Transport Geography Papers of PGS

2022, 25(4), 74-93

DOI 10.4467/2543859XPKG.22.023.17147

Otrzymano (Received): 17.11.2022
Otrzymano poprawioną wersję (Received in revised form): 28.12.2022
Zaakceptowano (Accepted): 28.12.2022
Opublikowano (Published): 31.12.2022

EFFECTS OF CHOICE OF DATA AGGREGATION METHOD TO A POINT ON WALKING ACCESSIBILITY RESULTS USING THE G2SFCA METHOD

Wpływ wyboru metody agregacji danych do punktu na wyniki dostępności pieszej uzyskane przy wykorzystaniu metody G2SFCA

Łukasz Lechowski

Institute of Urban Geography, Tourism Studies and Geoinformation, Faculty of Geographical Sciences, University of Lodz, Kopcińskiego 31, 90-142 Lodz, Poland

e-mail: lukasz.lechowski@geo.uni.lodz.pl

https://orcid.org/0000-0001-8919-2173

Citation: Lechowski Ł., 2022, Effects of choice of data aggregation method to a point on walking accessibility results using the G2SFCA method, *Prace Komisji Geografii Komunikacji PTG*, 25(4), 74-93.

Abstract: In spatial accessibility studies based on market areas, such as floating catchment area (FCA) family methods, it is crucial to identify the point to which weights are assigned, both on the demand and supply side. Bearing in mind that it is not always possible to work on disaggregated data, the aim of this paper was to investigate which method of determining a point, minimises bias in the estimation of walking accessibility. The research used the G2SFCA method, introduced by Dai, which has been employed several times to model walking accessibility. Results clearly show that point location methods for area units, based on disaggregating data to buildings, perform better at the scale of statistical districts or cadastral precincts, compared to those based on the centrally weighted mean. They also show that positional measures such as the Euclidean centrally weighted median can improve the results of analyses in units that are heterogeneous in terms of settlement network pattern.

Keywords: spatial accessibility, walking, GIS, 2SFCA, data aggregation

Introduction

Accessibility analysis, understood as the possibility of a relationship between two or more points (Śleszyński, 2014) requires the identification of source and destination points. Ideally, accessibility studies should be based on disaggregated data (Bryant, Delamater, 2019). Due to privacy concerns, it is common for data, such as population, to be aggregated to larger spatial units. The location of a point representing aggregated data directly affects the results of spatial accessibility analysis (Stepniak, Jacobs-Crisioni, 2017). Knowing how the method of determining the location of points representing aggregated data affects accessibility values can help minimize spatial accessibility estimation errors. This is particularly important in urban studies based, for example, on pedestrian traffic, in which even an insignificant change in the location of the aggregation point, at a low walking speed significantly affects the final accessibility result.

There are many methods for measuring spatial accessibility. Initially, accessibility analyses considered only the distance or travel time to the nearest service (Ikram et al., 2015). However, they did not consider weights representing the potentials of both the travel origins and destinations. Over time, cumulative accessibility began to be used, defined as the number of opportunities that can be reached within a given distance or travel time (Handy, Niemeier, 1997; Śleszyński, 2014). However, this method still does not consider the weights of individual destinations, and it also assumes that spatial accessibility is homogeneous across the assumed time or distance range (Rosik, 2021; Śleszyński, 2014). In the late 1950s, W. G. Hansen (1959) introduced gravity models to study potential accessibility. These models made it possible to differentiate accessibility according to the distance between source and destination using the distance decay function. Potential accessibility considers all the model's node linkages (origin-destination), the weight of individual destinations, and the distance decay function (Komornicki et al., 2015). Typically, potential accessibility studies use s-shaped, power, exponential, log-logarithmic or Gaussian functions (Bauer, Groneberg, 2016; Geurs, Ritsema van Eck, 2001; Rosik, 2012). A limitation of potential methods is that they still do not consider competition between market areas of supply-side locations, promote centrally located areas, and are sensitive to edge effects, between analysed units (Stępniak et al., 2017; Śleszyński, 2014). To address some of the above-mentioned limitations methods from the floating catchment area (FCA) family may be used to study spatial accessibility. They consider the supply and demand sides separately (Stępniak, 2013). The method proposed by (Peng, 1997) has undergone many modifications to account for the distance decay function, different extents of sub-market areas, multimodal transportation, spatio-temporal variation in accessibility (Bryant, Delamater, 2019; Park, Goldberg, 2021). Thanks to the advances made in access to spatial data, including the transportation network, and the development of GIS tools and techniques, methods from the FCA family have become extremely popular. One of their variations is the Gaussian two-step floating catchment area (G2SFCA) method proposed by D. Dai (2010). It is a variation of the two-step floating catchment area (2SFCA) approach, in which, a Gaussian function is used as the distance decay function (Dai, 2010). It prefers the selection of the nearest destination points to the health facility (Del Conte et al., 2022). Either in its initial or modified versions, it was used in studies of accessibility to health facilities or parks (Dai, 2010; Del Conte et al., 2022; Li et al., 2022).

There are many concepts of urban spatial development, based primarily on walking, bicycle, public transportation-based accessibility. This type of movement underlies the idea of compact, 15-minute, resilient cities (Abdelfattah et al., 2022; Moreno et al., 2021; Stevenson et al., 2016). It has a positive impact on the activity of seniors, thus extending their lifespan in good physical condition (Chen, Pan, 2020). Previous studies highlight the special role of pedestrian accessibility to primary health care services (Loo, Lam, 2012). It has been shown to significantly reduce hospitalizations reducing spending, needed to secure health services (Chen, Pan, 2020). It is usually noted that even 30 minutes of physical activity every day may lead to weight reduction or control, reducing the risk of overweight-related diseases and lead to a longer life in good physical condition (Daly et al., 2018; Rahman, Nahiduzzaman, 2019).

Spatial accessibility analysis is also applicable for research on social and spatial inequalities in access to services. It can be the basis for designating shortage areas for service provision (Daly et al., 2019; Lechowski, Jasion, 2021). Its results are often considered during the design of relevant health or environmental policies (Maarseveen et al., 2019; Tan et al., 2021). The FCA methods are also an efficient tool for modelling the effects of random events and natural disasters, to some extent indicating the resilience of cities, urbanised or rural areas to the occurrence of these events (Pregnolato et al., 2017; Wiśniewski et al., 2020). On their basis, decisions are made about which areas should be targeted first with programmes to minimise the impact of an occurrence (Wiśniewski et al., 2020). The health and lives of residents may depend on these decisions. Therefore, the results of accessibility analyses should reflect a repeatable, reality-based, and reliable accessibility of the study area to selected services. And this requires understanding how the location of the points themselves, representing aggregated units, affects the results of the analysis.

The question how data aggregation affects the results of accessibility analysis stems from the classic modifiable area unit problem (MAUP), which has been highlighted by S. Openshaw (1983), C. Jacobs-Crisioni et al. (2014), M. Stepniak and C. Jacobs-Crisioni (2017) or C. Wu et al. (2020), among others. It can refer both to scale and edge effects (Jacobs-Crisioni et al., 2014). Regarding spatial accessibility studies, it was addressed primarily in terms of potential accessibility (Stępniak, Rosik, 2015). In this view, the individual locations included in the study represent both sources of travel and destinations for other locations. Aggregation, if necessary, undergoes both sides of the model in the same way. For methods of the FCA family, destination locations are often disaggregated and do not coincide with source locations. There are just several works that address the impact of data aggregation on spatial accessibility results obtained using FCA methods. P. Apparicio et al. (2008) showed that a minimum of MAUP error is achieved by using the smallest aggregated spatial units possible. J. Bryant and P. L. Delamater (2019) noted that for the enhanced two-step floating catchment area (E2SFCA) method, the distance decay function adopted, and the change in shape of individual neighbouring spatial units at levels of aggregation, is important. To the author's knowledge, none of the previous work on the effect of data aggregation on spatial accessibility results obtained with the FCA family of methods has investigated how the point-determination approach for areal units itself interacts with these spatial accessibility results.

Attempting to use data compiled in this way in FCA surveys can lead to three types of biases:

A) Resulting from distance estimation from the aggregated point to services located outside the boundaries of the aggregated spatial unit.

B) Resulting from estimating the distance from an aggregated point to services located in the same spatial unit.

C) Related to the incorrect assignment of an establishment to an aggregated spatial unit (Stępniak et al., 2017).

M. Stepniak and C. Jacobs-Crisioni (2017) proved that when working on aggregate data, there are certain factors that affect the differences in travel times across the different spatial units studied. These factors contributed to the different access times of trip start and end points to the transport network, termed handicap or penalty in the literature (Stępniak et al., 2017). In that research, however, they were unable to identify clearly what these factors were. Thus, it was not possible to determine how to estimate the time to access the network from both the point of origin as well as the point of destination handicaps (Stępniak et al., 2017).

Bearing in mind that it is not always possible to work on disaggregated data, the aim of this paper was to investigate which of five tested methods of determining a point (mean centre, weighted mean centre, median centre, weighted median centre, central distance), minimises bias in the estimation of walking accessibility to primary health care (PHC) facilities. The research used the Gaussian two step floating catchment area method (G2SFCA), introduced by D. Dai (2010), which has been employed in its original or modified version to model, among other things, walking accessibility (Cao et al., 2022; Del Conte et al., 2022; Li et al., 2022).

The following research questions were posed for the purpose:

A) Do methods based on disaggregating data using finer auxiliary data and then subject to reaggregation produce better results than methods based solely on aggregating data using measures of central tendency?

B) Does assigning weights that represent population to points that and using them to measure walking accessibility yield better results?

C) Does the level of data aggregation affect the results achieved?

D) Whether the selection of the best method of point aggregation depends on the threshold for catchment area adopted?

This paper is structured as follows: The following chapter presents the theoretical background, discusses the state of the art of the FCA methods used so far, discusses the MAUP problem and analyses the state of the art of data aggregation to a point. The next part describes the research area and critically evaluates the data and methods used. The results of the research are then presented in reference to the research questions posed in the Introduction. In the following part, the results of the study are discussed with the state of the art, and paper findings are highlighted.

1. Theoretical background

1.1. Two-step floating catchment area methods (2SFCA)

Spatial accessibility, is a principal issue that has been addressed for decades in geographical research, especially related to transport geography. The very concept of accessibility, traditionally understood as a measure of the ease of interaction, has evolved many times, towards, for example, potential accessibility, which

Hansen defined as the potential of opportunities for interaction (Hansen, 1959). The dynamic growth of spatial accessibility applications in research and planning practice has been coupled with the development of GIS and the increasing availability of disaggregated data that can be used in transport network-based analyses (Neutens, 2015). The abundance of data, as well as the growth in computing power and the development of data processing algorithms, has led to the fact that today, spatial accessibility models often consider not only the location of origins and destinations, but also the modularity of transport, the spatiotemporal accessibility of the actors involved in the analysis and advanced transport network parameters like real time traffic flows, closed roads. As a result, the analysis can be presented not only statically but also temporally, considering both seasonality and daily fluctuations of the phenomenon (Del Conte et al., 2022; Du, Zhao, 2022; Park, Goldberg, 2021).

Methods from the FCA group, together with the potential quotient method, the doubly constrained spatial interaction model, are part of a group of methods for studying spatial accessibility that take competition effects into account (Rosik, 2021). Those derived from the 2SFCA method are special cases of the gravity model of spatial interaction (Rosik, 2021), and are commonly used in studies of spatial accessibility to health services (Stępniak, 2013). Hence, the concept of the 2SFCA method itself is worth explaining.

In its simplest form, the 2SFCA calculation procedure involves two steps (Stępniak, 2013).

Initially, for each service, e.g., a primary healthcare centre (PHC), the catchment area threshold is determined. This can be defined by any impedance, such as a time or distance unit. In each such market area, the potential of a given facility is divided by the total population located within the threshold.

$$D_j = \frac{A_j}{\sum_{i \in (d_{ij} < d_{max})} P_i}$$

where: A_j – the weight of a given service, Pi – population-weighted centroid located in each market area. In the second stage, a market area is set for each residence, in which all the shares of the facilities are summed up.

$$S_i = \sum_{j \in (d_{ij} < d_{max})} D_j$$

2SFCA methods are used to measure the spatial accessibility of customers to selected services (Wiśniewski et al., 2020). Unlike index-based measures of accessibility (e.g., indicators of infrastructure equipment, density indicators such as number of doctors per ten thousand people), it considers both the demand and supply side, which significantly increases the reliability of the results. First used by Z.-R. Peng (1997) to analyse labour market accessibility in Portland, Oregon, it has seen many modifications, mitigating its initial limitations. In the E2SFCA method, a decay function was introduced that differentiates the strength of the impact of each destination according to its distance to the origins under study (Huhndorf, Działek, 2018; Luo, Qi, 2009). Commonly used for this purpose are exponential, power or Gaussian distribution functions (Stępniak et al., 2017). A variation of 2SFCA based on the last function is the G2SFCA method (Dai, 2010). In the literature, one can additionally find 2SFCA methods based on Epanechnikov (e.g., KD2SFCA) or quartic function (enhanced KD2SFCA) (Dai, Wang, 2011; Polzin et al., 2014). The problem of a pre-defined size of the market area was solved in the 3SFCA method (Shatnawi et al., 2022; Wan et al., 2012) by introducing different ranges of market areas into the model. For each of these, a sub-accessibility was calculated and then added together to obtain the estimated total accessibility. In subsequent proposals, e.g., in the V2SFCA method (Du, Zhao, 2022) partial market areas up to the adopted boundary were created iteratively considering equal intervals. In recent years, with growing data availability and improved spatial accuracy, more and more attention has been paid to estimating spatial accessibility with respect to different transport modes (e.g., M2SFCA methods), temporal accessibility on the supply and demand side (PM2SFCA), and changes in mobility over time (Delamater, 2013; Luo et al., 2022). Variation in the location of origins over time is also considered to reflect the natural daily travel behaviour of residents (Park, Goldberg, 2021).

Methods based on 2SFCA have been widely used in studies of spatial accessibility to basic services, including medical services (Bauer, Groneberg, 2016; Chen, Pan, 2020; Dai, 2010; Huhndorf, Działek, 2018; Jin et al., 2019; Lechowski, Jasion, 2021; Stępniak et al., 2017), retail and shopping centres (Wiśniewski et al., 2020), parks (Cao et al., 2022; Dony et al., 2015; Luo et al., 2022), financial services (Langford et al., 2020).

1.2. Modifiable area unit problem (MAUP)

One of the issues addressed in the literature when applying FCA methods is the impact of data aggregation on spatial accessibility results (Bryant, Delamater, 2019); C. Jacobs-Crisioni et al. (2014); M. Stepniak and C. Jacobs-Crisioni (2017). As a rule, the use of non-aggregated data is recommended in this type of analysis (Bryant, Delamater, 2019). This is the only way to avoid the modifiable area unit problem (MAUP) studied by S. Openshaw (1983), G. Arbia (1989); C. Jacobs-Crisioni et al. (2014); M. Nalej (2018); M. Stepniak and C. Jacobs-Crisioni (2017).

The MAUP problem is commonly referred to two issues: the scale effect, and the zonation effect (Openshaw, 1983). The scale effect is the variability in results obtained for data aggregated to increasingly larger spatial units. It is caused by the uncertainty of how the data are aggregated to a certain number of surface spatial units (Fischer, Nijkamp, 2021; Openshaw, 1983). The edge effect refers to the situation where, variability in the results obtained is caused using alternative spatial units, while the number of subdivisions itself does not change (Fischer, Nijkamp, 2021; Openshaw, 1983). The MAUP problem indirectly results from the non-stationarity of the phenomenon under study and its heterogeneity within the spatial units adopted (Fischer, Nijkamp, 2021; Openshaw, 1983). MAUP is also a special case of 'ecological fallacy', which leads to aggregation error (Wu et al., 2020). In spatial accessibility analyses, the aggregation error arises from the need to use a point to represent spatial units that are area-based. While in the case of an area we assume that the population is evenly distributed and homogeneous, in the case of a point the entire population is assigned to a single location (Apparicio et al., 2008; Wu et al., 2020).

Typical for the scale effect is its influence on the average values and the variance of the results obtained. As a rule, due to central tendency levelling, as the area of analysed spatial units increases, the estimation error of the analysed characteristic between the values obtained for aggregated and disaggregated data should also increase. At the same time, the variance of this error should decrease. Previous research confirms this trend, although there are some deviations from the rule due to, among others, the spatial distribution of the population within the spatial unit, the characteristics of the transport network and the shape of the spatial unit itself (Bryant, Delamater, 2019).

In the literature, there are many ways of determining the location of points, representing surface units, which directly affect the size of the aggregation error. In spatial accessibility studies, two approaches for converting a polygon into a point dominate. The first is to generate the centroids of the polygons. While this approach can be consistent in terms of the overall direction of accessibility with disaggregated data, at the local level it can generate significant errors that can lead to misinterpretations of spatial accessibility (Bryant, Delamater, 2019). The second approach, called population weighted centroid, is more commonly recommended in the literature. It assumes the use of more accurate data representing the number of inhabitants in units with a higher level of detail to determine the centroid for spatial units aggregated to a larger unit (Jacobs-Crisioni et al., 2014). Even in this case, errors due to aggregation can be up to 10% compared to census tract data (Apparicio et al., 2008).

In classic location theories, such as Weber's model, the median centre is also often used. It was the basis for the determination of Weber's location equilibrium, which minimises the distance of a point determined in this way to all points included in the same spatial unit (Fischer, Nijkamp, 2021). So far, this method has not been verified for its impact on aggregation errors of spatial accessibility measures.

Another way to mitigate aggregation error is to disaggregate data in advance from more detailed data and use them for spatial accessibility analysis Disaggregation has so far been performed in a number of ways, such as using the dasymetric mapping technique Different datasets have also been used for disaggregation, e.g. address data, cadastral and household information (Wu et al., 2020), land use (Boone, 2008), built environment data (Pham et al., 2012), buildings (Logan et al., 2017). This approach reduces aggregation error, but due to the time required and hardware capabilities, it is difficult to carry out in surveys covering larger spatial units. n the latter case, it is often necessary to import surface data to represent a point (Stępniak et al., 2017). In the latter case, it is often necessary to import surface data to represent a point (Stępniak et al., 2017). This requires selecting the point in such a way that the influence of the MAUP effect on the results is as small as possible.

2. Data and methods

2.1. Study area

The study was conducted for area units located in Lodz, the fourth most populous city in Poland (~700,000 inhabitants) and the capital of the Łódzkie Voivodeship. The city is characterised by dense built-up areas in the centre, large green and recreational areas located in the northern and western parts of the city. It has no major natural spatial barriers of a linear nature. The central part of the city to the south-east and west is separated from the rest of the city by a railway track, which limits walking accessibility (Fig. 1).

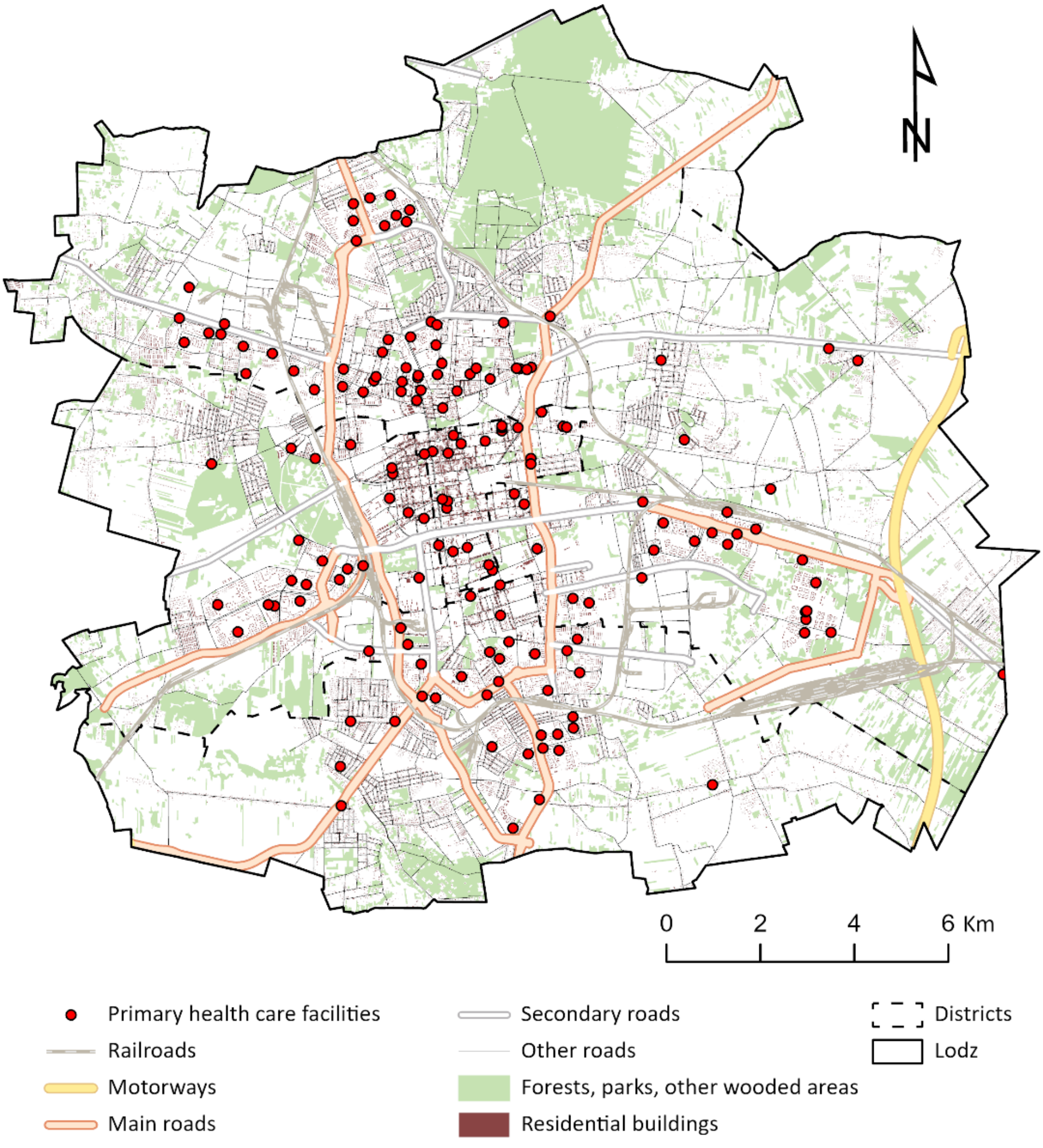

Fig. 1. Study area.

Source: own elaboration.

2.2. Data

To study the walking accessibility of Lodz's inhabitants (defined as the persons registered in Lodz) to the PHC facilities, disaggregated data were used: a) registered population, assigned to addresses, b) buildings, including information on their area, number of floors, function, c) detailed information on the PHC facilities contracted with the National Health Fund d) geometric and attribute information on the transport network (Tab. 1). Polygon layers containing the areas of census tracts, statistical districts, and registration precincts and districts were used to define the boundaries for data aggregation. In addition, the study used information from the 2021 national census to include the population of statistical localities directly adjacent to the study area.

Tab. 1. Spatial data used in the study.

Data used	Count	Data provider[1]	Application
Primary Healthcare facilities	270	NFZ	Destinations for OD Cost Matrix analyses, showing travel times from origin to each destination in the catchment area adopted
Registered population	38,849[2]	UMŁ	Setting of benchmark spatial accessibility, aggregation of population to census tracts, statistical districts, registration precincts, districts
Roads	16 691[3]	OSM	Network dataset design for G2SFCA analysis, OD Cost Matrix analyses
Dwellings	52,894	BDOT	Disaggregation of data from census tracts to dwellings, delimitation of points representing the units included in the different levels of data aggregation under examination
Census tracts	3,790	PRG	Basic level of population data aggregation, determining the location of points that represent the studied areal units
Statistical districts	731	PRG	Determining the effect of aggregation level on variation in aggregation errors
Registration precincts	215	PRG	Determining the effect of aggregation level on variation in aggregation errors
Districts	5	PRG	Determining the effect of aggregation level on variation in aggregation errors

[1]NFZ – National Health Fund, UMŁ – Lodz City Council, OSM – Open Street Map, BDOT – National Database of Topographic Objects, PRG - National Register of Boundaries

[2] Number of address points with at least one person registered at the address

[3] Total length of roads included in the analysis

Source: own elaboration.

Primary healthcare facilities (PHC)

The study was based on an analysis of the walking accessibility to PHC facilities by people registered in Lodz. A table containing addresses of health facilities was obtained using *Simple Object Access Protocol* (SOAP) services with the permission of the National Health Fund. This included all PHC facilities that, as of 2 February 2020, had a valid contract with the National Health Fund to provide healthcare services. Based on the address data, they were assigned to address point using the GeoPy package. The GeoPy package allows location assignment based on address data, using one of the geocoding web services such as Nominatim, ArcGIS or MapQuest (Lechowski, Jasion, 2021). To support the research, facilities located not only in Lodz, but also those in neighbouring towns and cities were selected. It included all PHC facilities accessible from places of residence (origins) within a 30-minute walk.

Population data

The data on the spatial distribution of people registered in Lodz, aggregated to addresses, were obtained from the Lodz City Council, up to date as of 31.12.2017. They served as the basis for determining the benchmark accessibility to PHC facilities for aggregated spatial units. These data were also geocoded, with the GeoPy package, using Python programming.

One should bear in mind that there is a 1.5-year time lag between the acquisition of data on the location of PHC facilities and the location of people registered in Lodz. The time gap may slightly affect the results obtained for spatial accessibility. However, it is too short a time to assume that this influence is significant. Population data based on the number of registered persons is an estimate and does not always correspond to reality. Not everybody fulfils their obligation to register. This applies both to people who rent flats and those who have moved in recent years.

As a result, as in the case of Central Statistical Office (CSO) statistics, this may lead to an overestimation of accessibility values in suburban areas and their underestimation in city centres (Śleszyński, 2011).

Data on the number of people registered in each location was collected only within the city of Lodz. The author did not have such detailed information for the boroughs neighbouring the city. Failure to consider the distribution of the population in the localities directly neighbouring Lodz could lead to an edge effect and a significant overestimation of the accessibility to the PHC facilities at the border of the study area (Stępniak et al., 2017). To mitigate this effect, population numbers were assigned to the statistical localities in the vicinity of Lodz, based on the 2021 census. This approach does not eliminate the errors caused by using two sets with populations collected in different time periods and defined differently. However, these distortions will still be much smaller, and they do not have such a significant impact on the main objective of the study. This is because the aggregated spatial units analysed were based on the same input data set.

Network dataset

Data from the OpenStreetMap (OSM) road layer was used to construct the transport network. Roads of type ('motorway_link', 'motorway', 'trunk', 'trunk_link') not accessible to pedestrian traffic were excluded from the analysis. All sections not linked to the network were also removed. The study assumed that a pedestrian travels at 4.7 km/h, which corresponds to the average speed of an apparently healthy person (Murtagh et al., 2021). This is in line with the most assumed walking speeds of between 4 and 4.8 km/h (Rakower et al., 2011).

Aggregation units

To identify the impact of the level of data aggregation on aggregation errors, the study used existing administrative and statistical divisions, suitable for different spatial scales of analysis.

Census tracts (n=3,790) were used as the basis for data aggregation. The finest spatial unit studied for accessibility to PHC was statistical districts (n=731). Surveys were also conducted at the level of registered precincts (n=251) and districts (n=5). The study also used statistical localities to estimate the number of residents of settlement units located outside the study area but within the market area of the respective PHCs. The delimitation of census tract boundaries, and registered precincts, is linked to the number of inhabitants. The legislator has assumed that there can be no more than two hundred dwellings and five hundred inhabitants in a single census tract, while for statistical districts, 2,700 persons and 999 dwellings respectively (Regulation 1998). Thus, census districts and statistical regions have a more homogeneous character in terms of the spatial structure of housing areas, compared to registered precincts or districts.

Disaggregation unit

The research also made use of the buildings layer recorded in the national database of topographic objects. Among the buildings, dwellings were distinguished, and the estimated number of registered persons was assigned to them. The estimation was based on CSO data presenting the average number of people living in a dwelling. The estimation was based on CSO data presenting the average number of people living in a residential unit. The method of assigning registered persons to a building is explained in the next subsection (Dai, 2010).

2.3. Methods

A fundamental limitation of the 2SFCA method proposed by W. Luo and F. Wang (2003) was the dichotomous approach to accessibility within market areas, regardless of the distance between origins and destinations. The distance decay functions introduced by (Luo, Qi, 2009) were a response to these limitations. Most commonly, gaussian, power or exponential functions are used for this purpose. Gaussian function was also applied in G2SFCA method, first used by (Dai, 2010) to assess racial inequalities in access to health services.

Like classic FCA methods, it consists of two stages. First, the market area on the supply is investigated. A Gaussian function is used as the distance decay function (Dai, 2010). It prefers the choices of the nearest destination points; hence it is well suited for the analysis of spatial accessibility to basic services such as health facilities, pharmacies (Dai, 2010; Del Conte et al., 2022). It is therefore useful especially when analysing accessibility for modes of transport characterised by low travel speeds, for walking, cycling, scooters, and public transport accessibility, where travel time or distance to the nearest facility may be of greater importance.

$$R_j = \frac{S_j}{\sum_{k\epsilon\{d_{kj}\le d_0} G(d_{kj}, d_0) P_k}$$

where: S_j – the weight of a given service, P_k – population-weighted centroid located in each market area, G – Gaussian distance decay function.

$$G\left(d_{kj}, d_0\right) = \begin{cases} \dfrac{e^{-\left(\frac{1}{2}\right)\times\left(\frac{d_{kj}}{d_0}\right)} - e^{-\left(\frac{1}{2}\right)}}{1 - e^{-\left(\frac{1}{2}\right)}} \\ 0 \end{cases}$$

In the second stage, for each P_k location all partial values of R_j multiplied by a Gaussian function are summed up. Its purpose is to differentiate the strength of the influence of individual locations, depending on the distance, between the demand and supply location sides.

$$A_i = \sum_{j\epsilon\{d_{kj}\leq d_0} R_j G(d_{kj}, d_0)$$

Benchmark data preparation

To determine the benchmark measure of spatial accessibility, disaggregated data was used, both in the case of address data, attributed to the number of registered persons, and on the supply side, in the form of the location of PHC facilities. the unit of accessibility in each study was the number of PHC facilities per person [facility/person]. Rather than decimals with exceedingly small values, the result was given in relation to ten thousand inhabitants. Subsequently, the results were aggregated using the arithmetic mean, as the indicator's invariant value (ratio) is its denominator (Hyndman, 2011).

Estimating the number of people registered in the building

Statistics Office data (CSO) and building attribute information – total building area, building function – were used to estimate the number of registered persons of individual buildings. Based on CSO[1] data it was found, that on average one dwelling had 2.17 inhabitants in 2017. The procedure for assigning weights to buildings was a three-stage process (Fig. 2). First, buildings with 1 or 2 dwellings were identified and, regardless of their area, the number of dwellings was multiplied by the average number of people living in the dwelling. People living in single-family houses and with two flats were subtracted from the total number of registered people. The remaining number of persons was used to calculate the average number of persons per square metre of housing in each census tract. This indicator was multiplied separately by the area of each multi-family building, thus obtaining the estimated number of persons per building in each census tract.

Methods of centroid localisation

Points representing area units were set based on measures of central tendency. For comparison, points were also determined as centroids of individual areal features. The study assumes that the centroids for statistical districts, cadastral precincts and districts will

[1] https://bdl.stat.gov.pl/bdl, accessed: 10.08.2022

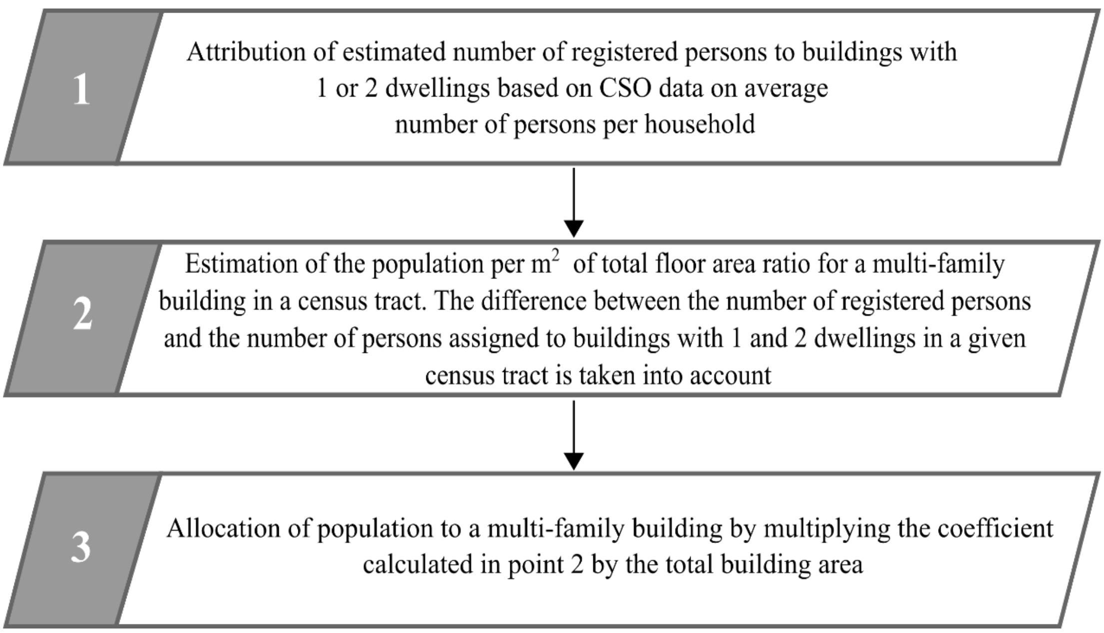

Fig. 2. Procedure for assigning the population to buildings.

Source: own elaboration.

be determined based on the centroids of the census tracts. Census tracts are the smallest aggregation unit of population data provided by the CSO. Their use allows the results obtained to be compared with other studies e.g., (Jacobs-Crisioni et al., 2014). It also has an application dimension, as the CSO does not share finer data. The number of registered persons was allocated to individual census tracts using population data provided by the Lodz City Council.

The study assessed the impact of point locations on spatial accessibility, determined by 5 different methods (Tab. 2). Central measures such as mean centre, median centre and central distance methods were used to determine the point. Kuenne can be used for this purpose (Burt et al., 2008). Using the median also to determine the focal point, it was tested whether minimising the distance between basic units improves the accessibility scores of the 2SFCA method. A similar measure to the median centre is the central distance method. The result of this method is the object for which the distance to all other features is minimum. This method was used to identify the most centrally located building in each of the spatial units studied. The OD matrix method available in ArcGIS software was used to perform the analysis. To make the analysis less time-consuming, only half of the buildings were selected from each spatial unit using a random number generator. The point representing the spatial unit was the building for which the total access time to the remaining houses was the lowest.

Tab. 2. Procedure for assigning the population to buildings.

Measure	Equation
Mean centre (MC)	$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}, \overline{Y} = \frac{\sum_{i=1}^{n} y_i}{n}$
Weighted mean centre (WMC)	$\overline{X} = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}, \overline{Y} = \frac{\sum_{i=1}^{n} y_i w_i}{\sum_{i=1}^{n} w_i}$
Median centre (MedC), central feature	MedC = $\min\{\sum_{i=1}^{n} \sqrt{(x_i - x_e)^2 + (y_i - y_e)^2}\}$
Weighted median centre (MedCw)	MedCw = $\min\{\sum_{i=1}^{n} w_i \sqrt{(x_i - x_{we})^2 + (y_i - y_{we})^2}\}$
Central distance (Cd)	Cd = $\sum^{n} \sum_{i=1}^{n-1} \sqrt{(x_i - x_{we})^2 + (y_i - y_{we})^2}$

Source: own elaboration.

Measures based on the arithmetic mean or weighted arithmetic mean are highly sensitive to outliers (Jaworska, Suchecka, 2014), hence the high heterogeneity of census tract areas within a registration precinct or district led to significant aggregation errors. Position measures based on the median, like measures based on the arithmetic mean, are independent of the transformation of coordinate systems (Jaworska, Suchecka, 2014) at the same time they minimise the distances between all, considered points (Kellerman, 1981; Vasiliev, 1996). Weighted median centre (known also as weighted Euclidean median), with coordinates (X_{we}, Y_{we}) also minimises the weighted distances between the set of points under consideration. There is no explicit mathematical method for determining both the central median and the weighted central median, but the algorithm proposed by Kuhn and

Two data sets were used in the study to determine mean centres and weighted mean centres. Separately, points were determined for the area units analysed based on census tracts and based on buildings with their assigned, estimated number of persons registered.

Network analysis

The analysis of accessibility was conducted on a network layer prepared in ArcGIS Pro. Due to the purpose and specificity of the study area – a city, only walking accessibility was analysed. The time taken to reach a service was used as the unit of spatial resistance. Based on the literature, it was assumed that healthcare facilities should be accessible within a maximum of 30 minutes; optimally, they should be located within 20 minutes. To investigate the extent to which the results are robust to the assumed market area threshold, spatial accessibility was also examined for market areas of 10 and 30 minutes. The network model is simplified in nature. Travel constraints, such as barriers to movement by people (e.g., high kerbs, quality of pavement, badly parked cars), traffic lights, important for crossing the street, have not been considered. The consequence is an overestimated coverage of the market areas. On the other hand, these extents are identical for all spot determination methods, thus putting less strain on the final analysis result.

Estimation of aggregation error

The basic characteristics of accessibility were presented using basic descriptive statistics tools. A measure of the root mean square error (RMSE) and symmetric mean absolute percentage error (sMAPE) were used to assess the best fit of the estimated values to the corresponding observed ones. The RMSE somewhat resembles the standard deviation of the sampled observations. It indicates the magnitude of the error regardless of its sign (Longley et al., 2015).

$$RMSE = \left[\sum \delta x^2/n\right]^{1/2}$$

where: is the difference between modelled and observed values for i sample, n – number of observations

The results were also compared using the Symmetric mean absolute percentage error. This is a measure that represents the error as a percentage, which makes it independent of scale and allows comparison between different data sets (Hyndman, 2011).

$$sMAPE = \frac{1}{n}\sum\left|\frac{A_t - E_t}{(A_t + \frac{E_t}{2})}\right|$$

where: – observed values, E_{t-} forecast values

sMape takes on values between 0 and 200% and allows the calculation of the error also in case the observed values in the surveyed spatial unit take the value 0 (Hyndman, 2011).

3. Results

In Lodz, a good, although spatially differentiated, accessibility to PHC facilities was observed (Fig. 3). As

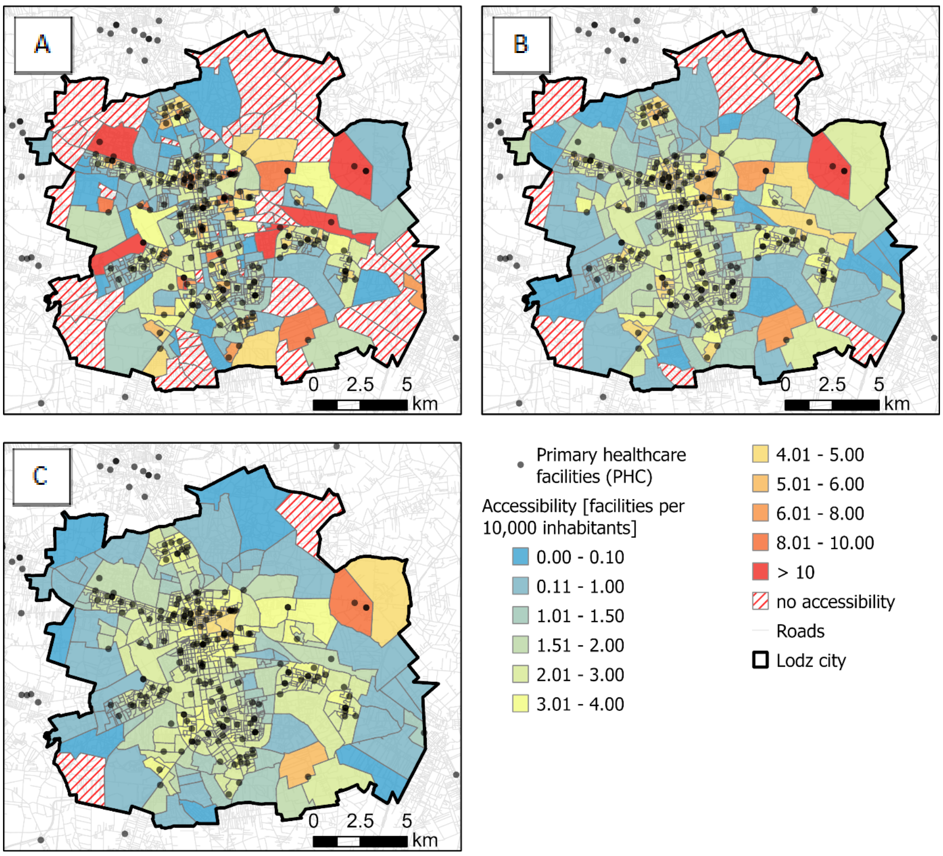

Fig. 3 Walking accessibility to primary healthcare facilities in Lodz in 2017 determined by the G2SFCA method for statistical districts with an assumed market area of a) 10 b) 20 and c) 30 minutes.
Source: own elaboration.

a rule, the inhabitants of the central part of the city and those living in large blocks of flats, such as Widzew or Retkinia, had a greater choice of health care centres within the reach of the market area than residents of the periphery. On the outskirts of the city, this accessibility often did not exceed one facility per 10,000 registered persons. As the market area increased, the number of individuals excluded decreased (n=55 for 10-minute coverage and n=3 for 30-minute coverage, respectively. An increase in market area simultaneously led to a decrease in the range between the maximum and minimum values of accessibility, leading to a decrease in the standard deviation for the community (Tab. 3). A statistically significant, negative relationship was observed between accessibility to PHCs and the characteristics describing the geometry of each spatial unit (its length and area), for market areas of 20 and 30 minutes. This relationship increased as the size of the market area increased. At the same time, the negative relationship between the number of registered persons and the walking accessibility to the PHCs in each time interval was small.

The means adopted to determine the points representing the statistical regions for the 20-minute market area had a slightly smoother distribution, compared to the baseline data (Tab. 4). Lower values were observed in the first quartile and higher values in the 3rd quartile for all measures discussed. At the same time, the methods based on arithmetic means had lower maximum values compared to the disaggregated value, while those based on the median had higher values. It was also tentatively found that, for statistical regions, measures based on population-weighted averages better represented observed values, as evidenced by both lower RMSE and sMAPE values. In addition, it was observed that disaggregating the data based on more detailed data, e.g., buildings, and then re-aggregating them can reduce estimation errors, equally well with or without the use of weights. Measures of location, minimising distances along the transport network to the remaining objects, however, did not yield the expected results, significantly exceeding the errors appearing in other methods.

Tab. 3. Relationship between walking accessibility to primary health care facilities and characteristics describing each statistical district (length of its border, area of district, number of people registered in each statistical district.

	Acc10	Acc20	Acc30	AREA	LEN	POP
Acc10	1.00					
Acc20	0.529	1.00				
Acc30	0.350	0.874	1.00			
AREA	0.011	-0.232	-0.304	1.00		
LEN	0.019	-0.290	-0.391	0.929	1.00	
POP	-0.062	-0.102	-0.131	0.128	0.190	1.00

where: Acc10, Acc20, Acc30 – Walking accessibility to services for the market area 10 20 and 30 minutes, AREA – area of statistical district, POP – persons registered in statistical district, p-value =0.01

Source: own elaboration.

Tab. 4. Descriptive statistics for spatial accessibility in statistical districts in the study area.

	count	min	max	std	avg	25%	50%	75%	RMSE	MAPE
Acc20	731.00	0.00	12.79	1.26	2.59	1.78	2.61	3.29	X	X
MC	731.00	0.00	12.12	1.45	2.60	1.66	2.58	3.39	0.66	14.74
CD	731.00	0.00	11.51	1.38	2.63	1.77	2.64	3.35	0.86	22.26
MED	731.00	0.00	14.16	1.47	2.62	1.74	2.59	3.42	0.59	12.48
MCB	731.00	0.00	12.28	1.42	2.60	1.68	2.62	3.36	0.53	11.37
WMC	731.00	0.00	11.30	1.39	2.61	1.72	2.60	3.39	0.54	12.79
WMCB	731.00	0.00	12.14	1.40	2.62	1.72	2.62	3.36	0.54	13.48
WMED	731.00	0.00	14.16	1.47	2.62	1.71	2.58	3.39	0.63	13.57

Acc20 – benchmark spatial accessibility, MC – mean centre-based accessibility, CD – central feature based on network dataset, MED – Euclidean Median Centre, MCB – residential building based mean centre, WMC – weighted mean centre, residential building based weighted mean centre, WMED – weighted Euclidean median centre
Source: own elaboration.

The spatial distribution of differences for the statistical regions shows the high variability of the values obtained in the individual models without population weighting (Fig. 4). The use of points based on positional measures (median, central distance) in the study led to a significant overestimation of accessibility values, relative to observed values. The points based on the central object determined using the transport network were the worst in this respect. The analysis also shows that large spatial units are much more likely to show a lack of walking accessibility, while accessibility, although small, does exist. The model based on the location of residential buildings better reflected spatial accessibility especially in those statistical areas where health facilities were located.

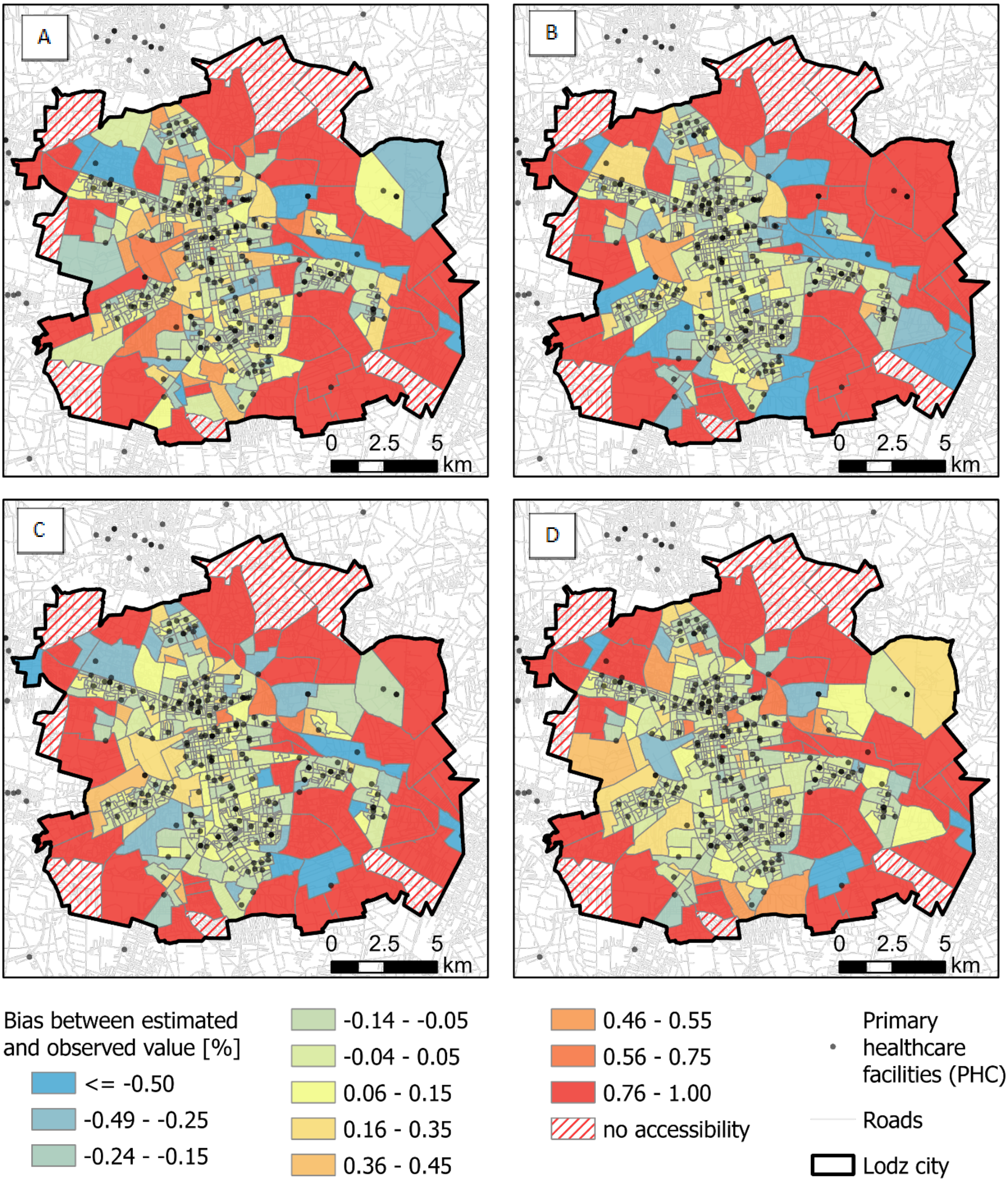

Fig. 4. Spatial distribution of differences between estimated and observed values for (a) RSMC (b) RSNetCF20, (c) RSMed20, (d) RSMCB20 (statistical districts).

Source: own elaboration.

The non-weighted accessibility scores achieved were compared with the accessibility measures that considered the weighting of the scores (Fig. 5). Few differences were observed between the classes with the largest and smallest differences in accessibility values. The improved results of the RMSE and sMAPE indices must therefore have been due more to the improved results for the intermediate classes, located mainly in the centre of the analysed area. Again, point determination methods based on positional measures led to the greatest overestimation of values. The use of building-based weights also often led to situations where accessibility was observed for census areas, but the analysed point generated based on buildings was outside the market area, or the point was much closer to the service than the analysis of the actual data would suggest. Such edge cases, however, were fewer than in the case of analyses based solely on the weighted average of the number of inhabitants in each census tract. At the same time, the RSWMC performed slightly better for small units with values close to the observed value.

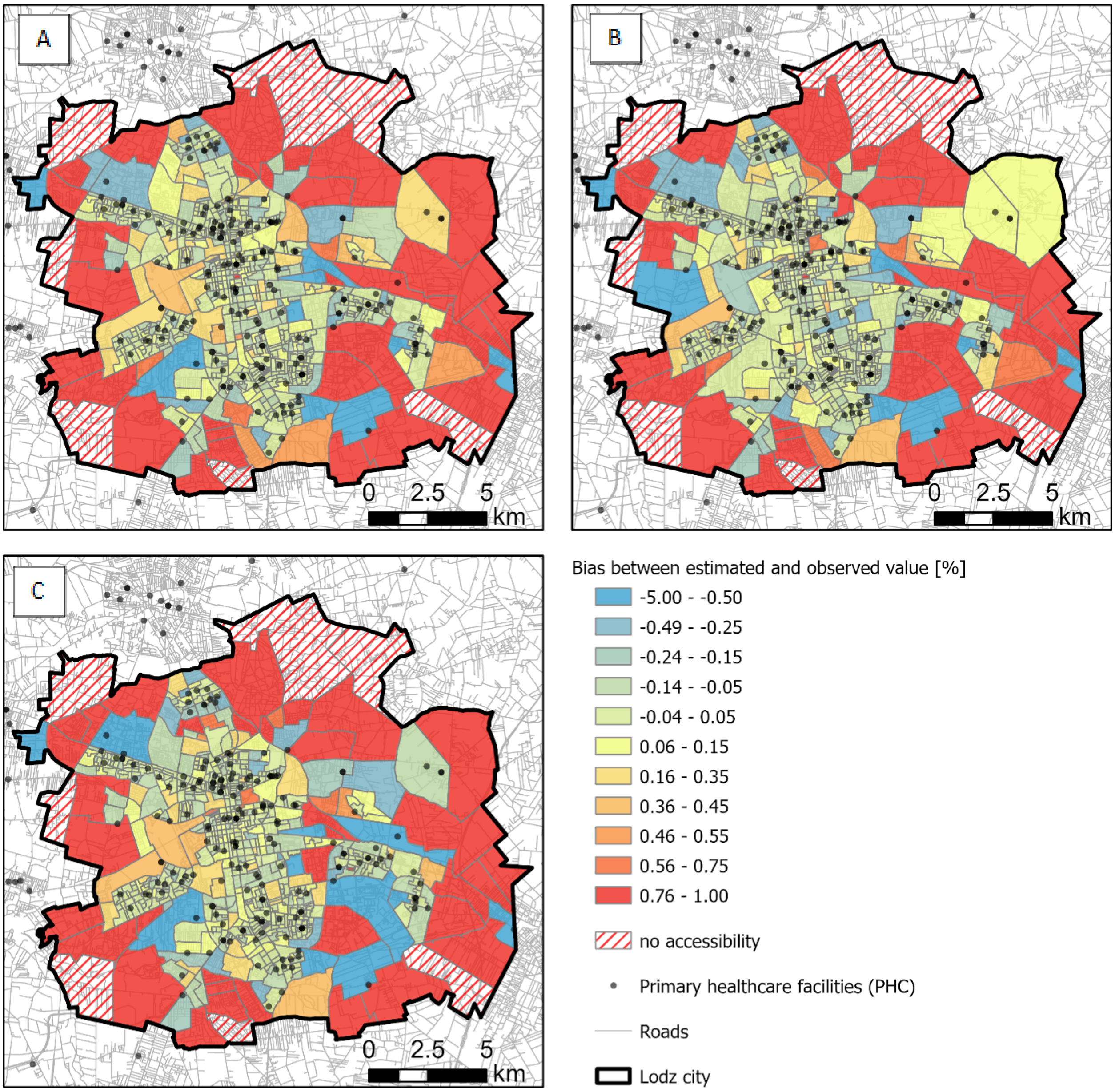

Fig. 5. Spatial distribution of differences between estimated and observed values for (a) weighted mean centre (WMC) (b) residential building based mean centre (WMCB) (c) weighted Euclidean median centre (WMED).

Source: own elaboration.

To find out to what extent the results obtained are dependent on the size of the market areas themselves, the RMSE and sMAPE results obtained for the tested methods of point generation for market areas of 10 and 30 minutes were compared (Tab. 5). The lowest, stable results for both RMSE and sMAPE were found for points generated based on the distribution of residential buildings (RSMCB). At the same time, comparably worst results were achieved for accessibility analysis based on the positional measures CD and WMED. They also showed that accessibility analysis based on the generation of the centroid of individual spatial units yields unstable RMSE and sMAPE results depending on the assumed size of the market area. When using point generation with weighted methods, the results obtained do not give a clear answer to what extent the use of weights improves the results obtained. The best performance is achieved by using weights to estimate the occupants of individual buildings, as evidenced by the relatively low RMSE and sMAPE values for the 10- and 30-minute areas. However, these values are not markedly different from those obtained using the unweighted spatial median.

had the lowest accessibility estimation errors relative to the benchmark data. The spatial median-based point generation method also had low values, especially for the 10 min interval. A highly overestimated score was observed for the 10 min area for the OEMCB. Analysis of the source data showed that this was due to the presence of one extremely high outlier. This was a precinct in which only nine people were registered (n=9 people) and to which two clinics were assigned, one of which, exclusive. Correcting the location of the points by introducing the weights of people registered in the buildings of the precincts surrounding the point in question, significantly changed its accessibility. Hence, significantly lower values were observed for the WMCB. The analysis showed that the averaged error values for the centroid of the units incorporating the weights (WMCB) were higher than the cases discussed earlier.

Different results were obtained by aggregating the data to the district level (Tab. 7). Low values of RMSE were accompanied by high values of sMAPE. It turned out that the most comparable results to the observed values were obtained when the points

Tab. 5. Comparison of the RMSE and MAPE results of accessibility to PHC facilities in statistical districts.

Time [min]	RMSE			sMAPE		
	10	20	30	10	20	30
MC	2.36	0.66	0.53	46.19	19.59	13.61
CD	2.21	0.86	1.36	34.62	17.52	11.90
MED	2.21	0.59	0.91	34.66	16.48	9.77
MCB	2.19	0.53	0.38	36.03	15.59	9.43
WMC	2.30	0.54	0.93	36.61	16.01	10.12
WMCB	2.19	0.54	0.94	34.66	15.29	9.61
WMED	2.24	0.63	0.93	36.47	17.30	10.15

MC – mean centre-based accessibility, CD – central feature based on network dataset, MED – Euclidean Median Centre, MCB – residential building based mean centre, WMC – weighted mean centre, residential building based weighted mean centre, WMED – weighted Euclidean median centre.
Source: own elaboration.

Literature on MAUP indicates that the size and shape of the units is a factor influencing the results. Therefore, the availability results achieved for the statistical districts were compared with the registration precincts (Tab. 6). As with the census districts, the average errors for both RMSE and MAPE decreased as the market area increased. At the same time, they were larger in the 10-minute and 20-minute ranges for the registration precincts than for the census districts. Characteristically, the RMSE errors for a market area of 30 min in the surveying precincts were slightly lower than for the census districts, but the sMAPE index was significantly higher for this area. Again, the measures based on disaggregated buildings (MCB and WMCB) were determined using positional measures such as central median, weighted central median. In contrast, poorer results compared to registration precincts were obtained both for the traditionally used population-weighted centroids and those based on the location of residential buildings. In the latter case, this can be interpreted by the dispersion of development around large compact green areas in the northern part of Lodz. Due to the dispersion of development around the forest, the point representing Baluty district was located inside this complex, which significantly deviated the compared pedestrian accessibility from the benchmark values.

Tab. 6. Comparison of the RMSE and sMAPE results of accessibility to PHC facilities by market area coverage for registration precincts.

Time [min]	RMSE			sMAPE		
	10	20	30	10	20	30
MC	18.04	4.30	0.58	72.29	69.12	48.07
NetCF	15.95	4.26	3.03	66.16	58.93	48.49
Med	18.03	3.62	0.58	69.27	64.60	45.26
MCB	77.05	3.55	0.50	69.53	56.48	38.54
WMC	37.42	6.94	0.94	78.99	77.86	54.48
WMCB	15.45	3.55	0.46	65.18	55.86	39.71
WMed	36.34	7.59	1.11	75.77	79.71	55.03

Source: own elaboration.

Tab. 7. Comparison of the RMSE and sMAPE results of accessibility to PHC facilities by market area coverage for districts.

Time [min]	RMSE			sMAPE		
	10	20	30	10	20	30
MC	2.61	1.60	1.44	179.72	131.48	95.62
NetCF	2.54	1.73	1.44	182.30	140.07	91.50
Med	2.40	1.21	1.05	160.46	74.78	46.24
MCB	2.70	1.69	1.49	198.04	144.87	94.80
WMC	2.63	1.22	1.09	185.40	84.52	50.45
WMCB	2.58	1.29	1.12	180.33	91.02	50.20
WMed	2.55	1.17	1.07	176.42	72.56	50.57

Source: own elaboration.

4. Discussion and conclusions

Methods from the FCA family are a valuable tool for assessing residents' accessibility to basic services. They work well for identifying shortage areas where action should be taken to improve access to key, quality-of-life services such as primary care facilities. However, these methods presuppose working on disaggregated and point data, which may not be feasible for reasons of equipment, organisation, or privacy concerns. Working on aggregated, population-based data, on the other hand, affects the results of accessibility studies, including FCA methods, as confirmed by P. Apparicio et al. (2017); M. Stepniak and C. Jacobs-Crisioni (2017); M. Stępniak and P. Rosik (2015); C. Wu et al. (2020) and this research. The MAUP effect, especially the scale effect, is clearly reflected in the increase in the relative error of the sMAPE analysed accessibility using the G2SFCA method, regardless of how the point is determined for areal units.

Studies have shown that reducing aggregated data, even at the level of census tracts to more precise elements, e.g., buildings, can reduce aggregation errors, provided that the settlement structure within a unit is homogeneous. Again, the results can be explained by MAUP and ecological fallacy effects (Wu et al., 2020). Minor aggregation errors apparent for data aggregated to buildings were observed for both statistical districts and registration precincts. When using districts, consisting of only five objects, disaggregation of the data, to determine the centroid from the data, is not applicable. If one considers that the absence of weights worked better for smaller units – statistical districts – and the use of weights for registration precincts, it can be concluded that modelling population for buildings only makes sense if the buildings within the unit are compact. Assigning weights to buildings also contributed to reducing the impact of outliers on the results, as shown in the study of the walking accessibility of statistical districts (compared to MCBs) with the threshold within 20 minutes. If, as in the case of districts (Fig. 7), it forms dispersed clusters of centroids it may be in a location far away from the services under study, underestimating the value of accessibility. Similar observations are found, for example, in the work of J. Bryant and P. L. Delamater (2019).

Based on the results obtained, it cannot be definitively concluded that the use of weights has a significant effect on reducing the aggregation error. This depended on the unit adopted and the method used for locating the point. Sometimes, on the contrary, it increases it. This is not consistent with the results obtained in other works, e.g., M. Stepniak and C. Jacobs-Crisioni (2017). However, one should consider that in the case of the potential availability both origins and destinations were subject to transformation and averaging. In the 2SFCA method, only the population was averaged. As a result, aggregation errors may depend on the type and distribution of amenities themselves, as pointed out by J. Hewko et al. (2002) among others.

The research also shows that the choice of the best method to generate a point from aggregated data depends on the level of aggregation. As a rule, it proved that for homogeneous or units consisting of only one cluster of buildings, it is better to rely on methods that disaggregate the units, e.g., to buildings. Of course, it would be best to use these data without further aggregation to a point, which is in line with other findings (Boone, 2008; Logan et al., 2017; Wu et al., 2020). If, however, this is not the case, the use of an average, or weighted average centre based on house distribution, may improve the results of the analysis. As the level of aggregation increases, smaller aggregation errors are obtained based on the generation of points by positional methods, those based on Euclidean median centre and weighted Euclidean median centre. Similar relationships were not found using the central distance method, based on a distance matrix generated along the transport network between all buildings located in the same area unit. This is an effect of the property of the measure, which minimises the Euclidean distances between the centroids of all census tracts, located cadastral precincts or districts.

The value of aggregation errors is also influenced by the size of the market area, which decreases as this size increases. The uncertainty and variation in the error results for a threshold of 10 minutes walking accessibility is clear. Based on the results obtained, it does not seem reasonable to make the choice of method dependent on the size of the threshold adopted – the size of the market area. Still, for statistical districts and census tracts, methods based on disaggregation of data to buildings were the best way to select points, while for districts, based on the Euclidean median centre.

However, the studies conducted have their limitations. First, they refer to a single, highly urbanised area and tested one chosen service in advance. J. Hewko et al. (2002) and F. Gao et al. (2017) found that aggregation error results may vary, depending on the amenity adopted. The adoption of a single study area, on the other hand, does not allow to verify whether the spatial, socio-economic structure of the analysed area can influence the aggregation errors obtained by different methods of locating points. Further studies should be made to verify whether the selection of units depends on the above-mentioned characteristics. The studies discussed here did not consider other distance decay functions used in accessibility studies. Consequently, they also did not assess whether the way centroid is placed should depend on the FCA family method used. The author considered the Gaussian function used in multimodal accessibility studies, including walking by (Cao et al., 2022; Del Conte et al., 2022; Li et al., 2022). The function favours destinations located at closer distances and assumed that up to a certain point, weight differences change little, then accelerate as distance increases. It is also relational in nature. It depends on the relationship between the distance to the service and the threshold. By adopting a higher threshold at the same distance from origin to destination, the value of the resistance function parameter will be higher. Since people always choose some PHC facility, but prefer those located at closer distances, it is appropriate for the study adopted. As the study is concerned with assessing the aggregation error, the question is whether the application of different functions causes a decrease or increase in the values and change the accuracy of the results about the geographical aggregation. Thus, future research should be extended to include the analysis of power and exponential functions (Rosik, 2012; Stepniak and Jacobs-Crisioni, 2017; Stępniak and Rosik, 2015). These studies also did not consider spatio-temporal variability on both the demand and supply side, which is important due to the daily life path of each actor in space (Park, Goldberg, 2021). In the case of the central distance-based method, due to hardware capabilities and analysis time, the method was tested on a random sample consisting of half of the residential buildings located in each aggregation unit. For small spatial units, this was sometimes only a few buildings. In such a small sample, the rejection of a large multi-family building could have led to a bias in the results of the method. However, since in large aggregation units such as districts the method also did not give the expected results, in the author's opinion, there are other more important factors that make it not the best choice of point location. Finally, it should be noted that there was a difference in the timing of the data acquisition of people's registrations and buildings. Consequently, new multi-family buildings commissioned between 2017 and 2019 contributed to a lower person per square meter ratio of total built area. Due to the lack of data, the model also does not consider rental housing.

Despite these limitations, the results clearly show that point location methods for area units, based on disaggregating data to buildings, perform better at the scale of statistical districts or cadastral precincts, compared to those based on the centrally weighted mean. They also show that positional measures such as the Euclidean centrally weighted median can improve the results of analyses in units that are heterogeneous in terms of settlement network pattern. Compared to the work of C. Wu et al. (2020) M. Stepniak and C. Jacobs-Crisioni (2017), it considers different point location techniques for surface objects. Both mentioned works recommended the population weighted technique, which gives smaller aggregation errors. The present study showed that this method based on mean centre is not necessarily the best. While techniques based on disintegrating the data and then weighting it with the mean work better for compact homogeneous residential areas, for larger areas it may be better to use point determination techniques based on positional measures like the median centre. To the author's knowledge, this is also the only paper that addresses this issue in terms of an accessibility study using a method from the FCA group, in which usually only one side of the analysis is aggregated.

Undoubtedly, further research should be conducted in this area. First, it should be verified whether spatial and socio-economic characteristics influence the choice of method for bringing surface units to a point. Moreover, the edge problem, as described by MAUP, should also be considered. Surveys based on census tracts or statistical districts naturally refer to population-weighted results. This is a direct result of their delimitation method. According to the Ordinance (1998), census districts may not have more than five hundred persons and two hundred dwellings and statistical regions 2,700 persons and 999 dwellings. Hence, in the future it should be checked, through the random method of delimitation, whether the results will be resistant to changes in their boundaries. It would also be necessary to compare the results obtained with outcomes obtained for other 2SFCA family methods, different distance decay functions and taking multimodal transport into account.

Abbreviations

2SFCA – two-step floating catchment area
CSO – Central Statistical Office
FCA – floating catchment area
OSM – Open Street Map
PHC – primary health care

References

Abdelfattah L., Desponte D., Fossa G., 2022, The 15-minute city: interpreting the model to bring out urban resiliencies, *Transport Research Procedia*, 60, 330-337.

Apparicio P., Abdelmajid M., Riva M., Shearmur R., 2008, Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues, *International Journal of Health Geographics*, 7, 1-14.

Apparicio P., Gelb J., Dubé A. S., Kingham S., Gauvin L., Robitaille É., 2017, The approaches to measuring the potential spatial access to urban health services revisited: distance types and aggregation-error issues, *International Journal of Health Geographics*, 16, 1-24.

Arbia G., 1989, *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*, Springer Dordrecht, Dordrecht.

Bauer J., Groneberg D. A., 2016, Measuring Spatial Accessibility of Health Care Providers – Introduction of a Variable Distance Decay Function within the Floating Catchment Area (FCA) Method, *PLoS One* 11(7), e0159148.

Boone C., 2008, Improving Resolution of Census Data in Metropolitan Areas using a Dasymetric Approach: Applications for the Baltimore Ecosystem Study, *Cities and Environment* (CATE) 1, 1-25.

Bryant J., Delamater P. L., 2019, Examination of spatial accessibility at micro- and macro-levels using the enhanced two-step floating catchment area (E2SFCA) method, *Annals of GIS,* 25, 219-229.

Burt J. E., Barber G. M., Rigby D. L. (eds.), 2008, *Elementary Statistics or Geographers*, 3 ed, The Guilford Press, New York, London.

Cao Y., Guo Y., Zhang M., 2022, Research on the Equity of Urban Green Park Space Layout Based on Ga2SFCA Optimization Method – Taking the Core Area of Beijing as an Example, *Land*, 11(8), 1323.

Chen T., Pan J., 2020, The effect of spatial access to primary care on potentially avoidable hospitalizations of the elderly: evidence from Chishui City, China, *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement,* 160(2), 645-665.

Dai D., 2010, Black residential segregation, disparities in spatial access to health care facilities, and late-stage breast cancer diagnosis in metropolitan Detroit, *Health Place,* 16, 1038-1052.

Dai D., Wang F., 2011, Geographic Disparities in Accessibility to Food Stores in Southwest Mississippi, *Environment and Planning B: Planning and Design*, 38, 659-677.

Daly M. R., Mellor J. M., Millones M., 2018, Do Avoidable Hospitalization Rates among Older Adults Differ by Geographic Access to Primary Care Physicians? *Health Services Research*, 53(Suppl 1), 3245-3264.

Daly M. R., Mellor J. M., Millones M., 2019, Defining Primary Care Shortage Areas: Do GIS-based Measures Yield

Different Results?, *The Journal of Rural Health*, 35, 22-34.

Del Conte D. E., Locascio A., Amoruso J., McNamara M. L., 2022, Modeling multimodal access to primary care in an urban environment, *Transportation Research Interdisciplinary Perspectives*, 13, 100550.

Delamater P. L., 2013, Spatial accessibility in suboptimally configured health care systems: A modified two-step floating catchment area (M2SFCA) metric, *Health and Place,* 24, 30-43.

Dony C. C., Delmelle E. M., Delmelle E. C., 2015, Re-conceptualizing accessibility to parks in multi-modal cities: A Variable-width Floating Catchment Area (VFCA) method, *Landscape and Urban Planning,* 143, 90-99.

Du M., Zhao S., 2022, An Equity Evaluation on Accessibility of Primary Healthcare Facilities by Using V2SFCA Method: Taking Fukuoka City, Japan, as a Case Study, *Land*, 11(5), 640.

Fischer M. M., Nijkamp P. (eds.), 2021, *Handbook of regional science*, 2 ed., Springer, Berlin, Heidelberg.

Gao F., Kihal W., Le Meur N., Souris M., Deguen S., 2017, Does the edge effect impact on the measure of spatial accessibility to healthcare providers? *International Journal of Health Geography*, 16, 46.

Geurs K. T., Ritsema van Eck J., 2001, *Accessibility measures: review and applications. Evaluation of accessibility impacts of land-use transport scenarios, and related social and economic impacts*, National Institute of Public Health and the Environment, Bilthoven, p. 265.

Handy S., Niemeier D., 1997, Measuring accessibility: an exploration of issues and alternatives*, Environment and Planning A*, 29, 1175-1194.

Hansen W. G., 1959, How Accessibility Shapes Land Use, *Journal of the American Institute of Planners*, 25, 73-76.

Hewko J., Smoyer-Tomic K. E., Hodgson M. J., 2002, Measuring Neighbourhood Spatial Accessibility to Urban Amenities: Does Aggregation Error Matter? *Environment and Planning A: Economy and Space*, 34, 1185-1206.

Huhndorf S., Działek J., 2018, Assessment of the walking accessibility of public libraries in Krakow using the E2SFCA approach, *Urban Development Issues*, 54, 5-14.

Hyndman R. J., 2011, Forecasting An Overview, [in:] M. Lovric (ed.), *International Encyclopedia of Statistical Science,* Springer.

Ikram S. Z., Hu Y., Wang F., 2015, Disparities in Spatial Accessibility of Pharmacies in Baton Rouge, Louisiana, *Geographical Review*, 105, 492-510.

Jacobs-Crisioni C., Rietveld P., Koomen E., 2014, The impact of spatial aggregation on urban development analyses, *Applied Geography*, 47, 46-56.

Jaworska R., Suchecka J., 2014, *Statystyka przestrzenna: metody analiz struktur przestrzennych*, Wydawnictwo C.H. Beck, Warszawa.

Jin M., Liu L., Tong Gong, Y., Liu Y., 2019, Evaluating the Spatial Accessibility and Distribution Balance of Multi-Level Medical Service Facilities, *International Journal of Environmental Research and Public Health*, 16(7), 1150.

Kellerman A., 1981, Centrographic measures in geography, *Geo Abstracts,* University of East Anglia, Norwich.

Komornicki T., Wiśniewski R., Baranowski J., Błażejczyk K., Degórski M., Goliszek S., Rosik P., Solon J., Stępniak M., Zawiska I., 2015, *Wpływ wybranych korytarzy drogowych na środowisko przyrodnicze i rozwój społeczno-ekonomiczny obszarów przyległych*, Instytut Geografii i Przestrzennego Zagospodarowania PAN, Warszawa.

Langford M., Higgs G., Jones S., 2020, Understanding Spatial Variations in Accessibility to Banks Using Variable Floating Catchment Area Techniques, *Applied Spatial Analysis and Policy*, 14(3), 449-472.

Lechowski L., Jasion A., 2021, Spatial Accessibility of Primary Health Care in Rural Areas in Poland, *International Journal of Environmental Research and Public Health*, 18(17), 9282.

Li Y., Xie Y., Sun S., Hu L., 2022, Evaluation of Park Accessibility Based on Improved Gaussian Two-Step Floating Catchment Area Method: A Case Study of Xi'an City, *Buildings*, 12(7), 871.

Logan T. M., Williams T. G., Nisbet A. J., Liberman K. D., Zuo C. T., Guikema S. D., 2017, Evaluating urban accessibility: leveraging open-source data and analytics to overcome existing limitations, *Environment and Planning B: Urban Analytics and City Science*, 46(5), 897-913.

Longley P. A., Goodchild M. F., Maguire D. J., Rhind D. W., 2015, *Geographic Information Science & Systems*, 4th ed., Wiley, United States of America.

Loo B. P. Y., Lam W. W. Y., 2012, Geographic accessibility around health care facilities for elderly residents in Hong Kong: a microscale walkability assessment, *Environment and Planning B: Planning and Design*, 39(4), 629-646.

Luo S., Jiang H., Yi D., Liu R., Qin J., Liu Y., Zhang J., 2022, PM2SFCA: Spatial Access to Urban Parks, Based on Park Perceptions and Multi-Travel Modes. A Case Study in Beijing, *ISPRS International Journal of Geo-Information*, 11(9), 488.

Luo W., Qi Y., 2009, An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians, *Health and Place*, 15(4), 1100-1107.

Luo W., Wang F., 2003, Measures of Spatial Accessibility to Healthcare in a GIS Environment: Synthesis and a Case Study in Chicago Region. Environ Planning B Planning and Design 30, 865-884.

Maarseveen M. V., Martínez-Martín J. A., Flacke J. (eds), 2019, *GIS in sustainable urban planning and management: a global perspective*, CRC Press, Boca Raton, London, New York.

Moreno C., Allam Z., Chabaud D., Gall C., Pratlong F., 2021, Introducing the "15-Minute City": Sustainability, Resilience and Place Identity in Future Post-Pandemic Cities, *Smart Cities*, 4(1), 93-111.

Murtagh E. M., Mair J. L., Aguiar E., Tudor-Locke C., Murphy M. H., 2021, Outdoor Walking Speeds of Apparently

Healthy Adults: A Systematic Review and Meta-analysis, *Sports Medicine*, 51, 125-141.

Nalej M., 2018, *Problem zmiennych jednostek odniesienia (MAUP) w badaniach pokrycia terenu. Przykład Łódzkiego Obszaru Metropolitalnego,* Praca doktorska, Wydział Nauk Geograficznych. Uniwersytet Łódzki, Łódź.

Neutens T., 2015, Accessibility, equity and health care: review and research directions for transport geographers, *Journal of Transport Geography*, 43, 14-27.

Openshaw S., 1983, The modifiable areal unit problem, *CATMOG (Concepts & Techniques in Modern Geography)* 38, 41 p.

Park J., Goldberg D. W., 2021, A Review of Recent Spatial Accessibility Studies That Benefitted from Advanced Geospatial Information: Multimodal Transportation and Spatiotemporal Disaggregation, *ISPRS International Journal of Geo-Information*, 10(8), 532.

Peng Z.-R., 1997, The Jobs-Housing Balance and Urban Commuting, *Urban Studies*, 34(8), 1215-1235.

Pham T.-T.-H., Apparicio P., Séguin A.-M., Landry S., Gagnon M., 2012, Spatial distribution of vegetation in Montreal: An uneven distribution or environmental inequity? *Landscape and Urban Planning*, 107(3), 214-224.

Polzin P., Borges J., Coelho A., 2014, An Extended Kernel Density Two-Step Floating Catchment Area Method to Analyze Access to Health Care, *Environment and Planning B: Planning and Design*, 41(4), 717-735.

Pregnolato M., Ford A., Wilkinson S. M., Dawson R. J., 2017, The impact of flooding on road transport: A depth-disruption function, *Transportation Research Part D: Transport and Environment*, 55, 67-81.

Rahman M. T., Nahiduzzaman K. M., 2019, Examining the Walking Accessibility, Willingness, and Travel Conditions of Residents in Saudi Cities, *International Journal Environmental Research and Public Health,* 16(4), 545.

Rakower R., Łabędzki J., Gadziński J., 2011, Konkurencyjność ruchu rowerowego w przestrzeni miejskiej, *Transport Miejski i Regionalny*, 02, 31-38.

Rosik P., 2012, *Dostępność lądowa przestrzeni Polski w wymiarze europejskim*, IGiPZ PAN, Warszawa.

Rosik P., 2021, Świat dostępności – metody i komponenty. Przykład analiz empirycznych przestrzeni Polski, *Prace Geograficzne*, 276, IGIPZ PAN, Warszawa.

Shatnawi N., Al-Omari A., Al-Mashaqbeh A., 2022, Comparative study of using E2SFCA and 3SFCA methods for selected healthcare resources in Jordan during COVID-19 pandemic, *International Journal of Healthcare Management,* 15(4), 347-356.

Stepniak M., Jacobs-Crisioni C., 2017, Reducing the uncertainty induced by spatial aggregation in accessibility and spatial interaction applications, *Journal of Transport Geography*, 61, 17-29.

Stevenson M., Thompson J., de Sá T. H., Ewing R., Mohan D., McClure R., Roberts I., Tiwari G., Giles-Corti B., Sun X., Wallace M., Woodcock J., 2016, Land use, transport, and population health: estimating the health benefits of compact cities, *The Lancet*, 388(10062), 2925-2935.

Stępniak M., 2013, Wykorzystanie metody 2SFCA w badaniach dostępności przestrzennej usług medycznych, *Przegląd Geograficzny*, 85(2), 199-218.

Stępniak M., Rosik P., 2015, The Impact of Data Aggregation on Potential Accessibility Values, [in:] I. Ivan, I. Benenson, B. Jiang, J. Horák, J. Haworth, T. Inspektor (eds.), *Geoinformatics for Intelligent Transportation*, Springer International Publishing, Cham, 227-240.

Stępniak M., Wiśniewski R., Goliszek S., Marcińczak S., 2017, *Dostępność przestrzenna do usług publicznych w Polsce*, IGiPZ PAN, Warszawa.

Śleszyński P., 2011, Oszacowanie rzeczywistej liczby ludności gmin województwa mazowieckiego z wykorzystaniem danych GUS, *Studia Demograficzne*, 160(2), 35-57.

Śleszyński, P., 2014, Dostępność czasowa i jej zastosowania, *Przegląd Geograficzny*, 86(2), 171-215.

Tan J., Wang X., Pan J., 2021, The effect of population distribution measures on evaluating spatial accessibility of primary health-care institutions: A case study from China, *Geospatial Health*, 16(1), 31-45.

Vasiliev I. R., 1996, Visualization of Spatial Dependence: An Elementary View of Spatial Autocorrelation, [in:] S. Arlinghau (ed.), *Practical Handbook of Spatial Statistics*, 1 ed. CRC Press, Boca Raton, 17-30.

Wan N., Zou B., Sternberg T., 2012, A three-step floating catchment area method for analyzing spatial access to health services, *International Journal of Geographical Information Science* 26(6), 1073-1089.

Wiśniewski S., Borowska-Stefańska M., Kowalski M., Sapińska P., 2020, Vulnerability of the accessibility to grocery shopping in the event of flooding, *Transportation Research Part D: Transport and Environment*, 87, 102510.

Wu C., Powe N. A., Copeland A., 2020, Minimizing aggregation errors when measuring potential access to services for social groups at the city scale, *Environment and Planning B: Urban Analytics and City Science*, 48(8), 2206-2220.