

UC Irvine

UC Irvine Previously Published Works

Title

Using Social Media to Identify Sources of Healthy Food in Urban Neighborhoods.

Permalink

<https://escholarship.org/uc/item/3rf5x9r2>

Journal

Journal of urban health : bulletin of the New York Academy of Medicine, 94(3)

ISSN

1099-3460

Authors

Gomez-Lopez, Iris N
Clarke, Philippa
Hill, Alex B
[et al.](#)

Publication Date

2017-06-01

DOI

10.1007/s11524-017-0154-1

Peer reviewed

Using Social Media to Identify Sources of Healthy Food in Urban Neighborhoods

Iris N. Gomez-Lopez · Philippa Clarke  · Alex B. Hill · Daniel M. Romero · Robert Goodspeed · Veronica J. Berrocal · V. G. Vinod Vydiswaran · Tiffany C. Veinot

Published online: 28 April 2017
© The New York Academy of Medicine 2017

Abstract An established body of research has used secondary data sources (such as proprietary business databases) to demonstrate the importance of the neighborhood food environment for multiple health outcomes. However, documenting food availability using secondary sources in low-income urban neighborhoods can be particularly challenging since small businesses play a crucial role in food availability. These small businesses are typically underrepresented in national databases, which rely on secondary sources to develop

data for marketing purposes. Using social media and other crowdsourced data to account for these smaller businesses holds promise, but the quality of these data remains unknown. This paper compares the quality of full-line grocery store information from Yelp, a crowdsourced content service, to a “ground truth” data set (Detroit Food Map) and a commercially-available dataset (Reference USA) for the greater Detroit area. Results suggest that Yelp is more accurate than Reference USA in identifying healthy food stores in urban areas. Researchers investigating the relationship between the nutrition environment and health may consider Yelp as a reliable and valid source for identifying sources of healthy food in urban environments.

I. N. Gomez-Lopez · D. M. Romero ·
V. G. Vinod Vydiswaran · T. C. Veinot
School of Information, University of Michigan, Ann Arbor, MI,
USA

P. Clarke (✉)
Institute for Social Research and Department of Epidemiology,
University of Michigan, 426 Thompson Street, Ann Arbor, MI
48104, USA
e-mail: pjclarke@umich.edu

A. B. Hill
Detroit Food Map Initiative, Detroit, MI, USA

R. Goodspeed
Taubman College of Architecture and Urban Planning, University
of Michigan, Ann Arbor, MI, USA

V. J. Berrocal
Department of Biostatistics, University of Michigan, Ann Arbor,
MI, USA

V. G. Vinod Vydiswaran
Department of Learning Health Sciences, University of Michigan,
Ann Arbor, MI, USA

Keywords Social media · Neighborhood · Food sources · Grocery stores · Yelp · Reference USA

Background

Limited access to nutritious food is associated with a higher risk of obesity and chronic disease [1–4]. Access to a supermarket, in particular, has been shown to be related to a reduced risk for obesity [5], especially among children and adolescents [6, 7]. Supermarkets, or “full-line grocery stores,” are defined as stores that carry higher quality, fresh food with a better selection, and lower cost compared to smaller food stores [8], thereby providing opportunities for better dietary choices [9].

However, there have been conflicting findings about the role of neighborhoods for residents' access to healthy foods (see Larson et al. ([10]) for a review). This may be due to a lack of accuracy in the measurements of retail food environments used in research studies, which often rely on secondary sources that are linked to the residential location of study subjects in a focal data set. Examples of secondary retail food data include business or market research databases (e.g., Reference USA, Nielsen TDLinx, Dun and Bradstreet), government food or agricultural registries, and local telephone directories [11]. Government food outlet registries [12] and Reference USA [13] have been shown to have the highest validity in identifying sources of retail food outlets [11]. However, the costs of purchasing commercial databases, such as Reference USA, can be prohibitive, and there remains limited evidence on the validity of these secondary sources for identifying full-line grocery stores in particular [14, 15].

Social media and other crowdsourced data hold promise as alternative, nonproprietary sources of secondary food outlet data. New data sources such as Yelp (<http://www.yelp.com>) and City Search (<http://www.citysearch.com>) have contributed to an increasing amount of georeferenced, crowdsourced data available for characterizing local areas. However, the validity of this information for identifying sources of healthy food has yet to be assessed. While increasing attention is being paid to the geographic footprint of social media to capture "social hotspots," or locations in which people generate the most social media content [16], little research has examined the potential of social media for characterizing nutrition resources in local areas. Additionally, recent research regarding fast food locations suggests that the overlap between Yelp and other secondary retail food data may be small [17].

In this paper, we investigate the quality (i.e., accuracy and coverage) of Yelp for identifying sources of healthy foods in urban neighborhoods. Yelp is a free, user-generated content service based on the sharing of user reviews and ratings of local businesses. For comparison, we also examine the validity of Reference USA (<http://www.referenceusa.com/>), a proprietary business database that has frequently been used to identify sources of healthy food in neighborhood research [13, 15]. We focus on neighborhoods in greater Detroit, a metropolitan area that has undergone substantial economic and structural changes in the past several decades, with consequences for residents' income levels

and the availability of healthy food [18]. Economic instability may lead to sudden changes in food availability as businesses close, affecting the validity of available data sources. Therefore, we assessed the quality of Yelp and Reference USA for identifying sources of healthy food in Detroit neighborhoods, by comparing them to a "ground truth" consisting of a comprehensive list of healthy food stores identified through the Detroit Food Map (<http://detroitfoodmap.com/>).

Methods

Regional Focus

Our area of study included the seven counties in Southeast Michigan (Wayne, Oakland, Macomb, Livingston, Washtenaw, Monroe, and St. Clair). These seven counties are part of the Greater Detroit Combined Statistical Area, a densely populated area in southeast Michigan, containing almost half of the total state population. These seven counties comprise the region served by the Southeast Michigan Council of Governments (SEMCOG) (<http://semcog.org/>).

Data

Detroit Food Map Initiative

Sources of healthy food were defined as "full-line" (sometimes known as "full service") grocery stores following the Michigan Department of Agriculture & Rural Development's (MDARD) definition as "a store selling fresh produce, fresh meat, fresh bread, and fresh dairy..." [19]. A list of full-line grocery stores in these counties came from the Detroit Food Map (DFM) initiative (<http://detroitfoodmap.com/>). DFM is a volunteer-driven, community-based initiative that assesses the quality of food stores as access points for nutritious and healthy food options in Metropolitan Detroit. DFM began with an enumerative list of 14,052 food establishments in the state of Michigan obtained via a Freedom of Information Act (FOIA) request (MDARD's Food Establishment License Application database accessed in December 2014) that was further restricted to 476 stores in the seven county area falling under six grocery retail type categories: supermarket (conventional), supercenter, supermarket (limited assortment), natural/gourmet foods, warehouse store, and

ethnic/specialty food/small grocery. Categories were drawn from the Trade Dimensions *Retail Site Database* via Policy Map (<https://www.policymap.com/data/our-data-directory/#Trade%20Dimensions:%20Grocery%20Retail%20Locations>).

Each store was then manually assessed as being full-line or not using a virtual verification method that used a standardized algorithm with Google Street View imagery. This method, piloted by DFM, considered stores to be full-line if (a) they were large (appeared to be 20,000 ft² or more in Google Street View) and (b) they did not have prominent signage advertising “liquor” or “lotto.” Field work conducted by DFM found multiple cases where so-called grocery stores were simply selling alcohol. These stores had liquor or lotto prominently displayed, and this was therefore used as a criterion for excluding these stores.

Using this approach, the 476 stores were labeled as (1) full-line ($n = 213$), (2) possibly full-line ($n = 30$), (3) not full-line ($n = 45$), (4) possibly closed ($n = 7$), (5) no usable image (i.e., poor image quality or older than June 2015) ($n = 179$), and (6) duplicates ($n = 2$). The reliability of this method was assessed by calling a random sample of 25% of the 258 stores labeled full-line or not full-line ($n = 64$) to determine if they met the full-line definition. Ninety percent of the stores in the random sample were correctly categorized using the Google Street View criteria.

A second phase of verification was then conducted via telephone to (i) determine whether the 30 stores that were deemed to be possibly full-line were actually full-line, (ii) verify the seven store closures, and (iii) obtain information for those 179 stores without a usable Google Street View image. These 216 stores were contacted by phone to determine the availability of fresh fruits, fresh vegetables, fresh dairy, and fresh bread in their stores. Of the total 280 stores contacted using this procedure (including the random sample of 64 described above), eight stores (2%) were identified as closed, and a total of 251 (90%) stores said “yes” to all of the above questions and were designated as full-line, while 21 (8%) did not carry one or more of these categories of food and were designated as not full-line.

In total, of the 474 grocery stores in the seven county area, 426 (90%) were determined to be full-line and 48 (10%) were determined to not be full-line or closed. These 426 full-line grocery stores from the DFM initiative were used as the ground truth to assess the validity of Yelp and Reference USA for identifying healthy food stores in this area.

Yelp

Yelp is a commercial website that provides user-contributed information and reviews of local businesses. Yelp was founded in 2004 to help people find local businesses and allows users to contribute different kinds of content, including reviews, rating scores, and photos. We used Yelp’s application program interface (API) to select all grocery stores within our seven county area using six Yelp business categories: “grocery stores,” “ethnic grocery stores,” “ethnic food markets,” “organic stores,” “health markets,” and “wholesale stores.” In September 2016, a total of 813 grocery stores were identified in the seven county area using Yelp’s API.

Reference USA

Reference USA is a proprietary listing of business establishments classified using North American Industry Classification System and Standard Industrial Classification System codes. We selected all businesses that fit within the category of full-line grocery stores under the MDARD definition, including primary NAICS or SIC codes for grocery stores, supermarkets, specialty food stores, and warehouse clubs/supercenters (see specific NAICS and SIC codes in Table 1). A total of 1631 grocery stores in the seven county area were identified using these NAICS and SIC codes with the Reference USA database (retrieval date September 2016).

All stores from all three sources (DFM, Reference USA, and Yelp) were geocoded and linked to the Census tract and Census block group levels to identify neighborhoods with healthy food. While the Census tract (a cluster of residential blocks drawn to encompass roughly 4000 people) is a typical spatial area used in studies of the local food environment [20–22], the smaller Census block group (averaging about 1500 people) captures more proximate food availability for residents. We identified the Census tracts and Census block groups in the seven county area where there was at least one grocery store according to each of the three data sources.

Analytic Strategy

Matching Stores Across Data Sources

We used a matching algorithm supplemented by manual checking to determine which grocery stores in Yelp

Table 1 Primary NAIC and SIC codes used to select grocery stores in Reference USA.

Category name	NAICS code	SIC code
Grocery stores	445,110 (supermarkets)	541,101 (food markets)
		541,105 (grocers)
		541,104 (food products retail)
		541,108 (grocers–health foods)
		539,901 (super markets and grocery stores)
Specialty food stores	445,299 (all other specialty food stores)	5499 (specialty food stores)
Warehouse clubs	452,910 (warehouse clubs and supercenters)	531,110 (wholesale clubs)

NAIC North American Industry Classification System, SIC Standard Industrial Classification System

matched with those in DFM and which stores in Reference USA matched with DFM. A fuzzy string matching algorithm was used to match the names and addresses of the businesses. This approach addresses typographical variations and misspellings in the format of store names and addresses across data sources. This is especially important in Yelp data since no naming conventions or protocols for storing names and address are followed. We used the Jaro-Winkler algorithm [23], which considers the number of common characters, the number of insertions, deletions, and transpositions, to compute a “string distance” measure between store names and addresses across two data sources. The algorithm gives a similarity score between 0 (no similarity) and 1 (exact match) to pairs of strings.

Pairs of stores with a score of 1 were considered to be the same store across sources. Pairs of stores with a score less than 1 and greater than 0.8 were considered to be probable matches and were manually verified. These included minor variations in store names between sources (e.g., Nino Salvaggio International Marketplace vs. Nino Salvaggio Intl Catering) and minor street number discrepancies. Pairs of stores with scores between 0.6 and 0.8 were considered ambiguous matches and were telephoned to determine whether they were a true match. Pairs of stores with scores below 0.6 were deemed nonmatches.

Assessing Validity

In order to determine the accuracy of Yelp and Reference USA for identifying sources of healthy food in urban neighborhoods, we assessed criterion validity using DFM data as the ground truth comparison. Criterion validity captures the accuracy of the secondary data source for identifying true full-line grocery stores and

the neighborhoods in which they are located. We first calculated the number of grocery stores in each secondary data source that were found in the DFM database (true positive rate; true grocery stores), as well as the number of grocery stores in each secondary data source that were not in the DFM database (false discovery rate; not true grocery stores). Similar to the sensitivity of a diagnostic tool in clinical practice for detecting patients with a disease, a higher true positive rate (equivalent to a low false negative rate or low miss rate) in a data source reflects a greater capacity to identify a higher proportion of true full-line grocery stores in the seven county area. Conversely, a high false discovery rate captures the lack of precision (low positive predictive value (PPV)) in the data source, translating into a lower chance that a store in Yelp or Reference USA is actually a full-line grocery store.

Because we were also interested in the spatial location of healthy food in these urban neighborhoods, we assessed the true positive rate (and miss rate) of Yelp and Reference USA for identifying neighborhoods with full-line grocers in both census tracts and census block groups. We also calculated the false positive rate, based on the number of neighborhoods with grocery stores according to the secondary data source that were not identified as neighborhoods with full-line stores in DFM. The converse to this is the true negative rate (or specificity), which captures the proportion of neighborhoods without a grocery store that were correctly identified as such by the secondary data source.

Results

Table 2 summarizes the number of grocery stores across all three data sources. According to the Detroit Food

Table 2 Summary of grocery stores in Michigan Seven-County Area by data source.

	Detroit Food Map	Reference USA	Yelp
Number of grocery stores	426	1631	813
Total number of census tracts with at least one Grocery store	340	847	524
Total number of block groups with at least one Grocery store	370	1121	628
Prevalence of grocery stores in census tracts	24%	60%	37%
Prevalence of grocery stores in block groups	9%	28%	16%

There are a total of 1414 census tracts and 4008 census block groups in the 7-county area

Map, 426 full-line grocery stores were located in the seven-county area. However, the prevalence of grocery stores was much higher according to the secondary data sources—almost twice the number of stores according to Yelp and almost four times the number of stores according to Reference USA. The seven-county metropolitan Detroit area encompasses 1414 census tracts and 4008 census block groups. According to DFM, 340 (24%) of these census tracts had at least one full-line grocery store, while 370 (9%) of the census block groups had at least one full-line store. Again, the estimated number of neighborhoods with healthy food was much higher in the secondary data sources, particularly Reference USA, where the prevalence of grocery stores in census tract and census block group neighborhoods was 60 and 28%, respectively.

Table 3 presents the true positive rate and precision for Yelp and Reference USA in comparison to DFM. A total of 314 of the 426 stores in the DFM list were identified in the Reference USA data (74% true positive rate, 26% miss rate). Yelp was less sensitive in identifying sources of healthy food: Just over half of the DFM stores were found in the Yelp data, yielding a true positive rate of 57% (43% miss rate).

Although Reference USA had a higher rate of coverage of all possible full-line stores (lower miss rate), it came at the cost of a higher false discovery rate. A total of 1317 of

the 1631 stores in the Reference USA database were not full-line stores according to DFM (81% false discovery rate). The Yelp database had higher precision than Reference USA (30 vs. 19%, respectively) with a false discovery rate of 70%.

When considering the spatial location of grocery stores within these Metro Detroit neighborhoods, Reference USA also had greater coverage of neighborhoods with full-line grocers compared to Yelp (Table 4). Of the 340 census tract neighborhoods with at least one full-line store according to DFM, 275 census tract neighborhoods were identified as such in the Reference USA database (81% true positive rate; 19% miss rate). In contrast, only 197 census tracts were identified as such in the Yelp data (58% true positive rate, 42% miss rate). The coverage for smaller census block group neighborhoods was slightly lower for both sources. Of the 370 census block groups with at least one full-line grocery store according to DFM, 289 (78%) were identified as such in Reference USA, while only 161 (56%) were identified as such in the Yelp data.

In contrast, the precision in correctly identifying neighborhoods with full-line grocery stores was higher in Yelp than in Reference USA. Yelp identified 524 census tracts as having at least one grocery store (197 correct, 38% precision, 62% false discovery rate). In contrast, Reference USA identified 847 census tracts as having at least one grocery store (275 correct, 32% precision, 68% false discovery rate). The false positive rate (incorrectly identifying a census tract as having a full-line grocery store) was 53% in the Reference USA database and only 30% in the Yelp database (true negative rates of 47 and 70%, respectively) (Table 4).

For both data sources, the precision was slightly lower when considering the availability of healthy food in the more proximate census block group neighborhoods, but Yelp still had higher data quality. Of the 628 block groups with grocery stores according to Yelp, 33% of them had full-line stores (67% false discovery rate). In contrast, of

Table 3 Is this a “full-line” grocery store?

	Yelp	Reference USA
True positive rate	57%	74%
Miss rate	43%	26%
Precision (PPV)	30%	19%
False discovery rate	70%	81%

Comparing grocery stores in Yelp and Reference USA with Detroit Food Map

PPV positive predictive value

Table 4 Is there a “full-line” grocery store in this neighborhood?

	Census Tract Neighborhoods		Census Block Group Neighborhoods	
	Yelp	Reference USA	Yelp	Reference USA
True positive rate	58%	81%	56%	78%
Miss rate	42%	19%	44%	22%
Precision (PPV)	38%	32%	33%	26%
False discovery rate	62%	68%	67%	74%
True negative rate	70%	47%	88%	77%
False positive rate	30%	53%	12%	23%

Comparing neighborhoods with at least one grocery store in Yelp and Reference USA with Detroit Food Map

PPV positive predictive value

the 1121 census block groups that Reference USA indicated had at least one grocery store, only 289 (26%) had full-line stores (74% false discovery rate) (Table 4).

Discussion

This paper is one of the first to examine the validity of using social media as a source of data on the urban food environment. We focused on full-line grocery stores in the greater Detroit area, a metropolitan area with a history of dramatic structural and economic changes and high poverty rates, where healthy food options can be scarce [18]. Although supermarkets and full-line stores constitute only a small segment of the city’s food system [24], they are an important source of healthy food in large metropolitan areas. Findings revealed that Reference USA overestimated the availability of grocery stores, while user-generated content from Yelp was more precise in identifying healthy food stores.

According to the Detroit Food Map, only 24% of the census tract neighborhoods in this area had at least one full-line grocery store. However, the prevalence of grocery stores in this area was dramatically overestimated by Reference USA. Fully, 60% of the census tracts in this area had a grocery store according to Reference USA. Yelp also overestimated the number of neighborhoods with grocery stores, but only by 7% compared to DFM. Thus, using secondary sources tends to overestimate the availability of healthy food, but user-generated content was more consistent with the ground truth source.

Compared to Reference USA, Yelp was somewhat less sensitive in identifying sources of healthy food: Just over half of the DFM stores were found in the Yelp data,

yielding a true positive rate of 57 vs. 74% for Reference USA. Similarly, the miss rate in Yelp was almost double that of Ref USA. However, although Reference USA had a lower miss rate, it came at the cost of a higher false discovery rate. Stores in the Yelp data had a higher chance of being a full-line grocery store than stores in the Reference USA database.

Similarly, with respect to location, Reference USA had a higher false positive rate than Yelp (incorrectly identifying a census tract as having a full-line grocery store when there was none). Previous research in South Carolina assessing the validity of secondary sources for identifying specific types of retail food establishments [15] also found a high overcount of supermarkets in Dun and Bradstreet data (61% false discovery rate), where only 43% of true supermarkets were correctly identified. False positive results have nontrivial consequences when researchers are using secondary data to identify areas where residents have access to healthy food. Specifically, compared to using Yelp, researchers using Reference USA would be more likely to determine that healthy food is available in a neighborhood when in fact it is not (higher false positives).

When considering social media versus proprietary data for identifying healthy food environments, researchers and practitioners should consider the trade-off between accuracy and coverage. While Reference USA had better coverage in our study (identifying more of the full-line grocery stores that were in the study area), that came at a cost of accuracy. Social media has the advantage that its data are derived from consumers, who essentially operate to vet the accuracy of the listing in a way that business listings do not. Furthermore, cost differences between the two data sources are striking, with Yelp being a free data source in comparison to the proprietary Reference USA.

Most of the stores in Reference USA that were not found in DFM were other food stores that did not match the definition of full-line grocers (e.g., convenience stores, other markets, ethnic food stores). However, there were also stores in the Reference USA database that were not food stores altogether (e.g., dollar stores, tire stores, garden supply stores, and clothing stores), suggesting that the NAICS and SIC codes in proprietary business databases are not screened for accuracy. Our results offer the intriguing possibility that free crowdsourced data may be more accurate than the costly alternatives that have been typically used in this field of research. This is promising not only for researchers but also for health care practitioners whose work touches upon food availability. Urban planners and public health officials, for example, regularly make decisions regarding land use and local programs for which local food availability information is an important input. As the healthcare system in the US increasingly moves to address the social determinants of health [25], there is growing interest in automated methods for identifying community resources to support referrals for disadvantaged patients. Our research suggests the possibility of using social media data for food access-related information, in contrast to expensive alternatives such as data available from commercial vendors.

Despite its innovation, this study had limitations that should be noted. We used the Detroit Food Map as our ground truth comparison, where the definition of full-line grocers was very specific. There was also a temporal mismatch (nearly a 2-year difference) in the data collected by DFM and the two secondary data sources. In a region where economic distress can lead to a short lifespan for businesses, many stores may have opened and closed in this 2-year period. While we used a telephone verification method to identify stores that closed since the MDARD list was generated in December 2014, new stores that opened since then (and found in Reference USA and Yelp) were potential nonmatches across sources.

There are also likely to be differences with respect to social media use in different areas. The adoption of broadband services is lower in rural compared to urban areas [26], which is likely to have consequences for the coverage of rural grocery stores in Yelp. Similarly, there may be notable differences in social media use in areas with different demographic and socioeconomic composition. Examining the validity of crowdsourced data by area characteristics is beyond the scope of this paper, but this is an important area of future research. There are a number of other crowdsourced data that could be used to assess

features of the urban environment in comparison to a ground truth source, including OpenStreetMap (<https://www.openstreetmap.us/>), Foursquare (<https://foursquare.com/>), MapMyRun (<http://www.mapmyrun.com/>)(for physical activity routes), and Zagat (<https://www.zagat.com/>)(for restaurants). Research using Foursquare to capture land use [27] is emerging in the urban planning and transportation literature as an alternative to the more costly traditional land use surveys. Other research has examined the utility of using crowdsourced data for characterizing the social and built environment of cities [16, 28]. We hope that further studies will pursue these research questions with other data in order to fully explore the potential of using crowdsourced data for urban health research.

In spite of these limitations, this study breaks new ground in considering the potential of social media for identifying sources of healthy food in urban neighborhoods. Researchers today can draw on an expanding set of information about urban neighborhoods, including novel sources such as commercial websites like Yelp, which rely on user contributions, in addition to more traditional private marketing databases. Data on full-line grocery stores are not readily available from many sources but are particularly important for research on food systems and urban health. Our findings suggest that researchers investigating the relationship between the nutrition environment and health can consider Yelp as a valid source for identifying sources of healthy food in urban environments.

Compliance with Ethical Standards

Funding Sources This study was funded by the University of Michigan Office of Research and the Rackham Graduate School Social Sciences Annual Institute; MCubed; the Alfred P Sloan Foundation Grant Number: 2014-5-05 DS; and the Gordon and Betty Moore Foundation through Grant GBMF3943 University of Michigan.

References

1. Giskes K, van Lenthe F, Avendano-Pabon M, Brug J. A systematic review of environmental factors and obesogenic dietary intakes among adults: are we getting closer to understanding obesogenic environments? *Obes Rev.* 2011;12(5): e95–e106.
2. Morland KB, Evenson KR. Obesity prevalence and the local food environment. *Health & Place.* 2009;15(2):491–5.

3. Auchincloss AH, Roux AVD, Brown DG, Erdmann CA, Bertoni AG. Neighborhood resources for physical activity and healthy foods and their association with insulin resistance. *Epidemiology*. 2008;19(1):146–57.
4. Auchincloss AH, Roux AVD, Mujahid MS, Shen M, Bertoni AG, Carnethon MR. Neighborhood resources for physical activity and healthy foods and incidence of type 2 diabetes mellitus: the multi-ethnic study of atherosclerosis. *Arch Intern Med*. 2009;169(18):1698–704.
5. Morland K, Roux AVD, Wing S. Supermarkets, other food stores, and obesity: the atherosclerosis risk in communities study. *Am J Prev Med*. 2006;30(4):333–9.
6. Powell LM, Auld MC, Chaloupka FJ, O'Malley PM, Johnston LD. Associations between access to food stores and adolescent body mass index. *Am J Prev Med*. 2007;33(4):S301–S7.
7. Liu GC, Wilson JS, Qi R, Ying J. Green neighborhoods, food retail and childhood overweight: differences by population density. *Am J Health Promot*. 2007;21(4s):317–25.
8. Mantovani R, Daft L, Macaluso T, Welsh J, Hoffman K. Authorized food retailers' characteristics and access study. US Department of Agriculture: Alexandria VA; 1997.
9. Glanz K, Basil M, Maibach E, Goldberg J, Snyder D. Why Americans eat what they do: taste, nutrition, cost, convenience, and weight control concerns as influences on food consumption. *J Am Diet Assoc*. 1998;98(10):1118–26.
10. Larson NI, Story MT, Nelson MC. Neighborhood environments: disparities in access to healthy foods in the U.S. *Am J Prev Med*. 2009;36(1):74–81.
11. Fleischhacker SE, Evenson KR, Sharkey J, Pitts SBJ, Rodriguez DA. Validity of secondary retail food outlet data: a systematic review. *Am J Prev Med*. 2013;45(4):462–73.
12. Lake AA, Burgoine T, Greenhalgh F, Stamp E, Tyrrell R. The foodscape: classification and field validation of secondary data sources. *Health & Place*. 2010;16(4):666–73.
13. Liese AD, Colabianchi N, Lamichhane AP, Barnes TL, Hibbert JD, Porter DE, et al. Validation of 3 food outlet databases: completeness and geospatial accuracy in rural and urban food environments. *Am J Epidemiol*. 2010;172(11):1324–33.
14. McKinnon RA, Reedy J, Morrisette MA, Lytle LA, Yaroch AL. Measures of the food environment. A compilation of the literature, 1990-2007. *Am J Prev Med*. [Review]. 2009;36(4 SUPPL):S124–S33.
15. Liese AD, Barnes TL, Lamichhane AP, Hibbert JD, Colabianchi N, Lawson AB. Characterizing the food retail environment: impact of count, type, and geospatial error in 2 secondary data sources. *J Nutr Educ Behav*. 2013;45(5):435–42.
16. Stefanidis A, Crooks A, Radzikowski J. Harvesting ambient geospatial information from social media feeds. *GeoJournal*. 2013;78(2):319–38.
17. Manduca R, Spielman SE, Folch D, editors. Fast food data: where user-generated content works and where it doesn't. Chicago, IL: Workshops on Big Data and Urban Informatics; 2014.
18. Zenk SN, Schulz AJ, Israel BA, James SA, Bao S, Wilson ML. Neighborhood racial composition, neighborhood poverty, and the spatial accessibility of supermarkets in metropolitan Detroit. *Am J Public Health*. 2005;95(4):660–7.
19. Michigan Department of Agriculture. Michigan's food & agriculture industry. 2012. Retrieved from http://www.michigan.gov/documents/mdard/1262-AgReport-2012_2_404589_7. Accessed 12 March 2016.
20. Morland K, Wing S, Diez Roux AV, Poole C. Neighborhood characteristics associated with the location of food stores and food service places. *Am J Prev Med*. 2002;22:23–9.
21. Morland K, Wing S, Diez Roux AV. The contextual effect of the local food environment on residents' diets. *Am J Public Health*. 2002;92:1761–7.
22. Franco M, Roux AVD, Glass TA, Caballero B, Brancati FL. Neighborhood characteristics and availability of healthy foods in Baltimore. *Am J Prev Med*. 2008;35(6):561–7.
23. Winkler WE. String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage. Washington DC: US Census Bureau, Division SR; 1990.
24. Taylor DE, Ard KJ. Food availability and the Food Desert frame in Detroit: an overview of the City's food system. *Environ Pract*. 2015;17(02):102–33.
25. National Academies of Sciences, Engineering, Medicine. A framework for educating health professionals to address the social determinants of health. Washington, DC: The National Academies Press; 2016.
26. LaRose R, Gregg JL, Strover S, Straubhaar J, Carpenter S. Closing the rural broadband gap. Sage – Thousand Oaks, CA: promoting adoption of the internet in rural America. *Telecommun Policy*. 2007;31(6–7):359–73.
27. Spyrtos S, Stathakis D, Lutz M, Tsinaraki C. Using Foursquare place data for estimating building block use. *Environment and Planning B*. Sage – Thousand Oaks, CA: Planning and Design. Article first published online: July 27, 2016. doi:10.1177/0265813516637607.
28. Spyrtos S, Stathakis D. Evaluating the services and facilities of European cities using crowdsourced place data. *Environment and Planning B: Urban Analytics and City Science*. Article first published online: January 2, 2017. doi:10.1177/0265813516686070