Health & Place 83 (2023) 103075

Contents lists available at ScienceDirect

Health and Place

journal homepage: www.elsevier.com/locate/healthplace

The quality of OpenStreetMap food-related point-of-interest data for use in epidemiological research

Maria Gabriela M. Pinho [a,b,c,*], Benjamin Flueckiger [d,e], Antonia Valentin [f], Maria-Iosifina Kasdagli [g], Kalliopi Kyriakou [h], Jeroen Lakerveld [b,c], Joreintje D. Mackenbach [b,c], Joline W.J. Beulens [b,d,i], Kees de Hoogh [d,e]

[a] Copernicus Institute of Sustainable Development, Department Environmental Sciences, Utrecht University, Utrecht, the Netherlands
[b] Amsterdam UMC, Location Vrije Universiteit Amsterdam, Epidemiology and Data Science, Amsterdam, the Netherlands
[c] Upstream Team, www.upstreamteam.nl, Amsterdam UMC, VU University Amsterdam, Amsterdam, the Netherlands
[d] Swiss Tropical and Public Health Institute, Allschwil, Switzerland
[e] University of Basel, Basel, Switzerland
[f] Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain
[g] Department of Hygiene, Epidemiology and Medical Statistics, School of Medicine, National and Kapodistrian University of Athens, Athens, Greece
[h] Institute for Risk Assessment Sciences (IRAS), Division of Environmental Epidemiology, Utrecht University, the Netherlands
[i] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, the Netherlands

ARTICLE INFO

Keywords:
OpenStreetMap
Google street view
Food environment
Quality assessment
Exposome

ABSTRACT

We assessed the quality of food-related OpenStreetMap (OSM) data in urban areas of five European countries. We calculated agreement statistics between point-of-interests (POIs) from OSM and from Google Street View (GSV) in five European regions. We furthermore assessed correlations between exposure measures (distance and counts) from OSM data and administrative data from local data sources on food environment data in three European countries. Agreement between POI data in OSM compared to GSV was poor, but correlations were moderate to high between exposures from OSM and local data sources. OSM data downloaded in 2020 seems to be an acceptable source of data for generating count-based food exposure measures for research in selected European regions.

1. Introduction

Food environments are changing rapidly and influencing our health (Pinho et al., 2020; James et al., 2017). For example, a high density of fast-food outlets (FFO) has been associated with cardiovascular disease risk (Poelman et al., 2018), and higher access to grocery stores and lower access to FFO has been associated to healthier diet:ry intake (Althoff et al., 2022). These associations are sometimes inconsistent, especially in Europe (Caspi et al., 2012). To better understand these relationships across countries, good quality and harmonized geospatial data are essential. One source of data is volunteered geographic information (VGI) such as OpenStreetMap (OSM). OSM data is licensed under the Open Data Commons Open Database License by the OpenStreetMap Foundation. Anyone is free to use, share and modify the data as long as credit is given to OSM and its contributors (OpenStreetMap. About OpenStreetMap, 2023). However, there are doubts about OSM data quality (Antoniou and Skopeliti, 2015). While data collection is harmonized across countries, OSM contributors do not necessarily follow rigid scientific principles. So, data quality needs to be assured to be suitable for use in academic research (Goodchild and Li, 2012). Despite developments on intrinsic quality assessment methods, such as the identification of errors and bugs and the evaluation of data completeness (Minghini and Frassinelli, 2019), there is no simple and standardized method to determine whether food-related OSM data are accurate, complete and reliable to be used in academic research, and across different regions. Therefore, we aimed to assess the quality of food environment data from OSM by assessing agreement with Google Street View point-of-interest (POI) data as trusted dataset in five European regions. In addition, we tested the correlation between exposures derived from OSM data and exposures derived using existing local trusted datasets in three of these regions.

* Corresponding author. Copernicus Institute of Sustainable Development, Department Environmental Sciences, Utrecht University, Utrecht, the Netherlands.
E-mail address: m.g.matiasdepinho@uu.nl (M.G.M. Pinho).

https://doi.org/10.1016/j.healthplace.2023.103075
Received 21 December 2022; Received in revised form 26 April 2023; Accepted 15 June 2023
Available online 14 July 2023

2. Methods

This study was part of the EU-funded EXPANSE project (Vlaanderen et al., 2021). Comparison of a trusted dataset (or "gold standard") with a test dataset is a common way to validate tools and datasets in epidemiological research. However, as is often the case (Antoniou and Skopeliti, 2015), a 'gold standard' dataset was not available. Thus, we developed a protocol for obtaining a trusted dataset using a virtual audit in Google Street View (GSV), and identified already existing trusted datasets where available, to be compared with OSM data. Even though field audits are typically regarded as the preferred method to obtain geo-spatial data in these settings, and we previously demonstrated that this can also be a virtual field audit (Bethlehem et al., 2014).

3. Open street map data

OSM data for Switzerland, Greece, Spain, the Netherlands and Poland were downloaded from Geofabrik (https://www.geofabrik.de) in June 2020 and POI features related to the food environment were selected. Using geo-spatial analysis, OSM food-related POI were linked to the street segments layer and the count of selected food retailer types in OSM (i.e., alcohol stores, beverage outlet, convenience store, supermarket, restaurant, FFO, café, bar, biergarten, and pub), was calculated per street segment. See Supplementary Table 1 for the list of OSM tags analysed and their definitions.

4. Reference data for comparison

4.1. Google Street View

Google Street View data were collected through virtual audits using Google earth. The virtual audit protocol was based on the existing SPOTLIGHT virtual audit tool (Bethlehem et al., 2014). The virtual audit was performed by a local researcher in random selections of street segments in Basel, Switzerland; Greater Athens Area, Greece; Barcelona, Spain; the Randstad region in the Netherlands and Lodz, Poland. For the random selection of street segments, the road network data from OSM were be used. To avoid selecting street segments in remote areas where we did not expect to find many food retailers (e.g., nature reserves), we selected the following land coverage categories from the CORINE 2018 database (European Union - European Environment Agency (EEA) - Copernicus Land Monitoring Service, 2020): continuous and discontinuous urban fabric; industrial or commercial units; road and rail networks and associated land; and spaces under construction development. Within these categories, only secondary, tertiary, residential, unclassified, and pedestrian road types were considered for street segment selection. Since virtually auditing all street segments in the study area would not be feasible, we selected a random sample of 5% of all eligible street segments in each city (i.e., segments falling within the selected land cover and road categories).

The selection of land cover, road types and random street segments sample were performed in QGIS. The selected street segments were imported into Google Earth for the virtual audit. Auditors virtually walked through the pre-selected street segments in Google Earth and collected the location and type of the food retailer. We present the total count of food retailers found via GSV in each study area, and used count of and distance to food retailer (types) per street segment for analysis.

4.2. Local data sources

In addition to GSV data collected through virtual audits using Google Earth, where available, we used local data sources to calculate exposure measures usually used in epidemiological studies focusing on the food environment. These exposures measures were then linked to a sample of residential address in the relevant study areas. We tested how count-based and distance-based exposures derived from OSM correlated with exposures derived from trusted sources using local data sources with a wider geographical coverage. Due to limited availability of local data sources, we performed this exposure-based assessment in regions of Netherlands and Switzerland, and in Barcelona city in Spain.

The local source for food-related POI used in the Netherlands was Locatus data from 2020 (LOCATUS, 2020), obtained via Lakerveld et al. (2020). Locatus is a commercial database validated for epidemiological research that performs regular field audits across the country and collects accurate information on location and type of several types of retailers (Canalia et al., 2020). In Switzerland, data from the General Classification of Economic Activities (NOGA) were used (the Swiss Confederation - Federal Statistical Office Business Registers Data, 2008). In Barcelona, data from Economic Activities Census on the ground floor of the city of Barcelona were used (Barcelona's City Hall Open Data Service - Department of Statistics and Data Dissemination of the Municipal Data Office, 2019).

4.3. Analysis of agreement based on POI locations: OSM vs. GSV

Using GSV data as the reference dataset, the percentage of specific agreement (i.e., percentage of positive and negative agreement) between OSM and GSV was calculated based on counts per street segment. In order to calculate the specific agreement, matching POI between the two datasets were identified. Thus, true positives (TP) were defined when the count of food retailers per street segment in OSM was equal to the count of food retailers per street segment in GSV, and different from zero. True negatives (TN) were defined when street segments in both OSM and GSV had no food retailers present. False positives (FP) were defined when the count per street segments in OSM was higher than the count per street segments in GSV. False negatives (FN) were defined when the count per street segments in OSM was lower than the count per street segments in GSV. Positive agreement is the proportion of observed OSM data points per street segment when these data points are also observed in GSV in the same street segment, and is calculated using the following formula 'positive agreement $= 2TP/(2\ TP + FN + FP)$'. Negative agreement is defined as proportion of street segment with no OSM data points when these data points are also missing in GSV in the same street segment, and is calculated using the following formula 'negative agreement $= 2TN/(2\ TN + FN + FP)$' (de Vet et al., 2013). Spearman correlations comparing the counts of food retailers per street segment were also calculated.

4.4. Agreement based on individual exposures to food retailers: OSM vs. local data sources

In the Netherlands, count of and distance to closest FFO, supermarkets and restaurants were calculated within an 800m-buffer around 1728 residential addresses in a cross-sectional survey designed to investigate interactions with the food environment among Dutch adults. In Switzerland count of and distance to closest restaurants (including FFO) and supermarkets were calculated within an 800m-buffer around 1000 randomly selected residential addresses across the country. In Barcelona, count of and distance to closest FFO and supermarkets were calculated within an 800m-buffer around 1459 randomly residential addresses across the city. According to a systematic review on food environment and diet, there is great variation of buffer sizes in the literature, with most studies presenting buffer sizes between 100 m and 2 miles (about 3219 m). Given this range in the literature, a Euclidian buffer of 800m was chosen because it is frequently used in the literature (Caspi et al., 2012), and is the equivalent of a 10-min walking distance, a plausible distance individuals are willing to walk in an European context.

These exposures were calculated both using OSM data and the local food environment data sources. In the Swiss NOGA database the restaurant category includes FFO, hence we compared this category to restaurants and FFO combined in OSM. Because one of our aims was to

Table 1
Descriptive statistics of the street segment sample.

Region	Total street segments sample	Street segments with GSV images		Street segments with missing GSV images	
	N	N	%	N	%
Randstad	2022	1917	94.8%	105	5.2%
Greater Athens Area	2981	2937	98.5%	44	1.5%
Basel	526	348	66.2%	178	33.8%
Lodz	1033	796	77.1%	237	22.9%
Barcelona	405	373	92.1%	32	7.9%

GSV: Google Street View.

test the correlation between exposures commonly used in epidemiological research as derived from OSM data local datasets, we present correlation analysis with the exposures operationalized in different ways, namely: count and distance as continuous variables; count in quartiles; count in pre-determined categories (i.e.; 0; 1–3; 4–9 and 10 or more food retailers within 800m buffer); and distance in pre-determined categories (0–500m from home to food retailer; 501–1000m; 1001–1500; 1501m or more). Recategorization of food environment variables is often done in epidemiological research due to their skewed nature.

Pearson correlations (for normally distributed data) and Spearman correlations (for non-normally distributed data) were used to compare the exposures derived from both datasets. Strength of correlations and level of agreement were evaluated as follows: <0.30 "poor"; 0.31–0.50 "fair"; 0.51–0.70 "moderate"; from 0.71 to 0.90 "high"; and >0.90 "very high".

5. Results

The total number of street segments ranged from 405 in Barcelona to 2981 in Greater Athens Area (Table 1). The study areas with relatively high percentage of missing GSV images on selected street segments were Basel (33.8%) and Lodz (22.9%). The total count of food retailers across countries was higher when assessed in GSV than in OSM in all instances (Table 2). However, in some areas this difference was larger than in others.

The positive agreement between OSM and GSV, and the correlation coefficients were poor across study areas for the totality food retailers and most individual categories (Table 3). Positive agreement and correlation were moderate (above 0.5) for supermarkets in the Randstad, and bars and FFO in Basel. The negative agreement, meaning the percentage of agreement on a negative rating (absence of a food retailer in the street segment) was very high, generally >90%, for almost all categories in all study areas with the exception of the totality of food retailers in Barcelona, which presented a moderate negative agreement (60.7%). This high percentage of negative agreement was due to the high amount of street segments with no food retailer present (Table 2), leading to a high number of true negatives and consequent high percentage of negative agreement.

Analysis based on exposure measures generally showed moderate to high correlations (Table 4). Overall, correlations were stronger for count-based measures than for distance-based measures. For example, Pearson correlation coefficients of 0.98 in Randstad, and 0.96 in Switzerland were found for continuous count of restaurants when comparing exposures derived from OSM and local data sources. Similarly, the correlation of counts of FFO in Barcelona was 0.88. On the other hand, distance to the closest restaurant was found to be 0.67 and 0.51 in Randstad and Switzerland, respectively, and distance to the closest FFO in Barcelona was 0.79.

6. Discussion

We tested the quality of food-related OSM data by evaluating the agreement between OSM and GSV POIs, and correlations between exposures derived with OSM data and local data sources. Across the five European study areas, positive agreement statistics between OSM and GSV POI were poor. However, correlations between exposure-based measures from a selection of three regions using OSM data and trusted local data sources from Switzerland, the Netherlands and Spain ranged from moderate to very high. Count-based measures showed higher correlations as compared to distance-based measures.

The fact that the total count of POIs in OSM was consistently lower than in GSV indicates that OSM has more incomplete data as compared to GSV. However, the high correlations between exposures derived from the different datasets indicates that areas with highest or lowest counts of food outlets tended to be the same in both datasets. A previous study interested in evaluating the spatial availability of alcohol outlets found OSM data to be 50% complete. The authors indicated that OSM, when used with caution and knowing its limitations, has the potential to advance the understanding between alcohol availability and health/social outcomes (Bright et al., 2018). Other studies analysing POI in general found that the quality of European OSM data is generally more complete than OSM data from countries in different continents (Zhou et al., 2022a, 2022b). In epidemiological research, the aggregated exposure measures are used more often than the POIs. Our results indicate that OSM data in 2020 may be useful as a source of harmonized food environment data, at least for the regions studied here. In the context of the EXPANSE project, OSM data from 2020 can be used to derive exposures measures based on count of food retailers. The current project only focused on urban areas and correlations between the datasets would likely be different in areas of different urbanization levels. In addition, whether OSM from earlier years could be used as

Table 2
Count of food retailers across street segments as found in OSM and GSV virtual audit.

Counts of food retailers across street segments	Randstad		Greater Athens		Basel		Lodz		Barcelona	
	GSV	OSM	GSV	OSM	GSV	OSM	GSV	OSM	GSV	OSM
Total[a]	612	427	900	608	63	32	57	36	608	600
Alcohol	15	0	14	0	1	0	5	0	6	0
Beverages	1	9	1	8	0	1	0	7	0	0
Convenience	23	10	116	62	7	4	17	8	35	16
Supermarket	22	21	54	32	4	5	2	0	53	13
Restaurant	180	140	123	92	19	5	8	9	178	111
Fast-food outlets	104	71	91	77	4	5	1	0	11	15
Café	91	60	133	139	8	4	1	0	64	52
Bar	59	37	13	29	4	4	1	0	123	35
Biergarten	0	0	0	0	0	0	0	0	0	0
Pub	13	59	1	0	0	1	0	2	0	10

OSM: Open Street Maps; GSV: Google Street View.
[a] Includes all food retailers, not just those listed below.

Table 3
Agreement statistics on counts of food retailers per street segment as found in OSM and GSV virtual audit.

Randstad							
Food retailer	TN	TP	FN	FP	Positive agreement	Negative agreement	Spearman correlation
TOTAL	1635	68	155	59	38.9%	93.9%	0.690
Alcohol	1902	0	15	0	0.0%	99.6%	N/A
Beverage	1907	0	1	9	0.0%	99.7%	−0.002
Convenience	1893	4	16	4	28.6%	99.5%	0.312
Supermarket	1891	13	7	6	66.7%	99.7%	0.700
Restaurant	1788	34	60	35	41.7%	97.4%	0.651
Fast-food outlets	1814	26	51	26	40.3%	97.9%	0.391
Café	1817	17	57	26	29.1%	97.8%	0.413
Bar	1868	5	31	13	18.5%	98.8%	0.348
Biergarten	1917	0	0	0	N/A	100.0%	N/A
Pub	1871	5	4	37	19.6%	98.9%	0.380
Greater Athens Area							
Food retailer	TN	TP	FN	FP	Positive agreement	Negative agreement	Spearman correlation
TOTAL	2218	111	435	193	26.1%	87.6%	0.405
Alcohol	2923	0	34	0	0.0%	99.4%	N/A
Beverage	2928	0	21	8	0.0%	99.5%	−0.001
Convenience	2783	17	113	44	17.8%	97.3%	0.166
Supermarket	2869	15	57	16	29.1%	98.7%	0.384
Restaurant	2774	23	102	58	22.3%	97.2%	0.262
Fast-food outlets	2820	15	82	40	19.7%	97.9%	0.185
Café	2738	21	110	88	17.5%	96.5%	0.226
Bar	2908	2	27	20	7.8%	99.2%	0.205
Biergarten	2937	0	20	0	0.0%	99.7%	N/A
Pub	2936	0	21	0	0.0%	99.6%	N/A
Basel							
Food retailer	TN	TP	FN	FP	Positive agreement	Negative agreement	Spearman correlation
TOTAL	311	8	25	4	35.6%	95.5%	0.635
Alcohol	347	0	3	0	0.0%	99.6%	N/A
Beverage	347	0	0	1	0.0%	99.9%	N/A
Convenience	341	4	3	0	72.7%	99.6%	0.753
Supermarket	340	1	3	4	22.2%	99.0%	0.213
Restaurant	331	1	15	1	11.1%	97.6%	0.373
Fast-food outlets	343	2	1	2	57.1%	99.6%	0.573
Café	340	2	5	1	40.0%	99.1%	0.563
Bar	342	2	2	2	50.0%	99.4%	0.494
Biergarten	348	0	2	0	0.0%	99.7%	N/A
Pub	347	0	0	1	0.0%	99.9%	N/A
Lodz							
Food retailer	TN	TP	FN	FP	Positive agreement	Negative agreement	Spearman correlation
TOTAL	737	7	33	19	21.2%	96.6%	0.297
Alcohol	790	0	5	1	0.0%	99.6%	N/A
Beverage	789	0	0	7	0.0%	99.6%	N/A
Convenience	773	1	15	7	8.3%	98.6%	0.075
Supermarket	794	0	2	0	0.0%	99.9%	N/A
Restaurant	783	3	4	6	37.5%	99.4%	0.371
Fast-food outlets	794	0	1	1	0.0%	99.9%	−0.001
Café	794	0	1	1	0.0%	99.9%	N/A
Bar	795	0	1	0	0.0%	99.9%	N/A
Biergarten	796	0	0	0	N/A	100.0%	N/A
Pub	794	0	0	2	0.0%	99.9%	N/A
Barcelona							
Food retailer	TN	TP	FN	FP	Positive agreement	Negative agreement	Spearman correlation
TOTAL	154	20	145	54	16.7%	60.7%	0.541
Alcohol	367	0	6	0	0.0%	99.2%	N/A
Beverage	373	0	0	0	N/A	100.0%	N/A
Convenience	331	2	28	12	9.1%	94.3%	0.093
Supermarket	324	7	37	5	0.2%	93.9%	0.298
Restaurant	237	17	81	38	0.2%	79.9%	0.359
Fast-food outlets	351	3	8	11	0.2%	97.4%	0.215
Café	299	9	36	29	21.7%	90.2%	0.295
Bar	270	6	77	20	11.0%	84.8%	0.122
Biergarten	373	0	0	0	N/A	100.0%	N/A
Pub	365	0	0	8	0.0%	98.9%	N/A

Spearman correlations coefficient are presented due to non-normally distributed data.
TN: True Negatives; TP: True Positives; FN: False Negatives; FP False Positives.
N/A: Not Applicable.

source of food environment data still needs to be analysed (Smith et al., 2021).

Our results should be viewed in light of some limitations. The exposure-based comparison could only be performed in three European areas due to the lack of a trustworthy comparison dataset in other regions. This dependency on other trusted datasets prevented a quality assessment with wider geographic and temporal coverage. Therefore, our results may not be extrapolated to other regions with different characteristics than those analysed in this study. Future studies could apply intrinsic methods of data quality evaluation, such as those that

Table 4
Correlations on individual exposure measures to food retailers – comparison of OSM and local data sources.

Count - Continuous variables	Pearson correlation coefficient[c]		
	Randstad - NL OSM *vs* Commercial database (Locatus)	Switzerland OSM *vs* Administrative database (NOGA)	Barcelona - ES OSM *vs* Administrative database (Census)
Fast-food outlet count	0.90	–	0.88
Restaurant count	0.98	0.96[d]	–
Supermarket count	0.92	0.87	0.91
Distance - Continuous variables			
Distance to closest fast-food outlet	0.68	–	0.79
Distance to closest restaurant	0.67	0.51[d]	–
Distance to closest supermarket	0.98	0.57	0.89
Count in quartiles	**Spearman Correlation coefficient** [c]		
Quartiles of fast-food outlet count	0.82	–	0.85
Quartiles of restaurant count	0.83	0.84[d]	–
Quartiles of supermarket count	0.90	0.83	0.88
Count in pre-determined categories[a]			
Four categories of fast-food outlet count	0.78	–	0.57
Four categories of restaurant count	0.85	0.88[d]	–
Four categories of supermarket count	0.85	0.80	0.49
Distance in pre-determined categories[b]			
Four categories of distance to fast-food outlet	0.66	–	0.64
Four categories of distance to restaurant	0.68	0.65[d]	–
Four categories of distance to supermarket	0.92	0.59	0.78

OSM = OpenStreetMap; NL = the Netherlands; ES = Spain; NOGA = General Classification of Economic Activities (NOGA).
[a] Categories are: 0; 1–3; 4–9 and 10 or more food retailers within 800m buffer.
[b] Categories are: 0–500 m from home to food retailer; 501–1000m; 1001–1500; 1501m or more.
[c] Spearman correlations coefficient are presented for to non-normally distributed data, and Pearson correlations are presented for normally distributed data.
[d] In Switzerland the restaurant category includes fast food outlets in both NOGA database and OSM.

checks the evolution of the data over time until completeness is observed, as such analyses do not require other trusted datasets. Another limitation is the mismatch between the date of GSV images and the collected OSM data. Although most analysed images present in GSV were taken from 2019, a certain degree of inaccuracy due to temporal mismatch cannot be disregarded as there were a few observations dating back to 2008. This has potentially had a significant impact on the percentage positive and negative agreement. Furthermore, we only focused on two commonly used measures of food retailer access in epidemiological studies, counts and distances, while different measures such as average distance to nearest five retailers may provide different results. As a strength of this study, we highlight the fact that we developed and applied a useful method for OSM data quality evaluation that is suitable for epidemiological and health research in general. In addition, we were able to include up to five European cities in the analysis and to identify good quality datasets to serve as reference data. This study adds to the literature as one of the first to evaluated the quality of VGI data using methods commonly used in epidemiological research. Future studies should consider the use of different buffer sizes. While varying buffer sizes may have little impact on exposure to food retailers in highly urbanized areas, the reliability of VGI data in other areas may depend on the chosen buffer size.

Despite the poor positive agreement based on absolute count of POI per street segment from OSM and GSV across five European regions, we found moderate to high correlations between count-based but not distance-based exposures derived with food-related OSM and local trusted data sources in Barcelona, Netherlands and Switzerland. In conclusion, OSM data downloaded in 2020 seems to be a good source of harmonized data for generating count-based exposure measures in epidemiological research in across these three European regions.

Funding source

This work is part of the EXPANSE project, funded from the European Union's Horizon 2020 research and innovation programme under grant agreement No 874627. Food retail data in the Netherlands were obtained via the Geoscience and Health Cohort Consortium (GECCO), which was financially supported by the Netherlands Organisation for Scientific Research (NWO), the Netherlands Organisation for Health Research and Development (ZonMw), and Amsterdam UMC.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.healthplace.2023.103075.

References

Althoff, T., Nilforoshan, H., Hua, J., Leskovec, J., 2022. Large-scale diet tracking data reveal disparate associations between food environment and diet. Nat. Commun. 13 (1).
Antoniou, V., Skopeliti, A., 2015. Measures and indicators of vgi quality: an overview. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences II-3/W5, 345–351.
Barcelona's City Hall Open Data Service - Department of Statistics and Data Dissemination of the Municipal Data Office, 2019. Economic Activities Census on the Ground Floor of the City of Barcelona. Available from: https://opendata-ajuntament.barcelona.cat/data/en/dataset/cens-activitats-comercials.
Bethlehem, J.R., Mackenbach, J.D., Ben-Rebah, M., Compernolle, S., Glonti, K., Bardos, H., et al., 2014. The SPOTLIGHT virtual audit tool: a valid and reliable tool to assess obesogenic characteristics of the built environment. Int. J. Health Geogr. 13, 52.
Bright, J., De Sabbata, S., Lee, S., Ganesh, B., Humphreys, D.K., 2018. OpenStreetMap data for alcohol research: reliability assessment and quality indicators. Health Place 50, 130–136.

Canalia, C., Pinho, M.G.M., Lakerveld, J., Mackenbach, J.D., 2020. Field validation of commercially available food retailer data in The Netherlands. Int. J. Environ. Res. Publ. Health 17 (6).

Caspi, C.E., Sorensen, G., Subramanian, S.V., Kawachi, I., 2012. The local food environment and diet: a systematic review. Health Place 18 (5), 1172–1187.

de Vet, H.C., Mokkink, L.B., Terwee, C.B., Hoekstra, O.S., Knol, D.L., 2013. Clinicians are right not to like Cohen's κ. BMJ 346.

European Union - European Environment Agency (EEA) - Copernicus Land Monitoring Service, 2020. Corine Land Cover 2018 [Version 2020_20u1:[Available from: https://land.copernicus.eu/pan-european/corine-land-cover/clc2018.

Goodchild, M.F., Li, L., 2012. Assuring the quality of volunteered geographic information. Spatial statistics 1, 110–120.

James, P., Seward, M.W., James O'Malley, A., Subramanian, S.V., Block, J.P., 2017. Changes in the food environment over time: examining 40 years of data in the Framingham Heart Study. Int. J. Behav. Nutr. Phys. Activ. 14 (1), 84.

Lakerveld, J., Wagtendonk, A., Vaartjes, I., Karssenberg, D., GECCO Consortium, 2020. Deep phenotyping meets big data: the Geoscience and hEalth Cohort COnsortium (GECCO) data to enable exposome studies in The Netherlands. Int. J. Health Geogr. 19.

LOCATUS, 2020. Locatus Retail Data. Available from: https://locatus.com/en/applications/locatus-retail-plans/.

Minghini, M., Frassinelli, F., 2019. OpenStreetMap history for intrinsic quality assessment: is OSM up-to-date? Open Geospatial Data. Software and Standards 4 (1).

OpenStreetMap. About OpenStreetMap [Available from: https://wiki.openstreetmap.org/wiki/About_OpenStreetMap.

Pinho, M.G.M., Mackenbach, J.D., den Braver, N.R., Beulens, J.J.W., Brug, J., Lakerveld, J., 2020. Recent changes in the Dutch foodscape: socioeconomic and urban-rural differences. Int. J. Behav. Nutr. Phys. Activ. 17 (1), 43.

Poelman, M., Strak, M., Schmitz, O., Hoek, G., Karssenberg, D., Helbich, M., et al., 2018. Relations between the residential fast-food environment and the individual risk of cardiovascular diseases in The Netherlands: a nationwide follow-up study. Eur J Prev Cardiol 25 (13), 1397–1405.

Smith, C.M., Kaufman, J.D., Mooney, S.J., 2021. Google street view image availability in the Bronx and San Diego, 2007-2020: understanding potential biases in virtual audits of urban built environments. Health Place 72, 102701.

the Swiss Confederation - Federal Statistical Office Business Registers Data, 2008. General Classification of Economic Activities. NOGA) [Available from: https://www.i14y.admin.ch/en/catalog/datasets/HCL_NOGA/description.

Vlaanderen, J., de Hoogh, K., Hoek, G., Peters, A., Probst-Hensch, N., Scalbert, A., et al., 2021. Developing the building blocks to elucidate the impact of the urban exposome on cardiometabolic-pulmonary disease: the EU EXPANSE project. Environ Epidemiol 5 (4), e162.

Zhou, Q., Wang, S.Z., Liu, Y.M., 2022a. Exploring the accuracy and completeness patterns of global land-cover/land-use data in OpenStreetMap. Appl. Geogr. 145.

Zhou, Q., Zhang, Y.H., Chang, K., Brovelli, M.A., 2022b. Assessing OSM building completeness for almost 13,000 cities globally. Int J Digit Earth 15 (1), 2400–2421.