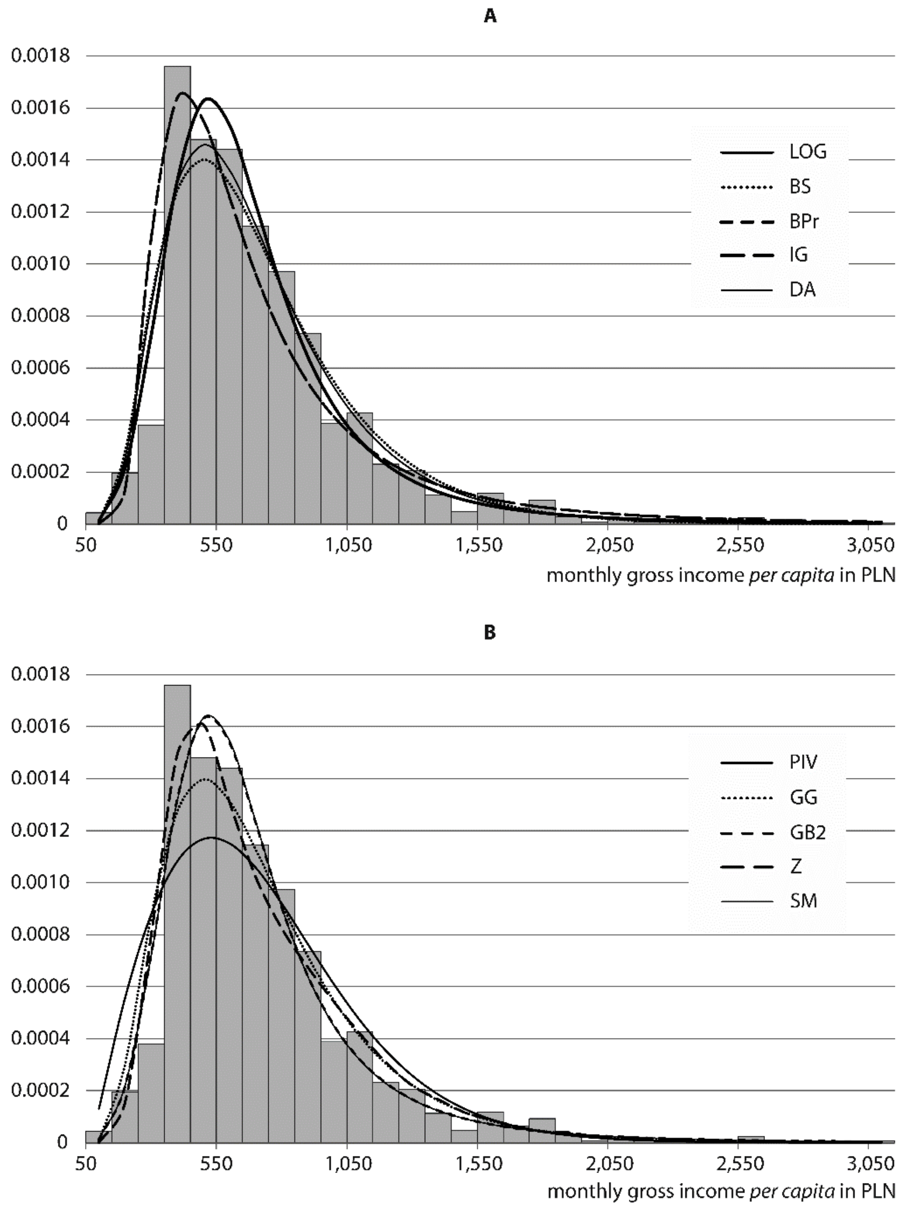
Figures 1 and 2 show histograms and estimated PDFs for the analysed models from Group I and Group II, respectively. In order to improve their legibility, only income values below PLN 3,000 are displayed.

Estimated PDFs for the BPr and IG models are almost identical (see Figure 1A). The situation is similar for the DA, SM and GB2 models (see Figure 1B). The above is also confirmed by the results presented in Tables 1 and 3.

The estimated PDFs for the ESGN1 and SSGN models are almost identical (see Figure 2B), which is also confirmed by the results provided in Tables 2 and 4.

When the estimated PDFs are very similar in shape, additional numerical measures are necessary. The first group of numerical measures consists of the values of information criteria $AIC$, $BIC$ and $HQIC$ (4). Tables 1 and 2 display values of the MLEs and the information criteria for the Group I and Group II of models, respectively. These models are sorted by $AIC$ values (the first three models in appropriate tables are in bold).

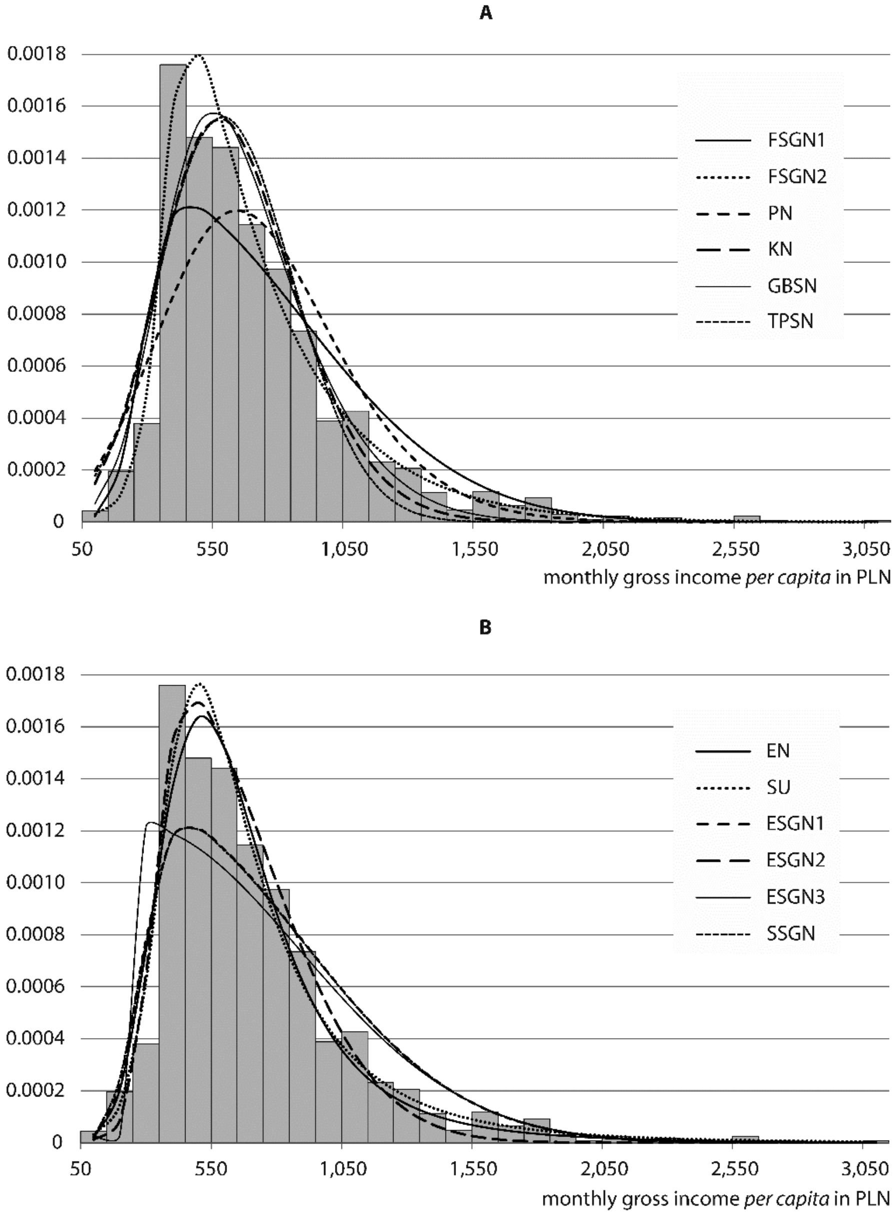
**Figure 1.** Histograms and probability density functions of distributions from Group I




Note. Distributions: LOG – lognormal, BS – Birnbaum-Saunders, BPr – beta prime, IG – inverse gamma, DA – Dagum, PIV – Pareto type IV, GG – generalised gamma, GB2 – generalised beta of the second kind, Z – Zenga, SM – Singh-Maddala.

Source: authors' work based on EU-SILC data.

**Figure 2.** Histograms and probability density functions of distributions from Group II




Note. Distributions: FSGN – flexible skew generalised normal, PN – power-normal, KN – Kumaraswamy-normal, GBSN – generalised Balakrishnan skew normal, TPSN – two-piece skew-normal, EN – expnormal, SU – SU Johnsona, ESGN – extended skew generalised normal, SSGN – shape skew generalised normal.

Source: authors' work based on EU-SILC data.

As the figures above indicate, the DA, SM and GB2 models stand out in the Group I distribution. The SU, FSGN2 and EN stand out in the Group II distribution. According to the information criteria, SU and FSGN2 (which are ND-derived distributions) produce better models of income distribution than DA, SM and GB2, which are typically used for this purpose. The $AIC$ ranking in both distribution groups is the same as the $BIC$ and $HQIC$ rankings.

**Table 1.** Values of maximum likelihood estimation and information criteria – Group I distributions

| Model | Estimates of $\boldsymbol{\Theta}$ | $AIC$ | $BIC$ | $HQIC$ |
|---|---|---|---|---|
| **DA** | $\hat{\sigma} = 550.3921,\ \hat{\alpha} = 3.2984,\ \hat{\beta} = 1.0904$ | 143,925.1 | 143,946.8 | 143,932.4 |
| **SM** | $\hat{\sigma} = 549.6120,\ \hat{\alpha} = 3.5085,\ \hat{\beta} = 0.9129$ | 143,925.4 | 143,947.1 | 143,932.8 |
| **GB2** | $\hat{\sigma} = 3.3767,\ \hat{\alpha} = 549.0722,\ \hat{\beta} = 1.0574,\ \hat{\gamma} = 0.9639$ | 143,926.9 | 143,955.8 | 143,936.7 |
| LOG | $\hat{b} = 0.4787,\ \hat{c} = 6.4631,\ \hat{a} = -58.4635$ | 144,226.5 | 144,248.1 | 144,233.8 |
| Z | $\hat{\beta} = 665.4616,\ \hat{\alpha} = 6.1377,\ \hat{\gamma} = 7.6694$ | 144,305.7 | 144,327.4 | 144,313.1 |
| GG | $\hat{\sigma} = 1.9195,\ \hat{\alpha} = 0.4740,\ \hat{\beta} = 15.4100$ | 144,465.5 | 144,487.2 | 144,472.8 |
| BS | $\hat{\alpha} = 0.4757,\ \hat{\sigma} = 674.6981,\ \hat{\mu} = -85.2903$ | 144,508.8 | 144,530.5 | 144,516.1 |
| PIV | $\hat{\mu} = 24.7489,\ \hat{\sigma} = 2{,}147.9490,\ \hat{\gamma} = 0.5191,\ \hat{\alpha} = 8.9888$ | 145,390.3 | 145,419.2 | 145,400.1 |
| BPr | $\hat{\alpha} = 2{,}219.1921,\ \hat{\beta} = 3.3177,\ \hat{\sigma} = 0.7331$ | 145,816.2 | 145,837.9 | 145,823.6 |
| IG | $\hat{\alpha} = 3.3021,\ \hat{\sigma} = 1{,}617.5991$ | 145,832.9 | 145,847.4 | 145,837.8 |

Note. As in Figure 1.
Source: authors' work based on EU-SILC data.

**Table 2.** Values of maximum likelihood estimation and information criteria – Group II distributions

| Model | Estimates of $\boldsymbol{\Theta}$ | $AIC$ | $BIC$ | $HQIC$ |
|---|---|---|---|---|
| **SU** | $\hat{c} = -1.5371,\ \hat{\delta} = 1.2759,\ \hat{a} = 275.6729,\ \hat{b} = 188.6396$ | 143,805.7 | 143,834.6 | 143,815.5 |
| **FSGN2** | $\hat{\mu} = 329.8192,\ \hat{\sigma} = 2{,}744.0519,\ \hat{\alpha} = 27.6787,\ \hat{\beta} = 97.1796,\ \hat{\gamma} = 7.7981$ | 143,825.0 | 143,861.1 | 143,837.2 |
| **EN** | $\hat{a}_1 = 1{,}249,\ \hat{b}_1 = 701.716,\ \hat{a}_2 = 10.747,\ \hat{b}_2 = -1.187,\ \hat{c} = 2.582$ | 144,159.0 | 144,195.1 | 144,171.2 |
| SSGN | $\hat{\mu} = 226.1337,\ \hat{\sigma} = 618.5770,\ \hat{\alpha} = 14.9750,\ \hat{\beta} = 2.1659,\ \hat{\gamma} = -0.1215$ | 146,141.2 | 146,177.3 | 146,153.4 |
| FSGN1 | $\hat{\mu} = 227.8910,\ \hat{\sigma} = 617.1702,\ \hat{\alpha} = 7.3362,\ \hat{\beta} = 1.8201,\ \hat{\gamma} = 6.2170$ | 146,142.6 | 146,178.7 | 146,154.8 |
| ESGN1 | $\hat{\mu} = 228.2867,\ \hat{\sigma} = 617.0445,\ \hat{\alpha} = 7.3236,\ \hat{\beta} = 0.0312,\ \hat{\gamma} = 0.0028$ | 146,142.6 | 146,178.7 | 146,154.9 |
| GBSN | $\hat{\mu} = 97.382,\ \hat{\sigma} = 480.465,\ \hat{\alpha} = 1.954,\ \hat{\beta} = 4.315,\ \hat{\gamma} = 0.015$ | 147,061.6 | 147,097.7 | 147,073.9 |
| ESGN3 | $\hat{\mu} = 200.929,\ \hat{\sigma} = 639.164,\ \hat{\alpha} = 98.292,\ \hat{\beta} = 2{,}468.553,\ \hat{\gamma} = -286.549$ | 147,164.7 | 147,200.8 | 147,176.9 |
| PN | $\hat{\mu} = -278.3487,\ \hat{\sigma} = 594.8178,\ \hat{\alpha} = 11.2871$ | 147,842.2 | 147,863.9 | 147,849.5 |
| KN | $\hat{\mu} = 109.364,\ \hat{\sigma} = 339.288,\ \hat{\alpha} = 4.928,\ \hat{\beta} = 0.703$ | 148,116.5 | 148,114.1 | 148,111.1 |
| ESGN2 | $\hat{\mu} = 278.806,\ \hat{\sigma} = 426.31,\ \hat{\alpha} = 5.819,\ \hat{\beta} = 1.076,\ \hat{\gamma} = 3.825$ | 149,350.3 | 149,386.4 | 149,362.6 |
| TPSN | $\hat{\mu} = 535.082,\ \hat{\sigma} = 270.331,\ \hat{\alpha} = 1.633,\ \hat{\beta} = 1.687$ | 153,299.6 | 153,328.5 | 153,309.4 |

Note. As in Figure 2.
Source: authors' work based on EU-SILC data.

Tables 3 and 4 present the values of the $W_p$ and $A_i$ $(i = 1,\ 2,\ 3)$ measures described in Section 4, with rankings for Group I and II distributions, respectively. The analysed models are sorted according to the values of final ranking $R$, which is based on the sum of sub-rankings $R()$.

As can be seen in Tables 3 and 4, models based on DA, GB2 and Z stand out in the Group I distribution. The SU, FSGN2 and EN stand out in the Group II distribution. According to the GoF measures, the best (highest) values of $W_p$ were obtained for the FSGN2, LOG and SU models. The best (lowest) values of $A_1$ are achieved for the LOG, FSGN2 and SU models. The best (lowest) values of $A_2$ are archived for the DA, GB2 and SM models. The best (lowest) values of $A_3$ are for the Z, BS and SU models.

**Table 3.** The final ranking $R$ of GoF measures – Group I distributions

| Model | $W_p$ | $R(W_p)$ | $A_1$ | $R(A_1)$ | $A_2$ | $R(A_2)$ | $A_3$ | $R(A_3)$ | $R$ |
|---|---|---|---|---|---|---|---|---|---|
| **DA** | 0.983 | 3 | 0.033 | 3 | 0.066 | 1 | 0.145 | 3 | 1 |
| **GB2** | 0.983 | 3 | 0.034 | 4 | 0.067 | 2 | 0.239 | 4 | 2 |
| **Z** | 0.983 | 3 | 0.034 | 4 | 0.134 | 6 | 0.024 | 1 | 3 |
| LOG | 0.994 | 1 | 0.013 | 1 | 5.863 | 7 | 0.766 | 8 | 4 |
| SM | 0.983 | 3 | 0.035 | 7 | 0.068 | 3 | 0.308 | 5 | 5 |
| BS | 0.983 | 3 | 0.034 | 4 | >1.000 | 10 | 0.043 | 2 | 6 |
| GG | 0.987 | 2 | 0.027 | 2 | 2.516 | 8 | 0.492 | 7 | 7 |
| BPr | 0.979 | 8 | 0.041 | 8 | 0.106 | 4 | 5.200 | 9 | 8 |
| IG | 0.979 | 8 | 0.042 | 9 | 0.107 | 5 | 5.290 | 10 | 9 |
| PIV | 0.968 | 10 | 0.063 | 10 | 391.900 | 9 | 0.440 | 6 | 10 |

Note. As in Figure 1.
Source: authors' work based on EU-SILC data.

**Table 4.** The final ranking $R$ of GoF measures – Group II distributions

| Model | $W_p$ | $R(W_p)$ | $A_1$ | $R(A_1)$ | $A_2$ | $R(A_2)$ | $A_3$ | $R(A_3)$ | $R$ |
|---|---|---|---|---|---|---|---|---|---|
| **SU** | 0.990 | 2 | 0.020 | 1 | 0.128 | 1 | 0.060 | 1 | 1 |
| **FSGN2** | 0.994 | 1 | 0.020 | 1 | 0.292 | 2 | 0.604 | 2 | 1 |
| **EN** | 0.973 | 3 | 0.049 | 3 | 2.029 | 5 | 4.910 | 5 | 3 |
| GBSN | 0.943 | 5 | 0.114 | 4 | 1.072 | 3 | 9.450 | 10 | 4 |
| SSGN | 0.927 | 6 | 0.146 | 5 | 1.478 | 4 | 6.940 | 6 | 5 |
| FSGN1 | 0.927 | 6 | 0.147 | 6 | >1.000 | 8 | 7.030 | 7 | 6 |
| ESGN1 | 0.926 | 8 | 0.147 | 6 | >1.000 | 8 | 7.070 | 8 | 7 |
| PN | 0.970 | 4 | 0.601 | 11 | >1.000 | 8 | 1.270 | 3 | 8 |
| ESGN3 | 0.916 | 10 | 0.153 | 8 | 20.204 | 6 | 4.480 | 4 | 9 |
| KN | 0.920 | 9 | 0.154 | 9 | >1.000 | 8 | 14.295 | 11 | 10 |
| TPSN | 0.865 | 11 | 0.266 | 10 | >1.000 | 8 | 19.060 | 12 | 11 |
| ESGN2 | 0.120 | 12 | 0.880 | 12 | 180.361 | 7 | 8.199 | 9 | 12 |

Note. As in Figure 2.
Source: authors' work based on EU-SILC data.

Tables 5 and 6 show numerical characteristics, i.e. the mean ($A$), the lower quartile ($Q_1$), the median ($Q_2$), the upper quartile ($Q_3$), the standard deviation ($SD$) and the coefficient of variation ($CoV$) calculated for the top five models according to the numerical measures presented in Tables 1 and 3 (Group I distributions) and in Tables 2 and 4 (Group II distributions). These empirical characteristics are compared with the theoretical ones using percentage relative errors $RE$. The analysed models are sorted according to the value of final ranking $R$ based on the sum of sub-rankings $R()$. The top three models are in bold.

The results presented in Tables 5 and 6 indicate that models based on DA, GB2 and SM stand out in the Group I distribution, while those based on SU, EN and FSGN2 stand out in the Group II distribution. The best model for the mean are Z, SU and DA, for the lower quartile are Z, LOG and SU, for the median are DA, GB2 and SM, for the upper quartile are EN, FSGN2 and SU, for the standard deviation are DA, SU and GB2, for the coefficient of variation are SU, DA and GB2.

Table 7 shows the values of the last GoF measure (10), specifically defined for this study, calculated for the models from Tables 6 and 7. The models are sorted according to the values of the proposed $QM$ measure. The models that stand out in relation to the $QM$ values are DA, GB2 and SM.

**Table 5.** Empirical ($ECH$) and theoretical characteristics with sub-rankings $R()$ and final ranking $R$ – Group I distributions

| Specification | $ECH$ | DA | GB2 | Z | SM | LOG |
|---|---|---|---|---|---|---|
| $A$ | 665.46 | 666.43 (0.14) | 667.06 (0.24) | 665.30 (0.02) | 667.52 (0.31) | 660.40 (0.76) |
| $R(A)$ | . | 2 | 3 | 1 | 4 | 5 |
| $Q_1$ | 400.53 | 413.63 (3.27) | 413.83 (3.32) | 404.05 (0.88) | 414.12 (3.39) | 405.69 (1.29) |
| $R(Q_1)$ | . | 3 | 4 | 1 | 5 | 2 |
| $Q_2$ | 570.76 | 570.51 (0.04) | 570.28 (0.08) | 572.02 (0.22) | 570.06 (0.12) | 582.58 (2.07) |
| $R(Q_2)$ | . | 1 | 2 | 4 | 3 | 5 |
| $Q_3$ | 801.06 | 791.34 (1.21) | 790.72 (1.29) | 842.23 (5.14) | 789.64 (1.43) | 826.88 (3.22) |
| $R(Q_3)$ | . | 1 | 2 | 5 | 3 | 4 |
| $SD$ | 435.26 | 452.94 (4.06) | 458.17 (5.26) | 372.25 (14.47) | 464.27 (6.66) | 364.81 (16.19) |
| $R(SD)$ | . | 1 | 2 | 4 | 3 | 5 |
| $CoV$ | 0.65 | 0.68 (4.61) | 0.69 (5.69) | 0.56 (13.85) | 0.70 (7.08) | 0.55 (15.08) |
| $R(CoV)$ | . | 1 | 2 | 4 | 3 | 5 |
| $R$ | . | 1 | 2 | 3 | 4 | 5 |

Note. Theoretical characteristic with $RE$ values (in %) in parentheses. $A$ – the mean, $Q_1$ – the lower quartile, $Q_2$ – the median, $Q_3$ – the upper quartile, $SD$ – the standard deviation, $CoV$ – the coefficient of variation.
Source: authors' work based on EU-SILC data.

**Table 6.** Empirical ($ECH$) and theoretical characteristics with sub-rankings $R()$ and the final ranking $R$ – Group II distributions

| Specification | $ECH$ | SU | EN | FSGN2 | GBSN | SSGN |
|---|---|---|---|---|---|---|
| $A$ | 665.46 | 664.99 | 643.64 | 660.90 | 608.83 | 715.09 |
| | | (0.08) | (3.28) | (0.69) | (8.51) | (7.45) |
| $R(A)$ | . | 1 | 3 | 2 | 5 | 4 |
| $Q_1$ | 400.53 | 413.15 | 420.52 | 413.49 | 413.19 | 423.10 |
| | | (3.16) | (5.00) | (3.24) | (3.17) | (5.64) |
| $R(Q_1)$ | . | 1 | 4 | 3 | 2 | 5 |
| $Q_2$ | 570.76 | 562.03 | 583.38 | 562.06 | 574.29 | 643.42 |
| | | (1.54) | (2.20) | (1.53) | (0.61) | (12.72) |
| $R(Q_2)$ | . | 3 | 4 | 2 | 1 | 5 |
| $Q_3$ | 801.06 | 792.82 | 795.83 | 794.45 | 769.52 | 937.81 |
| | | (1.03) | (0.66) | (0.83) | (3.94) | (17.06) |
| $R(Q_3)$ | . | 3 | 1 | 2 | 4 | 5 |
| $SD$ | 435.26 | 414.21 | 354.56 | 348.04 | 272.39 | 378.89 |
| | | (4.84) | (18.55) | (20.05) | (37.42) | (12.96) |
| $R(SD)$ | . | 1 | 3 | 4 | 5 | 2 |
| $CoV$ | 0.65 | 0.62 | 0.55 | 0.53 | 0.45 | 0.53 |
| | | (4.15) | (15.23) | (18.92) | (31.23) | (18.46) |
| $R(CoV)$ | . | 1 | 2 | 4 | 5 | 3 |
| $R$ | . | 1 | 2 | 2 | 4 | 5 |

Note. As in the Table 5.

Source: authors' work based on EU-SILC data.

**Table 7.** Values of the GoF measure $QM$ – Group I and Group II distributions

| Specification | DA | GB2 | SM | SU | FSGN2 | Z | LOG | EN | GBSN | SSGN |
|---|---|---|---|---|---|---|---|---|---|---|
| Group | I | I | I | II | II | I | I | II | II | II |
| $QM$ | 182.02 | 187.07 | 195.64 | 198.94 | 251.08 | 304.73 | 320.95 | 382.42 | 763.28 | 1,386.41 |

Note. As in Figure 1 and Figure 2.

Source: authors' work based on EU-SILC data.

The final collective model ranking, taking into account the results presented in Tables 1–7, is provided in Table 8. The final collective ranking of models is based on information criteria (4), GoF coefficients (6)–(8) and analysed characteristics including (9) and the measure proposed by the authors (10). It is worth noting that among the top three distributions are two ND-derived distributions, i.e. SU and FSGN2; the SU is slightly better than the Dagum distribution in terms of the analysed measures. ND-derived distributions (Group II), especially SU, can be a good alternative to well-known distributions, which are typically used to model income (Group I).

**Table 8.** Collective model ranking list for information criteria (IC), GoF coefficients (GOFC) and characteristics (CH) – Group I and Group II distributions

| Specification | Group | IC | GOFC | CH | Sum | *R* |
|---|---|---|---|---|---|---|
| SU | II | 1 | 3 | 2 | 6 | 1 |
| DA | I | 3 | 4 | 1 | 8 | 2 |
| FSGN2 | II | 2 | 2 | 4 | 8 | 2 |
| GB2 | I | 5 | 5 | 3 | 13 | 4 |
| Z | I | 8 | 1 | 6 | 15 | 5 |
| SM | I | 4 | 7 | 5 | 16 | 6 |
| LOG | I | 7 | 6 | 9 | 22 | 7 |
| EN | II | 6 | 9 | 7 | 22 | 7 |
| GBSN | II | 10 | 8 | 8 | 26 | 9 |
| SSGN | II | 9 | 10 | 10 | 29 | 10 |

Note. As in Figure 1 and Figure 2.
Source: authors' work based on EU-SILC data.

## 7. Conclusion

As far as the authors have been able to establish, this article is the first attempt to apply ND-derived distributions with real domain to the problem of model income.

The results obtained in this study confirm what has often been documented in the literature, namely that the Dagum model is particularly useful for income modelling (see e.g. Jędrzejczak, 1993, 2006; Trzcińska, 2020). However, as the authors have demonstrated with real income data and by applying different numerical measures, the family of income models with ND-derived distributions (Group II) can compete with distributions typically used for this purpose (Group I). As evidenced by the collective model ranking in Table 8, the SU Johnson distribution, representing the group of ND-derived distributions, can serve as an interesting alternative for income modelling. It is also worth emphasizing that the SU distribution can be an interesting model for variables with positive and negative values (e.g. corporate profits), which cannot be modelled with the common distributions (e.g. Dagum).

It is also worth noting that a longer series of income data from EU-SILC could help capture the dependencies between errors resulting from the use of certain families of distributions to approximate the characteristics of the empirical distribution depending on the phases of the business cycle. This could provide a more systematic assessment of a given family of distributions. Moreover the ND-derived distributions, especially SU and FSGN2, could be used to model income of more homogeneous groups, with a negligible distribution asymmetry, e.g. people with disabilities. These issues will be the subject of a study the authors plan to undertake in the nearest future.

## Acknowledgements

The work of Marcin Szymkowiak has been developed as part of a project entitled ‘Indirect estimation of disability in the 2011 census’, which was financed by Poland’s National Science Centre from a grant awarded by virtue of decision no. DEC-2013/11/B/HS4/01472.

## References

Aitchison, J., & Brown, J. A. C. (1957). *The Lognormal Distribution: with special reference to its uses in economics*. Cambridge University Press.

Arellano-Valle, R. B., Gómez, H. W., & Quintana, F. A. (2004). A new class of skew-normal distributions. *Communications in Statistics – Theory and Methods*, *33*(7), 1465–1480. https://doi.org/10.1081/STA-120037254.

Azzalini, A. (1985). A Class of Distributions which Includes the Normal Ones. *Scandinavian Journal of Statistics*, *12*(2), 171–178.

Bahrami, W., & Qasemi, E. (2015). A Flexible Skew-Generalized Normal Distribution. *Journal of Statistical Research of Iran JSRI*, *11*(2), 131–145. https://doi.org/10.18869/acadpub.jsri.11.2.131.

Behboodian, J. (1970). On the modes of a mixture of two normal distributions. *Technometrics*, *12*(1), 131–139. https://doi.org/10.2307/1267357.

Birnbaum, Z. W., & Saunders, S. C. (1969). A new family of life distributions. *Journal of Applied Probability*, *6*(2), 637–652. https://doi.org/10.2307/3212003.

Brzeziński, M. (2013). Parametric Modelling of Income Distribution in Central and Eastern Europe. *Central European Journal of Economic Modelling and Econometrics*, *5*(3), 207–230. https://doi.org/10.24425/cejeme.2013.119261.

Choudhury, K., & Abdul, M. M. (2011). Extended skew generalized normal distribution. *METRON*, *69*(3), 265–278. https://doi.org/10.1007/BF03263561.

Cordeiro, G. M., & de Castro, M. (2011). A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, *81*(7), 883–898. https://doi.org/10.1080/00949650903530745.

Dagum, C. (1977). A New Model of Personal Income Distribution: Specification and Estimation. *Économie Appliquée*, *30*(3), 413–436.

Gaddum, J. H. (1945). Lognormal distributions. *Nature*, *156*, 463–466. https://doi.org/10.1038/156463a0.

Główny Urząd Statystyczny. (2021). *Dochody i warunki życia ludności Polski – raport z badania EU-SILC 2019*. https://stat.gov.pl/en/topics/living-conditions/living-conditions/incomes-and-living-conditions-of-the-population-in-poland-report-from-the-eu-silc-survey-of-2019,1,12.html.

Gómez, H. W., Elal-Olivero, D., Salinas, H. S., & Bolfarine, H. (2011). Bimodal Extension Based on the Skew-Normal Distribution with Application to Pollen Data. *Environmetrics*, *22*(1), 50–62. https://doi.org/10.1002/env.1026.

Gupta, R. C., & Gupta, R. D. (2008). Analyzing skewed data by power normal model. *TEST*, *17*(1), 197–210. https://doi.org/10.1007/s11749-006-0030-x.

Jędrzejczak, A. (1993). Application of the Dagum distribution in the analysis of income distributions in Poland. *Acta Universitatis Lodziensis. Folia Oeconomica*, (131), 103–112.

Jędrzejczak, A. (2006). The characteristic of theoretical income distributions and their application to the analysis of wage distributions in Poland by regions. *Acta Universitatis Lodziensis. Folia Oeconomica*, (196), 183–198.

Jędrzejczak, A., & Pekasiewicz, D. (2020). *Teoretyczne rozkłady dochodów gospodarstw domowych i ich estymacja*. Wydawnictwo Uniwersytetu Łódzkiego.

Jędrzejczak, A., & Trzcińska, K. (2018). Application of the Zenga distribution to the analysis of household income in Poland by socio-economic group. *Statistica & Applicazioni*, *16*(2), 123–140. https://doi.org/10.26350/999999_000015.

Johnson, N. L. (1949). System of frequency curves generated by methods of translation. *Biometrika*, *36*(1/2), 149–176. https://doi.org/10.2307/2332539.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions* (vol. 2, 2nd ed.). John Wiley & Sons.

Kim, H.-J. (2005). On a class of two-piece skew-normal distributions. *Statistics. A Journal of Theoretical and Applied Statistics*, *39*(6), 537–553. https://doi.org/10.1080/02331880500366027.

Kordos, J. (1968). *Metody matematyczne badania i analizy rozkładów dochodów ludności*. Główny Urząd Statystyczny.

Kordos J. (1973). *Metody analizy i prognozowania rozkładów płac i dochodów ludności.* Państwowe Wydawnictwo Ekonomiczne.

Kot, S. M. (Ed.). (1999). *Analiza ekonometryczna kształtowania się płac w Polsce w okresie transformacji*. Wydawnictwo Naukowe PWN.

Kot, S. M. (2000). *Ekonometryczne modele dobrobytu*. Wydawnictwo Naukowe PWN.

Kumar, C. S., & Anusree, M. R. (2015). On an extended version of skew generalized normal distribution and some of its properties. *Communications in Statistics – Theory and Methods*, *44*(3), 573–586. https://doi.org/10.1080/03610926.2012.739251.

Kunte, S., & Gore, A. P. (1992). The paradox of large samples. *Current Science*, *62*(5), 393–395.

Lange, O. (1967). *Wstęp do ekonometrii* (4th ed.). Państwowe Wydawnictwo Naukowe.

Łukasiewicz, P., & Orłowski, A. (2004). Probabilistic Models of Income Distributions. *Physica A: Statistical Mechanics and its Applications*, *344*(1–2), 146–151. https://doi.org/10.1016/j.physa.2004.06.106.

Ma, Y., & Genton, M. G. (2004). Flexible Class of Skew-Symmetric Distributions. *Scandinavian Journal of Statistics*, *31*(3), 459–468.

McDonald, J. B. (1984). Some generalized functions for the size distribution of income. *Econometrica*, *52*(3), 647–663. https://doi.org/10.2307/1913469.

Mead, M., Nassar, M. M., & Dey, S. (2018). A Generalization of Generalized Gamma Distributions. *Pakistan Journal of Statistics and Operation Research*, *14*(1), 121–138. https://doi.org/10.18187/pjsor.v14i1.1692.

Metcalf, C. E. (1972). *An Econometric Model of Income Distribution*. Markham Publishing Company.

Nekoukhou, V., Alamatsaz, M. H., & Aghajani, A. H. (2013). A flexible skew-generalized normal distribution. *Communications in Statistics – Theory and Methods*, *42*(13), 2324–2334. https://doi.org/10.1080/03610926.2011.599003.

Ostasiewicz, K. (2013). Adekwatność wybranych rozkładów teoretycznych dochodów w zależności od metody aproksymacji. *Przegląd Statystyczny. Statistical Review*, *60*(4), 499–522.

Pareto, V. (1895). La legge della domanda. *Giornale Degli Economisti*, *10*(6), 59–68.

Pratesi, M. (Ed.). (2016). *Analysis of Poverty Data by Small Area Estimation*. John Wiley & Sons. https://doi.org/10.1002/9781118814963.

R Core Team. (2021). *The R Project for Statistical Computing*. https://www.R-project.org/.

Rasekhi, M., Hamedani, G. G., & Chinipardaz, R. (2017). A flexible extension of skew generalized normal distribution. *METRON*, *75*(1), 87–107. https://doi.org/10.1007/s40300-017-0106-2.

Rieck, J. R., & Nedelman, J. R. (1991). A Log-Linear Model for the Birnbaum-Saunders Distribution. *Technometrics*, *33*(1), 51–60. https://doi.org/10.2307/1269007.

Salamaga, M. (2016). Badanie wpływu metody estymacji teoretycznych modeli rozkładu dochodów na jakość aproksymacji rozkładu dochodów mieszkańców Krakowa. *Zeszyty Naukowe UEK*, *3*(951), 63–79. https://doi.org/10.15678/ZNUEK.2016.0951.0305.

Sharafi, M., & Behboodian, J. (2008). The Balakrishnan skew-normal density. *Statistical Papers*, *49*(4), 769–778. https://doi.org/10.1007/s00362-006-0038-z.

Singh, S. K., & Maddala, G. S. (1976). A Function for Size Distribution of Income. *Econometrica*, *44*(5), 963–970. https://doi.org/10.2307/1911538.

Stacy, E. W. (1962). A generalization of the gamma distribution. *The Annals of Mathematical Statistics*, *33*(3), 1187–1192.

Stacy, E. W., & Mihram, G. A. (1965). Parameter Estimation for a Generalized Gamma Distribution. *Technometrics*, *7*(3), 349–358. https://doi.org/10.2307/1266594.

Sulewski, P. (2019). Modified Lilliefors Goodness-of-fit Test for Normality. *Communications in Statistics – Simulation and Computation*, *51*(3), 1199–1219. https://doi.org/10.1080/03610918.2019.1664580.

Sulewski, P. (in press). New Members of The Johnson Family of Probability Distributions: Properties and Application. *REVSTAT – Statistical Journal*.

Trzcińska, K. (2020). Analysis of Household Income in Poland Based on the Zenga Distribution and Selected Income Inequality Measure. *Folia Oeconomica Stetinensia*, *20*(1), 421–436. https://doi.org/10.2478/foli-2020-0025.

Trzcińska, K. (2022). An Analysis of Household Income in Poland and Slovakia Based on Selected Income Models. *Folia Oeconomica Stetinensia*, *22*(1), 287–301. https://doi.org/10.2478/foli-2022-0014.

Venegas, O., Sanhueza, A. I., & Gómez, H. W. (2011). An extension of the skew-generalized normal distribution and its derivation. *Proyecciones. Journal of Mathematics*, *30*(3), 401–413. https://doi.org/10.4067/S0716-09172011000300007.

Vielrose, E. (1960). *Rozkład dochodów według wielkości*. Polskie Wydawnictwo Gospodarcze.

Wałęga, A., & Wałęga, G. (2021). Self-employment and over-indebtedness in Poland: Modelling income and debt repayments distribution. *EBER Entrepreneurial Business and Economics Review*, *9*(4), 51–65. https://doi.org/10.15678/EBER.2021.090404.

Wiśniewski, J. (1934). *Rozkład dochodów według wysokości*. Instytut Badania Koniunktur Gospodarczych i Cen.

Yadegari, I., Gerami, A., & Khaledi, M. J. (2008). A generalization of the Balakrishnan skew-normal distribution. *Statistics & Probability Letters*, *78*(10), 1165–1167. https://doi.org/10.1016/j.spl.2007.12.001.

Zenga, M. M. (2010). Mixture of Polisicchio's truncated Pareto distributions with beta weights. *Statistica & Applicazioni*, *8*(1), 3–25. https://www.vitaepensiero.it/scheda-articolo_digital/michele-zenga/mixture-of-polisicchios-truncated-pareto-distributions-with-beta-weights-999999_2010_0001_0002-151302.html.

Zenga, M. M., Pasquazzi, L., & Zenga, M. (2012). First applications of a new three-parameter distribution for non-negative variables. *Statistica & Applicazioni*, *10*(2), 131–149. https://statisticaeapplicazioni.vitaepensiero.it/scheda-articolo_digital/leo-pasquazzi-mariangela-zenga-michele-zenga/first-applications-of-a-new-three-parameter-distribution-for-non-negative-variables-999999_2012_0002_0037-151352.html.

## Appendix

Let $B(x,y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$ be the beta function defined by the gamma function and

$$v(x;\beta,k) = \begin{cases} 0.5\sqrt{\beta k}(1-k)^{-1}x^{-1.5} & x \in [\beta k, \beta/k] \\ 0 & \text{otherwise.} \end{cases}$$

The PDF or CDF for Group I distributions are as follows:

$$F_{LOG}(x;c,b,a) = \Phi\left[\frac{\ln(x-a)-b}{c}\right] \ (x > a;\ a,b \in R, c > 0);$$

$$F_{BS}(x;\alpha,\sigma,\mu) = \Phi\left[\frac{1}{\alpha}\left(\sqrt{\frac{x-\mu}{\sigma}} - \sqrt{\frac{\sigma}{x-\mu}}\right)\right] \ (x > \mu;\ \alpha > 0);$$

$$f_{BPr}(x;\alpha,\beta,\sigma) = \frac{\left(\frac{x}{\sigma}\right)^{\alpha-1}\left(1+\frac{x}{\sigma}\right)^{-\alpha-\beta}}{\sigma B(\alpha,\beta)} \ (x \geq 0;\ \alpha,\beta > 0);$$

$$f_{IG}(x;\alpha,\beta) = \frac{\beta^{\alpha}x^{-\alpha-1}}{\Gamma(\alpha)} \exp\left(-\frac{\beta}{x}\right) \ (x > 0;\ \alpha > 0, \beta > 0);$$

$$F_D(x;\sigma,\alpha,\beta) = \left[1 + \left(\frac{x}{\sigma}\right)^{-\alpha}\right]^{-\beta} \ (x > 0;\ \alpha,\beta > 0);$$

$$f_Z(x;\beta,\alpha,\gamma) = \int_0^1 v(x;\beta,k)\frac{k^{\alpha-1}(1-k)^{\gamma-1}}{B(\alpha,\gamma)}dk \ \left(x \geq 0;\ \beta,\alpha,\gamma > 0,\ k \in (0,1)\right);$$

$$f_{SM}(x;\sigma,\alpha,\gamma) = \frac{\alpha\gamma}{\sigma}\left(\frac{x}{\sigma}\right)^{\alpha-1}\left[1 + \left(\frac{x}{\sigma}\right)^{\alpha}\right]^{-\gamma-1} \ (x \geq 0;\ \alpha,\gamma > 0);$$

$$F_{PIV}(x;\mu,\sigma,\gamma,\alpha) = 1 - \left[1 + \left(\frac{x-\mu}{\sigma}\right)^{\frac{1}{\gamma}}\right]^{-\alpha} \ (x \geq \mu;\ \gamma,\alpha > 0);$$

$$f_{GG}(x;\sigma,\alpha,\beta) = \frac{\alpha}{\sigma\Gamma(\beta)}\left(\frac{x}{\sigma}\right)^{\alpha\beta-1}\exp\left[-\left(\frac{x}{\sigma}\right)^{\alpha}\right] \ (x \geq 0;\ \beta > 0);$$

$$f_{GB2}(x;\alpha,\sigma,\beta,\gamma) = \frac{\alpha}{\sigma B(\beta,\gamma)}\left(\frac{x}{\sigma}\right)^{\alpha\beta-1}\left[1+\left(\frac{x}{\sigma}\right)^{\alpha}\right]^{-(\beta+\gamma)} \ (x \geq 0;\ \alpha,\beta,\gamma > 0).$$

Special cases of PIV, GG and GB2 are presented in Tables A1–A3, respectively.

**Table A1.** Special cases of PIV

| Name | $\mu$ | $\sigma$ | $\gamma$ | $\alpha$ |
|---|---|---|---|---|
| Pareto type: I | $\sigma$ | . | 1 | . |
| II | . | . | 1 | . |
| III | . | . | . | 1 |
| Lomax | 0 | . | 1 | . |

Source: authors' work based on Jędrzejczak and Pekasiewicz (2020).

**Table A2.** Special cases of GG

| Name | $\sigma$ | $\alpha$ | $\beta$ |
|---|---|---|---|
| Exponential | . | 1 | 1 |
| Gamma | . | 1 | . |
| Weibull | . | . | 1 |
| Chi-square | 2 | 2 | $0.5n,\ n \in N$ |
| Chi | $\sqrt{2}$ | 2 | $0.5n,\ n \in N$ |
| Rayleigh | $\sigma\sqrt{2}$ | 2 | 1 |
| Maxwell-Boltzmann | $\sigma\sqrt{2}$ | 2 | 1.5 |

Source: authors' work based on Stacy and Mihram (1965).

**Table A3.** Special cases of GB2

| Name | $\sigma$ | $\alpha$ | $\beta$ | $\gamma$ |
|---|---|---|---|---|
| Singh–Maddala (Burr XII) | . | . | 1 | . |
| Dagum (Burr III) | . | . | . | 1 |
| Beta type II | . | 1 | . | . |
| Standard Burr XII | 1 | . | 1 | . |
| Standard Burr III | 1 | . | . | 1 |
| Standard Beta type II | 1 | 1 | . | . |
| Fisk (log-logistic) | . | . | 1 | 1 |
| Lomax (Pareto type II) | . | 1 | 1 | . |
| Inverse Lomax | . | 1 | . | 1 |
| Paralogistic | . | *h* | 1 | . |
| Inverse paralogistic | . | *a* | . | . |
| Fisher | *v/u* | *u/2* | *v/2* | 1 |

Source: authors' work based on Mead et al. (2018).

Let $a_1, a_2, b_1, b_2$ be multipurpose parameters, $c$ the semi-fraction parameter then

$$T(h,a)=\frac{1}{2\pi}\int_0^a \frac{e^{-0.5h^2(1+x^2)}}{1+x^2}dx \quad (h,a\in R).$$

The PDF or CDF for the distributions derived from ND are as follows:

$$f_{SN}(x;\mu,\sigma,\alpha)=\frac{2}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\alpha\frac{x-\mu}{\sigma}\right) \ \ (\alpha\in R,\ SN(\mu,\sigma,0)=N(\mu,\sigma);$$

$$f_{SGN}(x;\mu,\sigma,\alpha,\beta)=\frac{2}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\frac{\alpha(x-\mu)}{\sqrt{\sigma^2+\beta(x-\mu)^2}}\right) \ \ (\beta\geq 0,\ \alpha\in R),$$
$$SGN(\mu,\sigma,\alpha,0)=SN(\mu,\sigma,\alpha),\quad SGN(\mu,\sigma,0,0)=N(\mu,\sigma);$$

$$f_{ESGN1}(x;\mu,\sigma,\alpha,\beta,\gamma)=\frac{2}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)\Phi\left[\frac{\alpha\frac{x-\mu}{\sigma}}{\sqrt{1+\beta\left(\frac{x-\mu}{\sigma}\right)^2+\gamma\left(\frac{x-\mu}{\sigma}\right)^4}}\right] \ \ (\beta,\gamma\geq 0,\ \alpha\in R),$$

$$ESGN1(\mu,\sigma,\alpha,\beta,0)=SGN(\mu,\sigma,\alpha,\beta),\quad ESGN1(\mu,\sigma,0,\beta,\gamma)=N(\mu,\sigma),$$
$$ESGN1(\mu,\sigma,\alpha,0,0)=SN(\mu,\sigma,\alpha);$$

$$f_{ESGN2}(x;\mu,\sigma,\alpha,\beta,\gamma)=\frac{4}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)\int_{-\infty}^{\frac{\alpha(x-\mu)}{\sqrt{\sigma^2+\beta(x-\mu)^2}}}\varphi(t)\Phi\left(\frac{-\sqrt{\beta}\gamma t(x-\mu)}{\sqrt{\sigma^2+\beta(x-\mu)^2+\sigma^2\gamma^2}}\right)dt,$$
$$\alpha,\gamma\in R,\quad \beta\geq 0,\quad ESGN2(\mu,\sigma,0,\beta,0)=N(\mu,\sigma),$$
$$ESGN2(\mu,\sigma,\alpha,0,\gamma)=SN(\mu,\sigma,\alpha),\quad ESGN2(\mu,\sigma,\alpha,\beta,0)=SGN(\mu,\sigma,\alpha,\beta);$$

$$f_{ESGN3}(x;\mu,\sigma,\alpha,\beta,\gamma)=\frac{2}{\sigma(\gamma+2)}\varphi\left(\frac{x-\mu}{\sigma}\right)\left\{1+\gamma\Phi\left[\frac{\alpha(x-\mu)}{\sqrt{\sigma^2+\beta(x-\mu)^2}}\right]\right\},$$
$$\beta\geq 0,\gamma\geq -1,\quad \alpha\in R,\quad ESGN3(\mu,\sigma,\alpha,\beta,0)=ESGN3(\mu,\sigma,0,\beta,\gamma)=N(\mu,\sigma);$$

$$f_{SSGN}(x;\mu,\sigma,\alpha,\beta,\gamma)=\frac{2}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)\Phi\left[\frac{\alpha\frac{x-\mu}{\sigma}}{\sqrt{1+\beta\left|\frac{x-\mu}{\sigma}\right|^{2\gamma}}}\right] \ \ (\beta\geq 0,\ \gamma\neq 0,\ \alpha\in R),$$
$$\beta=0\Longrightarrow\gamma=1;\quad \alpha=0\Longrightarrow\beta=0,\ \gamma=1;\quad SSGN(\mu,\sigma,0,0,1)=N(\mu,\sigma),$$
$$SSGN(\mu,\sigma,\alpha,0,1)=SN(\mu,\sigma,\alpha),\quad SSGN(\mu,\sigma,\alpha,\beta,1)=SGN(\mu,\sigma,\alpha,\beta),$$
$$SSGN(\mu,\sigma,\alpha,\beta,2)=ESGN1(\mu,\sigma,\alpha,0,\gamma);$$

$$f_{SFN}(x;\mu,\sigma,\alpha,\beta)=\frac{1}{\sigma[1-\Phi(\beta)]}\varphi\left(\frac{|x-\mu|}{\sigma}+\beta\right)\Phi\left(\alpha\frac{x-\mu}{\sigma}\right) \ \ (\alpha,\beta\in R),$$
$$SFN(\mu,\sigma,0,0)=N(\mu,\sigma),\quad SFN(\mu,\sigma,\alpha,0)=SN(\mu,\sigma,\alpha);$$

$$f_{FSN}(x;\mu,\sigma,\alpha,\beta)=\frac{2}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)\Phi\left[\alpha\frac{x-\mu}{\sigma}+\beta\left(\frac{x-\mu}{\sigma}\right)^3\right] \ \ (\alpha,\beta\in R),$$
$$FSN(\mu,\sigma,0,0)=N(\mu,\sigma),\quad FSN(\mu,\sigma,\alpha,0)=SN(\mu,\sigma,\alpha);$$

$$f_{FSGN1}(x;\mu,\sigma,\alpha,\beta,\gamma)=\frac{2}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)\Phi\left[\frac{\alpha(x-\mu)+\frac{\gamma}{\sigma^2}(x-\mu)^3}{\sqrt{\sigma^2+\beta(x-\mu)^2}}\right] \ \ (\alpha,\gamma\in R,\ \ \beta\geq 0),$$
$$FSGN1(\mu,\sigma,0,0,0)=FSGN1(\mu,\sigma,0,\beta,0)=N(\mu,\sigma),$$

$$FSGN1(\mu,\sigma,\alpha,0,0) = SN(\mu,\sigma,\alpha), \quad FSGN1(\mu,\sigma,\alpha,\beta,0) = SGN(\mu,\sigma,\alpha,\beta),$$
$$FSGN1(\mu,\sigma,\alpha,0,\gamma) = FSN(\mu,\sigma,\alpha,\beta);$$

$$f_{FSGN2}(x;\mu,\sigma,\alpha,\beta,\gamma) = \frac{1}{\sigma[1-\Phi(\gamma)]}\varphi\left(\frac{|x-\mu|}{\sigma}+\gamma\right)\Phi\left[\frac{\alpha(x-\mu)}{\sqrt{\sigma^2+\beta(x-\mu)^2}}\right](\alpha,\gamma\in R,\ \beta\geq 0),$$
$$FSGN2(\mu,\sigma,\alpha,0,0) = SN(\mu,\sigma,\alpha), \quad FSGN2(\mu,\sigma,\alpha,\beta,0) = SGN(\mu,\sigma,\alpha,\beta),$$
$$FSGN2(\mu,\sigma,\alpha,0,\gamma) = SFN(\mu,\sigma,\alpha,\beta);$$

$$f_{KN}(x;\mu,\sigma,\alpha,\beta) = \frac{\alpha\beta}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)\left[\Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha-1}\left\{1-\left[\Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha}\right\}^{\beta-1} \ (\alpha,\beta>0),$$
$$|KN(\mu,\sigma,1,1) = N(\mu,\sigma),\ \ KN(\mu,\sigma,2,1) = SN(\mu,\sigma,1);$$

$$f_{BSN}(x;\mu,\sigma,\alpha,\beta) = \frac{\varphi\left(\frac{x-\mu}{\sigma}\right)\left[\Phi\left(\alpha\frac{x-\mu}{\sigma}\right)\right]^{\beta}}{\int_{-\infty}^{\infty}\varphi\left(\frac{x-\mu}{\sigma}\right)\left[\Phi\left(\alpha\frac{x-\mu}{\sigma}\right)\right]^{\beta}dx} \ \ (\alpha\in R,\ \ \beta=1,2,\dots),$$
$$BSN(\mu,\sigma,\alpha,0) = N(\mu,\sigma),\ BSN(\mu,\sigma,\alpha,1) = SN(\mu,\sigma,\alpha);$$

$$f_{GBSN}(x;\mu,\sigma,\alpha,\beta,\gamma) = \frac{\varphi\left(\frac{x-\mu}{\sigma}\right)\left[\Phi\left(\alpha\frac{x-\mu}{\sigma}\right)\right]^{\beta}\left[1-\Phi\left(\alpha\frac{x-\mu}{\sigma}\right)\right]^{\gamma}}{\int_{-\infty}^{\infty}\varphi\left(\frac{x-\mu}{\sigma}\right)\left[\Phi\left(\alpha\frac{x-\mu}{\sigma}\right)\right]^{\beta}\left[1-\Phi\left(\alpha\frac{x-\mu}{\sigma}\right)\right]^{\gamma}dx} \ (\alpha\in R,\ \ \beta,\gamma=1,2,\dots),$$
$$GBSN(\mu,\sigma,\alpha,\beta,0) = BSN(\mu,\sigma,\sigma,\alpha,\beta), \quad GBSN(\mu,\sigma,\alpha,0,0) = N(\mu,\sigma),$$
$$GBSN(\mu,\sigma,\alpha,1{,}0) = SN(\mu,\sigma,\alpha);$$

$$f_{TPSN}(x;\mu,\sigma,\alpha,\beta) = \frac{\phi\left(\frac{x-\mu}{\sigma},0,1\right)\Phi\left(\alpha\left|\frac{x-\mu}{\sigma}+\beta\right|,0,1\right)}{\sigma\left[I(\alpha>0)-2T\left(\frac{\alpha\beta}{\sqrt{1+\alpha^2}},\frac{1}{\alpha}\right)\right]} \ (\alpha,\beta\in R),$$
$$TPSN(\theta,\sigma,0{,}0) = N(\theta,\sigma);$$

$$F_{SU}(x;\alpha,\beta,\mu,\sigma) = \Phi\left[\alpha+\beta\operatorname{asinh}\left(\frac{x-\mu}{\sigma}\right);0,1\right],$$

$SU(0, 3.223, 0, 2.939)$, according to (1), is similar to the $N(0, 0.916)$ in 98.66%;

$$F_{SC}(x;\alpha,\mu,\sigma) = \Phi\left[\alpha+2\sinh\left(\frac{x-\mu}{\sigma}\right);0,1\right](\alpha\in R), SC(0,\mu,\sigma),$$

according to (1), is similar to the $N(\mu, 0.5\sigma)$ in 96.66%;

$$F_{EN}(x;a_1,b_1,a_2,b_2,c) =$$
$$= \Phi\left[c-\exp\left(\frac{a_1-x}{b_1}\right)+\exp\left(\frac{x-a_2}{b_2}\right)\right] \ \ (a_1,a_2\in R;\ \ \ b_1,b_2,\mathrm{c}>0),$$

$EN(a_1, b_1, a_1, b_1, 0)$, according to (1), is similar to the $N(a_1, 0.5b_1)$ in 96.66%,

$$EN(a_1,b_1,a_1,b_1,c) = SC(a_1,b_1,c).$$

Wiadomości Statystyczne. The Polish Statistician, 2023, *68*(6), 24–46 https://doi.org/10.59139/ws.2023.06.2
Statystyka w praktyce / Statistics in practice
Zgłoszony/submitted: 6.11.2022, poprawiony/revised: 27.04.2023, zaakceptowany/accepted: 22.05.2023

# Impact of a child's disability on the probability of the mother taking up paid employment

Olga Komorowska,[a] Arkadiusz Kozłowski[b]

**Abstract.** Performing paid work is beneficial in many ways, but not every person has equal access to it because of their social and economic situation. Discrepancies in this field are especially visible in the case of mothers. The aim of the study is to assess the impact of the child's disability on the probability of the mother taking up paid employment. The empirical analysis used a decomposition method derived from the Blinder and Oaxaca approach and the logistic regression. The analysis was based on individual household-level data from the representative Household Budget Survey for the years 2005–2020. When analysing the average from all the years, the employment rate of mothers of children without disabilities reached 70.9%, and that of mothers of children with disabilities only 40.2%. This gap was widening throughout the studied period. The lower employment rate among mothers of children with disabilities is caused in the most part directly by the child's disability (and the resulting factors). However, the variable that had the greatest impact on the economic activity in both groups of mothers was education.

**Keywords:** employment rate, mothers of children with disabilities, Blinder-Oaxaca decomposition, logistic regression
**JEL:** J21, C21, I38

# Wpływ niepełnosprawności dziecka na prawdopodobieństwo podjęcia pracy zawodowej przez matkę

**Streszczenie.** Praca zawodowa daje wiele korzyści, jednak nie wszyscy mają do niej równy dostęp ze względu na uwarunkowania społeczne i ekonomiczne. Nierówności w dostępie do pracy są szczególnie widoczne wśród matek. Celem badania omawianego w artykule jest ocena wpływu niepełnosprawności dziecka na prawdopodobieństwo podjęcia pracy zawodowej przez matkę. W analizie empirycznej wykorzystano metodę dekompozycji wywodzącą się z podejścia Blindera i Oaxaki oraz regresję logistyczną. Posłużono się jednostkowymi danymi za lata 2005–2020 dotyczącymi gospodarstw domowych, pochodzącymi z reprezentacyjnego badania budżetów gospodarstw domowych. Przeciętny wskaźnik zatrudnienia matek dzieci bez niepełnosprawności wynosił 70,9%, a matek dzieci z niepełnosprawnościami – 40,2%. Różnica jego wartości pomiędzy badanymi grupami matek zwiększała się w ciągu analizowanego okresu. Niższy wskaźnik zatrudnienia wśród matek dzieci z niepełnosprawnościami wynikał przede wszystkim z niepełnosprawności dziecka (i czynników, które są tego skutkiem). Zmien-

[a] Uniwersytet Gdański, Wydział Zarządzania, Polska / University of Gdańsk, Faculty of Management, Poland. ORCID: https://orcid.org/0000-0002-0305-8748. Autor korespondencyjny / Corresponding author, e-mail: olga.komorowska@ug.edu.pl.

[b] Uniwersytet Gdański, Wydział Zarządzania, Polska / University of Gdańsk, Faculty of Management, Poland. ORCID: https://orcid.org/0000-0001-6282-1494. E-mail: arkadiusz.kozlowski@ug.edu.pl.

ną, która miała największy wpływ na aktywność zawodową matek w obu grupach, było wykształcenie.

**Słowa kluczowe:** wskaźnik zatrudnienia, matki dzieci z niepełnosprawnościami, dekompozycja Blindera-Oaxaki, regresja logistyczna

## 1. Introduction

Work is an inseparable element of human life. Even though it involves effort on the part of the working person, it has substantial value in material, social and emotional terms. Work guarantees income and this, in most cases, makes it possible for the working person to satisfy the current and future needs of the household members (Giddens, 2008; Szyszka, 2016). The level of income marks the status of a household in society (Szymczak & Gawrycka, 2015). As regards individuals, work often gives them financial independency, increases the level of self-fulfilment, and improves the person's status in the family (Kuchařová, 1999; Sarkisian & Gerstel, 2004; Szyszka, 2016). In most cases, work also guarantees personal development, involves contacts with other people and helps to build self-identity and self-esteem (Giddens, 2008; Szyszka, 2016). Last but not least, work gives a feeling of safety through the provision of income (Gasińska, 2016).

Raising the share of women in the labour market and increasing the employment rate among women was one of the fundamental goals related to the labour market in the European Union. The 'Europe 2020' strategy set the target employment at 75% of the population aged 20–64 by 2020, which was possible to achieve by increasing the economic activity of women (European Commission, 2017). In Poland in 2005, the employment rate of women aged 25–49 was 66.1%, so one of the lowest in the EU. There were only four other countries where fewer women from this age group were employed, namely Spain, Greece, Italy and Malta. In 2020, the employment rate of women aged 25–49 grew in Poland to 76.7%, which promoted the country to 15th place among the 27 EU member states.

To compare, the three countries with the lowest employment rate among women aged 25–49 in 2005, i.e. Spain (64.5%), Greece (60.8%) and Italy (59.5%), still occupied the last three places 15 years later, with only a slight improvement in the case of Spain and Greece (68.0% and 58.7%, respectively), and a slight deterioration in the case of Italy (58.7%).

The economic activity of mothers is determined by a number of factors. An important group of such factors results from the culture of a given country and the traditional role of a woman in the family, as well as stereotypes related to women's work (Kurowska, 2012; Mosseri, 2021; Pufal-Struzik, 2017). Another group of determiners is directly related to the mother's resources, but also beliefs. Kurowska (2012) refers to the 'preference theory' (after C. Hakim), in which three groups of women are distinguished: children- and home-oriented, work-oriented, and the combination of the two. Pufal-Struzik (2017) points to internal barriers, such as the lack of faith in one's abilities or fear of social rejection. Since women perform most of the care-related tasks in families (Klaus & Vogel, 2019), the volume of care burden is an important determiner of their economic activity (Perry-Jenkins & Gerstel, 2020; Pollmann-Schult, 2015).

Another group of factors influencing women's economic activity relates to the broadly understood household, i.e. the number of children (Boushey, 2008), the partner's income (Kurowska, 2012), and the support from other family members, which is particularly important when institutional support is lacking (Amah, 2021). An important role in this context is also played by factors resulting from the availability of state aid and effective family policy (Hegewisch & Gornick, 2011; Kay, 2000; Windebank, 2012). Legal regulations on paternity leave, flexible working hours and the access to institutional care all have a positive impact on the economic activity of women raising children (Castro-García & Pazos-Moran, 2016; Haasa & Hwang, 2008; Hegewisch & Gornick, 2011). However, state support should be distributed in a balanced way, because, as Misra et al. (2011) indicate, too long parental leaves might discourage employers from hiring employees raising children, and too high benefits may stifle the beneficiaries' willingness to return to the labour market.

Studies on the economic activity of mothers of children with disabilities show that paid work has an additional significance in their case. It benefits not only the mother, but the child with a disability and the family as a social unit as well. Engaged in performing her job, the mother can take a break from her everyday problems. It is likely to help her gain some distance to her situation and a different perspective on life. The mother's paid employment contributes to the improvement of the quality of life of the entire family (Komorowska & Kozłowski, 2021).

According to studies carried out in the United States, the health-related quality of life of mothers of children with a disability who were economically active was higher than the health-related quality of life of such mothers who did not work (Bourke-Taylor et al., 2011). As shown by the studies on families with an adult intellectually disabled person (Komorowska & Kozłowski, 2021), the fact that the mother is engaged in work and has less time for other obligations increases the independence of her child, and this translates to the improved quality of life of the entire family. The simulation of the future retirement pension of a mother of a child with a disability who is economically active and the one who receives the nursing benefit[1] indicated that the mother who is economically inactive will receive a significantly

[1] The nursing benefit is the benefit granted to the child's carer when they give up paid employment to take care of the child (legal status as of 26.04.2023).

lower future retirement pension (Borski, 2019; Gierusz & Komorowska, 2022). The smaller retirement pension, in turn, is very likely to translate into the lower quality of life of the family members (Główny Urząd Statystyczny [GUS], 2020).

The aim of the study is to assess the impact of the child's disability on the probability of the mother taking up paid employment. The analysis covers the period of 2005–2020 and encompasses the study of the employment rate as such and factors affecting it, in order to distinguish the causes of discrepancies resulting from different distributions of independent variables from the ones resulting from the sole fact of the child's disability. Banaszkiewicz et al. (2019) already performed a similar study for Poland, but it was limited to a single year and involved slightly different methods of analysis.

## 2. Research method

The empirical analysis uses unit data from the Household Budget Survey (HBS) procured from Statistics Poland for the years 2005–2020.[2] We chose this particular survey because it collects data on all the members of households participating in the it including information about disability certificates and the economic activity[3] of household members. Additionally, the HBS shows kinship relations between household members. Among all the households participating in the HBS in a given year, the ones with at least one child aged 0–18 and his/her biological mother were selected for further analyses.

We studied the characteristics taken from the individual study of mothers and the characteristics of the entire household (also individual characteristics of other persons, if they were of significance to the mother's situation, e.g. the age of the youngest child). Another filtering factor was the mother's employment status: the study took into consideration the mothers who either were, or potentially could be, in paid employment, i.e. the persons who labelled themselves as 'working', 'working, but temporarily absent from work', 'unemployed', 'running the household', or

[2] The period of the analysis encompasses the first year of the COVID-19 pandemic (2020), when the performance of the HBS was slightly different than in previous years. First of all, the recruitment and interviews with the respondents, instead of being carried out in person, were made via telephone; not all of the sampled units were interviewed, and the resulting deficiencies were compensated by additional weighing (GUS, 2021). The results received for 2020 may thus differ from prior trends, and if this is the case, these divergences result from a change both in the method used by the HBS and in the overall life situation of the respondents, related to the pandemic. Therefore, the results for that year have to be treated with particular caution.

[3] It should be noted, however, that the HBS is not specifically designed for the labour market statistics and the question regarding economic activity is not the one recommended by the International Labour Organization. Therefore, the general employment rates mentioned in the Introduction, which are based on the Labour Force Survey, should not be compared with the employment rates estimated below. For the purpose of the comparative analysis in this study, however, the definitions used in the HBS are sufficient.

‘other’ (the first two groups consituted the ‘in paid employment’ category, while the remaining three made up the ‘not in paid employment’ category). The analysis excluded persons who labelled themselves as ‘pensioner, retired’, ‘student, pupil’ or ‘incapable of working’ (in total, there were 3.5% of such persons). The study unit was the child’s mother.

The studied variable (dependent variable) was the binary variable indicating whether the mother performed paid work (1 – yes, 0 – no). The grouping variable was the occurrence of a disability[4] in a child below 18 years of age. Based on the literature (Banaszkiewicz et al., 2019; Brown & Clark, 2017; Komorowska, 2017b; Komorowska & Kozłowski, 2021; Sapkota et al., 2017), the experience of one of the authors (cares for a child with a disability) and the availability of data, additional independent variables which were likely to affect the mother’s economic activity were singled out. These variables and their variants were:

- education level (lower secondary or below; basic vocational; upper secondary; tertiary);
- age of the youngest child (in 3-year intervals);
- age of the mother in years ([15, 25); [25, 35); [35, 45); [45, 55); 55+);
- class of locality (rural area; town with less than 20,000 inhabitants; town with 20,000–99,000 inhabitants; city with 100,000–199,000 inhabitants; city with 200,000–499,000 inhabitants; city with 500,000 and more inhabitants);
- number of children (1; 2; 3+);
- number of other adults in the household (0; 1; 2; 3+);
- spouse (partner) lives in the household (yes; no).

The main object of the analysis is the employment rate, defined for the needs of the study as the ratio of the number of mothers in paid employment to the number of all mothers in the study population. Based on the representative study, this rate may be estimated in the following way (Hájek estimator):

$$\hat{\bar{Y}} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i}, \tag{1}$$

where:

$y_i$ is the value of variable $y$ for the $i$-th individual ($y_i = 1$, if the mother is in paid employment, $y_i = 0$, if the mother is not in paid employment),

$w_i$ is the weight assigned to the $i$-th individual (resulting from the sampling design and potential adjustments),

$n$ is the sample size.

[4] In the HBS, the question about disability refers to whether a person has a certificate of disability (legal status).

The employment rate of mothers who are bringing up children with disabilities is different than that of mothers of children without disabilities. This discrepancy might be both caused by the sole fact of the child's disability (and the resulting specific situation of the mother, e.g. the necessity of providing continuous care to the child), as well as by other distributions of the independent variables (that affect the probability of the mother taking up work) in the two groups of mothers, which is not a direct consequence of the child's disability.

To estimate the share of a given source of variability in the total difference between the rates, we proposed a decomposition analogous to the one put forward independently by Blinder (1973) and Oaxaca (1973). The original proposals of Blinder and Oaxaca involved the decomposition of the difference in mean salaries of two groups of employees (separated on the basis of sex and skin colour), and its goal was to estimate the scale of discrimination against women and persons of colour. Since then, several other applications of this decomposition have been proposed and the methodology was generalised with respect to measures other than the mean. A review of such proposals can be found in Fortin et al. (2011).

While examining differences in the employment rates of mothers of children with disabilities and mothers of children without disabilities, the original decomposition method might be presented as follows:

$$\begin{gathered}\bar{Y}_{non} - \bar{Y}_{dis} = \bar{\boldsymbol{X}}_{non}^T\hat{\boldsymbol{\beta}}_{non} - \bar{\boldsymbol{X}}_{dis}^T\hat{\boldsymbol{\beta}}_{dis} = \\ \bar{\boldsymbol{X}}_{non}^T\hat{\boldsymbol{\beta}}_{non} - \bar{\boldsymbol{X}}_{dis}^T\hat{\boldsymbol{\beta}}_{dis} + \bar{\boldsymbol{X}}_{non}^T\hat{\boldsymbol{\beta}}_{dis} - \bar{\boldsymbol{X}}_{non}^T\hat{\boldsymbol{\beta}}_{dis} = \\ \left(\bar{\boldsymbol{X}}_{non}^T - \bar{\boldsymbol{X}}_{dis}^T\right)\hat{\boldsymbol{\beta}}_{dis} + \bar{\boldsymbol{X}}_{non}^T\left(\hat{\boldsymbol{\beta}}_{non} - \hat{\boldsymbol{\beta}}_{dis}\right),\end{gathered} \tag{2}$$

where:

$\bar{Y}_{non}, \bar{Y}_{dis}$ are the employment rates in the sample of mothers of children without disabilities and mothers of children with disabilities, respectively,

$\bar{\boldsymbol{X}}_{non}^T, \bar{\boldsymbol{X}}_{dis}^T$ are the vectors of mean values of independent variables in the sample of mothers of children without disabilities and mothers of children with disabilities, respectively (one of these variables is in the form of 1 for every individual to account for the constant term in the linear model),

$\hat{\boldsymbol{\beta}}_{non}, \hat{\boldsymbol{\beta}}_{dis}$ are the vectors of the estimated linear regression coefficients in the sample of mothers of children without disabilities and mothers of children with disabilities, respectively.

In the last row of equation (2), the first element $\left(\bar{\boldsymbol{X}}_{non}^T - \bar{\boldsymbol{X}}_{dis}^T\right)\hat{\boldsymbol{\beta}}_{dis}$ means the difference between the employment rates which results from different values of independent variables in the compared groups in the model adopted for the dependent variable. This is the part explained by different structures of the two populations which is called the composition effect or the endowment effect. The second part of the last row of equation (2), $\overline{\boldsymbol{X}}_{non}^{T}(\widehat{\boldsymbol{\beta}}_{non}-\widehat{\boldsymbol{\beta}}_{dis})$, means the difference between the employment rates which is solely the consequence of the fact of belonging to the first or the second group (and of the resulting different probabilities of employment, dependent on the variants of independent variables) and is called, among other names, the structure effect or the unexplained residual. In the original works of Blinder and Oaxaca, it was the measure of wage discrimination. However, it must be remembered that the size of the unexplained residual is affected both by the fact of belonging to a specific group (here: women bringing up children with or without a disability) and by differences in the values of independent variables, which were not included in the model.

Additionally, attention should be paid to the value of $\overline{\boldsymbol{X}}_{non}^{T}\widehat{\boldsymbol{\beta}}_{dis}$, which is the counterfactual employment rate of mothers of children with disabilities, i.e. the employment rate that would be achieved by those mothers if they had the same average values of independent variables as the mothers of children without disabilities.

In our study, the decomposition refers to differences in fractions, and a fraction is also a mean. Thus the application of the original proposals of Blinder and Oaxaca were possible. However, we did two modifications for the purpose of the analysis. First, we included the weights resulting from the fact that the data derive from a random sample (where the probability of individuals getting into a sample was different for each individual) in a similar way as Anastasiade and Tillé (2017). The second modification is the consequence of the fact that the studied variable is binary and thus the use of a linear model might not be appropriate. Because of that, the model of logistic regression was applied,[5] while the decomposition was made in the way similar to what Fairlie (2005) proposed:

$$\hat{\hat{Y}}_{non}-\hat{\hat{Y}}_{dis}=\left[\frac{\sum_{i=1}^{n_{non}}F\left(\boldsymbol{X}_{non,i}^{T}\,\widehat{\boldsymbol{\beta}}_{dis}^{(w)}\right)w_i}{\sum_{i=1}^{n_{non}}w_i}-\frac{\sum_{i=1}^{n_{dis}}F\left(\boldsymbol{X}_{dis,i}^{T}\,\widehat{\boldsymbol{\beta}}_{dis}^{(w)}\right)w_i}{\sum_{i=1}^{n_{dis}}w_i}\right]+ \\ +\left[\frac{\sum_{i=1}^{n_{non}}F\left(\boldsymbol{X}_{non,i}^{T}\,\widehat{\boldsymbol{\beta}}_{non}^{(w)}\right)w_i}{\sum_{i=1}^{n_{non}}w_i}-\frac{\sum_{i=1}^{n_{non}}F\left(\boldsymbol{X}_{non,i}^{T}\,\widehat{\boldsymbol{\beta}}_{dis}^{(w)}\right)w_i}{\sum_{i=1}^{n_{non}}w_i}\right], \tag{3}$$

where:

$\hat{\hat{Y}}_{non}, \hat{\hat{Y}}_{dis}$ are the estimates of the employment rate according to formula (1) for the population of mothers of children without disabilities and mothers of children with disabilities, respectively,

[5] For the sake of comparison, the entire analysis was performed by means of a linear model, and the results were very similar.

$\boldsymbol{X}_{non,i}^{T}, \boldsymbol{X}_{dis,i}^{T}$ are the vectors of the values of independent variables for the $i$-th individual in the sample of mothers of children without disabilities and mothers of children with disabilities, respectively,

$\widehat{\boldsymbol{\beta}}_{non}^{(w)}, \widehat{\boldsymbol{\beta}}_{dis}^{(w)}$ are the vectors of the estimated weighted logistic regression coefficients in the sample of mothers of children without disabilities and mothers of children with disabilities, respectively,

$n_{non}, n_{dis}$ are the sizes of samples of mothers of children without disabilities and mothers of children with disabilities, respectively, and

$F$ is the cumulative distribution function of the logistic distribution.

In the above equation, the summing is made for the individuals that belong to a given group, which is indicated by the summation limit ($n_{non}$ or $n_{dis}$). The counterfactual employment rate of mothers with children with disabilities is, in this case, calculated as $\frac{\sum_{i=1}^{n_{non}} F\left(\boldsymbol{X}_{non,i}^{T} \widehat{\boldsymbol{\beta}}_{dis}^{(w)}\right) w_i}{\sum_{i=1}^{n_{non}} w_i}$, i.e. it is the weighted average of the estimated probabilities of taking up paid work for a sample of mothers of children without disabilities with the use of the dependences observed in the sample of mothers of children with disabilities.

An additional analysis in the study involves the direct use of the logistic regression model (Harrell, 2015) to model the probability of the mother taking up paid work. In this case, given the relatively small size of the sample of mothers of children with disabilities, two approaches were adopted. The first one consisted in estimating two models using pooled data from all the studied years, one for mothers of children with disabilities and one for mothers of children without disabilities. The second involved estimating the models separately for each year but combining the set of mothers of children with disabilities with the set of mothers of children without disabilities, and adding an explanatory dummy variable indicating whether the particular mother's child has a disability. The weights resulting from the sampling design and subsequent calibration were also used in the models.

Some comparisons of the distributions of explanatory variables were made using the overlap coefficient $Z$ as a slightly modified version of Weitzman's measure (Weitzman, 1970). The formula for the coefficient for variable $p$ reads:

$$Z_p = \frac{\sum_k \min\left(f_{p,non,k}, f_{p,dis,k}\right)}{\sum_k \max\left(f_{p,non,k}, f_{p,dis,k}\right)}, \tag{4}$$

where:

$f_{p,non,k}$ is the proportion of mothers of children without disabilities with variant $k$ of variable $p$, and

$f_{p,dis,k}$ is the proportion of mothers of children with disabilities with variant $k$ of variable $p$.

The coefficient measures the similarity of (discrete) distributions of a variable between two groups (here: mothers of children with disabilities and mothers of children without disabilities). The coefficient is constrained to the range $[0, 1]$. $Z = 0$ means that the distributions are completely different, while $Z = 1$ indicates that the distributions are exactly the same.

## 3. Employment rate of mothers in the studied period

The estimated[6] number of mothers meeting the criteria presented above averaged 5.08 million, including 231,000 mothers who brought up children with disabilities. In the sample from the HBS, the number of mothers was systematically decreasing. In 2006 (the year with the largest number of mothers taking part in the HBS), there were 15,623 of them, and in 2020 only 10,675 (in 2020, the sample in the HBS was generally smaller than expected, yet the number of mothers with children studied in that year did not differ significantly from the negative trend observed in preceding years). These numbers also include mothers of children with disabilities, whose number decreased as well (670 on average between 2005 and 2007, 550 on average between 2017 and 2019, and only 407 in 2020).

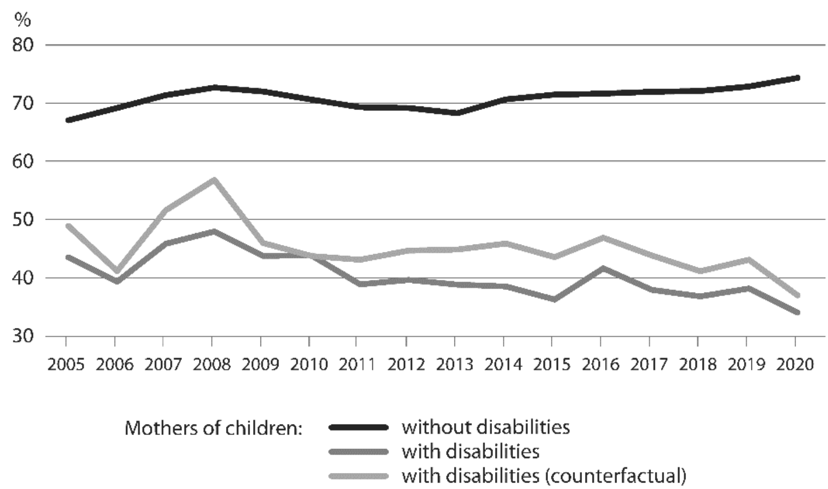
The employment rate of mothers was calculated as the estimated number of working mothers divided by the estimated number of mothers in the study population. There were two occurences of a change in the trend of the employment rate of mothers of children without disabilities in the studied period (cf. Figure). Until 2008, their employment rate was increasing (from 67.1% in 2005 to 72.7% in 2008). Subsequently, it started decreasing, and dropped to 68.3% in 2013. Later, it returned to an upward trend, in 2020 reaching 74.3%. Changes in the employment rate of mothers bringing up children in this period were analogous to the changes in the employment rate of women in general in the 25–34 and 35–44 age groups, while they differed slightly from the changes in the employment rate in general, in the case of which no significant drop after the outbreak of the crisis of 2008 was observed (GUS, 2022).

A different dynamic of changes throughout the studied period could be observed in the case of mothers bringing up children with disabilities. Disregarding any

[6] On the basis of the weights assigned to the study units in the HBS, it is possible to estimate the size of the studied population. Such weights are calculated by Statistics Poland and result both from the fact that a sample was selected by a composite sampling scheme (design weights), as well as from additional adjustments that make use of auxiliary information, aimed at reducing the sampling error and some non-random errors (calibration weights; GUS, 2018).

greater irregularities in the rate values, which result from a larger sampling error caused by a much smaller size of the sample, the overall tendency shows a definite decrease. In the best years of 2007 and 2008, the employment rate in this group of mothers exceeded 45%, while after 2011 it remained below 40%, with the exception of 2016. As a consequence, the difference in the employment rate of mothers of children with disabilities and mothers of children without disabilities, already significant at the beginning of the study period, widened even more over its course. Before 2010, this difference did not exceed 30 p.p.; between 2011 and 2016 it was approximately 30 p.p., and from 2017 to 2020 it exceeded 30 p.p. (in the latter year, it grew to over 40 p.p., but it was a specific year due to the pandemic, as mentioned above).

**Figure.** Employment rate of mothers, depending on whether their child has / does not have a disability




Source: authors' calculation based on the HBS data.

It should be mentioned that among mothers in paid employment, the percentage of those who worked part-time was gradually decreasing. As regards mothers of children without disabilities, 15.2% of them worked part-time in 2005, and only 7% in 2020. The same negative trend occurred among mothers of children with disabilities, but their part-time-working percentage was in most cases approximately 5–10 p.p. higher than that of mothers of children without disabilities (in 2005–2007, 22.6% on average, and in 2018–2020, 14.1%).

The employment rate differs depending on the characteristics of the mother and her overall life situation (independent variables). Changes observed over time might result both from the changes in the frequency of taking up work in various groups of mothers and from changes in the composition of the population of mothers. The greatest changes in the distribution of explanatory variables in the sample of all mothers throughout the study period were observed as regards the education level. Throughout the study period, the share of mothers with tertiary education grew sharply (from 15.2% in 2005 to 44.8% in 2020), which caused changes in the numbers of mothers with lower levels of education. The proportion of mothers with basic vocational education decreased in the studied period (from 32.7% in 2005 to 17.5% in 2020), as did the proportion of mothers with upper secondary education (from 38.5% in 2005 to 33.7% in 2020), and lower secondary or below education (from 12.7% in 2005 to 3.9% in 2020).

In the case of the mother's age, the share of mothers in the [35, 45] age range was continuously growing (from 38.7% in 2005 to 48.3% in 2020) at the expense of all other age ranges. As regards the age of the youngest child, an increase in the share of youngest children was observed (in the [0, 3) and [3, 6) age ranges) and a drop in the share of the oldest children (in the [12, 15) and [15, 18] ranges). For every age range, these were changes of several percentage points throughout the entire period studied.

With respect to the number of children, a continuous decrease in the number of mothers with three or more children was observed (the share of such mothers dropped from 17.3% in 2005 to 12.8% in 2020), whereas at the same time the number of mothers with two children increased (from 35.8% in 2005 to 40.5% in 2020). Similar tendencies occured in the case of the number of other adult persons in the household, apart from the mother. The share of households where there were additionally two or three or more other adults decreased, while at the same time the share of households where there was only one other adult person increased. In each case, the changes were regular, yet not very strong.

When comparing the distributions of independent variables among households of mothers of children with disabilities and households of mothers of children without disabilities, the greatest differences could be observed with respect to the number of children and the mothers' education level. The average overlap coefficients for these characteristics stood at 0.676 and 0.750, respectively. The difference in the number of children consisted in the fact that among mothers of children with disabilities, the share of mothers with three and more children was higher, while the share of mothers with one child was lower. These disparities came to be reduced over the years: at the beginning of the study period, they totalled 20 p.p., whereas at the end of it, they decreased to 15 p.p. As far as differences in the level of education are concerned, among mothers of children with disabilities, the share of those with tertiary education was smaller than the share of persons with tertiary education among mothers of children without disabilities. On the other hand, the proportion of persons with basic vocational or lower education among mothers of children with disabilities was larger than this proportion among mothers of children without disabilities (the difference in the 'tertiary' category in the majority of years exceeds 10 p.p.).

## 4. Decomposition of differences in the employment rate

In general, the differences in the distributions of independent variables between the two groups of mothers resulted in the observable employment rate among mothers of children with disabilities lower almost every year (except 2010) than the counterfactual index, i.e. the rate that would have been observed if the average levels of independent characteristics had been the same as for mothers of children without disabilities (cf. Figure). This means that the mothers of children with disabilities more often had the attributes associated with lower employment rates (e.g. lower level of education). The detailed results of the decomposition carried out in line with equation (3) are presented in Table 1. In the logistic regression models, explanatory variables listed earlier in the paper were used, except for the 'spouse (partner) in the household', which turned out to be statistically insignificant for both groups of mothers.

**Table 1.** Detailed results of the decomposition of differences in the employment rate

| Year | $\hat{\hat{Y}}_{non}$ | $\hat{\hat{Y}}_{dis}$ | $\hat{\hat{Y}}_{non} - \hat{\hat{Y}}_{dis}$ | Composition effect | Unexplained residual | $\hat{\hat{Y}}_{dis}$ (counterfactual) |
|---|---|---|---|---|---|---|
| 2005 | 0.671 | 0.435 | 0.236 | 0.054 | 0.182 | 0.489 |
| 2006 | 0.692 | 0.393 | 0.299 | 0.019 | 0.280 | 0.412 |
| 2007 | 0.714 | 0.458 | 0.255 | 0.057 | 0.198 | 0.516 |
| 2008 | 0.727 | 0.480 | 0.247 | 0.088 | 0.159 | 0.568 |
| 2009 | 0.720 | 0.438 | 0.282 | 0.022 | 0.260 | 0.460 |
| 2010 | 0.706 | 0.438 | 0.268 | −0.001 | 0.269 | 0.437 |
| 2011 | 0.693 | 0.389 | 0.304 | 0.042 | 0.262 | 0.431 |
| 2012 | 0.692 | 0.396 | 0.295 | 0.050 | 0.245 | 0.447 |
| 2013 | 0.683 | 0.388 | 0.295 | 0.060 | 0.234 | 0.448 |
| 2014 | 0.707 | 0.385 | 0.321 | 0.074 | 0.248 | 0.459 |
| 2015 | 0.715 | 0.362 | 0.352 | 0.073 | 0.279 | 0.435 |
| 2016 | 0.716 | 0.416 | 0.300 | 0.052 | 0.248 | 0.469 |
| 2017 | 0.719 | 0.379 | 0.340 | 0.059 | 0.281 | 0.438 |
| 2018 | 0.721 | 0.368 | 0.353 | 0.043 | 0.310 | 0.411 |
| 2019 | 0.729 | 0.382 | 0.347 | 0.049 | 0.298 | 0.431 |
| 2020 | 0.743 | 0.340 | 0.403 | 0.030 | 0.374 | 0.370 |
| Overall | 0.709 | 0.402 | 0.307 | 0.043 | 0.264 | 0.445 |

Note. $\hat{\hat{Y}}_{non}$ and $\hat{\hat{Y}}_{dis}$ are the estimates of the employment rate according to formula (1) for the population of mothers of children without disabilities and mothers of children with disabilities, respectively.

Source: authors' calculation based on the HBS data.

When analysing all the years together, the employment rate of mothers of children without disabilities was 70.9%, and that of mothers of children with disabilities 40.2%. The difference (30.7 p.p.) is predominantly the result of having a child with a disability (and potentially also the consequence of other explanatory variables, which were not included in the model), which altogether accounts for 26.4

p.p., i.e. 85.8% of the difference. The remaining 4.4 p.p., i.e. 14.2% of the difference, results from different distributions of explanatory variables. We can therefore see that the differences in the distributions of independent variables are responsible for the difference in the employment rates between the analysed groups of mothers only to a relatively small degree. It may thus be concluded that the much lower employment rate of mothers of children with disabilities is in the largest part caused directly by the child's disability and the resulting factors (see the 'Discussion' section).

## 5. Probability of the mother taking up paid work

With a view to assessing the type of impact of the socio-demographic characteristics of the mother and her environment on the probability of taking up paid work, the parameters of the logistic regression model were assessed, separately for mothers of children with disabilities and mothers of children without disabilities. Given the small size of the sample of mothers of children with disabilities, the results of the estimation of parameters in individual years were considerably varied. On average, however, no significant changes were observed in the study period in either group of mothers (apart from some exceptions, discussed below), so only the results for the models performed on the data for all the years together are presented in this work. Table 2 contains the evaluation of the impact of every variant of the independent variable on the probability of the mother taking up paid work. To illustrate this impact more clearly, it was shown in the form of points, as used in credit scoring (Jackowska & Wycinka, 2011; Thomas et al., 2002). The estimates of the parameters were transformed into points in such a way that their range fell between 0 and 100. The more points scored by a mother with a given attribute, the more probable it was that she would take up paid work. This way, the impacts of individual variants became easy to compare.

**Table 2.** Points for the characteristics in the logistic regression models for the probability of the mother taking up paid work

| Variable | Variant | Points (and standard errors in brackets) in the model for mothers of children | |
|---|---|---|---|
| | | without disabilities | with disabilities |
| Education level | lower secondary or below | 0 (1.3) | 0 (6.0) |
| | basic vocational | 24 (1.1) | 16 (5.4) |
| | upper secondary | 47 (1.0) | 40 (5.2) |
| | tertiary | 100 (1.0) | 100 (5.2) |
| Age of the mother in years | [15, 25) | 0 (1.2) | 0 (10.7) |
| | [25, 35) | 28 (0.8) | 23 (4.7) |
| | [35, 45) | 41 (1.0) | 41 (5.2) |
| | [45, 55) | 26 (1.2) | 35 (5.9) |
| | [55+) | 10 (2.3) | 26 (9.7) |

**Table 2.** Points for the characteristics in the logistic regression models for the probability of the mother taking up paid work (cont.)

| Variable | Variant | Points (and standard errors in brackets) in the model for mothers of children | |
|---|---|---|---|
| | | without disabilities | with disabilities |
| Age of the youngest child in years | [0, 3) | 0 (1.0) | 0 (5.2) |
| | [3, 6) | 31 (1.0) | 18 (4.7) |
| | [6, 9) | 49 (1.0) | 31 (4.5) |
| | [9, 12) | 59 (1.0) | 34 (4.4) |
| | [12, 15) | 67 (1.1) | 52 (4.7) |
| | [15, 18] | 70 (1.0) | 55 (4.6) |
| Class of locality | rural | 13 (1.0) | 30 (5.2) |
| | urban: ≥20k | 0 (1.1) | 0 (5.7) |
| | 20– 99k | 1 (1.1) | 8 (5.3) |
| | 100–199k | 1 (1.2) | 11 (6.2) |
| | 200–499k | 4 (1.2) | 9 (6.1) |
| | >500k | 18 (1.2) | 7 (6.2) |
| Number of children | 1 | 17 (1.0) | 3 (5.2) |
| | 2 | 10 (0.9) | 9 (4.6) |
| | 3+ | 0 (0.9) | 0 (4.2) |
| Number of other adults in the household | 0 | 13 (1.6) | 0 (6.5) |
| | 1 | 0 (1.0) | 6 (5.2) |
| | 2 | 2 (1.1) | 6 (5.7) |
| | 3+ | 2 (1.1) | 8 (5.8) |

Source: authors' calculation based on the HBS data.

The variable which had the greatest impact on the probability of the mother taking up paid work, both in the case of mothers of children without disabilities and mothers of children with disabilities, was the mother's education level. Only in the case of this variable the maximum of points (i.e. 100) was possible to achieve – for mothers with tertiary education. The basic vocational and upper secondary education in the case of mothers of children with disabilities translated into fewer points and thus increased their chances to take up paid work to a lesser degree (in comparison to the lower secondary or below education) than it was the case in the case of mothers of children without disabilities.

The age of the youngest child, irrespective of whether or not he/she had a disability, proved significant for the mother's economic activity. The dependence between the probability of taking up paid work by the mother and the age of her youngest child was positive: the older the child, the greater the probability that the mother would take up paid work (the points for subsequent age ranges of the child increase in a strictly monotonic way).

As regards the mother's age, the youngest mothers in the [15, 25) age range had the lowest chance for paid work, whereas those in the [35, 45) age range had the highest. When comparing the impact of this characteristic on our two groups of mothers, it can be noted that the mothers of children with disabilities have a slightly lower chance of taking up work at a younger age [25, 35) than the mothers of children without disabilities, but a significantly higher chance when they are aged 45 and older.

The remaining variables in the model have a significantly lower impact on the probability of taking up paid work, which is reflected in the scope of points which might be received for the extreme variants of a given variable. However, all the explanatory variables are statistically significant, based on Wald $\chi^2$ statistics, with the exception of the 'number of other adults in the household' for the mothers of children with disabilities. Additionally, Table 3 presents the quality measures of the models, which indicate that both models in total have a statistically significant impact on the dependent variable, but the degree of their fit to data is moderate.

**Table 3.** Quality measures of logistic regression models for the probability of the mother taking up paid work

| Specification | Model for mothers of children | |
|---|---|---|
| | without disabilities | with disabilities |
| Number of observations | 203,465 | 9,391 |
| Nagelkerke $R^2$ | 0.20 | 0.16 |
| Area under the curve (AUC) | 0.73 | 0.69 |
| Likelihood ratio test $\chi^2_{v=22}$ | 30,894.628<br>$p < 0.0001$ | 1,150.148<br>$p < 0.0001$ |

Source: authors' calculation based on the HBS data.

To demonstrate the changes in time in the impact of individual independent variables, separate models, analogous to the ones presented above, were prepared from samples from every year. However, because of a small size of the sample of mothers of children with disabilities, only one model for each year was made per the sample in total, where a binary explanatory variable was added, indicating whether the mother was raising a child with or without a disability. When interpreting the results of these models using the same convention of points for individual variants of variables (range 0–100), it can be seen that, in the majority of cases, the individual variants of variables retained their relative significance. Certain fluctuations occurred, but they did not initiate longer trends.

Nevertheless, there were two exceptions to this rule. The first was the binary characteristic indicating if the mother had or did not have a child with a disability. The range of points for the variants of this variable, and thus its significance, increased over the years. Each year within the studied period, bringing up a child with a disability added 0 points, while bringing up a child without a disability contributed a growing number of points (36 points in 2005 and 76 points in 2020). As regards the odds ratio, the values amounted to 2.45 in 2005 and 5.46 in 2020. This means that the odds of paid work for a mother of a child without a disability was

2.54 times higher in 2005 and 5.46 times higher in 2020 than the odds of paid work for a mother of a child with a disability in the same period.

The second exception was the 'class of locality' variable and its 'rural' variant. The chance of getting paid work for the residents of rural areas dropped sharply in the study period. Between 2005 and 2007, this variant of the place of residence was able to contribute 27–28 points (most from all variants of this variable), while at the end of the study period, between 2019 and 2020, it became the least advantageous variant for which no points could be received. Therefore, even though the score for the 'rural' variant, presented in Table 2, shows its relatively positive impact on the mother's chance of taking up paid work, it has to be remembered that this is the averaged value for all years. At the same time, we have to bear in mind that the place of residence was one of the least significant variables in the model.

## 6. Discussion

A positive diagnosis of a child's disability brings changes of the emotional, professional and financial character to the entire life of his or her family. The additional obligations related to the child's rehabilitation, treatment and care resurface. It is often the case that the expenses on the medications and supplementary classes or treatment supporting the child's development are higher than those borne by the parents of children without disabilities (Komorowska, 2017b). Often, one of the carers – most often the mother[7] – either gives up her professional activity entirely or limits it for the sake of caring for the child (Komorowska, 2017b; Komorowska et al., 2019). In the case of children with intellectual disability, many mothers leave the labour market for an extended period of time, sometimes permanently (Chou et al., 2018; Gomez Mandic et al., 2017; Komorowska & Kozłowski, 2021). Household chores, caring for a child with a disability and paid work become hard to reconcile (Einam & Cuskelly, 2002; Morris, 2014; Working Families and Unum, 2018).

In the case of mothers who bring up children with disabilities, there are also additional factors that affect their economic activity, such as: the child's age and his/her condition, the mother's mental and emotional condition, assistance or its lack in caring for and rehabilitating the child (including the availability of institutional care for the child), the mother's own perception of her role in the workplace, her education and the amount of allowances received from the state (Banaszkiewicz et al., 2019; Brown & Clark, 2017; Gogoi et al., 2016; Komorowska, 2017a; Komorowska & Kozłowski, 2021; Sapkota et al., 2017).

[7] In Poland, it is predominantly women who take up the majority of obligations related to childcare. Women usually have to reconcile various roles and tasks in the family. They become the 'managers of disability'.

The nursing benefit is one of the allowances that might discourage the economic activity of mothers bringing up children with disabilities. It is because in order to get it, the carer has to give up their economic activity completely for the sake of taking care of the child. In 2009, the income criterion related to the eligibility for this allowance was waived, and the amount of the allowance grew significantly later in the study period. In 2010, it was PLN 520, in 2014 PLN 1,000, and in 2020 PLN 1,830 (cf. Table 4). Since 2016, families have also been receiving a child benefit, PLN 500 per child, in the framework of the 'Family 500 Plus' programme which, in combination with the nursing benefit, generated the amount of fixed income which was relatively attractive compared to the potential wages, especially in the case of carers with lower professional qualifications or lower education level.[8]

**Table 4.** Amount of the nursing benefit and the minimum wage in Poland

| Year | Minimum gross wage | Nursing benefit |
|---|---|---|
| | in PLN | |
| 2009 | 1,126 | 420[a]; 520 |
| 2010 | 1,317 | 520[b] |
| 2011 | 1,386 | 520; 620[c] |
| 2012 | 1,500 | 620 |
| 2013 | 1,600 | 620[d]; 720[e]; 820[f] |
| 2014 | 1,680 | 820; 1,000[g] |
| 2015 | 1,750 | 1,200 |
| 2016 | 1,850 | 1,300[h] |
| 2017 | 2,000 | 1,406[i] |
| 2018 | 2,100 | 1,477 |
| 2019 | 2,250 | 1,583[j] |
| 2020 | 2,600 | 1,830 |

a The amount of allowance from 1st May 2004 to 31st Oct. 2009. b The waiver of the income criterion with respect to the eligibility for the nursing benefit. c As of 1st Nov. 2011, the nursing benefit in the amount of PLN 520.00 increased by an extra benefit of PLN 100.00. d Protests of parents in the Sejm aimed at urging the authorities to increase the amount of the nursing benefit. e Between 1st April 2013 and 30th June 2013, the nursing benefit consisted of the base benefit (PLN 520.00) plus an extra benefit (PLN 200.00). f Between 1st July 2013 and 30th April 2014, the nursing benefit consisted of the base benefit (PLN 620.00) plus an extra benefit (PLN 200.00). g Between 1st May 2014 and 31st Dec. 2014, the nursing benefit consisted of the base benefit (PLN 800.00) plus an extra benefit of PLN 200.00. h The 'Family 500 Plus' programme for the second and subsequent child without an income threshold, and for the first child with a higher income threshold when there is a child with a disability in the family. i The introduction of the mechanism of the annual valuation of the nursing benefit by the percentage of the increase in the minimum salary. j The 'Family 500 Plus' programme for the first child without an income threshold.

Source: authors' study based on Infor (2022).

[8] In 2020, the minimum net salary in Poland was PLN 1,920. A carer of a disabled child who did not work in order to take full care of the child received the child benefit in the framework of the 'Family 500 Plus' programme per each child plus the nursing benefit in the amount of PLN 1,830, so in total PLN 2,330 net (in the variant of having just one child), which was higher than the minimum salary by PLN 410. In addition, carers of children with a disability were also often entitled to the 'attendance allowance' of PLN 215.84.

These benefits were not included as variables in the analyses. However, the increase in the amount of benefits together with a decrease in the employment rate of mothers bringing up children with disabilities (also mothers of children without disabilities with lower levels of education, which is often associated with lower earnings) indicates a correlation between the state policy in the field of social benefits and the economic activity of mothers. Further research is necessary to establish the causal relationship between the level of allowances and the employment rate.

In recent years, changes aimed at supporting parenthood in Poland have been introduced to the Polish labour law,[9] e.g. the parental leave, by the amendment from 2013.[10] Over the years 2022–2023, thanks to two European Union directives,[11] further changes to the labour law were introduced, the purpose of which was to foster work and family life balance (with a special focus on employees who are parents or guardians of children). These changes included parental leave extension by 9 weeks, to 41 weeks, when one child is born, and to 43 weeks in the case of a multiple pregnancy, as well as an increase to 65 weeks in the case of the birth of a child with a severe or irreversible disability or an incurable disease (67 weeks in the case of a multiple pregnancy), a care leave of 5 days a year, and the possibility of remote work. The provisions mentioned above and the proposals of changes are characterised by parents as helpful and are generally appreciated. They might also boost the economic activity of mothers of children with disabilities.

## 7. Conclusions

Between 2005 and 2020, an upward trend in the employment rate of mothers of children without disabilities prevailed. This was mainly the consequence of a rise in the proportion of mothers with tertiary education in the Polish society and the increasing trend of having children at a later age. In the same period, on the other hand, there was a downward trend in the employment rate of mothers of children with disabilities. The analysis demonstrates that the gap between the employment rates of these two groups of mothers, large and continuously widening, mainly results from the sole fact of the child's disability and its direct consequences (and above all, from the fact that more time and resources are needed to care for

---

[9] Act of 26th June 1974 – the Labour Code (Pol. Ustawa z dnia 26 czerwca 1974 r. – Kodeks pracy).

[10] Act of 28th May 2013 amending the Labour Code and some other laws (Pol. Ustawa z dnia 28 maja 2013 r. o zmianie ustawy – Kodeks pracy oraz niektórych innych ustaw).

[11] Directive (EU) 2019/1158 of the European Parliament and of the Council of 20th June 2019 on work-life balance for parents and carers and repealing Council Directive 2010/18/E and Directive (EU) 2019/1152 of the European Parliament and of the Council of 20th June 2019 on transparent and predictable working conditions in the European Union.

a child with a disability than for a child without such a condition). One of the potential contributors to the lack of increase in the scale of the economic activity of mothers of children with disabilities is the principle of granting the nursing benefit, which requires the mother taking care of a child with a disability to completely give up her economic activity.

Out of all explanatory variables considered in the analysis, education level had the strongest impact on the probability of a mother taking up paid employment, with tertiary education offering the highest probability of such an occurence, and lower education levels yielding monotonically lower probabilities. The age of the youngest child turned out to be another variable with strong predictive power – the older the child, the more probable it was that the mother would be employed. The result which came as a slight surprise was the lack of any considerable impact of the variables connected with the number of children, class of locality, and the number of other adults in households.

It is necessary to mention that the HBS data do not make it possible to determine whether economically inactive mothers would take up paid employment if they came across such a possibility. These are not panel data, and hence there is no way of checking how the employment status of a specific woman changes over time. The HBS does not provide information about the type of the child's disability, which is informative in the context of the amount of time that has to be assigned to the care and rehabilitation of such a child.

In the situation where the employee market in Poland is shrinking, it seems worthwhile to seriously consider including the economically inactive mothers of children with disabilities in the labour market. Not only changes in legislation (e.g. in the provisions of granting the nursing benefit) are required, but also a more extensive support system for mothers of children with disabilities, in the form of e.g. increased availability of institutional care and psychological assistance. The latter need results from the fact that mothers of children with disabilities often see themselves as full-time carers for their children, and devote themselves fully to this role.

## Bibliography

Amah, O. E. (2021). Managing the negative effects of work-to-family and family-to-work conflicts on family satisfaction of working mothers' in Nigeria: the role of extended family support. *Community, Work & Family*, *24*(3), 257–271. https://doi.org/10.1080/13668803.2019.1697646.

Anastasiade, M.-C., & Tillé, Y. (2017). Decomposition of gender wage inequalities through calibration: Application to the Swiss structure of earnings survey. *Survey Methodology*, *43*(2), 211–234. https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2017002/article/54887-eng.pdf?st=aMgj51fu.

Banaszkiewicz, D., Komorowska, O., & Kozłowski, A. (2019). Różnice w stopie zatrudnienia matek wychowujących dziecko z niepełnosprawnością i bez niepełnosprawności. *Studia Oeconomica Posnaniensia*, *7*(1), 7–24. https://doi.org/10.18559/soep.2019.1.1.

Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*, *8*(4), 436–455. https://doi.org/10.2307/144855.

Borski, M. (2019). Aktywizacja zawodowa opiekunów osób z niepełnosprawnościami jako ważne zadanie państwa – wybrane zagadnienia. *Studia z Zakresu Prawa Pracy i Polityki Społecznej*, *26*(2), 153–165. https://doi.org/10.4467/25444654SPP.19.011.10275.

Bourke-Taylor, H., Howie, L., & Law, M. (2011). Barriers to maternal workforce participation and relationship between paid work and health. *Journal of Intellectual Disability Research*, *55*(5), 511–520. https://doi.org/10.1111/j.1365-2788.2011.01407.x.

Boushey, H. (2008). Opting out? The Effect of Children on Women's Employment in the United States. *Feminist Economics*, *14*(1), 1–36. https://doi.org/10.1080/13545700701716672.

Brown, T. J., & Clark, C. (2017). Employed Parents of Children with Disabilities and Work Family Life Balance: A Literature Review. *Child & Youth Care Forum*, *46*(6), 857–876. https://doi.org/10.1007/s10566-017-9407-0.

Castro-García, C., & Pazos-Moran, M. (2016). Parental Leave Policy and Gender Equality in Europe. *Feminist Economics, 22*(3), 51–73. https://doi.org/10.1080/13545701.2015.1082033.

Chou, Y.-C., Kröger, T., & Pu, C.-Y. (2018). Underemployment among mothers of children with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*, *31*(1), 152–158. https://doi.org/10.1111/jar.12336.

Directive (EU) 2019/1152 of the European Parliament and of the Council of 20th June 2019 on transparent and predictable working conditions in the European Union (O.J. EU L 2019.7.11).

Directive (EU) 2019/1158 of the European Parliament and of the Council of 20th June 2019 on work-life balance for parents and carers and repealing Council Directive 2010/18/E (O.J. EU L 2019.7.12).

Einam, M., & Cuskelly, M. (2002). Paid employment of mothers and fathers of an adult child with multiple disabilities. *Journal of Intellectual Disability Research*, *46*(2), 158–167. https://doi.org/10.1046/j.1365-2788.2002.00383.x.

European Commission. (2017). European semester thematic factsheet women in the labour market. https://commission.europa.eu/system/files/2020-06/european-semester_thematic-factsheet_labour-force-participation-women_en_0.pdf.

Eurostat. (2022). *Employment rates by sex, age and citizenship (%)* [Data set]. Retrieved July 13, 2022, from https://ec.europa.eu/eurostat/databrowser/view/LFSA_ERGAN__custom_242510/default/table?lang=en.

Fairlie, R. W. (2005). An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models. *Journal of Economic and Social Measurement*, *30*(4), 305–316. https://doi.org/10.3233/JEM-2005-0259.

Fortin, N., Lemieux, T., & Firpo, S. (2011). Decomposition Methods in Economics. In O. C. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (vol. 4; pp. 1–102). https://doi.org/10.1016/S0169-7218(11)00407-2.

Gasińska, M. (2016). Dochody gospodarstw domowych w Polsce – wybrane obiektywne i subiektywne ujęcia i dane. *Zeszyty Naukowe Uczelni Vistula,* (5), 100–142. https://i.vistula.edu.pl/media/docs/CA7I9I7z9SWG2f1YDnErQ3LDP.pdf.

Giddens, A. (2008). *Socjologia* (A. Szulżycka, Trans.). Wydawnictwo Naukowe PWN.

Gierusz, A., & Komorowska, O. (2022). Szacunek wysokości przyszłej emerytury matki dziecka z niepełnosprawnością aktywnej zawodowo i emerytury matki pobierającej świadczenie pielęgnacyjne. *Polityka Społeczna*, (7), 10–16. https://doi.org/10.5604/01.3001.0015.9687.

Główny Urząd Statystyczny. (2018). *Zeszyt metodologiczny. Badanie budżetów gospodarstw domowych.* https://stat.gov.pl/obszary-tematyczne/warunki-zycia/dochody-wydatki-i-warunki-zycia-ludnosci/zeszyt-metodologiczny-badanie-budzetow-gospodarstw-domowych,10,2.html.

Główny Urząd Statystyczny. (2020). *Jakość życia osób starszych w Polsce*. https://stat.gov.pl/obszary-tematyczne/warunki-zycia/dochody-wydatki-i-warunki-zycia-ludnosci/jakosc-zycia-osob-starszych-w-polsce,26,2.html.

Główny Urząd Statystyczny. (2021). *Budżety gospodarstw domowych w 2020 roku*. https://stat.gov.pl/obszary-tematyczne/warunki-zycia/dochody-wydatki-i-warunki-zycia-ludnosci/budzety-gospodarstw-domowych-w-2020-roku,9,15.html.

Główny Urząd Statystyczny. (2022). *Aktywność ekonomiczna ludności Polski – 1 kwartał 2022 r.* https://stat.gov.pl/obszary-tematyczne/rynek-pracy/pracujacy-bezrobotni-bierni-zawodowo-wg-bael/aktywnosc-ekonomiczna-ludnosci-polski-i-kwartal-2022-roku,4,46.html.

Gogoi, R. R., Kumar, R., & Pereira Deuri, S. (2016). Anxiety, depression, and quality of life in mothers of children with intellectual disability. *Open Journal of Psychiatry & Allied Sciences*, *8*(1), 71–756. https://doi.org/10.5958/2394-2061.2016.00046.X.

Gomez Mandic, C., Johaningsmeir, S., Corden, T. E., Earle, A., Acevedo-Garcia, D., & Gordon, J. B. (2017). Impact of caring for children with medical complexity on parents' employment and time. *Community, Work & Family*, *20*(4), 444–458. https://doi.org/10.1080/13668803.2016.1202195.

Haasa, L., & Hwang, C. P. (2008). The impact of taking parental leave on fathers' participation in childcare and relationships with children: Lessons from Sweden. *Community, Work & Family*, *11*(1), 85–104. https://doi.org/10.1080/13668800701785346.

Harrell, F. E. Jr. (2015). *Regression modeling strategies. With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis* (2nd ed.). Springer. https://link.springer.com/book/10.1007/978-3-319-19425-7.

Hegewisch, A., & Gornick, J. C. (2011). The impact of work-family policies on women's employment: a review of research from OECD countries. *Community, Work & Family*, *14*(2), 119–138. https://doi.org/10.1080/13668803.2011.571395.

Infor. (2022). *Świadczenia opiekuńcze*. Retrieved October 10, 2022, from https://www.infor.pl/wskazniki/prawo-pracy-i-ubezpieczen-spolecznych/zasilki-i-swiadczenia/80,674,Swiadczenia-opiekuncze.html.

Jackowska, B., & Wycinka, E. (2011). Zastosowanie scoringu do oceny ryzyka ubezpieczeniowego. *Zeszyty Naukowe / Uniwersytet Ekonomiczny w Poznaniu*, (182), 225–235.

Kay, T. (2000). Leisure, Gender and Family: The Influence of Social Policy. *Leisure Studies*, *19*(4), 247–265. https://doi.org/10.1080/02614360050118823.

Klaus, D., & Vogel, C. (2019). Unbezahlte Sorgetätigkeiten von Frauen und Männern im Verlauf der zweiten Lebenshälfte. In C. Vogel, M. Wettstein, & C. Tesch-Römer (Eds.), *Frauen und Männer in der zweiten Lebenshälfte* (pp. 91–112). Springer VS. https://doi.org/10.1007/978-3-658-25079-9_6.

Komorowska, O. (2017a). Czynniki wpływające na aktywność zawodową rodziców wychowujących dzieci z niepełnosprawnością. *Polityka Społeczna*, (3), 27–32. https://polityka-spoleczna.ipiss.com.pl/resources/html/article/details?id=201734.

Komorowska, O. (2017b). Sytuacja dochodowa rodzin wychowujących dziecko z niepełnosprawnością. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (483), 68–79. https://doi.org/10.15611/pn.2017.483.06.

Komorowska, O., & Kozłowski, A. (2021). *Jakość życia rodzin opiekujących się dorosłą osobą z niepełnosprawnością intelektualną*. Oficyna Wydawnicza Impuls.

Komorowska, O., Kozłowski, A., & Słaby, T. (2019). Comparative Analysis of Poverty in Families with a Disabled Child and Families with Non-Disabled Children in Poland in the Years 2014 and 2016. *Statistics in Transition new series*, *20*(3), 97–117. https://doi.org/10.21307/stattrans-2019-026.

Kuchařová, V. (1999). Women and Employment. *Czech Sociological Review, 7*(2), 179–194. https://doi.org/10.13060/00380288.1999.35.12.08.

Kurowska, A. (2012). Modelowanie decyzji o prokreacji i aktywności zawodowej. *Polityka Społeczna*, (10), 1–11. https://polityka-spoleczna.ipiss.com.pl/resources/html/article/details?id=200548.

Misra, J., Budig, M., & Boeckmann, I. (2011). Work-family policies and the effects of children on women's employment hours and wages. *Community, Work & Family*, *14*(2), 139–157. https://doi.org/10.1080/13668803.2011.571396.

Morris, L. A. (2014). The Impact of Work on the Mental Health of Parents of Children with Disabilities. *Family Relations*, *63*(1), 101–121. https://doi.org/10.1111/fare.12050.

Mosseri, S. E. (2021). Finding middle ground: the relationship between cultural schemas and working mothers' work-family strategies. *Community, Work & Family*, *24*(3), 331–356. https://doi.org/10.1080/13668803.2019.1682968.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, *14*(3), 693–709. https://doi.org/10.2307/2525981.

Perry-Jenkins, M., & Gerstel, N. (2020). Work and Family in the Second Decade of the 21st Century. *Journal of Marriage and Family*, *82*(1), 420–453. https://doi.org/10.1111/jomf.12636.

Pollmann-Schult, M. (2015). Familie, Erwerbsarbeit, Einkommen. In P. B. Hill, & J. Kopp (Eds.), *Handbuch Familiensoziologie* (pp. 613–639). Springer VS. https://doi.org/10.1007/978-3-658-02276-1_20.

Pufal-Struzik, I. (2017). Aktywność zawodowa współczesnych kobiet – trudności w realizacji nowych ról i tradycyjnych obowiązków. *Polskie Forum Psychologiczne*, *22*(2), 242–257. https://doi.org/10.14656/pfp20170204.

Sapkota, N., Pandey, A. K., Deo, B. K., & Shrivastava, M. K. (2017). Anxiety, depression and quality of life in mothers of intellectually disabled children. *Journal Psychiatrists' Association of Nepal*, *6*(2), 28–35. https://doi.org/10.3126/jpan.v6i2.21757.

Sarkisian, N., & Gerstel, N. (2004). Explaining the Gender Gap in Help to Parents: The Importance of Employment. *Journal of Marriage and Family*, *66*(2), 431–451. https://doi.org/10.1111/j.1741-3737.2004.00030.x.

Szymczak, A., & Gawrycka, M. (2015). Praca jako dobro indywidualne i społeczne. *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, (214), 276–287.

Szyszka, M. (2016). Aktywność zawodowa w opinii pracujących kobiet. *Annales Universitatis Mariae Curie-Skłodowska Lublin – Polonia*, *41*(2), 103–122. https://doi.org/10.17951/i.2016.41.2.103.

Thomas, L. C., Edelman, D. B., & Crook, J. N. (2002). *Credit Scoring and Its Applications*. Society for Industrial and Applied Mathematics.

Ustawa z dnia 26 czerwca 1974 r. Kodeks pracy (Dz.U. 1974 nr 24, poz. 141).

Ustawa z dnia 28 maja 2013 r. o zmianie ustawy – Kodeks pracy oraz niektórych innych ustaw (Dz.U. 2013 poz. 675).

Weitzman, M. S. (1970). *Measures of Overlap of Income Distributions of White and Negro Families in the United States*. U.S. Bureau of the Census.

Windebank, J. (2012). Social Policy and Gender Divisions of Domestic and Care Work in France. *Modern & Contemporary France, 20*(1), 21–35. https://doi.org/10.1080/09639489.2011.631699.

Working Families and Unum. (2018). *2018 Off Balance – Parents of disabled children and paid work*. https://www.workingfamilies.org.uk/publications/off-balance-2018/.

Wiadomości Statystyczne. The Polish Statistician, 2023, *68*(6), 47–66 https://doi.org/10.59139/ws.2023.06.3
Spisy powszechne – problemy i wyzwania / Issues and challenges in census taking
Zgłoszony/submitted: 4.11.2022, poprawiony/revised: 27.05.2023, zaakceptowany/accepted: 31.05.2023

# Use of new technologies and evidence-based decisions: key factors in the strategy for the 2020 Population and Housing Census in Mexico in the context of the COVID-19 pandemic

Edgar Vielma Orozco[a]

**Abstract.** The development of population and housing censuses implies a challenge regarding the necessary resources, logistics, and operations for National Statistical Offices (NSOs). At the global level, this challenge was more significant in the 2020 census round due to the COVID-19 health contingency. This led Mexico's National Institute of Statistics and Geography (Instituto Nacional de Estadística y Geografía – INEGI) to analyze possible scenarios in the face of the pandemic and rethink some activities in order to adapt to the contingency. As a result of these measures, the 2020 Population and Housing Census in Mexico turned out a success.

The aim of the paper is to share INEGI's solutions implemented before, during, and after the COVID-19 contingency, which favored the development and conclusion of the census in Mexico and which other countries might find useful. Among the key elements of the Mexican census strategy was the incorporation of technologies for data collection, e.g. migrating from paper questionnaires to the use of mobile computing devices (MCDs), which reduced the time of capture and processing of information. In addition, a brief analysis of the behavior of past pandemics facilitated the decision-making. The main lessons learned from the Mexican experience include: the importance of maintaining the generation of official statistics in crisis contexts, the need for NSOs to have a robust risk management system that contemplates all types of scenarios and allows them to act in any contingency, and the need to implement innovative data collection methods and extend the use of Information and Communication Technologies (ICT).

**Keywords:** 2020 Population and Housing Census, COVID-19, strategy, mobile computing devices

**JEL:** J19, O33, D81

[a] National Institute of Statistics and Geography (do 1.02.2023 r. / until 1.02.2023); National Bank of Public Works and Services, S.N.C., Meksyk / Mexico. ORCID: https://orcid.org/0000-0002-0762-9165.
E-mail: edgar.vielma@banobras.gob.mx; vielmov@yahoo.com.mx.

# Wykorzystanie nowych technologii i decyzje oparte na dowodach: kluczowe elementy strategii prowadzenia Spisu Powszechnego Ludności i Mieszkań 2020 w Meksyku w kontekście pandemii COVID-19

**Streszczenie.** Z prowadzeniem spisów powszechnych ludności i mieszkań wiąże się wiele wyzwań, z którymi muszą się mierzyć urzędy statystyczne, zarówno w zakresie logistyki, jak i niezbędnych zasobów i środków. Runda spisów 2020 okazała się szczególnym wyzwaniem z powodu wybuchu pandemii COVID-19. W tej sytuacji meksykański Narodowy Instytut Statystyki i Geografii (Instituto Nacional de Estadística y Geografía – INEGI) przeanalizował różne scenariusze działań i dostosował procedury spisowe do warunków pandemicznych. To pozwoliło na sprawne przeprowadzenie w Meksyku Powszechnego Spisu Ludności i Mieszkań w 2020 r.

Celem artykułu jest przedstawienie rozwiązań wypracowanych przez INEGI i wdrożonych w okresie poprzedzającym pandemię, w jej trakcie oraz po jej ustaniu, dzięki którym badanie zostało z powodzeniem zrealizowane i które mogą okazać się przydatne dla innych krajów. Jednym z kluczowych elementów przyjętej strategii było zastosowanie nowoczesnych technologii gromadzenia danych i rezygnacja z kwestionariuszy papierowych na rzecz mobilnych urządzeń liczących (ang. *mobile computing devices*), co skróciło czas uzyskiwania i przetwarzania informacji. Ponadto analiza przebiegu wcześniejszych pandemii ułatwiła podejmowanie decyzji. Doświadczenia wyniesione ze spisu w Meksyku wskazują m.in. na wielką wagę ciągłości opracowywania oficjalnych statystyk nawet w sytuacjach kryzysowych, potrzebę posiadania przez urząd statystyczny stabilnego systemu zarządzania ryzykiem, który uwzględnia wiele scenariuszy i pozwala działać w każdej sytuacji, oraz konieczność wdrażania innowacyjnych metod gromadzenia danych i szerszego wykorzystania technologii informacyjno-komunikacyjnych.

**Słowa kluczowe:** Spis Powszechny Ludności i Mieszkań 2020, COVID-19, strategia, mobilne urządzenia liczące

## 1. Introduction

At the international level, 53 countries participated in the United Nations (UN) 2020 census round; in many countries, this round of national censuses coincided with the onset of the pandemic due to COVID-19 and, due to health restrictions, only eight countries managed to conduct their census, including Mexico, which was also among the first to publish the census results.

Evidence-based decision-making was crucial for the National Statistical Office (NSO) of Mexico to continue and conclude the census activities planned for 2020. The analysis of statistical and historical information from pandemics such as SARS-CoV-1 was key to estimating the COVID-19 behavior, influencing the decision that the census activities could continue.

Conducting a census in Mexico generally implies a challenge because the entire national territory has to be traveled through and almost 44 million dwellings visited, in addition to mobilizing approximately 200,000 people, including interviewers, supervisors, and operational personnel.

The completion of the 2020 Census was successful due to the incorporation of current technological advances to perform various census processes, specifically, the use of mobile computing devices (MCDs) for the integration of digital applications in the various census processes such as streamlining the data collection and the incorporation of the essential validation criteria simultaneously with the collection of information. For this, it was necessary to develop different applications linked to the census operation and staff training. For the management of the devices and the preservation of the security of information, computer communication systems were structured.

The backup mechanism was paper-based in places where MCDs could not be used. Additionally, two complementary methods of collecting information were established: self-enumeration (through the NSO web page) and telephone-assisted interviewing. The series of efforts and decisions made within the NSO allowed the preliminary results of the 2020 Census to be published on 25th January 2021.

The objective of this article is to share the experiences and strategies implemented before, during, and after the COVID-19 contingency, which favored the development and conclusion of the census, despite adverse circumstances.

## 2. Pandemic: measures taken and guidelines

In Mexico, carrying out a census represents a significant challenge because the survey has to cover 2 million square kilometers of the national territory, and, as mentioned before, almost 44 million dwellings have to be visited as well as approximately 200,000 people (interviewers, supervisors, and operational personnel) have to be mobilized. This means a considerable human, financial, technological, and material investment (Instituto Nacional de Estadística y Geografía [INEGI], 2021b). In most countries, censuses are carried out every 10 years. In Mexico, they are carried out in years ending in zero, and the most recent census coincided with the start of the COVID-19 pandemic (INEGI, 2021a).

The key element of the strategy for the 2020 Census in Mexico, realized through the National Institute of Statistics and Geography (INEGI), was evidence-based decision-making. Statistical and historical information from previous pandemics (such as the 1918 influenza (H1N1) and the SARS-CoV-1 outbreak in 2003) was analyzed, as shown in Figures 1 and 2.

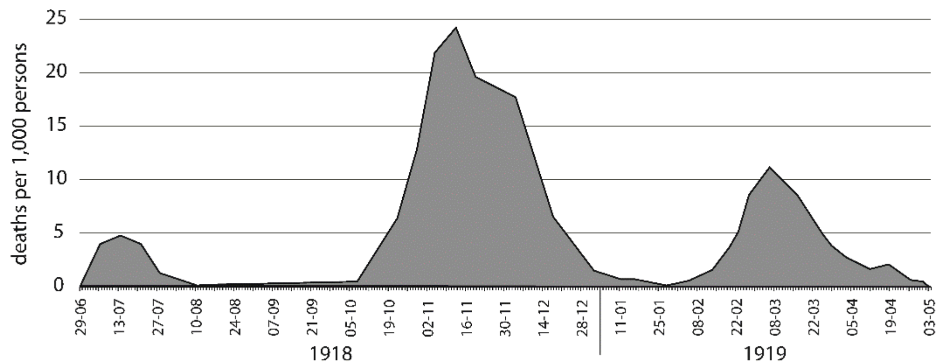
In addition, the knowledge about the 1889–1890 flu pandemic was utilized, even though month by month data on that pandemic's behavior was not available. It was important for the decision-making, however, to remember it occurred and for how long. This was the first pandemic in a fully interconnected world.

The H1N1 virus, during 1918–1920, affected mainly young adults between 20 and 40 years of age; there were about 500 million people infected, and at least 50 million died (Taubenberger & Morens, 2006). The statistical and historical analysis of the H1N1 pandemic shows that the first wave (from March to May 1918) was not very deadly, the second wave (from September to December 1918) was more violent and deadly, the third wave (from February to March 1919) was less outspread than the previous two waves, and the fourth wave (from January to April 1920) had less impact than the preceding ones. The SARS-CoV-1 pandemic showed, on the other hand, how quickly an infection spreads in an interconnected world. However, despite the speed of contagions, just over 8,000 people became ill, and approximately 1,000 died (World Health Organization [WHO], 2003).

Information from previous pandemics was useful for determining that their stages take place according to a wave pattern; that is, infections and deaths do not occur with the same frequency and intensity every day, but there are blocks in which these phenomena increase or decrease. Furthermore, it was possible to identify that the peaks of infections and deaths do not occur during the first days or weeks of the virus outbreak (Vielma Orozco, 2022b). The brief analysis of the existing data of past pandemics, which was conducted in less than a week, helped to estimate the pattern that this new pandemic would have. After this analysis, it was determined that the activities of the 2020 Census could continue as planned. One of the most important reasons why it was possible to collect information was that the census coincided with the early stage of the contingency (named Phase 1 by the health authorities), where infections were still relatively rare and there were no strict measures to follow.

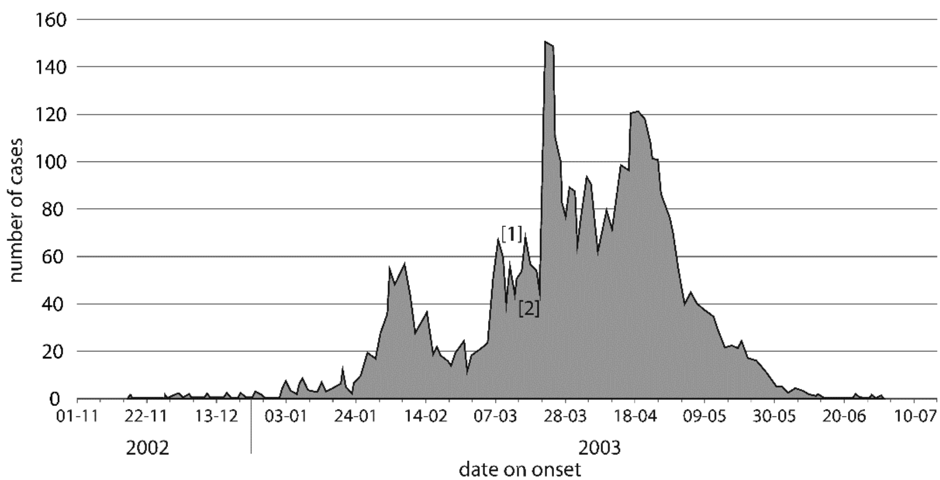
Here it is very important to highlight that it was not the depth of the data analysis, but rather the identification of existing information related to previous pandemics, which allowed the best decisions to be taken in less than a week, and additionally the knowledge that postponing the census for six months or a year would take us to the most critical moment of the pandemic, as it happened to most NSOs in other countries. Also, we anticipated that if we postponed the census, a year later (most of the pandemics analyzed lasted for two years, plus slight additional recurrences), due to the effects of the pandemic, financial and human resources would be limited, as in fact happened during 2021 and 2022, mainly due to the considerable limitation of the economic activity and the purchase of supplies to eradicate the pandemic (vaccines and intensive medical care).

**Figure 1.** Three pandemic waves: influenza and pneumonia mortality combined weekly, United Kingdom




Source: adapted from Taubenberger and Morens (2006).

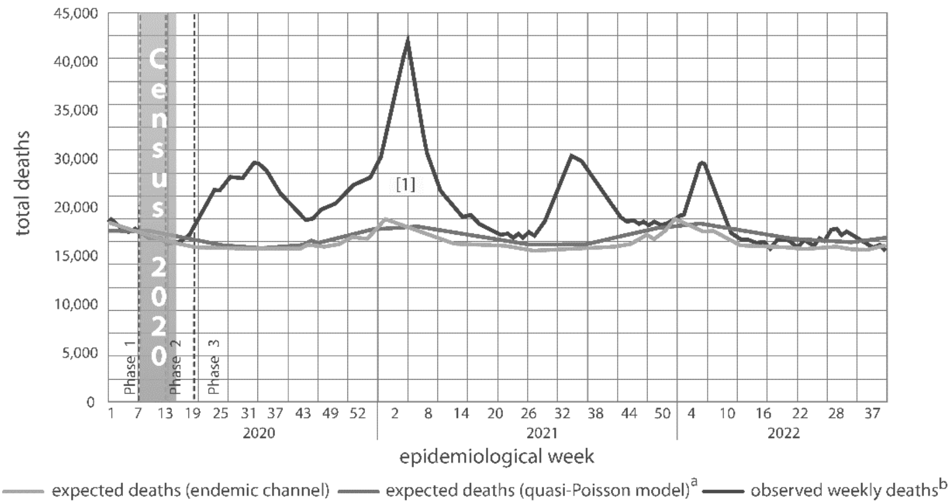
**Figure 2.** Probable SARS cases by week of onset worldwide




Note. $n$ = 5,910. [1] WHO issues global alert: 12th March. [2] WHO issues first travel advisory: 15th March.
Source: adapted from the WHO (2003).

As can be seen in Figure 3, the decision taken by the INEGI regarding the 2020 Census proved to be the right one, since practically at the end of the enumeration in the field, the government indicated the entry to Phase 2 of the pandemic, referring to the beginning of the community transmission of the virus, and weeks later to Phase 3, an epidemiological stage where the virus had already reached thousands of people in various localities (Gobierno de México, 2020a).

**Figure 3.** COVID-19 effect. Excess mortality in Mexico, 2020–2022




a Generalized linear model with the generalized estimating equations (GEE) method was fitted to estimate the expected value of the number of weekly deaths by state, age group, and sex during 2015–2019.
b Preliminary information.
Note. The filled box on the graph indicates the 2020 Census survey period, shown in the context of the weeks and phases of the pandemic. [1] Compared to the "expected deaths" line, the "observed weekly deaths" line establishes the excess mortality during the COVID-19 pandemic.
Source: INEGI (2023).

It should be noted that, at the international level, 53 countries made up the 2020 census round of the United Nations (UN). Still, due to health restrictions, only eight states managed to carry out their censuses, including Mexico, which was also one of the first to publish its results (INEGI, 2021a).

The complexity that characterizes censuses requires several years of planning, and in the case of the 2020 Census in Mexico, the preparations began three years prior to the survey, as seen in Figure 4. In 2017, a public consultation was held to define the theme and methodology of the questionnaires. Then, in 2018 and 2019, different preparation activities were carried out, such as the pilot test, personnel recruitment, and the start of the communication campaign.

**Figure 4.** Development of the 2020 Census activities

| 2017 | 2018 | 2019 | 2020 |
|---|---|---|---|
| • Planning of census scenarios | • Planning and execution of the thematic tests | • Statistical design of the census sample | • Capture of the characteristics the urban environment and localities with less than 2,500 inhabitants |
| • Public consultation for the content and the methodology of questionnaires | • Definition of the conceptual framework and the collection instruments | • Cartographic update in rural areas of Basic Geostatistics | • Start of the census survey |
| • Definition of the census project | • Pilot census | • Statistical design of the post-enumeration survey | • Sample survey for targeted coverage estimation |
| • Testing of operating procedures | • Bidding for the purchase of Mobile Computing Devices | • Recruitment and selection of personnel | • Information processing |
| | | • Training of operating personnel | • Generation of results |
| | | • Start of the communication campaign | • Publication of results (January 2021) |

Source: author's work.

Data collection, their processing, and the generation and publication of the results were scheduled for 2020. In these stages, the planning and technification of the census favored the process's speed, even in the context of the pandemic (INEGI, 2021b).

The global spread of the SARS-CoV-2 virus occurred rapidly during the first months of 2020: the first confirmed death from the virus was reported on 11th January in Wuhan, China (El Mundo, 2020). On 22th January, the Mexican Ministry of Health issued an epidemiological alert in response to the new coronavirus (Comisión Nacional de Vigilancia Epidemiológica, 2020).

During February, the INEGI surveyed the characteristics of the urban environment and localities with less than 2,500 inhabitants, and it was not until the 27th of that month that the first case of COVID-19 infection in Mexico was reported (Guzmán Aguilar, 2021).

The enumeration of inhabitants was carried out from 2nd to 27th March, and relevant events related to the health contingency continued to occur during that period. On 11th March, the WHO declared the COVID-19 pandemic (Organización Panamericana de la Salud, 2020). On 18th March, the first death due to the new virus was registered in Mexico (Secretaría de Salud, 2020a). As of this date, the INEGI allocated resources for acquiring face masks, antibacterial gel, and liquid soap for the staff use. The enumeration of dwellings continued, and on 20th March, additional measures were requested from the INEGI personnel with fieldwork, such as the prohibition of entering homes, avoiding any physical contact between personnel, maintaining a minimum distance of 1.5 meters between people, the confinement of personnel suspected of contagion, and the prohibition of fieldwork by personnel belonging to groups at risk,[1] all the above to protect interviewers and informants (INEGI, 2021a).

On 23rd March, social distancing began, and the first mobility restrictions were made official; basic prevention measures were issued, non-essential activities were suspended, and mass events were rescheduled (Gobierno de México, 2020b). As of this date, the INEGI employees began to work from home, except for the census and survey personnel, who had to comply with the measures of using masks, healthy distance, and the use of antibacterial gel for the last days of the data collection.

The census operations took place in the atmosphere of expectation, mainly due to different reactions of the informants to the field staff. Around the third week of the survey, an increase in the level of aversion and refusal to answer the interview in dwellings was observed among the population, which just coincided with the number of the confirmed positive COVID-19 cases starting to multiply (INEGI, 2021a). There were even media occurrences where public figures and officials requested the suspension of the 2020 Census activities. This situation negatively influenced some sectors of the population (El Dictamen, 2020). In isolated cases, events of aggression towards the census personnel were reported.

It should be noted that from the beginning of the enumeration, invitation letters were distributed so that inhabitants could answer remotely, accessing the 2020 Census online platform or by telephone; this was an option both if the respondents wished so and when it had not been possible to collect the information after three visits to the dwelling. During the field operation, 800,000 letters were delivered, and another 800,000 were sent by mail (INEGI, 2021a).

On 30th March, the General Health Council of Mexico published the "Agreement declaring the epidemic of disease caused by the SARS-CoV-2 virus as a sanitary emergency due to force majeure" in the Official Gazette of the Federation (Secretaría de Gobernación, 2020a). The following day, the Ministry of Health declared a sanitary emergency with 6,620 positive cases, with which non-essential activities were suspended, including the census and surveys in the national territory (Secretaría de Gobernación, 2020b); as a result, the INEGI officially announced the suspension of the face-to-face data collection.

The enumeration phase had been completed several days earlier, but the verification phase had yet to be finished. The post-enumeration survey had to be carried out; the former was postponed, and the latter was cancelled. By that time, the number of infected INEGI personnel had been minimal; however, a follow-up

[1] People with comorbidities associated with COVID-19, such as: diabetes, hypertension, and obesity, as well as elderly and pregnant women.

identification was created for all the personnel to prevent and mitigate the spread of COVID-19.

On 1st June, the reopening of activities began following the criteria of the Epidemiological Risk[2] (Secretaría de Salud, 2020b). As a result, on 15th June the Institute was able to resume activities in 16 states with orange status in the stoplight, following the "Guide for the mitigation and prevention of COVID-19", elaborated by INEGI (2020a).

On 17th July, the Ministry of Health published the "Agreement by which all censuses and surveys to be carried out in the national territory are resumed" in the Official Gazette of the Federation (Secretaría de Gobernación, 2020c), and at the same time, thanks to the collaboration between this federal agency and the INEGI, "General guidelines for the mitigation and prevention of COVID-19 in the generation of statistical and geographic information" was published (INEGI, 2020b), in which the minimum performance criteria were defined to resume field activities.

## 3. Mobile computing devices

For most of the 20th century and the beginning of the 21st century, using paper questionnaires was the usual way of conducting population and housing censuses. This traditional method involves several processes during the collection phase, including printing, distributing copies to various application sites, filling by hand during collection, and returning to the regional centers. All the above results in an extensive collection of printed documents that need to be captured manually in computers, which implies a significant amount of human and technological resources, in addition to the work hours required for this task.

On the other hand, during the second decade of the 21th century, the miniaturization of computer components reached such an extreme that it is now possible to have in the palm of one's hand a device that can match the technological performance of a basic computer, as is the case of tablets and cell phones.

Given the above, the INEGI proposed incorporating current technological advances to carry out various tasks of the 2020 Census, specifically using MCD for integrating digital applications in different processes. In addition to speeding up data collection, these devices made it possible to incorporate essential validation criteria simultaneously with data collection, such as question passes and the review of

[2] The Epidemiological Risk is an indicator that measures the level of epidemiological risk of COVID-19 in the states and municipalities of the Mexican Republic. It comprises information on hospital occupancy, SARS-CoV-2 positivity, the trend of hospitalized cases, and the trend of the COVID-19 syndrome. The stoplight classification is: red (maximum risk), orange (high risk), yellow (medium risk), and green (low risk).

congruence between related variables and valid values, which reduced the omission of data and improved the quality of information.

An institutional projection estimated that using the traditional method for the 2020 Population and Housing Census in Mexico would mean processing 35 million paper questionnaires (Vielma, 2022a). Furthermore, a comparative scenario analysis determined that the collection of the information on paper would generate a cost of over 1.5 billion Mexican pesos, while data collection by means of MCDs would cost only 300 million Mexican pesos (without the purchase of the devices); therefore, the remaining 1.2 billion Mexican pesos could be used for the purchase of the devices (INEGI, 2020c).

Considering that the 2020 Census would require slightly over 185,000 devices, it was estimated that in order to keep the cost of the survey carried out by means of MCDs at the same level as the cost of the traditional census (on paper), each device should cost slightly less than USD 346 (at the exchange rate in pesos at that date). However, if devices were acquired at a lower price, this would generate savings.

The actual cost per unit was less than USD 94, to which professional services, materials, and furniture expenses had to be added. At the end of the comparison, the estimated cost of the survey with MCDs, compared to the paper-based scenario, showed a saving of almost 60%; it should be noted that this only reflects the difference in the direct cost of capturing the information using paper or mobile devices (INEGI, 2020c).

After deciding to implement MCDs for the 2020 Census, several minimum characteristics of the devices were determined. Some of them had to have: an adequate processor and memory, an Android 6.0 or higher operating system, a screen with good resolution and brightness to be used outdoors, 32 GB internal storage to house the operating system, information capture applications, cartography, training material, satellite images, and the collected data; a battery that would guarantee operation for a full day of work; connectivity through the Universal Mobile Telecommunications System (UMTS or 3G network), as it is the most widespread in the country; and GPS to support the location of the personnel in the field (INEGI, 2019). Another important detail was that it could be operated just with one hand for more comfort for the interviewers.

Tests and field exercises were conducted before the purchase of the devices. Therefore, in 2017, the Operational Strategy Test was performed with various digital devices meeting the above mentioned requirements, in selected areas of Baja California, Hidalgo, Tabasco, and Yucatan (INEGI, 2021d). Five types of tablets of different brands, models, sizes, capacities, and operating systems were analyzed. More specifically, what was tested was their viability of use in the interviews, the performance of the device and applications in different climates, its connectivity, ability to send information from the survey field, and data backup.

With the results of the above-mentioned test, the MCD characteristics were selected, and a limited number of units were purchased to be used during 2018 in the 2020 Population and Housing Census pilot test, thus subjecting them to an exercise in actual operating conditions. At the same time, various requirements for the bidding and purchase of the equipment were analyzed. In addition, in 2019, the selected MCD was combined with the established operating procedures and the selected technological and communications tools to make the final adjustments to the IT (Information Technology) infrastructure.

In the rules of the public bidding process for the purchase, it was established that the devices should be new, not remanufactured or discontinued, and have warranty coverage of at least one year. It was also stipulated that suppliers should participate in technical performance and endurance testing of the device, be responsible for transportation and delivery, and install customized software upon delivery.

Finally, it was stipulated that upon the delivery of the total number of devices, a sample of them would be reviewed. The shipment would be regarded undelivered if more than 1.5% of this sample did not meet the expected performance standard and characteristics. The procurement process lasted almost a year, and towards the end of 2019, a visit was made to the manufacturing plant.

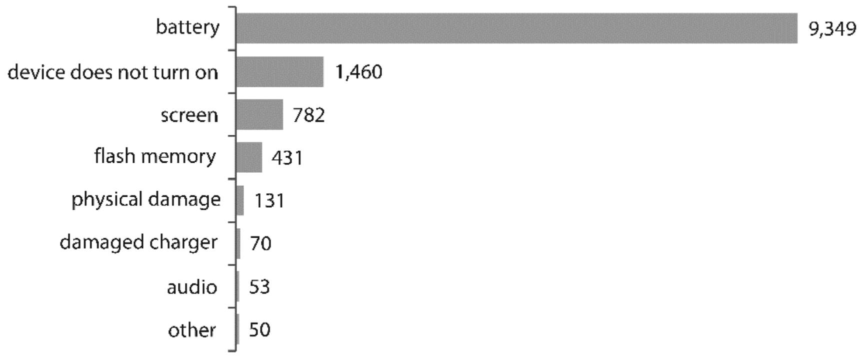
The MCD selected in the bidding process to be used by the INEGI in the 2020 Census had the following technical parameters:

- an MT6580 quad-core processor;
- Android 9.0 operating system;
- a 5.99 inch (15.21 cm) and 1440 by 720 pixels IPS display;
- brightness above 400 nits;
- capacity of 2 GB of RAM and 32 GB of internal storage;
- a 3,900 mAh battery that allows more than 8 hours of continuous use;
- connectivity with USB memory via OTG cable.

The inventory of the 185,824 devices was carried out at the Institute's central offices, and during the process, defective devices were identified and returned to the supplier for replacement. Once the inventory was completed, the devices were distributed to 34 local offices in the states of the country; this task was mainly carried out by vehicles and personnel of the Institute, taking various precautions for the safety of transportation.

As for warranties, they were attended directly by the supplier in each of the 34 local offices, with a maximum resolution time of 24 hours. In 2020, 6.6% of the equipment failed in some way. There were 12,326 failures, mostly related to the battery, as shown in Figure 5.

**Figure 5.** Main failures of MCDs




Note. $n$ = 12,326 failures during 2020, occurring in 6.6% of the MCDs used in the census.
Source: author's work.

## 4. Technology and developed applications

It should be noted that the technology used in the 2020 Census (and applied to such a task for the first time) was not limited to the incorporation of MCDs to the survey process. The devices were the most visible, but without the efficient software, no digital hardware would be helpful. Therefore, to carry out the census, different applications related to the census operation and staff training had to be developed within the INEGI. Likewise, computer communication systems were structured to manage the devices and preserve the security of information.

This way, the following applications were created and installed in the MCD: Kiosko INEGI, which limited the usability of the device to census-related functions; Census Administrator, which had tools for capturing information; Census Cartographic Module (Spanish acronym MCC) that was used to update cartographic records; and the application for the training of operational figures (Spanish acronym CAAP).

As mentioned before, the Kiosko INEGI restricted the functions of the MCDs only to applications developed for the 2020 Census (INEGI, 2020c). With user credentials and an enablement pattern from the MCD, the device started the application for the interviewer or supervisor of interviewers. Due to the restrictions, the main features/functions available on the device were: device battery indicator, brightness level, making calls between operating figures, enabling or disabling the Wi-Fi, 3G and location.

Likewise, the camera was used to read the QR codes of the labels, invitations for self-enumeration, and printed questionnaires. Due to security policies, the installation and uninstallation of applications, factory reset, notifications, access to settings, USB debugging, and connection to any PC were disabled.

The work area access module was in the Census Administrator, which is the beginning of data registration for urban blocks and dwellings, necessary to carry out census interviews. The GPS provided the location of the supervisor and interviewer's work areas while visualizing and managing workloads. It also helped transfer cartographic information from the supervisor and interviewer and provided the tracking module which controlled the progress of the survey in their work areas.

Likewise, the Census Administrator was the computer tool that allowed the application of the questionnaires during direct interviews (INEGI, 2020c). This included: the Basic Questionnaire and the Expanded Questionnaire for private inhabited dwellings, the Urban Environment Questionnaire, the Locality Questionnaire, and the Capture Module for the registration of each property (List of Properties).

The MCC made it possible to register cartographic updates detected during the surveillance in urban and rural areas, which contributed to the comprehensive census coverage and reliability of information during the survey. In addition, this application made it possible to create and delete urban blocks (areas), change the names of objects and even locations, and modify block fronts.

A total of 6,624 cartographic packages in GEOJSON format, one for each area manager, were generated and installed in the devices of the interviewer and supervisor. The information layers contained: urban and rural blocks, specific localities, road axes, block fronts, and satellite images. Due to a large size of files, there were minor setbacks in loading and visualization of the cartographic layers derived from the number of polygons that make up an area (especially in rural areas).

On the other hand, the CAAP training module was a software application designed for training the operational staff. It integrated the learning contents in infographics of the installed tools, the device care, standard solutions to problems, audio and video on each learning topic, maintenance of accessories and USB data backup devices, and self-evaluation.

Using the CAAP for personnel preparation was beneficial, as it allowed the trainee to have autonomous learning while receiving immediate feedback and evaluation of their verification activities, as well as being able to follow up on their performance. It also allowed the trainee to review audiovisual content at any time or place and to exercise the use of the device's systems.

The information captured in the MCDs was then sent in two ways: via the Internet (the Interviewer transferred the information collected to his/her Supervisor who integrated it into the central database through the OPERA platform), and through mobile data. A prepaid SIM (Subscriber Identity Module) card was installed in each MCD. With this, the interview data were sent at the time of completion. If there was no mobile data service, the data were saved for sending later when the device detected a signal.

For the 2020 Census, the OPERA platform was used, which showed which devices were enrolled, and in addition their incidences (INEGI, 2020c). OPERA is an institutional multi-operational technological platform that provides services through modules or computer tools to meet the needs of the INEGI's census projects at almost all stages. This system has been developed on the basis of a modular design, allowing high scalability and integration with the requirements of different institutional projects.

It is composed of the following main modules: the operational monitoring of the main tasks; logistical monitoring to control the provision of material resources to field personnel; monitoring contingencies and operational incidents to support their prevention and ensure their registration; IT support and management of the MCD; recruitment and selection of personnel; data integration (to transfer the information captured to the central database); and remote distribution of digital and operational materials.

Together, the modules provided tools for executing and supporting the operational personnel's activities. They also informed managers and directors of the situation and aspects that arose throughout the project, enabling them to make decisions that would contribute to improving the quality of data collection.

## 5. Data collection

For most of the 33.6 million interviews in the 2020 Population and Housing Census, the information was collected through digital questionnaires that were viewed and uploaded through the MCDs (CAMI[3] method). The devices streamlined the conduct of face-to-face interviews and made it easier to automate the follow-up and monitoring tasks (INEGI, 2021c).

A small percentage of the data collection was done traditionally, on paper, mostly in high-risk (e.g. crime-ridden) areas, or when there was a technical failure with the device (mainly due to problems with the battery, screen or ignition). This way, paper-based collection functioned as a backup mechanism. Each questionnaire had a unique quick response code (QR), which made them unrepeatable and controlled their distribution among the personnel responsible for the data collection. On the

[3] Computer-assisted mobile interviewing, sometimes just known as computer-assisted personal interviewing (CAPI).

whole, around 700,000 questionnaires were filled out on paper and captured manually (INEGI, 2021d).

Additionally, two complementary methods were defined. The first was self-enumeration, in which the informant provided data on the dwelling and its residents through the INEGI web page, on the basis of the invitation letter delivered to them by the interviewer (CAWI[4] method). The second method was an assisted interview by telephone (CATI[5] method), in which the informant called the number indicated in the invitation letter to provide data on the dwelling and its inhabitants. Due to the COVID-19 health contingency, it was necessary to use the mailing of self-enumeration invitations to complete another small percentage of interviews.

Furthermore, the application used in the MDC, CAWI and CATI to capture information was used on mobile devices such as netbooks to collect information from the population living in collective housing (nursing homes, prisons, shelters, etc.).

At the end of the 2020 Census, 97.7% of the data on dwellings and persons were captured on digital devices, 2% on paper, and 0.3% via the web or telephone. A total of 43,903,443 dwellings were visited, of which 35,233,462 (80.3%) were inhabited, and 8,669,981 (19.7%) were uninhabited or intended for temporary use. Of the total number of inhabited dwellings, private and collective, 95.5% coverage was achieved (INEGI, 2021c).

Thanks to the above-mentioned efforts and decisions, it was possible to publish the basic results of the 2020 Population and Housing Census in Mexico on 25th January 2021. According to them, there were 126,014,024 inhabitants in Mexico, of whom 64,540,634 were women (51.2%), and 61,473,390 men (48.8%). The information was also generated for each state and municipality in the country, disaggregated at the urban block and locality level, providing the main socio-demographic characteristics of the population, such as age and place of residence.

This information was an essential input for the country's decision-makers in the distribution of vaccines against COVID-19, as it made it possible to geographically locate and prioritize vulnerable populations (Vielma Orozco, 2022b).

[4] Computer-assisted web interviewing.
[5] Computer-assisted telephone interviewing.

## 6. Main differences between the PAPI and CAMI (supported by CAWI and CATI) method for the 2010 and 2020 censuses

In this section, we present a brief comparison of the information on the IT technologies and field strategies used in the 2010 and 2020 Mexican censuses (Table).

**Table.** Comparison of strategies between the 2010 and 2020 censuses

| Strategy stage | 2010 Census | 2020 Census |
|---|---|---|
| Input development | The content of both the didactic materials and the census questionnaires was developed. | |
| | All census material was printed (manuals, questionnaires, cartography). | All census material was developed on a digital platform (manuals, questionnaires, cartography) that was then stored on mobile computing devices (MCDs). |
| Selection and hiring | The selection and hiring process of all personnel was carried out on a digital platform. | The selection and hiring process of all personnel was carried out on a digital platform. For 2020, there was a greater integration of the selection modules. |
| Interviewer training | The training was carried out in classrooms, led by a tutor who explained the thematic contents.<br>Trainees received their printed material, which was abundant (it was not possible to guarantee that they would read all the material). After the training, they were evaluated. | The training was organized on MCDs using CAAP, with audio, image and reading content, for greater attention from the interviewers.<br>The interviewers listened to the content through headphones and read the information which appeared in the MCD. At the end of each section, they were evaluated on the same device.<br>The role of the tutor was only advisory for the correct use of the device. |
| Cartographic update | The cartographic update was carried out on paper maps and took weeks to capture and validate. | The cartographic update was carried out on the MCD and any adjustment was identified and applied in real time. |
| Interview methodology | Face-to-face interview with paper-and-pencil (PAPI). | Face-to-face computer-assisted interview (CAMI), with the support of the CAWI and CATI methods. |
| Information capture platform | A data information capture platform was developed for paper questionnaires, once they were used and sent to the local offices. The information questionnaires arrived in the INEGI central offices three to four weeks after. | A data information capture platform was developed for the questionnaires applied through the MCDs. Once these were used, they were sent in real time to the computer servers of the INEGI central offices. |
| Monitoring of the census operation in the field | The follow-up of the census operation in the field took a week to know the progress of the first day.<br>The leaders of the central offices took a week to make decisions. | The follow-up of the census operation in the field took five hours to know the progress of the first day.<br>Every day at 7:00 a.m., the leaders of the central offices reviewed the specific progress throughout the national territory and sent instructions at the beginning of the day of operations to the coordinator in different country localities. |

**Table.** Comparison of strategies between the 2010 and 2020 censuses (cont.)

| Strategy stage | 2010 Census | 2020 Census |
|---|---|---|
| Geographical identification of the dwellings | Through a paper map. | Through a coordinate emitted by the MCD GPS.<br>The label of the dwellings had a QR code that associated the number with the preloaded cartography of the MCD. |
| Information validation | The information was first registered and then validated, which typically took two to three weeks.<br>In many cases, it was difficult to return to the field to correct some faulty information. | The MCD identified basic inconsistencies in real time, which made it possible to consult the interviewee about what was detected at that precise moment. |
| Data security | As the information was captured in paper questionnaires, anyone could access what a household had answered. | As the information was captured on an MCD, nobody could access the information, once the information was sent in real time. |
| Safety of the interviewing staff | Staff whereabouts were known once they returned to their local offices (in rural areas this could take days). | The MCD sent the geolocation signals of the location of the interviewing staff. |
| Data production time | After the census data collection, the effective time of processing the data took 9 months. | After the census data collection, the effective time of processing the data took 6 months. |

Source: author's work.

The benefits of using MCDs instead of paper questionnaires included a reduction in the operating cost by approximately 47 million dollars, or 60%, compared to the previous census. But the greatest benefit was the improved timeliness of data, as well as its better quality.

## 7. Conclusions

The activity connected to the 2020 Housing and Population Census in Mexico began long before the COVID-19 pandemic, so when the SARS-CoV-2 alert was issued, the census work was already well underway.

It is essential to highlight that the analysis of the behavior of previous pandemics facilitated evidence-based decision-making, since it was foreseen that the pandemic would not last months but years; it was also possible to observe that, in its first days, infections and deaths from COVID-19 would be few and that as the weeks pass, the situation would become more serious.

Therefore, it was decided that the field collection of the 2020 Census data would continue, as it is an indispensable socio-demographic statistical activity. The organization of the census-related activities over a relatively long period, the digitalization of the interview process, and the review of historical information from the past pandemics constituted the determining factors to the fact that most of the census stages were carried out on time and only some final activities were rescheduled.

Likewise, using applications and systems through MCDs made the capture of information more straightforward and agile, generating savings in human and material resources, time, and money simultaneously. If MCDs had not been used, data processing would have taken months, and the information from the paper questionnaires would have required putting many people inside a confined space for its capture, which, in the context of the COVID-19 pandemic, is unthinkable. Therefore, the opportunity to present relevant data for the health emergency would have been lost.

Also, we learned some important lessons following the COVID-19 pandemic, so other institutions and statistical offices can use this information to make decisions in the face of eventualities.

The first piece of useful knowledge drawn from the Mexican census experience is the great importance of providing continuity in the generation of official statistics, since the usual indicators, as well as the indicators derived from the current situation, are an essential input to the decision-making in critical circumstances. The second "lesson" is the significance of maintaining adequate communication to build a relationship of trust between the national statistical office, the population, and government authorities, to be ready for emerging changes in activities. The third valuable experience points to the vital importance of having robust risk management that considers all types of scenarios, even those that seem unlikely, to aid decision-making, have various strategies and define specific courses of action. The fourth "lesson" is realizing the need to implement innovative or uncommon information-gathering methods in developing countries, such as telephone-based information gathering and self-enumeration.

The fifth piece of knowledge gathered from the 2020 Census in Mexico is the necessity to further increase the use of Information and Communication Technologies (ICT) to perform tasks related to collecting and processing information in a more agile and efficient way, which will lead to savings in economic and human resources. And the sixth and final "lesson" learned from our census is that it is necessary to strengthen inter-agency collaboration through officials with a high level of general and managerial knowledge of events that may impact mass statistical programs, who know both the activities of their organization and those of others, and who know other subjects (in this case health and history) to analyze the behavior of pandemics. If we draw appropriate conclusions from the 2020 Census in Mexico, we will be able to prepare better for making effective decisions in the future difficult situations.

## References

Comisión Nacional de Vigilancia Epidemiológica. (2020, January 21). *Casos de infección respiratoria asociados a Nuevo Coronavirus (2019-nCoV).* https://www.gob.mx/cms/uploads/attachment/file/527027/AE-Nuevo_Coronavirus_2019_nCoV.pdf.

El Dictamen. (2020, March 18). *Sherlyn le cierra la puerta al Inegi por miedo al coronavirus.* https://www.eldictamen.mx/espectaculos/sherlyn-le-cierra-la-puerta-al-inegi-por-miedo-al-coronavirus/.

El Mundo. (2020, January 11). *Primera muerte en China por un nuevo virus misterioso.* https://www.dw.com/es/primera-muerte-en-china-por-un-nuevo-virus-misterioso/a-51964444.

Gobierno de México. (2020a). *Fases o escenarios de contingencia y nivel de propagación del COVID-19.*

Gobierno de México. (2020b). *Jornada Nacional de Sana Distancia.* https://www.gob.mx/cms/uploads/attachment/file/541687/Jornada_Nacional_de_Sana_Distancia.pdf.

Grupo Interinstitucional para la Estimación del Exceso de Mortalidad por todas las causas. (2021). Resultados. Exceso de mortalidad nacional en México. *Boletín estadístico sobre el exceso de mortalidad por todas las causas durante la emergencia por COVID-19*, (10), 11. https://coronavirus.gob.mx/wp-content/uploads/2021/02/Boleti%CC%81n_X_Exceso_Mortalidad_SE53_MX_15Feb2021.pdf.

Guzmán Aguilar, F. (2021, February 25). COVID-19: A un año del paciente cero en México. *Gaceta UNAM*. https://www.gaceta.unam.mx/covid-19-a-un-ano-del-paciente-cero-en-mexico/.

Instituto Nacional de Estadística y Geografía. (2019). *Convocatoria a la licitación pública internacional mixta abierta a la participación de cualquier interesado, celebrada bajo la cobertura de los tratados internacionales, número LA-040100992-E11-2019, para la adquisición de Dispositivos de Cómputo Móvil para el Censo de Población y Vivienda 2020.*

Instituto Nacional de Estadística y Geografía. (2020a). *Guía para la mitigación y prevención de COVID-19.*

Instituto Nacional de Estadística y Geografía. (2020b). *Lineamiento general para la mitigación y prevención del COVID-19 en la generación de información estadística y geográfica.*

Instituto Nacional de Estadística y Geografía. (2020c, April 7–8). *Revisión de Ficha de Sistematización de Experiencias para Países de América Latina y el Caribe* [ideoconference]. Twentieth meeting of the Executive Committee of the Statistical Conference of the Americas of ECLAC, Santiago.

Instituto Nacional de Estadística y Geografía. (2021a). *Así se contó México*. https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825007046.

Instituto Nacional de Estadística y Geografía. (2021b). *Censo de Población y Vivienda 2020: Marco conceptual.* https://www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825197520.pdf.

Instituto Nacional de Estadística y Geografía. (2021c). *Censo de Población y Vivienda 2020: Presentación de resultados: Estados Unidos Mexicanos.* https://www.inegi.org.mx/contenidos/programas/ccpv/2020/doc/Censo2020_Principales_resultados_EUM.pdf.

Instituto Nacional de Estadística y Geografía. (2021d). *Censo de Población y Vivienda 2020: Síntesis metodológica y conceptual.* https://www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825197537.pdf.

Instituto Nacional de Estadística y Geografía. (2023, February 27). *Estadísticas de Defunciones Registradas (EDR) enero a septiembre 2022 (preliminar).* https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2023/EDR/EDR2022_3t.pdf.

Organización Panamericana de la Salud. (2020, March 11). *La OMS caracteriza a COVID-19 como una pandemia*. https://www.paho.org/es/noticias/11-3-2020-oms-caracteriza-covid-19-como-pandemia.

Secretaría de Gobernación. (2020a, March 30). *Acuerdo por el que se declara como emergencia sanitaria por causa de fuerza mayor, a la epidemia de enfermedad generada por el virus SARS-CoV2 (COVID-19).* https://www.dof.gob.mx/nota_detalle.php?codigo=5590745&fecha=30/03/2020#gsc.tab=0.

Secretaría de Gobernación. (2020b, March 31). *Acuerdo por el que se establecen acciones extraordinarias para atender la emergencia sanitaria generada por el virus SARS-CoV2*. https://www.dof.gob.mx/nota_detalle.php?codigo=5590914&fecha=31/03/2020#gsc.tab=0.

Secretaría de Gobernación. (2020c, July 17). *Acuerdo por el que se reanudan todos los censos y encuestas a realizarse en el territorio nacional.* https://www.dof.gob.mx/nota_detalle.php?codigo=5596930&fecha=17/07/2020.

Secretaría de Salud. (2020a, March 18). *Hoy en México falleció la primera persona con COVID-19* [Tweet]. Twitter. https://twitter.com/SSalud_mx.

Secretaría de Salud. (2020b). *Semáforo de riesgo epidemiológico COVID-19: Indicadores y metodología*. https://coronavirus.gob.mx/wp-content/uploads/2020/06/Lineamiento_Semaforo_COVID_05Jun2020_1600.pdf.

Taubenberger, J. K., & Morens, D. M. (2006). 1918 Influenza: the mother of All Pandemics. *Emerging Infectious Diseases*, *12*(1), 15–22. https://doi.org/10.3201/eid1201.050979.

Vielma Orozco, E. (2022a). *Historia de los censos de población en México. Siglo XIX* (Vol. I) [Unpublished book]. INEGI.

Vielma Orozco, E. (2022b). *Lessons from crisis management by COVID-19 in the generation of sociodemographic statistics* [Manuscript submitted for publication]. INEGI.

World Health Organization. (2003). *Consensus document on the epidemiology of severe acute respiratory syndrome* (*SARS*). https://apps.who.int/iris/handle/10665/70863.

Wiadomości Statystyczne. The Polish Statistician, 2023, *68*(6), 67–68
Dyskusje. Recenzje. Informacje / Discussions. Reviews. Information

# WYDAWNICTWA GUS. MAJ 2023
# PUBLICATIONS OF STATISTICS POLAND. MAY 2023

W ofercie wydawniczej Głównego Urzędu Statystycznego z ubiegłego miesiąca warto zwrócić uwagę na następującą publikację:
Among Statistics Poland's last month's publications, we would like to recommend:

***Zużycie energii w gospodarstwach domowych w 2021 r.***
***Energy consumption in households in 2021***

Najnowsza edycja publikacji zawierającej wyniki przeprowadzanego co trzy lata badania zużycia paliw i energii w gospodarstwach domowych, które stanowi moduł badania budżetów gospodarstw domowych.


**Język:** polski, angielski
**Language:** Polish, English
**Seria:** Analizy statystyczne
**Series:** Statistical analyses
**Dostępne wersje:** drukowana i elektroniczna z tablicami w formacie Excel
**Available in:** printed and electronic form with Excel tables

W opracowaniu szczegółowo przeanalizowano różne aspekty wykorzystywania energii w gospodarstwach domowych, z uwzględnieniem zmian, jakie nastąpiły w latach 2002–2021. Podano informacje o zużyciu energii, w tym pozyskiwanej ze źródeł odnawialnych, w Polsce i innych krajach Unii Europejskiej. Zwrócono uwagę m.in. na sposoby użytkowania energii, wyposażenie gospodarstw domowych w urządzenia ją pobierające i czynniki strukturalne mające wpływ na wielkość jej zużycia, a także zaprezentowano dane o zużyciu paliw przez samochody osobowe w gospodarstwach domowych. Wykorzystano m.in. wyniki badania budżetów gospodarstw domowych w zakresie dochodów i wydatków oraz dane dystrybutorów ciepła i gazu ziemnego.

W przygotowaniu publikacji wzięli udział pracownicy Urzędu Statystycznego w Rzeszowie, Agencji Rynku Energii oraz Ministerstwa Klimatu i Środowiska.

W maju br. ukazały się ponadto:
- „Biuletyn statystyczny" nr 4/2023;
- *Ceny robót budowlano-montażowych i obiektów budowlanych (marzec 2023 r.)*;

- *Koniunktura w przetwórstwie przemysłowym, budownictwie, handlu i usługach 2000–2023 (maj 2023)*;
- *Polska w liczbach 2023*;
- *Produkcja ważniejszych wyrobów przemysłowych w kwietniu 2023 r.*;
- „Przegląd Statystyczny. Statistical Review" nr 4/2022;
- *Skup i ceny produktów rolnych w 2022 r.*;
- *Sytuacja społeczno-gospodarcza kraju w kwietniu 2023 r.*;
- „Wiadomości Statystyczne. The Polish Statistician" nr 4/2023;
- „Wiadomości Statystyczne. The Polish Statistician" nr 5/2023;
- *Wyniki finansowe przedsiębiorstw niefinansowych 01–12 2022 r.*;
- *Zeszyt metodologiczny. Koszt całkowity usług publicznych świadczonych przez jednostki samorządu terytorialnego*;
- *Zielone Płuca Polski w 2021 r.*;
- *Zmiany strukturalne grup podmiotów gospodarki narodowej na obszarach przygranicznych na terenie Polski w latach 2021–2022.*

**Joanna Sadowy**
Główny Urząd Statystyczny, Departament Opracowań Statystycznych, Polska
Statistics Poland, Statistical Products Department, Poland

Wszystkie publikacje GUS w wersji elektronicznej są dostępne na stronie stat.gov.pl/publikacje/publikacje-a-z. Wersje drukowane (jeśli zostały wydane) można zamawiać pod adresem: zws-sprzedaz@stat.gov.pl.
All the publications of Statistics Poland available in electronic form can be accessed at stat.gov.pl/en/publications. Printed versions (if available) may be ordered at: zws-sprzedaz@stat.gov.pl.

# DLA AUTORÓW
FOR THE AUTHORS
(for the English translation of the information given below, please visit ws.stat.gov.pl/ForAuthors)

W „Wiadomościach Statystycznych. The Polish Statistician" („WS") zamieszczane są artykuły o charakterze naukowym poświęcone teorii i praktyce statystycznej, które prezentują wyniki oryginalnych badań teoretycznych lub analitycznych wykorzystujących metody statystyki matematycznej, opisowej bądź ekonometrii. Ukazują się również artykuły przeglądowe, recenzje publikacji naukowych oraz inne opracowania informacyjne. W czasopiśmie publikowane są prace w języku polskim i angielskim.

Od 2007 r. „WS" znajdują się na liście czasopism naukowych MEiN. Zgodnie z komunikatem Ministra Edukacji i Nauki z dnia 1 grudnia 2021 r. w sprawie wykazu czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych „WS" otrzymały 70 punktów.

„Wiadomości Statystyczne. The Polish Statistician" są udostępniane w następujących bazach, repozytoriach, katalogach i wyszukiwarkach: Agro, BazEkon, Biblioteka Nauki, Central and Eastern European Academic Source (CEEAS), Central and Eastern European Online Library (CEEOL), Central European Journal of Social Sciences and Humanities (CEJSH), Directory of Open Access Journals (DOAJ), EBSCO Discovery Service, European Reference Index for the Humanities and Social Sciences (ERIH Plus), Exlibris Primo, Google Scholar, ICI Journals Master List, ICI World of Journals, Norwegian Register for Scientific Journals and Publishers (The Nordic List) oraz Summon.

Za publikację artykułów na łamach „WS" autorzy nie otrzymują honorariów ani nie wnoszą opłat.

## 1. Zgłaszanie artykułów

Prace przeznaczone do opublikowania w „WS" należy przesyłać za pośrednictwem platformy Editorial System: www.editorialsystem.com/ws.

Zgłaszany artykuł powinien być zanonimizowany, tj. pozbawiony informacji o autorze/autorach (również we właściwościach pliku), podziękowań i informacji o źródłach finansowania, a także innych informacji wskazujących na afiliację lub umożliwiających zidentyfikowanie autora. Jeżeli w pracy występują tablice, wykresy lub mapy, powinny być umieszczone w treści artykułu. Materiały graficzne, razem z danymi do nich, należy ponadto załączyć jako osobny plik / osobne pliki, najlepiej w formacie Excel. **Prosimy o niestosowanie stylów i ograniczenie formatowania do wymogów redakcyjnych.** Więcej informacji w pkt 4 *Wymogi redakcyjne*.

Razem z artykułem należy przesłać skan/zdjęcie oświadczenia o oryginalności pracy i niezłożeniu jej w innym wydawnictwie. **Załączenie oświadczenia jest warunkiem poddania pracy ocenie wstępnej i skierowania do recenzji.**

Zgłoszenie artykułu do opublikowania w „WS" oznacza zgodę na jego udostępnienie na licencji Creative Commons Uznanie autorstwa – Na tych samych warunkach 4.0 (CC BY-SA 4.0).

Autorzy mają prawo do samodzielnego umieszczania w wybranych przez siebie repozytoriach artykułu w wersji zarówno zgłoszonej do „WS", jak i zaakceptowanej do opublikowania oraz opublikowanej, z zastrzeżeniem wymogu niezwłocznego podania w repozytorium informacji o numerze „WS", w którym praca się ukazała, wraz z linkiem do niej (DOI).

## 2. Przebieg prac redakcyjnych

Zgłoszony artykuł jest oceniany i opracowywany w czteroetapowym procesie:

1. **Ocena wstępna**, dokonywana przez redakcję. Polega na weryfikacji naukowego charakteru artykułu oraz jego struktury i zawartości pod kątem wymogów redakcyjnych, a także zgodności tematyki z profilem czasopisma. Autor uzupełnia i poprawia artykuł stosownie do uwag redakcji, a w przypadku nieuwzględnienia danej uwagi uzasadnia swoje stanowisko. Warunkiem skierowania pracy do recenzji jest potwierdzenie oryginalności tekstu uzyskane za pomocą systemu antyplagiatowego. W przypadku wykrycia znacznego podobieństwa do innych prac artykuł zostanie odrzucony.
2. **Ocena recenzentów**, dokonywana przez specjalistów w danej dziedzinie. Artykuł oceniają dwaj recenzenci spoza jednostki naukowej, przy której afiliowany jest autor; w przypadku pracy w języku angielskim co najmniej jeden recenzent jest afiliowany przy jednostce zagranicznej. W razie sprzecznych opinii dwóch recenzentów powoływany jest trzeci recenzent. Recenzenci kierują się kryteriami oryginalności i jakości opracowania zarówno w odniesieniu do treści, jak i formy artykułu.

   Autorzy artykułów, które otrzymały pozytywne oceny, wprowadzają poprawki zalecane przez recenzentów i przesyłają zmodyfikowaną wersję pracy. Jeśli pojawi się różnica zdań dotycząca zasadności proponowanych zmian, autorzy są zobligowani do uzasadnienia swojego stanowiska.
3. **Ocena Kolegium Redakcyjnego** (KR), decydująca o przyjęciu pracy do publikacji. Jest dokonywana na podstawie recenzji, z uwzględnieniem opinii redaktorów tematycznego i merytorycznego. Polega m.in. na weryfikacji dokonania przez autora zmian w artykule stosownie do uwag recenzentów. KR ocenia artykuł pod względem poprawności i spójności merytorycznej oraz zaleca autorowi wprowadzenie poprawek, jeśli są one konieczne, aby praca spełniała wymogi czasopisma. Autorowi przysługuje prawo do odwołania od decyzji o niepublikowaniu artykułu. W takim przypadku powinien on skontaktować się z redakcją „WS" i przedstawić uzasadnienie. Ostateczna decyzja w tej sprawie należy do redaktora naczelnego.

   **W „WS" publikowane są wyłącznie te artykuły, które otrzymają pozytywną ocenę na każdym z wymienionych etapów i zostaną poprawione przez autora zgodnie z otrzymanymi uwagami (chyba że autor przedstawi argumenty uzasadniające nieuwzględnienie danej uwagi).**

   Artykuły przyjęte przez KR do publikacji są zamieszczane na stronie internetowej czasopisma w zakładce Early View, gdzie znajdują się do czasu opublikowania w konkretnym wydaniu „WS".
4. **Opracowanie redakcyjne, autoryzacja i korekta.** Artykuł zakwalifikowany do druku jest poddawany opracowaniu merytorycznemu i językowemu. Redakcja zastrzega sobie prawo do zmiany tytułu i śródtytułów, modyfikowania tablic, wykresów i innych elementów graficznych oraz przeredagowania treści bez naruszenia zasadniczej myśli autora.

   Po opracowaniu redakcyjnym artykuł jest przesyłany do autoryzacji. Tekst zatwierdzony przez autora, po składzie i łamaniu, jest poddawany korekcie i rewizji (II korekcie).

Autor dokonuje korekty autorskiej tekstu na etapie rewizji. Wykresy i inne materiały graficzne są opracowywane na podstawie plików i danych przekazanych przez autora i poddawane korekcie i rewizji. Autor dokonuje ich akceptacji na etapie rewizji.

W przypadku odkrycia błędów w opublikowanym artykule zamieszcza się na łamach „WS" sprostowanie, a artykuł w wersji elektronicznej jest poprawiany i umieszczany na stronie internetowej „WS" ze stosownym wyjaśnieniem.

## 3. Zasady etyki publikacyjnej COPE

Redakcja „WS" podejmuje wszelkie starania w celu utrzymania najwyższych standardów etycznych zgodnie z wytycznymi Komitetu ds. Etyki Publikacyjnej (COPE), dostępnymi na stronie internetowej www.publicationethics.org, oraz wykorzystuje wszystkie możliwe środki mające na celu zapobieżenie nadużyciom i nierzetelności autorskiej. Przyjęte zasady postępowania obowiązują autorów, Radę Naukową, Kolegium Redakcyjne, redakcję, pracowników Wydziału Czasopism Naukowych GUS, recenzentów i wydawcę.

### 3.1. Odpowiedzialność autorów

1. Artykuły naukowe kierowane do opublikowania w „WS" powinny zawierać precyzyjny opis badanych zjawisk i stosowanych metod oraz autorskie wnioski i sugestie dotyczące rozwoju badań i analiz statystycznych. Autorzy powinni wyraźnie określić cel artykułu oraz jasno przedstawić wyniki przeprowadzonej analizy. Prezentacja efektów badań statystycznych zaprojektowanych i przeprowadzonych przez autorów wymaga opisania zastosowanej w nich metodologii. W przypadku nowatorskich metod analizy pożądane jest podanie przykładu ilustrującego ich zastosowanie w praktyce statystycznej. Autorzy ponoszą odpowiedzialność za treści prezentowane w artykułach. W razie zgłaszania przez czytelników zastrzeżeń odnoszących się do tych treści autorzy są zobligowani do udzielenia odpowiedzi za pośrednictwem redakcji.
2. Na autorach spoczywa obowiązek zapewnienia pełnej oryginalności przedłożonych prac. Redakcja nie toleruje przejawów nierzetelności naukowej autorów, takich jak:
   - duplikowanie publikacji – ponowne publikowanie własnego utworu lub jego części;
   - plagiat – przywłaszczenie cudzego utworu lub jego fragmentu bez podania informacji o źródle;
   - fabrykowanie danych – oparcie pracy naukowej na nieprawdziwych wynikach badań;
   - autorstwo widmo (*ghost authorship*) – nieujawnianie współautorów, mimo że wnieśli oni istotny wkład w powstanie artykułu;
   - autorstwo gościnne (*guest authorship*) – podawanie jako współautorów osób o znikomym udziale lub niebiorących udziału w opracowywaniu artykułu;
   - autorstwo grzecznościowe (*gift authorship*) – podawanie jako współautorów osób, których wkład jest oparty jedynie na słabym powiązaniu z badaniem.

   Autorzy deklarują w stosownym oświadczeniu, że zgłaszany artykuł nie narusza praw autorskich osób trzecich, nie był dotychczas publikowany i nie został złożony w innym wydawnictwie oraz że jest ich oryginalnym dziełem, i określają swój wkład w opracowanie artykułu. Jeżeli doszło do zaprezentowania podobnych materiałów podczas konferencji lub sympozjum naukowego, to podczas składania tekstu do publikacji w „WS" autorzy są zobowiązani poinformować o tym redakcję.

3. Autorzy są zobowiązani do podania w treści artykułu wszelkich źródeł finansowania badań będących podstawą pracy.
4. Główną odpowiedzialność za rzetelność przekazanych informacji, łącznie z informacją na temat wkładu poszczególnych współautorów w powstanie artykułu, ponosi zgłaszający artykuł.
5. Autorzy zgłaszający artykuły do publikacji w „WS" biorą udział w procesie recenzji double-blind peer review, dokonywanej przez co najmniej dwóch niezależnych ekspertów z danej dziedziny. Po otrzymaniu pozytywnych recenzji autorzy wprowadzają zalecane przez recenzentów poprawki i dostarczają redakcji zaktualizowaną wersję opracowania wraz z pisemnym poświadczeniem uwzględnienia poprawek. Jeśli pojawi się różnica zdań co do zasadności proponowanych zmian, należy wyjaśnić, które poprawki zostały uwzględnione, a w przypadku ich nieuwzględnienia – uzasadnić swoje stanowisko.
6. Jeżeli autorzy odkryją w swoim maszynopisie lub tekście już opublikowanym błędy, nieścisłości bądź niewłaściwe dane, powinni niezwłocznie poinformować o tym redakcję w celu dokonania korekty, wycofania tekstu lub zamieszczenia sprostowania. W przypadku korekty artykułu już opublikowanego jego nowa wersja jest zamieszczana na stronie internetowej „WS" wraz ze stosownym wyjaśnieniem.

### 3.2. Odpowiedzialność Rady Naukowej, Kolegium Redakcyjnego i Wydziału Czasopism Naukowych GUS

1. Rada Naukowa (RN) kształtuje profil programowy czasopisma, określa kierunki jego rozwoju i konsultuje jego zakres merytoryczny.
2. Kolegium Redakcyjne (KR) podejmuje decyzję o publikacji danego artykułu z uwzględnieniem ocen recenzentów oraz opinii zespołu redakcyjnego. W swoich rozstrzygnięciach członkowie KR kierują się kryteriami merytorycznej oceny wartości artykułu, jego oryginalności i jasności przekazu, a także ścisłego związku z celem i zakresem tematycznym „WS". Oceniają artykuły niezależnie od płci, rasy, pochodzenia etnicznego, narodowości, religii, wyznania, światopoglądu, niepełnosprawności, wieku lub orientacji seksualnej ich autorów.
3. Zespół redakcyjny, wyodrębniony z KR, tworzą redaktor naczelny i jego zastępca, redaktorzy tematyczni i redaktor merytoryczny. Członkowie zespołu redakcyjnego weryfikują nadsyłane artykuły pod względem merytorycznym, oceniają ich zgodność z celem i zakresem tematycznym „WS" oraz sprawdzają spełnienie wymogów redakcyjnych i przestrzeganie zasad rzetelności naukowej. Ponadto wybierają recenzentów w taki sposób, aby nie wystąpił konflikt interesów, i dbają o zapewnienie uczciwego, bezstronnego i terminowego procesu recenzowania.
4. Za sprawny przebieg procesu wydawniczego, poinformowanie wszystkich jego uczestników o konieczności przestrzegania obowiązujących zasad i przygotowanie artykułów do publikacji odpowiadają pracownicy Wydziału Czasopism Naukowych (WCN) GUS. W celu uzyskania obiektywnej oceny oryginalności nadsyłanych artykułów przed skierowaniem ich do recenzji WCN wykorzystuje system antyplagiatowy. Informacje dotyczące

artykułu mogą być przekazywane przez WCN wyłącznie autorom, recenzentom, członkom RN i KR oraz wydawcy.

5. Zmiany dokonane w tekście na etapie przygotowania artykułu do publikacji nie mogą naruszać zasadniczej myśli autorów. Wszelkie modyfikacje o charakterze merytorycznym są z nimi konsultowane.
6. W przypadku podjęcia decyzji o niepublikowaniu artykułu nie może on zostać w żaden sposób wykorzystany przez wydawcę lub uczestników procesu wydawniczego bez pisemnej zgody autorów. Autorzy mogą się odwołać od decyzji o niepublikowaniu artykułu. W tym celu powinni się skontaktować z redaktorem naczelnym lub sekretarzem redakcji „WS" i przedstawić stosowną argumentację. Odwołania autorów są rozpatrywane przez redaktora naczelnego.
7. Członkowie RN i KR ani pracownicy WCN nie mogą pozostawać w jakimkolwiek konflikcie interesów w odniesieniu do artykułów zgłaszanych do publikacji. Przez konflikt interesów należy rozumieć sytuację, w której jakiekolwiek interesy lub zależności (służbowe, finansowe lub inne) mogą mieć wpływ na ocenę artykułu lub decyzję o jego publikacji.
8. W celu przeciwdziałania nierzetelności naukowej wymagane jest złożenie przez autorów oświadczenia, w którym deklarują, że zgłaszany artykuł nie narusza praw autorskich osób trzecich, nie był dotychczas publikowany i jest ich oryginalnym dziełem, a także określają swój wkład w opracowanie artykułu.
9. W celu zapewnienia wysokiej jakości recenzji wymagane jest złożenie przez recenzentów oświadczenia o przestrzeganiu zasad etyki recenzowania COPE i niewystępowaniu konfliktu interesów.
10. W przypadku uzasadnionego podejrzenia na jakimkolwiek etapie procesu wydawniczego, że autorzy dopuścili się nierzetelności naukowej (zob. pkt 3.1. Odpowiedzialność autorów), zespół redakcyjny skrupulatnie zbada sprawę ewentualnego nadużycia. Jeśli nierzetelność autorów zostanie udowodniona, to zgłoszony przez nich artykuł zostanie odrzucony przez KR, a autorzy otrzymają informację o podjętej decyzji wraz z jej uzasadnieniem.
11. Czytelnicy, którzy mają wobec autorów opublikowanego artykułu uzasadnione podejrzenia o nierzetelność naukową, powinni powiadomić o tym redaktora naczelnego lub sekretarza redakcji. Po zbadaniu sprawy ewentualnego nadużycia czytelnicy zostaną poinformowani o rezultacie przeprowadzonego postępowania. W przypadku potwierdzenia nadużycia, na łamach czasopisma zostanie zamieszczona stosowna informacja.

### 3.3. Odpowiedzialność recenzentów

1. Recenzenci przyjmują artykuł do recenzji tylko wtedy, gdy uznają, że:
   - posiadają odpowiednią wiedzę w określonej dziedzinie, aby rzetelnie ocenić pracę;
   - zgodnie z ich stanem wiedzy nie istnieje konflikt interesów w odniesieniu do autorów, przedstawionych w artykule badań i instytucji je finansujących, co potwierdzają w oświadczeniu;
   - mogą wywiązać się z terminu ustalonego przez redakcję, aby nie opóźniać publikacji.
2. Recenzenci są zobligowani do zachowania obiektywności i poufności oraz powstrzymania się od osobistej krytyki. Zawsze powinni uzasadnić swoją ocenę, przedstawiając stosowną argumentację.

3. Recenzenci powinni wskazać ważne dla wyników badań opublikowane prace, które w ich ocenie powinny zostać przywołane w ocenianym artykule.
4. W razie stwierdzenia wysokiego poziomu zbieżności treści recenzowanej pracy z innymi opublikowanymi materiałami lub podejrzenia innych przejawów nierzetelności naukowej recenzenci są zobowiązani poinformować o tym redakcję.
5. Po ukończeniu recenzji przechowywanie przesłanych przez redakcję materiałów (w jakiejkolwiek formie) oraz posługiwanie się nimi przez recenzentów jest niedozwolone.

### 3.4. Odpowiedzialność wydawcy

1. Materiały opublikowane w „WS" są chronione prawem autorskim.
2. Wydawca udostępnia pełną treść wszystkich artykułów w internecie w trybie otwartego dostępu, tj. bezpłatnie i bez technicznych ograniczeń, od 1 stycznia 2022 r. na licencji Creative Commons Uznanie autorstwa – Na tych samych warunkach 4.0 (CC BY-SA 4.0). W przypadku artykułów zgłoszonych do „WS" od 2022 r. dozwolone jest dzielenie się artykułem (kopiowanie i rozpowszechnianie go w dowolnym medium i formacie) oraz adaptowanie go (w dowolnym celu, także komercyjnym) na warunkach określonych w tej licencji. Z pozostałych artykułów zamieszczonych w czasopiśmie można korzystać w ramach otwartego dostępu, zgodnie z ustawą o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego.
3. Wydawca deklaruje gotowość do opublikowania poprawek, wyjaśnień oraz przeprosin.

## 4. Wymogi redakcyjne

Zgodnie z wymogami czasopisma omawiany w artykule problem badawczy powinien być jednoznacznie zdefiniowany oraz istotny dla oceny zjawisk społecznych lub gospodarczych. Artykuł powinien zawierać wyraźnie określony cel badania, precyzyjny opis badanych zjawisk i stosowanych metod, uzyskane wyniki przeprowadzonej analizy oraz autorskie wnioski.

### 4.1. Struktura i zawartość artykułu

Wymagane elementy artykułu:
1. Tytuł.
2. Dane autora: imię/imiona i nazwisko, afiliacja w języku polskim i angielskim, ORCID, wkład w powstanie artykułu, adres e-mail. Wśród autorów artykułu wieloautorskiego należy wskazać autora korespondencyjnego.
3. Streszczenie (zalecana objętość – do 1200 znaków ze spacjami, forma bezosobowa). W przypadku artykułu opisującego badanie empiryczne powinno zawierać: cel, przedmiot, okres i metodę badania, źródła danych i najważniejsze wnioski z badania. W przypadku artykułów o innym charakterze należy podać co najmniej cel pracy, jej przedmiot i najważniejsze wnioski.

   **Streszczenie to podstawowe źródło informacji o artykule, warunkujące też decyzję czytelnika o zapoznaniu się z całą pracą. Dlatego powinno być przygotowane ze szczególną starannością i dbałością o umieszczenie w nim wszystkich wymaganych elementów.**

4. Słowa kluczowe – najistotniejsze pojęcia lub wyrażenia użyte w pracy (nie mniej niż trzy). Powinny być zawarte w streszczeniu i/lub tytule.
5. Kod/kody z klasyfikacji Journal of Economic Literature (JEL).
6. Tłumaczenie tytułu, streszczenia i słów kluczowych (na język angielski w przypadku artykułu napisanego w języku polskim, a na język polski w przypadku artykułu napisanego w języku angielskim).
7. W artykule opisującym badanie empiryczne wymagane są następujące części:
   - wprowadzenie, zawierające syntetyczne przedstawienie zagadnień teoretycznych, uzasadnienie podjęcia danego problemu badawczego, cel badania i krytyczne odniesienie do literatury przedmiotu. W wyjątkowych przypadkach, kiedy istotne dla podjętego tematu jest obszerniejsze przedstawienie dyskusji toczącej się w literaturze, przegląd literatury może stanowić odrębną część artykułu;
   - metoda badania, uwzględniająca przedmiot i okres badania, źródła danych i zastosowane metody badawcze, w tym uzasadnienie ich wyboru;
   - wyniki badania – analiza danych oraz interpretacja wyników i odniesienie ich do rezultatów wcześniejszych badań (dyskusja). W uzasadnionych przypadkach dyskusja może stanowić odrębną część artykułu;
   - podsumowanie, które powinno być zwięzłe i odzwierciedlać istotę problemu badawczego przedstawionego w artykule, bez podawania danych liczbowych; końcowe wnioski powinny odnosić się do treści artykułu, a w szczególności do celu badania.

   Wszystkie części powinny być opatrzone numerami.
8. Bibliografia, zawierająca pełny wykaz prac i materiałów przywołanych w artykule, przygotowana zgodnie z wymogami czasopisma.

## 4.2. Przygotowanie artykułu

1. Artykuł powinien być utrzymany w formie bezosobowej.
2. Tekst należy zapisać alfabetem łacińskim. Nazwy własne, tytuły itp. oryginalnie zapisane innym alfabetem powinny być poddane transliteracji.
3. Nie należy stosować stylów; formatowanie należy ograniczyć do wymogów redakcyjnych.
4. Objętość artykułu łącznie ze streszczeniem, słowami kluczowymi, bibliografią, tablicami, wykresami i innymi materiałami graficznymi nie powinna być mniejsza niż 10 stron maszynopisu ani przekraczać 20 stron.
5. Edytor tekstu: Microsoft Word, format *.doc lub *.docx.
6. Krój czcionki:
   - Arial – tytuł, autor, streszczenie, słowa kluczowe, kody JEL, śródtytuły, elementy graficzne (tablice, zestawienia, wykresy, schematy), przypisy;
   - Times New Roman – tekst główny, bibliografia.
7. Wielkość czcionki:
   - 14 pkt – tytuł, autor, śródtytuły wyższego rzędu;
   - 12 pkt – tekst główny, śródtytuły niższego rzędu;
   - 10 pkt – pozostałe elementy.
8. Marginesy – 2,5 cm z każdej strony.

9. Interlinia – 1,5 wiersza; tablice i przypisy – 1 wiersz; przed tytułami rozdziałów i podrozdziałów oraz po nich – pusty wiersz.
10. Wcięcie akapitowe – 0,4 cm; bibliografia – bez wcięcia, wysunięcie 0,4 cm.
11. Przy wyliczeniach należy posłużyć się listą punktowaną z punktorami w postaci kropek (wysunięcie 0,4 cm, wcięcie 0 cm); wiersze (oprócz ostatniego) zakończone średnikiem.
12. Strony ponumerowane automatycznie.
13. Tablice i elementy graficzne (wykresy, mapy, schematy) muszą być przywołane w tekście.
14. Wykresy, mapy i schematy należy zamieścić w tekście głównym. Wykresy powinny być edytowalne (optymalnie wykonane w programie Excel; w przypadku wykonania w programie graficznym powinny mieć postać wektorową). Wykresy i inne materiały graficzne należy przekazać osobno, najlepiej w pliku programu Excel lub innym edytowalnym w pakiecie Microsoft Office.
15. Tablice muszą być edytowalne. Nie należy stosować rastrów, cieniowania, pogrubiania czy też podwójnych linii itp.
16. Wskazówki dotyczące opracowywania map znajdują się w publikacji *Mapy statystyczne. Opracowanie i prezentacja danych*, dostępnej na stronie internetowej GUS.
17. Pod tablicami i każdym elementem graficznym należy podać źródło opracowania, a także objaśnić użyte w nich skróty i symbole.
18. Literowe symbole liczb i innych wielkości niezłożonych należy zapisywać małą lub dużą literą i pismem pochyłym (np. $a$, $A$, $y(x)$, $a_i$); wektorów – pismem pochyłym i pogrubionym (np. $\boldsymbol{a}$, $\boldsymbol{A}$, $\boldsymbol{w}$, $\boldsymbol{y}(x)$, $\boldsymbol{w}_i$); macierzy – pismem prostym i pogrubionym (np. $\mathbf{A}$, $\mathbf{a}$, $\mathbf{M}$, $\mathbf{Y}(x)$, $\mathbf{M}_i$).
19. Objaśnienia znaków umownych i zapisów w tablicach: kreska (–) – zjawisko nie wystąpiło; zero (0) – zjawisko istniało w wielkości mniejszej od 0,5; (0,0) – zjawisko istniało w wielkości mniejszej od 0,05; kropka (.) – brak informacji, konieczność zachowania tajemnicy statystycznej, wypełnienie pozycji jest niemożliwe lub niecelowe; „w tym" – oznacza, że nie podaje się wszystkich składników sumy.
20. Stosowane są następujące skróty: tablica – tabl., wykres – wykr.
21. Wszystkie zawarte w artykule informacje, dane i stwierdzenia wykraczające poza wiedzę powszechną – np. wyniki badań innych autorów, zarówno o charakterze empirycznym, jak i koncepcyjnym – muszą być opatrzone przypisem bibliograficznym. Przez wiedzę powszechną należy rozumieć informacje ogólnie znane i niebudzące wątpliwości ani kontrowersji w danej grupie społecznej, np. utworzenie GUS w 1918 r. lub powstanie UE w 1993 r. na podstawie traktatu z Maastricht. Natomiast dane statystyczne udostępniane lub publikowane np. przez GUS lub Eurostat nie należą do takich informacji. Charakteru wiedzy powszechnej nie mają również stwierdzenia odnoszące się do idei, zjawisk i procesów społecznych, politycznych czy gospodarczych. Nawet pozornie zdroworozsądkowe idee zmieniają bowiem swój sens w zależności od kultury, języka lub dyscypliny naukowej, a także bywają w rozmaity sposób konceptualizowane, jak np. pojęcie poznania w naukach społecznych.

    **Podanie źródła jest konieczne niezależnie od tego, czy informacje lub stwierdzenia są ujęte w ramy cytatu, czy przedstawione bez dosłownego przytoczenia, np. w formie parafrazy. Jeżeli stwierdzenie może budzić jakiekolwiek wątpliwości odbiorców, autor powinien wskazać stosowne źródło podawanej informacji.**

22. Przypisy rzeczowe, słownikowe lub informacyjne należy umieszczać na dole strony. Przypisy bibliograficzne, zgodnie ze standardem APA (American Psychological Association), należy podawać w tekście głównym.
23. Bibliografię należy przygotować zgodnie ze standardem APA.

## 4.3. Zasady przywoływania publikacji w treści artykułu

| Wyszczególnienie | Przykład przywołania | |
|---|---|---|
| | w odsyłaczu | w treści zdania |
| **Autor indywidualny** | | |
| Jeden autor | (Iksiński, 2001) | Iksiński (2001) |
| Dwóch autorów | (Iksiński i Nowak, 1999) | Iksiński i Nowak (1999) |
| Trzech autorów lub więcej | (Jankiewicz i in., 2003) | Jankiewicz i in. (2003) |
| **Autor instytucjonalny** | | |
| Nazwa funkcjonuje jako powszechnie znany skrótowiec: pierwsze przywołanie w tekście | (International Labour Organization [ILO], 2020) | International Labour Organization (ILO, 2020) |
| kolejne przywołanie | (ILO, 2020) | ILO (2020) |
| Pełna nazwa | (Stanford University, 1995) | Stanford University (1995) |
| **Typ publikacji** | | |
| Publikacja bez ustalonego autorstwa | (*Skrócony tytuł …*, 2015) | *Pełny tytuł* (2015) |
| Publikacja bez roku wydania | (Iksiński, b.r.) | Iksiński (b.r.) |
| Akt prawny | (Pełny tytuł) | Pełny tytuł |
| Strona internetowa / Zbiór danych: znana data publikacji | (Iksiński, 2020) / (Nazwa instytucji, 2020) | Iksiński (2020) / Nazwa instytucji (2020) |
| nieznana data publikacji | (Iksiński, b.r.) / (Nazwa instytucji, b.r.) | Iksiński (b.r.) / Nazwa instytucji (b.r.) |
| **Rodzaj przywołania** | | |
| Przywoływanie kilku prac (porządek prac – chronologiczny, porządek autorów – alfabetyczny) | (Iksiński, 1997, 1999, 2004a, 2004b; Nowak, 2002) | Iksiński (1997, 1999, 2004a, 2004b) i Nowak (2002) |
| Przywoływanie publikacji za innym autorem (uwaga: w bibliografii należy wymienić tylko pracę czytaną) | (Nowakowski, 1990, za: Zieniecka, 2007) | Nowakowski (1990, za: Zieniecka, 2007) |

Źródło: opracowanie na podstawie: American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th edition). https://doi.org/10.1037/0000165-000.

## 4.4. Przykłady opisu bibliograficznego

Bibliografia powinna być zamieszczona na końcu opracowania. Prace należy uszeregować alfabetycznie według nazwiska pierwszego autora. W przypadku dwóch lub więcej prac tego samego autora / tych samych autorów trzeba je uporządkować chronologicznie według roku publikacji. Jeśli kilka prac tego samego autora / tych samych autorów zostało opublikowanych w tym samym roku, należy podać je w kolejności alfabetycznej według tytułu i odpowiednio oznaczyć literami a, b, c itd.

| Typ publikacji | Przykład opisu bibliograficznego |
|---|---|
| **Artykuł w czasopiśmie** | |
| W wersji drukowanej | Nazwisko, X. (rok). Tytuł artykułu. *Tytuł czasopisma*, *rocznik* (zeszyt), strona początku–strona końca. |
| Dostępny w internecie, z DOI | Nazwisko, X., Nazwisko 2, Y. (rok). Tytuł artykułu. *Tytuł czasopisma*, *rocznik*(zeszyt), strona początku–strona końca. https://doi.org/xxx. |
| Dostępny w internecie, bez DOI | Nazwisko, X., Nazwisko 2, Y., Nazwisko 3, Z. (rok). Tytuł artykułu. *Tytuł czasopisma*, *rocznik*(zeszyt), strona początku–strona końca. https://xxx. |
| **Maszynopis** | |
| Niepublikowany / przygotowywany przez autora / zgłoszony do publikacji, ale jeszcze niezaakceptowany | Nazwisko, X. (rok). *Tytuł* [maszynopis niepublikowany / w przygotowaniu / zgłoszony do publikacji]. |
| Zaakceptowany do publikacji | Nazwisko, X. (w druku). Tytuł artykułu. *Tytuł czasopisma*. |
| Opublikowany nieformalnie (np. na stronie internetowej autora) | Nazwisko, X., Nazwisko 2, Y. (rok). *Tytuł artykułu*. https://xxx. |
| Opublikowany w trybie online first (przed włączeniem do zeszytu) | Nazwisko, X. (rok). Tytuł artykułu. *Tytuł czasopisma*. Online first. https://xxx. |
| **Książka** | |
| W wersji drukowanej | Nazwisko, X. (rok). *Tytuł książki*. Wydawnictwo. |
| Dostępna w internecie, z DOI | Nazwisko, X. (rok). *Tytuł książki*. Wydawnictwo. https://doi.org/xxx. |
| Dostępna w internecie, bez DOI | Nazwisko, X. (rok). *Tytuł książki*. Wydawnictwo. https://xxx. |
| W przekładzie | Nazwisko, X. (rok). *Tytuł książki* (tłum. Y. Nazwisko). Wydawnictwo. |
| Wydanie wielotomowe:<br>tom zatytułowany | Nazwisko, X. (rok). *Tytuł książki: nr tomu. Tytuł tomu*. Wydawnictwo. |
| tom niezatytułowany | Nazwisko, X. (rok). *Tytuł książki* (nr tomu). Wydawnictwo. |
| Kolejne wydanie | Nazwisko, X. (rok). *Tytuł książki* (nr wydania). Wydawnictwo. |
| Pod redakcją: w języku polskim | Nazwisko, X. (red.). (rok). *Tytuł książki*. Wydawnictwo. |
| w języku angielskim | Nazwisko, X. (Ed.). (rok). *Tytuł książki*. Wydawnictwo. |
| Rozdział w pracy zbiorowej | Nazwisko, X. (rok). Tytuł rozdziału. W: Y. Nazwisko, Z. Nazwisko 2 (red.), *Tytuł książki* (s. strona początku–strona końca). Wydawnictwo. https://doi.org/xxx lub https://xxx. |
| **Inne prace** | |
| Raport: autor indywidualny | Nazwisko, X. (rok). *Tytuł raportu*. Wydawnictwo. |
| autor instytucjonalny | Nazwa instytucji. (rok). *Tytuł raportu*. Wydawnictwo. |
| Working Papers | Nazwisko, X. (rok). *Tytuł pracy* (nazwa serii i numer). https://doi.org/xxx lub https://xxx. |
| Sesja konferencyjna / prezentacja / referat | Nazwisko, X. (rok, dzień i miesiąc). *Tytuł pracy* [typ wystąpienia, np. referat]. Nazwa konferencji, miejsce konferencji. |
| Rozprawa doktorska: nieopublikowana | Nazwisko, X. (rok). *Tytuł pracy* [niepublikowana rozprawa doktorska]. Nazwa instytucji nadającej tytuł doktorski. |
| opublikowana | Nazwisko, X. (rok). *Tytuł pracy* [rozprawa doktorska, nazwa instytucji nadającej tytuł doktorski]. https://xxx. |
| Akt prawny | Pełny tytuł aktu prawnego wraz z datą publikacji w dzienniku urzędowym. |

| Typ publikacji | Przykład opisu bibliograficznego |
|---|---|
| **Strona internetowa** | |
| Znana data publikacji, zawartość strony się nie zmienia | Nazwisko, X. (rok, dzień i miesiąc). *Tytuł*. https://xxx. |
| Nieznana data publikacji, zawartość strony się zmienia | Nazwa instytucji. (b.r.). *Tytuł*. Pobrane dzień, miesiąc i rok pobrania z https://xxx. |
| **Zbiór danych** | |
| Surowe dane nieopublikowane | Nazwisko, X. (rok wydania pracy, w której dane są wykorzystywane) [opis danych, np. surowe dane nieopublikowane dotyczące…]. Źródło danych (np. nazwa uniwersytetu). |
| Dane opublikowane:<br>znana data publikacji, zawartość zbioru się nie zmienia | Nazwisko, X. (rok). *Nazwa zbioru danych* [zbiór danych]. Wydawca. https://xxx. |
| nieznana data publikacji, zawartość zbioru się zmienia | Nazwa instytucji. (b.r.). *Nazwa zbioru danych* [zbiór danych]. Wydawca. Pobrane dzień, miesiąc i rok pobrania z https://xxx. |

Źródło: opracowanie na podstawie: American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th edition). https://doi.org/10.1037/0000165-000.

**Praca przygotowana w sposób niezgodny z powyższymi wskazówkami będzie odesłana do autora z prośbą o dostosowanie formy artykułu do wymogów redakcyjnych.**

# DZIAŁY „WS" – TEMATYKA ARTYKUŁÓW
## *WS* SECTIONS – TOPICS OF THE ARTICLES

Pełny opis zakresu tematycznego działów: ws.stat.gov.pl/AimScope
Description of the topics covered in each section: ws.stat.gov.pl/AimScope

### Studia metodologiczne / Methodological studies

- Oryginalne teoretyczne rozwiązania metodologiczne ze wskazaniem ich praktycznej użyteczności
- Prace przeglądowe i porównawcze oraz dotyczące etyki w statystyce, które wnoszą pionierski wkład poznawczy do obecnego stanu wiedzy

### Statystyka w praktyce / Statistics in practice

- Nowatorskie zastosowania narzędzi i modeli statystycznych oraz analiza i ocena statystyczna zjawisk społeczno-ekonomicznych i innych, prowadzona w szczególności na danych z zasobów statystyki publicznej
- Wykorzystanie narzędzi informatycznych do uzyskiwania i przetwarzania informacji statystycznych, naliczania danych wynikowych, ich prezentacji i rozpowszechniania
- Projektowanie badań statystycznych, uzyskiwanie, integracja i przetwarzanie danych oraz generowanie wynikowych informacji statystycznych i kontrola ich ujawniania

### Studia interdyscyplinarne. Wyzwania badawcze / Interdisciplinary studies. Research challenges

- Wyzwania badawcze wynikające z rosnących potrzeb użytkowników danych statystycznych i wymagające zaangażowania znacznych środków oraz rozwiązań z różnych dziedzin nauki i techniki
- Wykorzystanie technologii informacyjnych i komunikacyjnych, innowacyjność, przetwarzanie i analiza zagadnień związanych z data science i big data
- Wyniki badań prowadzonych przez przedstawicieli dyscyplin innych niż statystyka z wykorzystaniem metod statystycznych

### Spisy powszechne – problemy i wyzwania / Issues and challenges in census taking

- Propozycje rozwiązań – zarówno organizacyjnych, jak i metodologicznych – możliwych do zastosowania w spisach oraz rezultaty analiz danych spisowych
- Praktyczne aspekty związane z gromadzeniem i udostępnianiem danych ze spisów, w tym dotyczące obciążenia odpowiedzi i ochrony tajemnicy statystycznej

### Edukacja statystyczna / Statistical education

- Metody i efekty nauczania statystyki oraz popularyzacja myślenia statystycznego i rzetelnego posługiwania się informacjami statystycznymi
- Problemy związane z kształceniem w zakresie umiejętności stosowania statystyki na wszystkich poziomach edukacji, a także dotyczące wykorzystywania nowoczesnych koncepcji i metod dydaktycznych oraz pomocy naukowych w nauczaniu statystyki

### Z dziejów statystyki / From the history of statistics

- Historia prowadzenia obserwacji statystycznych oraz rozwoju ich metodologii i narzędzi
- Życie i osiągnięcia zawodowe wybitnych statystyków, jak również działalność najważniejszych instytucji i organizacji statystycznych w Polsce i za granicą

### In memoriam

- Nekrologi i artykuły wspomnieniowe

### Informacje. Recenzje. Dyskusje / Discussions. Reviews. Information

- Teksty nierecenzowane i niemające charakteru artykułów naukowych: sprawozdania z konferencji naukowych i innych wydarzeń dotyczących statystyki, recenzje książek, omówienia nowości wydawniczych GUS, polemiki i dyskusje