

## Short Communication

## Validation of presence of supermarkets and fast-food outlets in Copenhagen: case study comparison of multiple sources of secondary data

Chalida M Svastisalee<sup>1,2,\*</sup>, Bjørn E Holstein<sup>1,2</sup> and Pernille Due<sup>1</sup><sup>1</sup>National Institute of Public Health, University of Southern Denmark, Øster Farimagsgade 5A, 2nd Floor, Copenhagen K, Denmark 1399; <sup>2</sup>Department of Public Health, University of Copenhagen, Copenhagen, Denmark

Submitted 9 October 2011: Final revision received 8 February 2012: Accepted 16 February 2012: First published online 22 March 2012

**Abstract***Objective:* We examined the quality of food outlet addresses provided by secondary sources and determined whether they could be physically located in the field.*Design:* Addresses of food outlets in fourteen school districts in the northern part of Copenhagen were obtained from multiple business locators. We geocoded 202 addresses using a geographic information system and cross-referenced the sources against each other using a validation grid. Physical presence was determined via street survey. We applied gamma statistics and calculated positive predictive value, sensitivity and percentage agreement to assess the overall correspondence between our test of physical presence and each source of secondary information.*Setting:* The study took place within city boundaries of Copenhagen, Denmark.*Subjects:* Food outlets within fourteen school districts within Copenhagen.*Results:* Positive predictive value between field results and secondary sources indicated good to excellent correspondence (range: 0.81–0.98), comparable with other studies. Gamma coefficients indicated low to high positive correspondence (range: 0.23–0.98).*Conclusions:* Despite moderately high correspondence between secondary sources of address information and field observation, the findings illustrate that the use of combined sources is recommended.**Keywords**  
Food environment  
Geographic information system  
Validation  
Food outlets  
Europe  
Secondary sources

Investigations examining the relationship between the food environment and health depend on secondary sources of information to supply characteristics of local food access. Relatively little is published on the validity and reliability of these sources<sup>(1)</sup>.

In a recent review by Kelly and colleagues<sup>(2)</sup>, methods used to validate the location of food outlets obtained from secondary data range from list cross-referencing<sup>(3–6)</sup> to direct observation<sup>(7–14)</sup>. Correspondence between sources was reported as percentage match or agreement<sup>(6,7,9,10,13,15)</sup>, sensitivity<sup>(11–13)</sup> or positive predictive value (PPV)<sup>(12,13,16)</sup>. Validation studies reporting correspondence statistics have generally been performed in North America<sup>(3,4,6,9–12,14,15,17)</sup> or the UK<sup>(5,7,13,16)</sup>. To our knowledge, there are no other published validation studies of the food environment elsewhere. We are interested in investigating the quality of secondary lists in Denmark, as there tends to be a strong tradition in using public registry information, but less emphasis on its validation. Therefore, our aim was to evaluate to what extent food outlet addresses from various sources can be confirmed by direct observation.

**Methods**

Data were collected as part of a study examining food outlet exposure in Copenhagen<sup>(18)</sup>. We obtained a combined total of 202 addresses of all food outlets trading in 2006 within fourteen Copenhagen school districts. Addresses were initially obtained from the Danish Central Business Registry (CVR)<sup>(19)</sup>, which is the national tax registry (*n* 159) used to list all legally operating businesses. Food outlets were selected within postal codes of fourteen school districts in Copenhagen, and by business categories drawn by European business codes or Nomenclature des Activités Economiques (NACE) codes<sup>(20)</sup>. Additionally, we cross-referenced the tax registry for absence or presence using several sources of address information obtained for or referenced for 2006: (i) Teledanmark<sup>(21)</sup>, a telephone company (*n* 189); (ii) Google Maps Denmark (<http://www.googlemaps.dk>; *n* 172); (iii) Krak Denmark (<http://www.krak.dk>; *n* 171), a web-based business search engine; and (iv) Stockmann Company<sup>(22,23)</sup> (*n* 109), a commercial address supplier that publishes annual information on large

\*Corresponding author: Email chsv@niph.dk

chain supermarkets and food kiosks (single-owned and chain stationary convenience stores). Addresses were geocoded and categorized according to NACE descriptions.

We conducted an on-site food outlet inventory in fourteen Copenhagen school districts by using a combined address list derived from five sources of information. Gamma statistics (Goodman–Kruskal's gamma) were performed to determine a summary level of concordant and discordant pairs between secondary sources and field data, establishing statistical significance at the 0.05 level. We also calculated the PPV, sensitivity and percentage agreement<sup>(24)</sup> in order to maintain comparability with other studies. We use the same categorization for PPV as Paquet and colleagues<sup>(11)</sup>: <0.30 (poor), 0.31–0.50 (fair), 0.51–0.70 (moderate), 0.71–0.90 (good) and >0.91 (excellent). Gamma coefficients were evaluated according to Kreiner<sup>(25)</sup>: >0.30 (strong), 0.15–0.30 (moderate) and <0.15 (weak).

## Results

A total of 147 of 202 food outlets were present, while fifty-five were absent from field survey (Table 1). Approximately 25% of kiosks and 27% of fast-food outlets (hot dog stands and grill bars) were found in the field, while another 26% of kiosks and 27% of fast-food outlets were absent.

**Table 1** Frequency and percentage of food outlet classifications in fieldwork, Copenhagen, Denmark, 2006

Category	Total present	%	Total absent	%
Retail bakery	5	3.4	5	9.1
Service station with kiosk	1	0.7	0	0
Small grocer	13	8.8	7	12.7
Kiosk	36	24.5	14	25.5
Supermarket	2	1.4	0	0
Discount grocer	16	10.9	1	1.8
Green grocer	1	0.7	0	0
Hypermarket	9	6.1	7	12.7
Chocolatier	2	1.4	0	0
Restaurant	20	13.6	5	9.1
Hot dog stand, grill bar	40	27.2	15	27.3
Café/coffee bar	2	1.4	1	1.8
Total	147	100.0	55	100.0

Compared with an ideal PPV (1.0), all data sources scored above 0.80 (Table 2), indicating a high proportion of addresses positively identified on the street. There was strong and significant positive concordance between the field and Google Maps ( $\gamma = 0.87$ ) and Krak.dk ( $\gamma = 0.89$ ), with Stockmann having the highest correspondence with field survey ( $\gamma = 0.98$ ). Sensitivity values were highly significant, ranging between 0.74 (Teledanmark) and 0.93 (Stockmann), indicating moderate to high concordant pairs. In correspondence with field findings, the tax registry had the lowest percentage of agreement (0.64), while Stockmann had the highest (0.91).

## Discussion

Findings suggest that levels of correspondence between secondary data sources and field analysis were good to excellent and comparable with coefficients reported elsewhere<sup>(11,13,16)</sup>. Hoehner and Schootman<sup>(24)</sup> reported overall percentage agreement of 32.0% using two databases of food outlet data in St. Louis, MO, USA, while we report higher values. Differences in percentage agreement between the studies may be attributed to study size and scope, and are thus not directly comparable. We included gamma statistics in the present analysis because we wanted to obtain robust measures of agreement, which allows us to examine the association of two groups containing a large number of ties. Gamma statistics were low when percentage agreement fell below 0.75, indicating that scores may be sensitive to distributional changes or low cell number, further illustrating the importance of examining frequency distributions of concordant and discordant pairs.

There are also several explanations for differences found between sources of address information. One difference may be related to why each source of address information is constructed in the first place. The tax registry for instance, as it is designed to list tax revenue, likely includes extraneous information such as a home address instead of a business address, if the operator owns a small business. Additionally, the registry chronicles all businesses ever having a tax number, but does not

**Table 2** Comparability of sources of food outlet data against fieldwork, by type and by agreement statistics, Copenhagen, Denmark, 2006

Statistic	Teledanmark (n 189)	Tax registry (n 159)	Google Maps (n 172)	Krak.dk (n 128)	Stockmann (n 109)
Gamma	0.27	0.23	0.87*	0.89*	0.98*
95% CI	−0.27, 0.81	−0.11, 0.57	0.76, 0.98	0.79, 0.99	0.95, 1.00
PPV	0.95*	0.81*	0.95*	0.90*	0.98*
95% CI	0.90, 0.98	0.74, 0.87	0.92, 0.97	0.83, 0.96	0.93, 1.00
Sensitivity	0.74*	0.75*	0.81*	0.88*	0.93*
95% CI	0.67, 0.80	0.67, 0.81	0.75, 0.87	0.79, 0.93	0.86, 0.97
Percentage agreement	0.71	0.64	0.78	0.80	0.91

PPV, positive predictive value.

\*Statistic (95% CI) significant at <0.0001 level.

provide information on whether the business is actively operating. Characteristics based solely on the tax registry may result in an overestimation in the number of food outlets due to inclusion of either obsolete businesses or home addresses of owners<sup>(26)</sup>. In our case, tax registry information would have led to under-reporting of the food environment. We noted higher correspondence values for other secondary sources of address information (the telephone company, Krak.dk and Google Maps), while acknowledging that results could be affected by subscription status, leading to under-reporting in the number of food outlets if the vendor withdraws advertising services during sampling of addresses. However, secondary sources of business information such as the telephone company are still significant sources of food outlet information<sup>(26)</sup>. Lastly, we found very high correspondence using addresses purchased from Stockmann, which may indicate the reliability of purchased information. However, the information provided represents only high-revenue supermarkets as well as outlets listed as food kiosks. Thus, a considerable number of small independent supermarkets or fast-food outlets will not be represented.

Study limitations include the use of a small study sample, rather than a city-wide census, and we have not considered other potential contexts such as rural or urban differences. While area deprivation may affect percentage agreement as shown by Cummins and Macintyre<sup>(16)</sup>, we were unable to stratify the analysis due to sample size limitations. Additionally, we cannot rule out misclassification bias. Business categorizations tend to be broader within the tax registry and may not reflect actual items sold by food outlets, making it even more necessary to conduct field surveys. Overall strengths include an examination of a wide variety of food outlets, providing a varied representation of the spatial distribution of the neighbourhood food environment. Our study used a variety of sources of secondary information, providing us with a reliable estimation of food outlets in the field.

## Conclusion

The current study demonstrates that address information used to characterize the food environment should be used with caution and correspondence between secondary address sources and field observations vary. While there may not be a single source used to characterize food environments in Denmark, the use of combined sources is recommended.

## Acknowledgements

This work was made possible through grants funded by the Danish Graduate School in Public Health Science (GRASPH) at the University of Copenhagen; the Danish

Health Fund (grant number 2007B057); the Danish Heart Foundation (grant number 07-4-B195-A1547-22388F); and the Nordea Fund. The authors declare that there are no conflicts of interest. C.M.S., B.E.H. and P.D. were responsible for conception of the study, study design and setup. C.M.S. carried out data collection, analyses and drafting of the manuscript. B.E.H. and P.D. assisted with drafting the manuscript and provided additional commentary.

## References

1. McKinnon RA, Reedy J, Morrissette MA *et al.* (2009) Measures of the food environment: a compilation of the literature, 1990–2007. *Am J Prev Med* **36**, 4 Suppl., S124–S133.
2. Kelly B, Flood VM & Yeatman H (2011) Measuring local food environments: an overview of available methods and measures. *Health Place* **17**, 1284–1293.
3. Maddock J (2004) The relationship between obesity and the prevalence of fast food restaurants: state-level analysis. *Am J Health Promot* **19**, 137–143.
4. Wang MC, Gonzalez AA, Ritchie LD *et al.* (2006) The neighborhood food environment: sources of historical data on retail food stores. *Int J Behav Nutr Phys Act* **3**, 15.
5. Burgoine T, Lake AA, Stamp E *et al.* (2009) Changing foodscapes 1980–2000, using the ASH30 Study. *Appetite* **53**, 157–165.
6. Jilcott SB, McGuirt JT, Imai S *et al.* (2010) Measuring the retail food environment in rural and urban North Carolina counties. *J Public Health Manag Pract* **16**, 432–440.
7. Cummins S & Macintyre S (1999) The location of food stores in urban areas: a case study in Glasgow. *Br Food J* **101**, 545–553.
8. Cummins S & Macintyre S (2002) A systematic study of an urban foodscape: the price and availability of food in Greater Glasgow. *Urban Studies* **39**, 2115–2130.
9. Creel JS, Sharkey JR, McIntosh A *et al.* (2008) Availability of healthier options in traditional and nontraditional rural fast-food outlets. *BMC Public Health* **8**, 395.
10. Sharkey JR & Horel S (2008) Neighborhood socioeconomic deprivation and minority composition are associated with better potential spatial access to the ground-truthed food environment in a large rural area. *J Nutr* **138**, 620–627.
11. Paquet C, Daniel M, Kestens Y *et al.* (2008) Field validation of listings of food stores and commercial physical activity establishments from secondary data. *Int J Behav Nutr Phys Act* **5**, 58.
12. Daniel M, Paquet C, Auger N *et al.* (2010) Association of fast-food restaurant and fruit and vegetable store densities with cardiovascular mortality in a metropolitan population. *Eur J Epidemiol* **25**, 711–719.
13. Lake AA, Burgoine T, Greenhalgh F *et al.* (2010) The foodscape: classification and field validation of secondary data sources. *Health Place* **16**, 666–673.
14. Sharkey JR, Johnson CM & Dean WR (2010) Food access and perceptions of the community and household food environment as correlates of fruit and vegetable intake among rural seniors. *BMC Geriatr* **10**, 32.
15. Sharkey JR (2009) Measuring potential access to food stores and food-service places in rural areas in the US. *Am J Prev Med* **36**, 4 Suppl., S151–S155.
16. Cummins S & Macintyre S (2009) Are secondary data sources on the neighbourhood food environment accurate? Case-study in Glasgow, UK. *Prev Med* **49**, 527–528.
17. Sharkey JR, Horel S, Han D *et al.* (2009) Association between neighborhood need and spatial access to food stores and fast food restaurants in neighborhoods of colonias. *Int J Health Geogr* **8**, 9.

18. Svastisalee CM, Nordahl H, Glumer C *et al.* (2011) Supermarket and fast-food outlet exposure in Copenhagen: associations with socio-economic and demographic characteristics. *Public Health Nutr* **14**, 1618–1626.
19. Experian Denmark (2006) CD Direct Online Market Guide. <https://www.kob.dk> (accessed December 2008).
20. European Commission (2005) List of NACE codes. [http://europe.eu.int/comm/competition/mergers/cases/index/nace\\_all.html](http://europe.eu.int/comm/competition/mergers/cases/index/nace_all.html) (accessed March 2005).
21. Teledanmark (2006) Teledanmark Marketing Data (Teledanmark Markeds Data). <http://www.nnmarkedsdata.dk> (accessed March 2006).
22. Stockmann Group (2007) *Supermarket Yearbook*. Asnæs, Denmark: Retail Institute Scandinavia.
23. Stockmann Group (2007) *Kiosk Yearbook*. Asnæs, Denmark: Retail Institute Scandinavia.
24. Hoehner CM & Schootman M (2010) Concordance of commercial data sources for neighborhood-effects studies. *J Urban Health* **87**, 713–725.
25. Kreiner S (1999) Korrelations koefficienter (Correlation coefficients). In *Secondary Statistik problemløsning. Præmisser, teknik og analyse. (Statistical Problemsolving. Premises, Techniques and Analyses)*, pp. 189–238. Copenhagen: Jurist og Økonomforbundets Forlag (Publisher of the Association of Lawyers and Economists).
26. Burgoine T (2010) Collecting accurate secondary foodscape data. A reflection on the trials and tribulations. *Appetite* **55**, 522–527.