Geographical accessibility to healthcare by point-of–interest data from online maps: a comparative study

Heng-Qian Huang-Fu,[1] Nan Zhang,[2] Li Wang,1 Hui-Juan Liang,[2] Ben-Song Xian,[2] Xiao-Fang Gan,[2] Ying-Si Lai[1-5]

[1]Department of Medical Statistics, School of Public Health, Sun Yat-sen University, Guangzhou; [2]Faculty of Health Management, Inner Mongolia Medical University, Hohhot; [3]Sun Yat-sen Global Health Institute, Sun Yat-sen University, Guangzhou, China; [4]Health Information Research Center, Guangdong Key Laboratory of Medicine, School of Public Health, Sun Yat-sen University, Guangzhou; [5]Guangzhou Joint Research Center for Disease Surveillance, Early Warning and Risk Assessment, Guangzhou, China

Abstract

Geographical accessibility is important for promoting health equity, and calculating it requires the locations of all existing healthcare facilities in a region. Authoritative location data collected by governments is accurate but mostly not publicly available, while Point-Of-Interest (POI) data from online sources, such as Baidu Maps and AutoNavi Maps are easily accessible. However, the accuracy of the latter has not been thoroughly analyzed. Taking Baotou, a medium-sized city in China, as an example, we assessed the suitability of using POI data for measuring geographic accessibility to healthcare facilities.We computed the difference of geographic accessibility calculated based on POI data and that on authoritative data. Logistic regression and a multiple linear regression model was applied to identify factors related to the consistency between the two data sources. Compared to authoritative data, POI data exhibited discrepancies, with completeness of 54.9% and accuracy of 63.7%. Geographic accessibility calculated based on both data showed similar patterns, with good consistency for hospitals and in urban areas. However, large differences (>30 minutes) were shown in rural areas for primary healthcare facilities. The differences were small regarding to population-weighted average accessibility (with slight underestimation of 3.07 minutes) and population coverage across various levels of accessibility (with differences less than 1% of the population) for the entire area. In conclusion, POI data can be considered for use in both urban areas and at the level of entire city; however, awareness should be raised in rural areas.

Correspondence: Ying-Si Lai, Department of Medical Statistics, School of Public Health, Sun Yat-sen University, Guangzhou, China. Tel.: +86.20.8733.0730
E-mail:laiys3@mail.sysu.edu.cn.

Key words: geographic accessibility; healthcare institution; POI data China

Conflict of interest: the authors declare no potential conflict of interest, and all authors confirm accuracy.

Availability of data and materials: all data generated or analyzed during this study are included in this published article.

Acknowledgements: we thank the Health Commission of our study city for providing the data. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality.

Received: 18 June 2024.
Accepted: 21 November 2024.

Geospatial Health 2024; 19:1322
doi:10.4081/gh.2024.1322

Introduction

Health equality receives worldwide attention, and according to the World Health Organization (WHO) increasing access to healthcare services is essential for reducing health inequities (World Health Organization, 2023). Its five dimensions: availability, accessibility, accommodation, affordability and acceptability (Penchansky & Thomas 1981) are key components that refer to the difficulty or ease of physically moving from a potential user's location to that of the service provider. In previous studies, geographic accessibility has been widely utilized to assess healthcare equity and this provides valuable information for resource allocation (Cao *et al.* 2021; Geldsetzer *et al.* 2020; Kotavaara *et al.* 2021; Rader *et al.* 2020). A common measure of geographic accessibility is the travel time to the nearest healthcare provider (Miller 2018; Weiss *et al.* 2020). Therefore, population distribution, the locations of healthcare facilities and travel times between them are needed to assess geographic accessibility. Population distribution data can be obtained from open-access data sources, while travel times can be calculated based on road networks or friction surfaces from different open-access databases. However, obtaining accurate and comprehensive locations of healthcare facilities across a large geographic area is a challenge.

Authoritative data of healthcare facilities are often difficult for researchers to obtain, as they are typically collected and managed by government health departments, but the degree of openness and the level of detail in this data vary across countries and regions (Maina *et al.* 2019). For example, the most authoritative data on geographiclocations of healthcare facilities in China are collected by local health agencies and reported by the Ministry of Health of the People's Republic of China (MoH) through the "National Health Statistics Network Direct Reporting System" platform (Ministry of Health of the People's Republic of China 2012). The data include name, type and detailed address of healthcare facilities (National Bureau of Statistics 2021), but this information is usually not publicly available. National health authorities in other countries, such as the Centers for Medicare & Medicaid Services (CMS) in the United States and the Federal Joint Committee (G-BA) in Germany, collect similar data, which are also not fully open to the public. Field survey is another way to obtain comprehensive data on the location and type of healthcare facilities. For example, Deng *et al.* (2015), conducted a field survey to collect locations of healthcare institutions in Shizhu County, Chongqing, China, and assessed the spatial accessibility of healthcare services. However, this method is time-consuming and labour-intensive, so it is generally only feasible for limited research areas.

Many researchers tend to choose publicly available point-of-interest (POI) data from online maps [e.g., Google Maps (https://maps.google.com/), OpenStreetMap (https://www.openstreetmap.org/), Baidu Maps (https://map.baidu.com/), AutoNavi Maps (https://www.amap.com/) etc.], which include healthcare facilities and consist of point locations with geographic coordinates and additional attributes such as name and type (Psyllidis *et al.* 2022).These data can be accessed through public websites and can also be directly retrieved via their Application Programming Interfaces (APIs).Due to their convenience of access and extensive spatial coverage, these data are commonly utilized in analyzing accessibility to healthcare (Jia *et al.* 2022; Khazi-Syed *et al.* 2023; Zhang *et al.* 2020). Weiss et al. (2020) utilized POI data from Google Maps and OpenStreetMap as the major source of geographical information on healthcare facilities, and, created the global maps of geographic accessibility to healthcare based on this information. When conducting studies on geographic accessibility in China, researchers often rely on Baidu Mapsand AutoNavi Maps as data sources to obtain geographic information on healthcare facilities (Hu *et al.* 2023; Jia *et al.* 2022; Liu *et al.* 2019; Wang *et al.* 2020). Jia *et al.* (2022) utilized POI data from AutoNavi Maps to create a map of spatial accessibility to primary healthcare in China.

As POI data from online maps are usually collected from crowd-sourced users and workers employed by map companies (Sohu News 2016; Wuhan University 2018), some actual locations may not be included, while others may be incorrectly labelled as healthcare facility POIs, such as long-closed facilities, general wellness centres or pharmacies. Thus, the use of POI data to calculate geographic accessibility may differ from the actual situation and incomplete information can underestimate accessibility, while places incorrectly labelled as healthcare facility POIs may lead to overestimation. The impact of these inaccuracies varies by region: in areas with a high density of healthcare facilities, the effect of incomplete or mislabelled data is limited (Ni *et al.*, 2016), while in regions with sparse healthcare coverage, such issues can have a greater impact on accessibility estimates. Steiniger *et al.* (2016) studied the OpenStreetMap POI data to evaluate urban accessibility (accessibility to day-to-day activities such as grocery stores, restaurants, and shopping) finding that the completeness of the OpenStreetMap POI data varied from 7% to 73% and POI data tended to underestimate accessibility in most areas, while remaining consistent in certain areas, with no clear linear relationship between the number of POIs and the accessibility score. However, the difference between healthcare facility POI data from online maps and authoritative data, as well as the suitability of utilizing POI data for measuring geographic accessibility of healthcare facilities, remains unexplored. In addition, previous studies have identified several factors associated with the consistency between POI data and reference data, including urban-rural divisions, road distribution, population distribution, economic status and the distribution of crowd-sourced data contributors (Borkowska and Pokonieczny 2022; Mullen *et al.* 2015; Yang *et al.* 2018).

In this study, we compared POI data from online data sources with authoritative data to assess the suitability of using the former to measure the geographic accessibility to healthcare. We assessed the consistency of POI data obtained from Baidu Maps and AutoNavi Maps with authoritative data provided by Baotou's Health Commission and used logistic regression to identify associated factors. We computed the geographic accessibility and quantified the difference between POI data and authoritative data. Multiple linear regression was applied to identify factors related to the difference.

Materials and Methods

Study area

Baotou, a medium-sized city located in northern China, was selected as the study area. With a population of 2.71 million and an area of 27,768 km^2, the city comprises 81 township-level administrative units (Appendix 11). Baotou includes both densely populated, economically developed urban areas and sparsely populated rural areas, providing a balanced representation for our study. The moderate scale of the city allows for comprehensive cross-checking and detailed comparison of all healthcare facilities and map POIs.

Data sources

We considered two types of healthcare facilities that are typically used in accessibility analyses: hospitals and primary healthcare (PHC) facilities. We further classified PHC facilities into three sub-groups based on the context in China (General Office of the State Council of the People's Republic of China 2015): i) healthcare facilities providing comprehensive healthcare services primarily for local residents, including Township Health Centres (THCs) and Community Health Centres (CHCs); ii) facilities offering basic healthcare services primarily for local residents, including Community Health Stations (CHSs) and Village Clinics (VCs), which operate at one level below the CHCs and THCs, respectively (Li *et al.* 2017); and iii) other healthcare facilities providing basic healthcare services for specific populations in particular settings, such as private clinics, school health clinics and worksite medical rooms.

We collected authoritative data and POI data of healthcare facilitieswithin the study area. The authoritative data for 2019 were obtained from the city's Health Commission and included all types of healthcare facilities within the area, such as names, types and detailed addresses (e.g., district, county, street and house number). The POI corresponding data were obtained from AutoNavi Maps and Baidu Maps during the period from 13 June to 8August, 2021. Similarly to the authoritative data, we further classified POIs into hospitals and PHC facilities since facility names in the Chinese healthcare system typically indicate their types, which allows straightforward identification between these two types of facilities (Details given in Appendix 2).

Population counts at the spatial resolution of 1×1 km^2 were obtained from the WorldPop gridded population estimate dataset (WorldPop 2020).Friction surfaces, as explained in the Malaria Atlas Project (https://malariaatlas.org/),were downloaded at a 1×1 km² resolution for 2020 from that website. They integrate multiple road data sources to represent travel speeds and were used here to calculate travel times between locations and health facilities. Road networks, including all types of roads and the Gross Domestic Product (GDP) at a 1×1 km² resolution, were obtained from various open-access data sources (Table 1 and Appendix 3 for details).

Methodology

Completeness and accuracy of POI data

We employed completeness and accuracy (Batini *et al.* 2009) as indicators to assess the consistency of POI data with authoritative data.We cross-checked healthcare facilities in the POI data and authoritative data by matching names and locations, primarily to determine whether they are within the same township. Completeness was defined as the extent to which POI data covered real-world healthcare facilities based on authoritative data. Accuracy was defined as the extent to which POI data correctly identified the existence of these facilities, meaning that data incorrectly labelled as healthcare facility POIs wereconsidered inaccurate.Completeness was expressed as:

$$\frac{\text{Number of healthcare facilities in POI data identical to those in authoritative data}}{\text{Number of healthcare facilities in authoritative data}}$$

and accuracy as:

$$\frac{\text{Number of healthcare facilities in POI data identical to those in authoritative data}}{\text{Number of healthcare facilities in POI data}}$$

Measurement of geographic accessibility

Previous studies often mapped geographic accessibility at the grid cell level (*e.g.*, 1 km) and at lower administrative levels (Weiss *et al*., 2020). Therefore, we analyzed geographic accessibility at both grid cell and township levels. While the grid cell level provides fine granularity for accessibility analysis, the township level is appropriate for providing insights into healthcare resource allocation for policymakers and investors. In this way, we had a range covering a high-resolution (1×1 km²) approach denoting residential areas, and a coarser level representing the smallest publicly available administrative boundary in China. The grid was overlaid onto the study area, resulting in 42,888 cells, where cells with a population density of 0.25 or higher were selected as potential residential locations, resulting in a total of 33,398 cells. At this level, geographic accessibility for each grid cell was assessed by calculating the travel time from the corresponding pixel to the nearest healthcare facility. Travel times were calculated using Dijkstra's algorithm (Dijkstra 1959) based on a friction surface (https://malariaatlas.org/)with the ideal travel situation, *i.e.* based on themotorized mode, such as cars or motorcycles (Weiss *et al.* 2018). At the township level,we calculated the population-weighted average geographic accessibility by averaging the cell-level travel time weighted by the corresponding population count calculated as follows:

$$A_k = \frac{\sum_i (T_i \times P_i)}{\sum_i P_i} \quad \text{(Eq. 1)}$$

where k represents the township under study; A_k the population-weighted average geographic accessibility at township k; i the index each grid cell within township k; T_i the travel time for grid cell i; and P_i thepopulation count for grid cell i.

The population-weighted average geographic accessibility and the population coverage at different scores of geographic accessibilities are commonly used metrics for assessing an entire region (Cao *et al.* 2021; Weiss *et al.* 2020). To assess the geographic accessibility for Baotou, we calculated these two metrics. The population coverage was done by summing the population counts within each level of geographic accessibility and dividing by the total population as follows:

$$C_j = \frac{\sum_{n \epsilon j} P_n}{\sum_n P_n} \quad \text{(Eq. 2)}$$

where j represents the level of geographic accessibility categorizing the grid cells based on the travel time required to reach a healthcare facility (*e.g.*, 0–15 min or 15–30 min; the population coverage at the geographic accessibility level j; n the index representing each individual grid cell within the study area; and P_n: the population count for grid cell n. To assess the consistency of geographic accessibility derived from the two data sources, we computed the differences in geographic accessibility at the grid cell and township levels as well as for the entire study area, by subtracting the values based on authoritative data from those calculated using the POI data.

Exploration of associated factors

Based on previous studies on factors associatedwith the quality of POI data (Borkowska and Pokonieczny 2022; Mullen *et al.* 2015; Yang *et al.* 2018), we considered road distribution, population density, GDP level and urban-rural divisions as potential factors associated with POI completeness and accuracy as well as the consistency of geographic accessibility calculated based on POI data and that based on authoritative data. In analyzing POI completeness and accuracy, as well as differences in geographic accessibility at the grid cell level, road distribution was measured as the Euclidean distance from the cell centre to the nearest road. At the township level, it was assessed as road density, calculated as the total length of roads per unit area. Based on an initial exploratory analysis, we found that Baotou's population density distribution is non-linear. To better capture this characteristic, we classified population density into three categories that align with the city's actual distribution patterns: low (<20 people/km²), medium (21-500 people/km²) and high (>500 people/km²).

Logistic regression was done to identify associated factors on POI completeness and accuracy. To assess the consistency of geographic accessibility calculated using POI data compared to that calculated with authoritative data, multiple linear regression was done to identify factors associated with log-transformed differences in geographic accessibility at both grid cell and township level. To normalize their distribution, the differences of geographic accessibility were transformed on a Log10 scale, and to prevent collinearity, we conducted a variable selection process. We found that although the model variables had low Variance Inflation Factor (VIF) values, population density and the GDP level were highly correlated (coefficient >0.8). Since the spatial prediction of the GDP level was related to population density, the model retained population density and excluded the GDP level.

Results

Completeness and accuracy of POI data

According to the authoritative data, there were 1,915 healthcare institutions (103 hospitals and 1,812 PHC facilities) in the study area. In comparison, a total of 1,651 healthcare facilities (127 hospitals and 1,524 PHC facilities) were obtained from online maps as POI data (Appendix 4). Of these, 1,052 facilities were found to be identical for both sources, with the majority distributed in the south-western urban part of the city. The lower number of online POIs data was particularly evident in the southern part. On the other hand, POIs incorrect labelled as healthcare facilities were mainly found in densely populated areas. Overall, the POI data showed a moderate completeness (55%) and accuracy (64%), with notable variations between urban and rural areas as well as among different types of healthcare facilities (Table 2). Urban areas showed a higher level of completeness (67%) compared to rural areas (24%) and hospitals had higher completeness (90%) compared to primary healthcare facilities (53%). Among PHCfacilities, the completeness in the rural areas was the lowest (3%). Similar accuracy was found in both urban (64%) and rural (63%) areas. Hospitals showed higher accuracy (73%), while PHC facilities were lower: 63% and 63%, respectively, while other facilities had the lowest accuracy (59%).

Geographic accessibility

Geographic accessibility based on POI data and based on authoritative data showed a similar pattern (Figure 1). Generally, urban areas in the south-western part of the city exhibited higher accessibility, while rural areas showed lower levels with an uneven distribution. For hospitals, the consistencywas good, except in the eastern areas (Figure 1a, c). For PHC facilities, the differences were mainly seen in the rural areas, particularly in the northern and eastern parts of the city that has grasslands as major landcover and where relatively fewer residents are located. Here, time differences exceeding 30 minutes were seen (Figure 1b,d).

Scatterplots comparing geographic accessibility based on POI data and that on authoritative data were produced at both grid cell and township levels (Figure 2). In urban areas, most scatter points distributed around the 1:1 line (Figure 2a, b) suggesting similar values of geographic accessibility resulting from the two data sources for both hospitals and PHC facilities. There were only a few points above the 1:1 line at the grid cell level for e PHC facilities indicating that underestimation existed by POI data in a few grid cells; however, this type of underestimation was not obvious at the township level. In the rural areas, there were quite a number of points above the 1:1 line, particularly for the PHC facilities (Figure 2c, d), which told us that the POI data is responsible for the underestimation in rural areas.

For the whole study area, the population-weighted average geographic accessibility was found to be 13.53 minutes to hospitals and 5.33 minutes to the PHC facilities based on authoritative data, while the POI data gave a slight underestimation, i.e.14.08 and 8.39 min, respectively(Table 3). This was mostly be accounted for by the rural areas as there were differences there of 13.11 and 3.22 minutes for PHC facilities and for hospitals, respectively. In urban areas, the difference was less than 1 minute. The population coverage for hospitals based on POI data was similar to that based on authoritative data, and the differences across all ranges of accessibility were less

Table 1. General data information.

Type of data	Year	Data source
Administrative units	2021	AutoNavi Map
Friction surfaces	2020	Malaria Atlas Project[a]
Population	2020	WorldPop[b]
Road networks	2021	Service for geographic information[c]
Gross domestic product (GDP)	2019	Resource and Environment Science and Data Center[d]

[a]Global motorized friction surface(https://malariaatlas.org/). Accessed 2022-8-23; [b]National catalogue of population counts /unconstrained individual countries 2000-2020 – UN-adjusted 1 km^2 resolution(https://www.worldpop.org/). Accessed 2022-5-25; [c]Nationalcatalogue service for geographic information, 1:1 million public version of basic geographic information data 2020 (https://www.webmap.cn/). Accessed 2022-6-20; [d]China's GDP spatial distribution km grid data set(https://www.resdc.cn). Accessed 2022-8-20,

Table 2. Completeness and accuracy of POI data.

	Completeness (%)			Accuracy (%)			
	Rural	Urban	Total	Rural	Urban	Total	
Hospitals		84	92	90	76	73	73
PHC facilities							
THCs and CHCs	55	86	72	85	85	85	
CHSs and VCs	3	51	23	77	82	81	
Other PHC facilities	55	68	66	51	59	59	
All PHC facilities	21	66	53	61	63	63	
All facilities		23	67	55	62	64	64

PHC, primary healthcare; THC, township health centre; CHC, community health centre; CHS, community health station; VC, village clinic.

than 1%. For the PHC facilities, there was a slight underestimation by the POI data, though within 3% (Figure 3).

Associated variables

For the completeness of healthcare facility POIs, we found that population density and the urban/rural division were significantly associated with POI completeness (Table 4). Using Odds Ratio (OR) and Confidence Intervals (CI), healthcare facilities located in areas with medium and high population density showed significantly higher completeness compared to those in low-density areas: OR=1.261, 95% CI=1.040-1.528, $p = 0.018$ and OR=2.449, 95% CI=2.005-2.991, $p<0.001$, respectively. Facilities in urban areas also showed significantly higher POI completeness compared to those in rural areas (OR=1.658, 95%CI=1.450-1.895, $p<0.001$). This suggests that healthcare facilities situated in areas with higher population density, such as in urban areas tend to have more complete POI data.

Regarding the likelihood of incorrect labelling, we found that distance to the nearest road and population density was significantly associated with POI data accuracy (Table 4). POIs in areas with medium (OR=0.463, 95%CI=0.291-0.738, p=0.033) and high population density (OR=0.610, 95%CI=0.387-0.961, p=0.001) exhibited significantly lower accuracy compared to those in low-density areas. This finding suggests that POI accuracy tends to be lower in areas with both medium and high-density compared to low-density areas. Distance to the nearest road was negatively correlated with POI accuracy; for each 1 km crease in distance, the odds of POI accuracy decreased by a factor of 0.914 (OR = 0.914, 95% CI = 0.853-0.980, p =0.011), with POIs located closer to roads tending to have lower accuracy.

As shown in Table 5, population density, road distribution and the urban/rural division were significant associated with the difference between the geographic accessibility calculated based on POI data and that based on authoritative data both at the grid cell level and the township level. At the township level, areas with medium population density (with a 0.303-fold decrease, 95%*CI*=0.126-0.730, $p<0.001$) and high population density (with a 0.067-fold decrease, 95%*CI*=0.024-0.200, $p<0.001$) showed significantly smaller differences of accessibility compared to low-density areas, suggesting that geographic accessibility calculated using POI data is more accurate in areas with denser populations. Road density was negatively correlated with the difference since each 1 km road per km² increase corresponded to a 0.470 decrease (95%*CI*= -0.45- -0.205, p=0.001) for each $\log_{10}$ unit of difference in accessibility indicating higher accuracy in areas with greater road density. Geographic accessibility in urban areas showed lower differences compared to that in rural areas (with a 0.126-fold decrease, 95%*CI*=0.050-0.291, $p<0.001$, implying that POI-based accessibility is more accurate in urban settings. At the grid cell level, similar results were found: areas with high population density, closer distance to the nearest road and urban locations tended to show smaller accessibility differences indicating greater accuracy in POI-based measures. The results were consistent with the observations depicted in Figure 1.

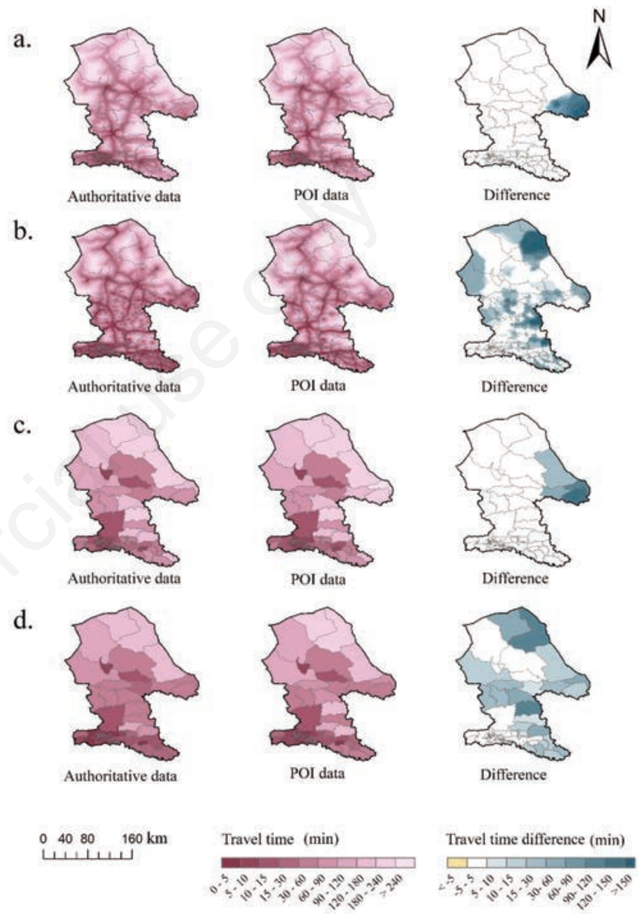

Figure 1. Travel time to healthcare facilities at the grid cell and township levels. **a**) travel time to hospitals at the grid cell level; **b**) travel time toprimary healthcare facilities at the grid cell level; **c**) travel time to hospitals at the township level; **d**) travel time toprimary healthcare facilities at the township level.

Table 3. Population-weighted average travel time.

Area	Data	Hospitals (min)	PHC (min)
Rural	Authoritative	52.91	20.98
	POI data	56.13	34.09
	Difference	3.22	13.11
Urban	Authoritative	2.08	0.77
	POI data	1.85	0.92
	Difference	-0.23	0.15
Total	Authoritative	13.53	5.33
	POI data	14.08	8.39
	Difference	0.55	3.07

Discussion

POI data are commonly used to analyze the geographic accessibility of healthcare facilities (Cao *et al.* 2021; Hu *et al.* 2023; Weiss *et al.* 2020), but there is scant evidence supporting the rationale for this approach. In this study, we added an exploration of the quality of healthcare facilities based on POI data by comparison with authoritative data. Our results can thus provide a reference for applying POI data to improve the accuracy and reliability of accessibility studies. We found that these data are not entirely consistent with authoritative ones. Due to slow data collection and updating procedures in regions with limited transportation and fewer map

Table 4. Factors correlated to completeness and accuracy in logistic regression analysis.

Variable	Non-standardized coefficient[a] B[c]	S.E.[d]	Standardized coefficientb Beta[e]	p
Completeness				
Distance to the nearest road	-0.015	0.011	-0.030	0.156
Population density				
Low (<20 people/km^2)	ref	ref	ref	ref
Medium (21-500 people/km^2)	0.101	0.043	0.087	0.018*
High (>500 people/km^2)	0.389	0.044	0.365	<0.001*
Area type				
Rural	ref	ref	ref	ref
Urban	0.220	0.030	0.196	<0.001*
Accuracy				
Distance to the nearest road	-0.039	0.015	-0.066	0.011*
Population density				
Low (<20 people/km^2)	ref	ref	ref	ref
Medium (21-500 people/km^2)	-0.334	0.103	-0.247	0.033*
High (>500 people/km^2)	-0.215	0.101	-0.165	0.001*
Area type				
Rural	ref	ref	ref	ref
Urban	0.036	0.039	0.024	0.359

*[a]Variables in their original units; [b]normalized variables for comparison; [c]non-standardized regression coefficient; [d]standardized regression coefficient; [e]standard error of the non-standardized coefficients. REF refers to reference category used as baseline for comparison;*indicates p<0.05.*

Table 5. Factors correlated to the difference of geographic accessibility in multiple linear regression analysis.

Variable	Non-standardized coefficient[a] B[c]	S.E.[d]	Standardized coefficientb Beta[e]	p
Township level				
Road distribution				
Road density	-0.470	0.133	-0.226	0.001*
Population density				
Low (<20 people/km^2)	ref	ref	ref	ref
Medium (21-500 people/km^2)	-0.518	0.192	-0.255	<0.001*
High (>500 people/km^2)	-1.164	0.234	-0.588	<0.001*
Area type				
Rural	ref	ref	ref	ref
Urban	-0.921	0.193	-0.464	<0.001*
Grid cell level				
Road distribution				
Distance to the nearest road	0.023	0.002	0.047	<0.001*
Population density				
Low (<20 people/km^2)	ref	ref	ref	ref
Medium (21-500 people/km^2)	-0.123	0.017	-0.036	<0.001*
High (>500 people/km^2)	-0.535	0.054	-0.049	<0.001*
Area type				
Rural	ref	ref	ref	ref
Urban	-1.009	-0.023	-0.220	<0.001*

*[a]Variables in their original units; [b]normalized variables for comparison; cnon-standardized regression coefficient; [d]standardized regression coefficient; [e]standard error of the non-standardized coefficients. REF refers to reference category used as baseline for comparison; *indicates p<0.05.*

users, some actual locations are not included in POIs, and the presence of incorrectly labelled healthcare facility POIs is an inherent limitation, as confirmed by research in fields such as transport infrastructure and retail (Klinkhardt *et al.* 2023; Wuhan University 2018). Our regression analysis results indicate that healthcare facilities in higher population density and urban areas tend to have more complete POI data, consistent with previous findings suggesting that healthcare facilities in these areas are more visible and therefore more likely to be marked as POIs (Ather 2010; Mooney *et al.* 2013; Mullen *et al.* 2015). POI accuracy was also associated with population density and distance to the nearest road, which may be due to the higher number of other types of POIs, such as pharmacies and beauty salons in these areas increasing the likelihood of mislabelling.

The impact of incorrectly labelled healthcare POIs on the calculation of geographic accessibility was minimal both at the grid cell and township levels in urban areas, as they were concentrated in areas with dense healthcare facilities. However, due to the omission of certain healthcare facilities during POI data collection, particularly those not located near other healthcare facilities, the geographic accessibility calculated from POI data had a tendency to underestimate accessibility at varying degrees across most areas. On the other hand, using POI data to estimate hospital accessibility in both urban and rural areas, as well as to assess geographic accessibility for primary healthcare facilities in urban areas, appears viable in the absence of authoritative data, with minimal underestimation in only a few areas. However, caution should be exercised with respect to assessments of PHC facilities in rural areas, where data completeness is often lower leading to greater underestimation of accessibility. The results from multiple linear regression also suggested that areas characterized by dense road networks, higher population density and urban areas tended to yield more accurate results when using POI data to calculate geographic accessibility. This finding aligns with previous studies regarding factors affecting POI data quality (Borkowska and Pokonieczny 2022; Mullen *et al.* 2015; Yang *et al.* 2018).

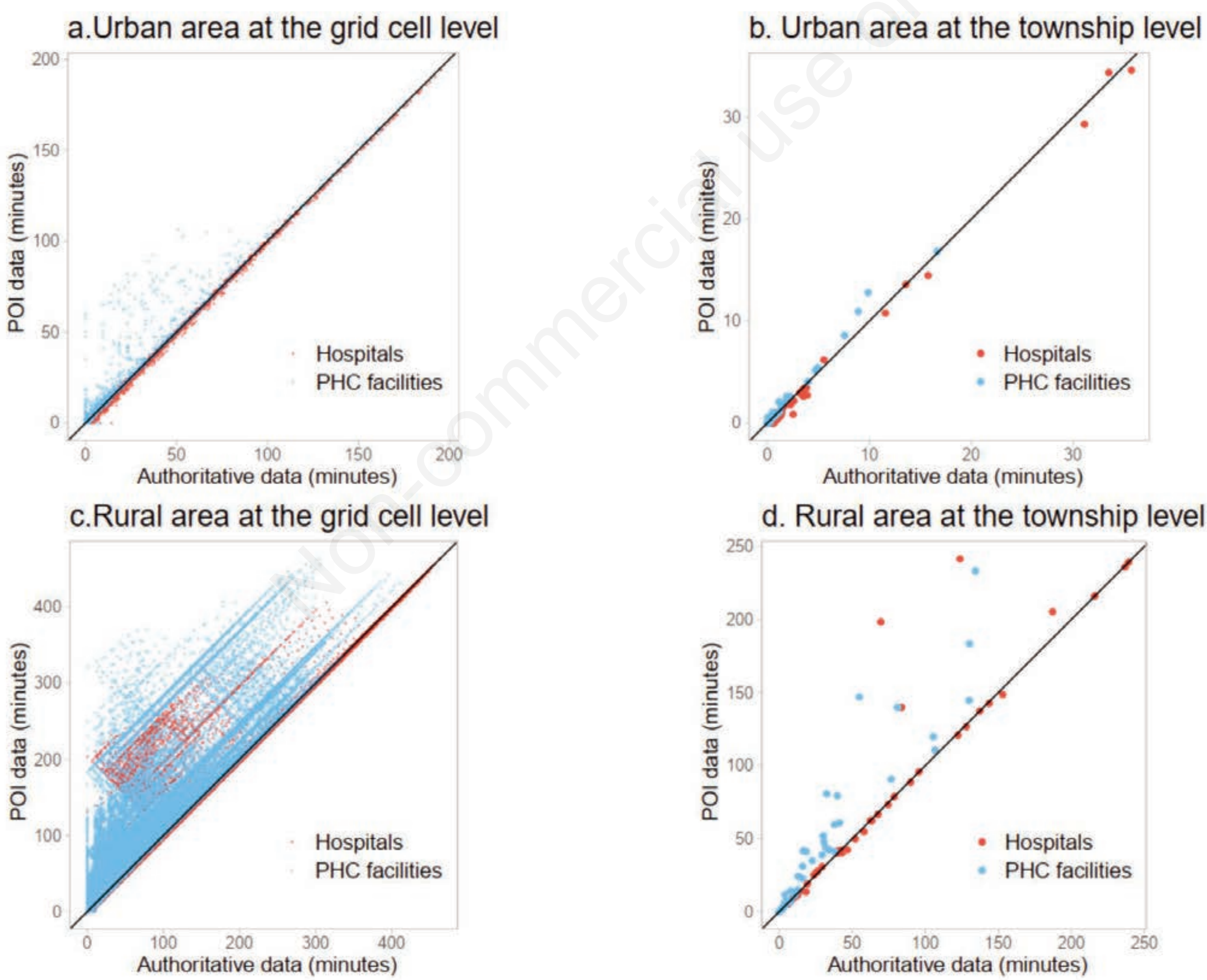

Figure 2. Scatterplots of travel time based on POI data and on authoritative data. PHC, primary healthcare. Each data point corresponds to the geographic accessibility based on POI data (the horizontal axis) compared to that based on authoritative data (the vertical axis) at each grid cell or town. The situation of points above the line indicates a higher travel time and a lower geographic accessibility, compared to authoritative data, while situations below the line means the opposite.

When evaluating geographic accessibility across the entire study area, we calculated population coverage and the population-weighted average geographic accessibility by combining population data, so both scores were influenced by the population distribution (Linard *et al.* 2012). Around 80% of the total population live in urban areas, where differences of geographical accessibility were small. In contrast, areas with larger differences of geographic accessibility had sparse population; thus, their influence on the entire study area was relatively minor. Therefore, calculating population coverage and the population-weighted average accessibility, combining population counts when using POI data as a data source, yields relatively accurate geographic accessibility assessments for the entire area investigated.

This study has several limitations that should be acknowledged. First, the POI data was collected in July 2021, while the authoritative data was collected in December 2019, during which time increases and/or decreases in the number of healthcare facilities might have occurred. Second, our study focused on a medium-sized city in China. Nevertheless, the exploration of associated factors provides insights for using POI data in areas with different backgrounds. However, cautions should be made when extrapolating the findings to other types of cities, such as provincial capitals and poverty-stricken areas. In addition, the quality of POI data provided by various map suppliers may vary in different regions or countries. Third, the calculation of travel time was performed using a friction surface with motorized modes of transportation, without accounting for other factors that could influence the travel time, such as modes of transportation, weather effects or road conditions. However, the primary objective of this study was to compare POI data with authoritative data rather than to achieve precise travel time estimates. Fourth, we only focused on the geographic accessibility, which is a simple measure of accessibility. More advanced metrics are common, e.g., considering the spatial impedance and the capacity of healthcare facilities (number of health workers, hospital beds, etc.). For example, a few studies calculated spatial accessibility by employing POI data together with data of healthcare resources obtained from official hospital websites (Wang *et al.* 2020). However, the difficulty in obtaining resource data for primary healthcare facilities limits the effectiveness of this approach. Finally, while our study explored the use of POI data as a substitute across multiple spatial scales, this analysis serves only as a preliminary reference for using POI data in geographic accessibility studies. Currently, there is no clear method to determine the conditions under which geographic accessibility measurements based on POI data achieve acceptable accuracy standards.

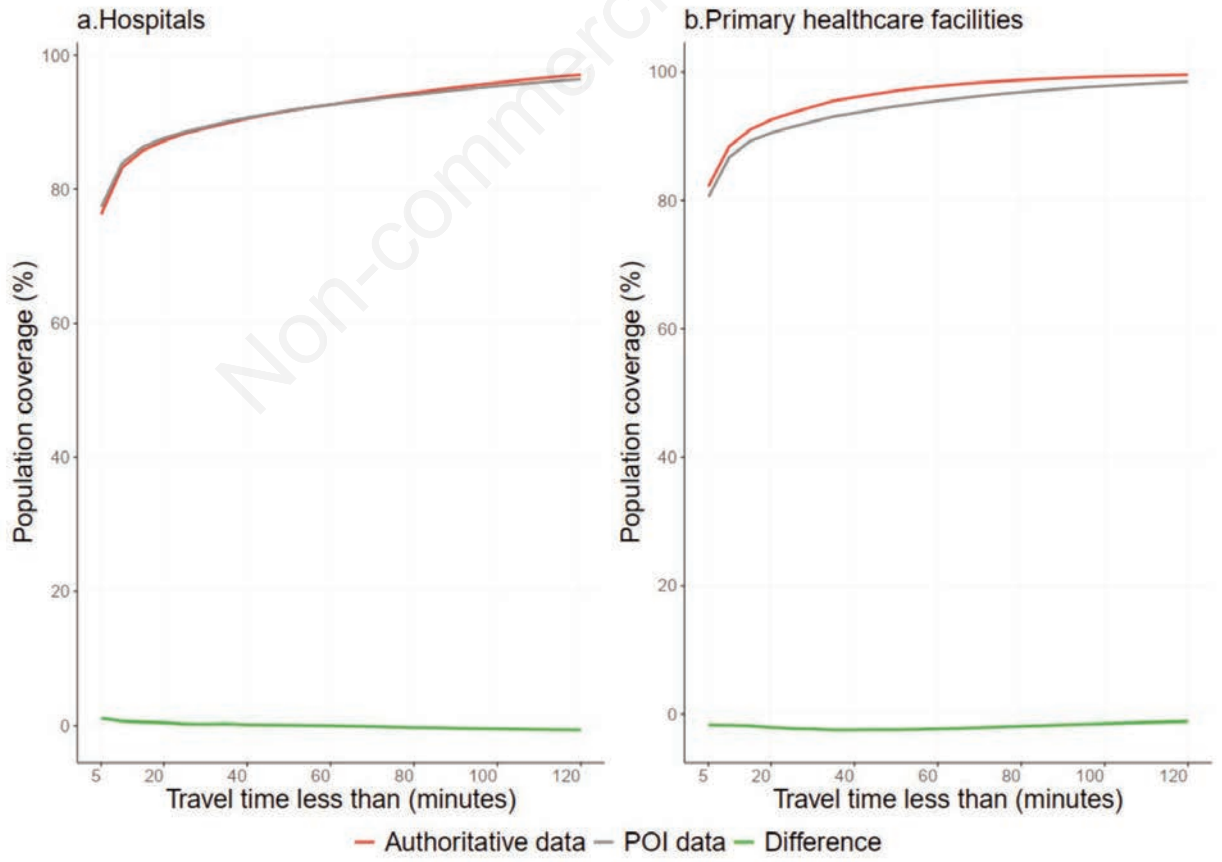

Figure 3. Population coverage of travel time to the nearest health care facility.

Conclusions

Factors associated with POI quality were examined and compared to authoritative data for the calculation of geographic accessibility. There were discrepancies in POI data, leading to underestimations in geographic accessibility at both grid cell and township levels, particularly in the rural areas. However, accessibility differences were small when evaluating the geographic accessibility of the entire area by incorporating population counts. POI data should therefore be considered to assess geographic accessibility to healthcare facilities for both urban areas and the entire city-level area weighted by population; however, awareness should be raised with regard to rural areas, particularly for PHC facilities at the township and grid cell levels. The findings presented here would assist researchers in effectively utilizing healthcare institution POI data for healthcare service analysis, further supporting evidence-based health planning and resource allocation.

References

Ather AA. 2010. How many volunteers does it take to map an area well? The validity of linus' law to volunteered geographic information. Cartogr J 47:315-322.

Batini C, Cappiello C, Francalanci C, Maurino A. 2009. Methodologies for data quality assessment and improvement. ACM Comput Surv 41:1-52.

Borkowska S, Pokonieczny K. 2022. Analysis of openstreetmap data quality for selected counties in poland in terms of sustainable development. Sustainability 14:3728.

Cao W-R, Shakya P, Karmacharya B, Xu DR, Hao Y-T, Lai Y-S. 2021. Equity of geographical access to public health facilities in nepal. BMJ Glob Health 6:e006786.

Centers for Medicare & Medicaid Services. 2022. Centers for medicare & medicaid services (cms). Accessed 25 May 2022. Available from: https://www.cms.gov

Deng L, Shao Ja, Guo Y, Xu X. 2015. Spatial accessibility of medical services in mountainous regions based on modified two-step floating catchment area method:A case study of shizhu county, chongqing Prog Geogr 34:716-725.

Dijkstra EW. 1959. A note on two problems in connexion with graphs. Numer Math.

dos Anjos Luis A, Cabral P. 2016. Geographic accessibility to primary healthcare centers in mozambique. Int J Equity Health 15:1-13.

Federal Joint Committee. 2022. Federal joint committee. Accessed 22 May 2022. Available from: https://www.g-ba.de/

Geldsetzer P, Reinmuth M, Ouma PO, Lautenbach S, Okiro EA, Bärnighausen T, Zipf A. 2020. Mapping physical access to health care for older adults in sub-saharan africa and implications for the covid-19 response: A cross-sectional analysis. Lancet Healthy Longev 1:e32-e42.

General Office of the State Council of the People's Republic of China. 2015. National health service system planning outline (2015-2020). Chin Pract J Rural Dr 22:1-11.

Hu S, Zhao R, Cui Y, Zhang D, Ge Y. 2023. Identifying the uneven distribution of health and education services in china using open geospatial data. Geogr Sustain 4:91-9.

Jia P, Wang Y, Yang M, Wang L, Yang X, Shi X, Yang L, Wen J, Liu Y, Yang M, Xin J, Zhang F, Jiang L, Chi C, Zhang L, Ma X, Ma X, Zhao L, Li W. 2022. Inequalities of spatial primary healthcare accessibility in china. Soc Sci Med 314:115458.

Khazi-Syed A, Pecherskiy M, Krambeck H. 2023. Analyzing access to health facilities by road using unconventional data sources. Transp Res Rec 2677:169-177.

Klinkhardt C, Kühnel F, Heilig M, Lautenbach S, Wörle T, Vortisch P, Kuhnimhof T. 2023. Quality assessment of openstreetmap's points of interest with large-scale real data. Transp Res Rec 2677:661-674.

Kotavaara O, Nivala A, Lankila T, Huotari T, Delmelle E, Antikainen H. 2021. Geographical accessibility to primary health care in finland–grid-based multimodal assessment. Appl Geogr 136:102583.

Linard C, Gilbert M, Snow RW, Noor AM, Tatem AJ. 2012. Population distribution, settlement patterns and accessibility across africa in 2010. PLoS One 7:e31743.

Li X, Lu J, Hu S, Cheng KK, De Maeseneer J, Meng Q, Mossialos E, Xu DR, Yip W, Zhang H. 2017. The primary health-care system in china. Lancet 390:2584-2594.

Liu S, Qin Y, Xu Y. 2019. Inequality and influencing factors of spatial accessibility of medical facilities in rural areas of china: A case study of henan province. Int J Environ Res Public Health 16:1833.

Maina J, Ouma PO, Macharia PM, Alegana VA, Mitto B, Fall IS, Noor AM, Snow RW, Okiro EA. 2019. A spatial database of health facilities managed by the public health sector in sub saharan africa. Sci Data 6:134.

Miller EJ. 2018. Accessibility: Measurement and application in transportation planning. Transp Rev 38:551-555.

Ministry of Health of the People's Republic of China. 2012. Notice of the ministry of health on the implementation of five systems including the national health resources and medical services survey system. Accessed 20 November 2022. Available from: http://www.nhc.gov.cn/zwgkzt/ppxxhjs1/201301/d09bc944bb57409f8522f94874694e5c.shtml

Mooney P, Corcoran P, Ciepluch B. 2013. The potential for using volunteered geographic information in pervasive health computing applications. J Ambient Intell Humaniz Comput 4:731-745.

Mullen WF, Jackson SP, Croitoru A, Crooks A, Stefanidis A, Agouris P. 2015. Assessing the impact of demographic characteristics on spatial error in volunteered geographic information features. GeoJournal 80:587-605.

National Bureau of Statistics. 2021. National health resources and medical service statistical survey system. Accessed 13 October 2023. Available from: http://www.stats.gov.cn/tjfw/bmdcxmsp/bmzd/202201/t20220121_1826888.html

Palk L, Okano JT, Dullie L, Blower S. 2020. Travel time to healthcare facilities, mode of transportation, and hiv elimination in malawi: A geospatial modelling analysis. Lancet Glob Health 8:e1555-e1564.

Penchansky R, Thomas JW. 1981. The concept of access: Definition and relationship to consumer satisfaction. Med Care 19:127-140.

Psyllidis A, Gao S, Hu Y, Kim E-K, McKenzie G, Purves R, Yuan M, Andris C. 2022. Points of interest (poi): A commentary on the state of the art, challenges, and prospects for the future. Comput Urban Sci 2:20.

Rader B, Astley CM, Sy KTL, Sewalk K, Hswen Y, Brownstein JS, Kraemer MUG. 2020. Geographic access to united states sars-cov-2 testing sites highlights healthcare disparities and may

bias transmission estimates. J Travel Med 27:taaa076.
Sohu News. 2016. Conversation with gu weihao from baidu maps: Data technology shapes the future of maps. Accessed October 23 2023. Available from: https://www.sohu.com/a/109243758_449147
Steiniger S, Poorazizi ME, Scott DR, Fuentes C, Crespo R. Can we use openstreetmap pois for the evaluation of urban accessibility? In: Proceedings of the GIScience 2016, 2016. Montreal, Canada, Vol. 1.
Wang K, Bai J, Dang X. 2020. Spatial difference and equity analysis for accessibility to three-level medical services based on actual medical behavior in shaanxi, china. Int J Environ Res Public Health 18:112.
Weiss DJ, Nelson A, Gibson HS, Temperley W, Peedell S, Lieber A, Hancher M, Poyart E, Belchior S, Fullman N. 2018. A global map of travel time to cities to assess inequalities in accessibility in 2015. Nature 553:333-336.
Weiss DJ, Nelson A, Vargas-Ruiz CA, Gligorić K, Bavadekar S, Gabrilovich E, Bertozzi-Villa A, Rozier J, Gibson HS, Shekel T, Kamath C, Lieber A, Schulman K, Shao Y, Qarkaxhija V, Nandi AK, Keddie SH, Rumisha S, Amratia P, Arambepola R, Chestnutt EG, Millar JJ, Symons TL, Cameron E, Battle KE, Bhatt S, Gething PW. 2020. Global maps of travel time to healthcare facilities. Nat Med 26:1835-1838.
World Health Organization. 2023. Universal health coverage (uhc). Accessed 20 October 2023. Available from: https://www.who.int/news-room/fact-sheets/detail/universal-health-coverage-(uhc)
WorldPop. 2020. Population counts / unconstrained individual countries 2000-2020 un adjusted (1km resolution). Accessed 25 May 2022. Available from: https://www.worldpop.org/
Wuhan University. 2018. geoscience café sharing of cutting-edge technologies in the production of autonavi map data. Accessed 5 October 2023. Available from: https://liesmars.whu.edu.cn/info/1063/3912.htm
Yang S, Shen J, Konečný M, Wang Y, Štampach R. Study on the spatial heterogeneity of the poi quality in openstreetmap. In: Proceedings of the the 7th International Conference on Cartography and GIS, 2018. Sozopol, Bulgaria, 18-23.
Zhang Y, Li W, Deng H, Li Y. 2020. Evaluation of public transport-based accessibility to health facilities considering spatial heterogeneity. J Adv Transp 2020:1-10.

Online supplementary materials

Appendix 1.

Figure S1. Study area.

Figure S2. General data information. (a) Township level administrative divisions; (b) Friction surface; (c) GDP; (d) Population density and road distribution

Appendix 2. Point-of-Interest (POI) data retrieval and processing overview

Table S1. POI Data Search Keywords and Final Results

Appendix 3. Comparison between POI data and authoritative data.

Figure S3. Comparison between POI data and authoritative data. (a) Authoritative data; (b) POI data; (c) Identical data; (d) Missing POI data; (e) Redundant POI data.

Appendix 4. Geographic accessibility at the township level.

Figure S4. Geographic accessibility at the township level.