Field Validation of Secondary Data Sources for Enumerating Retail Tobacco Outlets in a State without Tobacco Outlet Licensing

Heather D'Angelo[a,*], **Sheila Fleischhacker**[b], **Shyanika W. Rose**[a], and **Kurt M. Ribisl**[a,c]

[a]Department of Health Behavior, University of North Carolina (UNC) Gillings School of Global Public Health, Chapel Hill, North Carolina, USA

[b]National Institutes of Health, Division of Nutrition Research Coordination, Washington, DC, USA

[c]UNC Lineberger Comprehensive Cancer Center, Chapel Hill, North Carolina, USA

Abstract

Identifying tobacco retail outlets for U.S. FDA compliance checks or calculating tobacco outlet is difficult in the 13 states without tobacco retail licensing or where licensing lists are unavailable for research. This study uses primary data collection to identify tobacco outlets in three counties in a non-licensing state and validate two commercial secondary data sources. We calculated sensitivity and positive predictive values (PPV) to validate the secondary data sources, and conducted a geospatial analysis to determine correct allocation to census tract. ReferenceUSA had almost perfect sensitivity (0.82) while Dun & Bradstreet (D&B) had substantial sensitivity (0.69) for identifying tobacco outlets; combined, sensitivity improved to 0.89. D&B identified fewer "false positives" with a PPV of 0.82 compared to 0.71 for ReferenceUSA. ReferenceUSA geocoded over 90% of outlets to the correct census tract. Combining two commercial data sources resulted in enumeration of nearly 90% of tobacco outlets in a three county area. Commercial databases appear to provide a reasonably accurate way to identify tobacco outlets for enforcement operations and density estimation.

Keywords

Validation study; Tobacco outlets; Commercial secondary data

Introduction

Access to supermarkets, convenience stores, and recreational facilities has been associated with smoking (Henriksen et al., 2008), obesity (Lovasi, 2009), and physical activity (Gordon Larsen, 2006) and may create an environment that either enhances or diminishes a resident's

*Correspondence to: 1535 Ann St, Beaufort, NC 28516. Tel.: +413 214 2687; fax: 9198 43 3418. hdangelo@email.unc.edu.

ability to make health promoting choices. The number, types and locations of retail outlets are often proxies for access to tobacco products, (Henriksen et al., 2008), food (Larson et al., 2009), or places to be physically active (Boone-Heinonen et al., 2009). For example, studies have found youth living in communities with comparatively higher retail tobacco outlet density were more likely to use tobacco and living near tobacco outlets made it more difficult for adults to quit smoking (Henriksen et al., 2008, Novak et al., 2006, Reitzel et al., 2011). Lower income and racial/ethnic minority neighborhoods have disproportionately higher exposure to retail tobacco outlets (Hyland et al., 2003, Fakunle et al., 2010, Peterson et al., 2005, Rodriguez et al., 2012, Schneider et al., 2005), potentially contributing to higher tobacco use among these groups (Frieden, 2011). To further our understanding of how the tobacco retail environment influences tobacco use valid data sources are needed to enumerate tobacco outlets and to accurately identify areas with increased exposure to tobacco products.

In the United States (US), tobacco retail licensing data is often used to calculate tobacco outlet density (Fakunle et al., 2010, Henriksen et al., 2008, Lipperman-Kreda et al., 2012, Hyland et al., 2003, Peterson et al., 2005, Schneider et al., 2005). Yet, licensing lists may be unavailable to researchers, and 13 States do not require tobacco retailer licensing, making such estimation difficult (CDC, 2012). The quality of the sampling list used for US Food and Drug Administration (FDA) compliance checks or to enforce youth tobacco access laws determines whether and how many tobacco outlets will be missed. A state without a licensing list as a starting point may create a sampling frame from state or local business lists, statewide retail license/permit lists or statewide liquor license/permit lists.

Over the last decade, obesity researchers have increasingly relied on secondary data sources (e.g., ReferenceUSA or government food registries) to enumerate retail food and recreational environments. They have linked information on the location of food and activity resources to neighborhood characteristics to understand the impact on weight status and disparities in obesity faced by lower income, certain racial/ethnic groups and rural communities. (Powell et al., 2011, Fleischhacker et al., 2011)

Although primary data collection is the most accurate approach, (Hosler and Dharssi, 2010, Sharkey, 2009) it is resource intensive. "Ground-truthing," or identifying outlets through a systematic field canvass of a targeted study area without using secondary data sources, may be feasible in small cities or counties, but is daunting in larger areas (Fleischhacker et al., 2013). Researchers or state-level staff may need to rely on secondary data sources to enumerate larger study areas. For example, tobacco outlets may be located anywhere on the over 40,000 miles of primary and secondary roads in Kentucky or Virginia, neither of which has tobacco retailer licensing. Commercial secondary data sources have several benefits compared with primary data collection: they can be searched by establishment type (e.g., convenience stores), provide telephone numbers and addresses to aid in the verification process, and are typically less expensive than primary data collection.

Since grocery and convenience stores also sell tobacco products (Hosler and Kammer, 2012), similar methods could potentially help identify tobacco outlets. While there have been numerous studies examining the validity for enumerating food outlets (Fleischhacker et

al., 2013), no studies, to our knowledge, have examined the validity of secondary data sources for enumerating tobacco outlets. One study estimated tobacco outlet density by gathering primary data (Novak et al., 2006) and another identified 88% of the outlets on Washington State's licensing list using a secondary data source without conducting primary data collection (Rodriguez et al., 2012).

The purpose of this study is to provide evidence-informed guidance on whether secondary data sources are a reasonable alternative to primary data collection in order to enumerate the tobacco retail environment. A second purpose is to examine whether secondary data sources allocate outlets to the correct census tract, and to compare tobacco outlet density calculated by primary and secondary sources, particularly in jurisdictions that do not have a comprehensive list of tobacco outlets.

Methods

Study Area

The study area described previously (Rose et al., 2013) included three geographically diverse counties in North Carolina (NC), USA, a state without tobacco retail licensing. Buncombe County, including the Asheville, NC Metro Area, has a median household income of $44,190, 6.4% of the population is African American, and encompasses 656.7 square miles in Appalachia. Durham County is more urban and includes the Durham-Chapel Hill, NC Metro Area, has a median household income of $49,894, 38.0% of the population is African American and encompasses 286.0 square miles. New Hanover County includes the coastal Wilmington, NC Metro Area, has a median household income of $48,553, 14.8% of the population is African American, and encompasses 191.5 square miles. (US Census Bureau, 2012)

Identifying Probable Tobacco Outlets Using Secondary Data Sources

We searched ReferenceUSA (Infogroup, Inc.) in May 2011 using primary North American Industry Classification System (NAICS) codes (U.S. Census Bureau, 2007) for the top ten retail industries that sell tobacco products (U.S. Census Bureau, 2010)(Table 1). We selected NAICS codes that represented approximately 98% of all tobacco sales in 2007 and used them to produce a list of probable tobacco retail outlets in the three study counties (D'Angelo et al., 2012). Unlike retail food outlets that can be identified by NAICS codes directly (e.g., convenience stores), the tobacco product category is sold at a variety of outlet types and not every outlet of a particular type sells tobacco. The one exception is "tobacco stores," that account for only 10% of tobacco product sales, while convenience and gas/convenience stores account for over 50% of tobacco product sales(U.S. Census Bureau, 2010). Therefore, we first identified *probable tobacco outlets*, outlets that are a store type listed in Table 1 and then secondarily verified them in the field as *actual tobacco outlets* if they sold tobacco products. To estimate the validity of an additional business list, we obtained a list from Dun & Bradstreet, Inc. (Dun & Bradstreet, 2005) (D&B) in November 2011 after primary data collection using the same 10 NAICS codes.

We cleaned the ReferenceUSA and D&B lists first, by sorting the lists by NAICS code and excluding those not on our inclusion list (e.g., Food Health Supplement Stores). Next, we excluded chains that do not sell tobacco products (e.g. Target™ (Target, 2009)). Given a high rate of non-retail outlets in the pharmacy category, we called all non-chain retail pharmacies, pharmaceutical companies or labs identified in the Pharmacy and Drug Store NAICS category to verify whether tobacco was sold and, if not, excluded them. After exclusions were applied, we sorted the lists by address and eliminated exact duplicates by name and address within each data source separately. We flagged listings with the same address but different name for field verification.

Identifying Actual Tobacco Outlets through Primary Data Collection

Eight trained observers in teams of two conducted primary data collection from June to September 2011. Teams drove all primary and secondary roads in each county using 2010 TIGER/Line roads data from the Census Bureau. Shopping centers were included, but office parks and industrial parks were not. We used ArcGIS Version 10.0 to create driving routes for each county. Primary data collection covered 1,622 miles in Durham County, 1,330 miles in Buncombe County and 522 miles in New Hanover County. Teams located and verified each outlet listed and spotted any tobacco retail outlets that fell into one of the ten NAICS codes listed in Table 1 on the route that were not listed. Each outlet was assigned one of the following dispositions: (1) Sells tobacco to consumers, in business; (2) Does not sell tobacco to consumers, in business; (3) Out of business; (4) Could not locate; or (5) Duplicate record.

Teams verified the sale of tobacco products from the exterior by observing the presence of tobacco product advertisements or "We card" signs. If neither of these were visible, a data collector entered the outlet to determine whether it sold tobacco products. An outlet was classified as *out of business* if it appeared to be closed permanently (e.g. empty store front). If an outlet was closed temporarily and we could not confirm the disposition, it was revisited or called to verify whether tobacco was sold. Outlets were classified as *could not locate* if either the address was not found or the outlet was not at the address listed. We eliminated duplicate records from the final list of tobacco outlets. Observers recorded a Global Positioning System (GPS) waypoint at the front door of each outlet using a Garmin GPSMap 60Cx and took a photograph of the outlet.

The D&B list was not verified on-site, but was matched with the final list of outlets from primary data collection. We considered outlets with the exact address as a match. We assigned the disposition and outlet type identified during primary data collection to all matched outlets. We mapped outlets to determine whether any new outlets observed were duplicates of those on the D&B list. We called outlets listed by D&B and not observed to determine their disposition.

Statistical Analyses

We calculated sensitivity and positive predictive value (PPV) for each secondary data source, and for both combined. Sensitivity measures how well the data source captures the actual number of outlets, or "true positives". For example, if ReferenceUSA identified 50 of

100 actual outlets, its sensitivity would be 0.50. PPV gives an understanding of the number of "false positives". For example, if we found 50 tobacco outlets during primary data collection out of 200 outlets identified by ReferenceUSA, the PPV would be 0.25. Sensitivity and PPV were calculated for probable and actual tobacco outlets because the purpose of commercial secondary data sources is to identify all outlets, not just those that sell tobacco. We calculated the following: (1) *probable tobacco outlet sensitivity*: the proportion of outlets observed during primary data collection that were listed by the secondary data source; (2) *actual tobacco outlet sensitivity*: the proportion of outlets selling tobacco observed during primary data collection that were listed by the secondary data source; (3) *probable tobacco outlet PPV*: the proportion of outlets listed by the secondary data source that were observed during primary data collection; and (4) *actual tobacco outlet PPV*: the proportion of outlets listed by the secondary data source that were observed as selling tobacco during primary data collection. We used the following criteria adapted from Altman (1991) to interpret sensitivity and PPV: poor (0–0.2), fair (0.21–0.4), moderate (0.41–0.6), good (0.61–0.8) and very good (0.81–1.0). We calculated measures and standard errors overall and by outlet type. We assigned the field verified outlet type rather than the outlet type assigned by the data source; for outlets closed or not located, we used the outlet type assigned by the data source. We conducted analyses in June 2012 using SAS Software, Version 9.2. We used the GPS coordinate, ArcGIS and data from the 2010 U.S. Census to determine allocation to census tract. Finally, we calculated tobacco outlet density from each source separately, and both combined, and examined correlations between estimated and actual mean tobacco outlet density (number of outlets per 1000 people).

Results

Primary data collection identified 662 tobacco outlets (Table 2). Teams added 73 of those outlets in the field because they were not identified by either secondary data source. Convenience stores with gas stations were the most common type of tobacco outlet (44.9%), followed by convenience stores (15.4%), supermarkets (15.1%), pharmacies (11.9%) and tobacco stores (5.7%).

ReferenceUSA identified 971 probable tobacco outlets; 761 remained after cleaning the lists and applying exclusions(i.e. wrong NAICS codes, non-tobacco chains, duplicates). D&B identified 704 probable tobacco outlets; 553 remained after exclusions. Among the 761 probable outlets identified by ReferenceUSA, 86.5% were open and 71.2% sold tobacco, while 13.5% were out of business or could not be located. Of the 553 probable outlets identified by D&B, 90.6% were open and 82.3% sold tobacco, while 9.4% were out of business or could not be located (Table 2).

ReferenceUSA had a higher sensitivity overall for finding both probable and actual tobacco outlets compared to D&B (Tables 3 and 4). ReferenceUSA had very good sensitivity (0.82) and D&B had good sensitivity (0.69) for identifying actual tobacco outlets. Combining sources improved actual tobacco outlet sensitivity to 0.89; in other words, nearly 90% of tobacco outlets were identified in the study area by combining ReferenceUSA and D&B lists.

Across outlet types, probable and actual tobacco outlet sensitivity ranged from moderate to very good for ReferenceUSA and from poor to very good for D&B (Tables 3 and 4). ReferenceUSA had the highest actual tobacco outlet sensitivity for beer, wine and liquor stores (1.0), warehouse clubs and supercenters (0.92) and supermarkets (0.91) (Table 4). Actual tobacco outlet sensitivity for D&B was very good for warehouse clubs and supercenters (0.83), supermarkets (0.83) and pharmacies and drug stores (0.82). Combining secondary data sources improved both probable and actual tobacco outlet sensitivity to good or very good for all outlet types except discount department stores (Tables 3 and 4).

PPV for probable tobacco outlets was very good for both data sources, while actual tobacco outlet PPV was good for ReferenceUSA and very good for D&B. That is, 82% of outlets identified by D&B sold tobacco products while only 71% identified by ReferenceUSA sold tobacco products (Table 3). Both secondary data sources had the highest actual tobacco outlet PPV for convenience stores with gas stations and warehouse clubs. Overall, 71% of outlets identified by both secondary data sources were open and sold tobacco products.

ReferenceUSA allocated 90.7% of outlets to the correct census tract. Outlets incorrectly allocated by ReferenceUSA were located in census tracts with a lower percentage of non-Hispanic White residents (p<.05), and a higher percentage of Hispanic residents (p<.05) compared to outlets correctly allocated. Differences in percentages of Non-Hispanic Black residents and median household income did not reach significance.

We compared actual retail outlet density with density computed only from reliance on commercial lists. Actual mean tobacco outlet density at the census tract level was 1.02 outlets per 1000 people (Table 5). We calculated density based on each list after applying exclusions (i.e. "cleaned" list), and before and after field verification. Density calculated from cleaned, but not field verified lists simulates what practitioners or researchers could use to calculate tobacco retailer density in the absence of field canvassing or having a licensing list. The cleaned ReferenceUSA list estimated slightly higher density (1.35) and D&B was similar to field verified estimates (0.95). We estimated combined density at 1.40 outlets per 1000 people (Table 5). The number of retailers per tract followed a similar pattern. Actual density was significantly and positively correlated with density estimated from ReferenceUSA (r=0.29, p=0002), D&B (r=0.41, p<.0001) and both lists combined (r=0.38, p<.0001) (data not shown).

Discussion

We examined the evidence for validity reported for two commercial secondary tobacco outlet data sources using primary data collection to ascertain their utility in identifying tobacco outlets in non-licensing states. Combined, ReferenceUSA and D&B identified nearly 90% of the 662 tobacco outlets in the study area. ReferenceUSA had a higher sensitivity than D&B at identifying both probable and actual tobacco outlets. In states without tobacco retail licensing, combining ReferenceUSA and D&B could be an alternative approach to identifying tobacco outlets for enforcement operations and advocacy purposes. While both secondary data sources over-counted tobacco outlets, D&B listed fewer false positives, or outlets that turned out not to sell tobacco products and most closely estimated

actual tobacco outlet density. Therefore D&B may provide a more accurate estimate, when the goal is estimating tobacco outlet density for research and practice efforts without primary data collection.

Prior studies have reported mixed results for the evidence for validity for secondary retail food outlet data (Fleischhacker et al., 2013). Liese et al. (2010) conducted on-site verification of ReferenceUSA and D&B to enumerate retail food outlets in rural and urban counties in South Carolina. Sensitivity for locating food outlets using D&B was 0.63, similar to this study. Sensitivity using ReferenceUSA, however, was 0.61, lower than found in this study. Two studies using ground-truthing found that D&B had moderate sensitivity (Powell et al., 2011, Fleischhacker et al., 2012) and ReferenceUSA had either fair (Lucan et al., 2013), good (Powell et al., 2011) or very good (Gustafson et al., 2012, Fleischhacker et al., 2012) sensitivity. With the exception of Fleischhacker (2012), previous sensitivities for ReferenceUSA are lower than both the probable (0.84) and actual (0.82) tobacco outlet sensitivity found in this study. Exclusion criteria may account for these differences. A validation study of an unnamed commercial database found moderate sensitivity for identifying physical activity resources (0.54) (Boone et al., 2008) lower than the sensitivity found here. Similar to this study, Liese et al.(2010) found that combining ReferenceUSA and D&B improved sensitivity, indicating that each secondary data source contributes unique outlets. Geospatial analyses revealed that ReferenceUSA allocated over 90% of outlets to the correct census tract, similar to other findings at the census tract (Liese et al., 2010) and at the block group level (Boone et al., 2008).

Combining ReferenceUSA and D&B improved tobacco outlet sensitivity to very good for all outlet types except tobacco stores and discount department stores. Warehouse clubs and supermarkets had higher tobacco outlet sensitivity than other outlet types, which may indicate lower turnover rates at these outlets. Conversely, both secondary data sources frequently missed tobacco stores. We searched ReferenceUSA by company name and location for tobacco stores identified during primary data collection that were not in secondary data source lists. Tobacco stores were variously categorized as Gift, Novelty & Souvenir Stores, Other Miscellaneous Nondurable Goods Merchant Wholesalers, Full-Service Restaurants, and Farm Supplies Merchant Wholesalers. The undercounting of tobacco stores may cause them to be missed during compliance checks, which is problematic because tobacco stores are more likely than other outlet types to sell tobacco to minors (Widome et al., 2012) and may be more likely to be non-adherent with FDA tobacco point-of-sale provisions (Rose et al., 2013). As a result, it might be worthwhile to supplement commercial sources with Internet searches for tobacco shops. Despite commerical secondary data sources over- and under-counting store types, the density estimates by each data source were significantly, although modestly, correlated to the actual density of tobacco outlets in the study area. On average, the D&B density estimate was 7% lower and ReferenceUSA was 32% higher than actual tobacco oulet density. Although tobacco outlet desnity estimated by the secondary data sources was higher than actual density, they correlated will with actual density and are likely to still be valide when examining associations between density and behaivoral outcomes, like smoking.

The majority of the ten NAICS codes were useful in identifying tobacco outlets. One exception was newsstands. No newsstands that sold tobacco were identified in this study area; nevertheless, the inclusion of any outlet type should be informed by the specific retail environment assessed. Newsstands selling tobacco are more common in large urban areas than in the mid-sized cities examined in this study. Another problematic NAICS outlet type is pharmacies and drug stores which includes non-retail outlets such as laboratories. Phone screening of non-chain pharmacies excluded outlets that did not sell tobacco and improved sensitivity above what would have been found using a raw or un-edited list. Including only chain pharmacies known to sell tobacco (e.g. Rite Aid) may be an important step when utilizing secondary data sources to enumerate tobacco outlets. Before estimating density in certain locations, future studies should take into account that several cities ban the sale of tobacco products in retail pharmacies (Katz, 2013) and that CVS recently announced that they will stop selling tobacco products in their stores by October 2014.

This study has several limitations. First, our study was conducted in only three counties in one southeastern state in the US; thereby, our findings might have limited generalizability to other US and non-US jurisdictions. Second, we only identified and verified 10 types of tobacco outlets; we may have missed tobacco outlets that were not one of these 10 types. Third, investigators were guided in the field by a list and map of probable tobacco outlets and were not blinded to the secondary data source in the field or during the post-hoc analysis, which could contribute towards bias, although there were no established hypotheses or assumptions made about either source. Finally, we narrowed business lists before primary data collection to increase the likelihood of identifying tobacco outlets. Since we excluded outlets known not to sell tobacco prior to analysis, we could not calculate overall sensitivity of either secondary data source. For example, we excluded supermarket chains that did not sell tobacco (e.g., Whole Foods); therefore, this study cannot estimate the total number of supermarkets in the study area. Yet, applying exclusions helped avoid visiting outlets unnecessarily and, in practice, when using a secondary data source without on-site verification, eliminating known non-tobacco outlets should make tobacco outlet enumeration more effective and efficient.

Conclusions

To our knowledge, this is the first study to report evidence for validity of secondary data sources for identifying probable and actual tobacco outlets using primary data. Although ReferenceUSA and D&B undercounted the true number of tobacco outlets, combining the two secondary data sources resulted in the enumeration of nearly 90% of all tobacco outlets in the study area. Both lists were correlated with actual tobacco outlet density. In North Carolina and perhaps other non-licensing states, commercial secondary data sources may be a useful way to identify tobacco outlets to aid in enforcement operations and estimate tobacco outlet density.

References

Altman, DG. Practical statistics for medical research. Vol. 12. CRC Press; 1991.

Boone-Heinonen J, Evenson K, Song Y, Gordon-Larsen P. Built and socioeconomic environments: patterning and associations with physical activity in US adolescents. Int J Behav Nutr Phys Act. 2009; 7:45–45. [PubMed: 20487564]

Boone JE, Gordon-Larsen P, Stewart JD, Popkin BM. Validation of a GIS facilities database: quantification and implications of error. Ann Epidemiol. 2008; 18:371–377. [PubMed: 18261922]

Centers for Disease Control and Prevention. State Tobacco Activities Tracking and Evaluation (STATE) System. 2012

D'Angelo, H. Secondary Source Validity for Enumerating Tobacco Retailers in Non-licensing States. National Conference on Tobacco or Health; August 15–17, 2012; 2012.

Diez Roux A, Merkin SS, Hannan P, Jacobs DR, Kiefe CI. Area characteristics, individual-level socioeconomic indicators, and smoking in young adults the coronary artery disease risk development in young adults study. Am J Epidemiol. 2003; 157:315–26. [PubMed: 12578802]

Dun and Bradstreet. The DUNS right quality process: the power behind quality information. Dun and Bradstreet; Waltham, Mass: 2005.

Difranza JR. State and federal compliance with the synar amendment: federal fiscal year 1998. Arch Pediatr Adolesc Med. 2001; 155:572–578. [PubMed: 11343500]

Fakunle D, Morton CM, Peterson NA. The importance of income in the link between tobacco outlet density and demographics at the tract level of analysis in New Jersey. J Ethn Subst Abuse. 2010; 9:249–59. [PubMed: 21161808]

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):462–473. [PubMed: 24050423]

Fleischhacker SE, Evenson KR, Rodriguez DA, Ammerman AS. A systematic review of fast food access studies. Obes Rev. 2011; 12:e460–71. [PubMed: 20149118]

Fleischhacker SE, Rodriguez DA, Evenson KR, Henley A, Gizlice Z, Soto D, Ramachandran G. Evidence for validity of five secondary data sources for enumerating retail food outlets in seven American Indian Communities in North Carolina. Int J Behav Nutr Phys Act. 2012; 9:137. [PubMed: 23173781]

Frieden T. Forward: CDC Health Disparities and Inequalities Report-United States, 2011. Morbidity and mortality weekly report. Surveillance summaries (Washington, DC: 2002). 2011; 60:1.

Gordon-Larsen P, Nelson MC, Page P, Popkin BM. Inequality in the built environment underlies key health disparities in physical activity and obesity. Pediatrics. 2006; 117:417. [PubMed: 16452361]

Gustafson A, Lewis S, Wilson C, Jilcott-Pitts S. Validation of food store environment secondary data source and the role of neighborhood deprivation in Appalachia, Kentucky. BMC Public Health. 2012; 12:688. [PubMed: 22914100]

Henriksen L, Feighery EC, Schleicher NC, Cowling DW, Kline RS, Fortmann SP. Is adolescent smoking related to the density and proximity of tobacco outlets and retail cigarette advertising near schools? Prev Med. 2008a; 47:210–4. [PubMed: 18544462]

Hosler AS, Dharssi A. Identifying retail food stores to evaluate the food environment. Am J Prev Med. 2010; 39:41–44. [PubMed: 20537845]

Hosler AS, Kammer JR. Point-of-purchase tobacco access and advertisement in food stores. Tob Control. 2012; 21:451–2. [PubMed: 22411730]

Hyland A, Travers MJ, Cummings KM, Bauer J, Alford T, Wieczorek WF. Tobacco outlet density and demographics in Erie County, New York. Am J Public Health. 2003; 93:1075–6. [PubMed: 12835184]

Katz MH. Tobacco-free pharmacies: can we extend the ban? Tobacco control. 2013; 22(6):363–4. [PubMed: 24052546]

Larson NI, Story MT, Nelson MC. Neighborhood environments: disparities in access to healthy foods in the U.S. Am J Prev Med. 2009; 36:74–81. [PubMed: 18977112]

Liese AD, Colabianchi N, Lamichhane AP, Barnes TL, Hibbert JD, Porter DE, Nichols MD, Lawson AB. Validation of 3 food outlet databases: completeness and geospatial accuracy in rural and urban food environments. Am J Epidemiol. 2010; 172:1324–33. [PubMed: 20961970]

Lipperman-Kreda S, Grube JW, Friend KB. Local tobacco policy and tobacco outlet density: associations with youth smoking. J Adolesc Health. 2012; 50:547–52. [PubMed: 22626479]

Lovasi GS, Hutson MA, Guerra M, Neckerman KM. Built environments and obesity in disadvantaged populations. Epidemiologic reviews. 2009; 31:7–20. [PubMed: 19589839]

Lucan SC, Maroko AR, Bumol J, Torrens L, Varona M, Berke EM. Business list vs ground observation for measuring a food environment: saving time or waste of time (or worse)? J Acad Nutr Diet. 2013; 113(10):1332–1339. [PubMed: 23871107]

Novak SP, Reardon SF, Raudenbush SW, Buka SL. Retail tobacco outlet density and youth cigarette smoking: a propensity-modeling approach. Am J Public Health. 2006; 96:670–6. [PubMed: 16507726]

Peterson NA, Lowe JB, Reid RJ. Tobacco outlet density, cigarette smoking prevalence, and demographics at the county level of analysis. Subst Use Misuse. 2005; 40:1627–35. [PubMed: 16253931]

Powell LM, Han E, Zenk SN, Khan T, Quinn CM, Gibbs KP, Pugach O, Barker DC, Resnick EA, Myllyluoma J, Chaloupka FJ. Field validation of secondary commercial data sources on the retail food outlet environment in the U.S. Health Place. 2011; 17:1122–1131. [PubMed: 21741875]

Infogroup, Inc. [Accessed May 2011] ReferenceUSA database. n.d. www.referenceusa.com

Reitzel LR, Cromley EK, Li Y, Cao Y, Dela Mater R, Mazas CA, Cofta-Woerpel L, Cinciripini PM, Wetter DW. The effect of tobacco outlet density and proximity on smoking cessation. Am J Public Health. 2011; 101:315–320. [PubMed: 21164089]

Rodriguez D, Carlos HA, Adachi-Mejia AM, Berke EM, Sargent JD. Predictors of tobacco outlet density nationwide: a geographic analysis. Tobacco Control. 2012

Rose SW, Myers AE, D'Angelo H, Ribisl KM. Retailer adherence to Family Smoking Prevention and Tobacco Control Act, North Carolina, 2011. Prev Chronic Dis. 2013; 10:E47. [PubMed: 23557638]

Schneider JE, Reid RJ, Peterson NA, Lowe JB, Hughey J. Tobacco outlet density and demographics at the tract level of analysis in Iowa: implications for environmentally based prevention initiatives. Prev Sci. 2005; 6:319–25. [PubMed: 16163568]

Sharkey JR. Measuring potential access to food stores and food-service places in rural areas in the U.S. Am J Prev Med. 2009; 36:S151–5. [PubMed: 19285206]

Target. [Accessed January 25 2013] Target Corporate Responsibility Report [Online]. 2009. Available: https://corporate.target.com/_ui/pdf/2009_full_report.pdf

U.S. Census Bureau. [Accessed July 16 2012] State and County QuickFacts. Data derived from Population Estimates, American Community Survey, Census of Population and Housing, State and County Housing Unit Estimates, County Business Patterns, Nonemployer Statistics, Economic Census, Survey of Business Owners, Building Permits, Consolidated Federal Funds Report [Online]. 2012. Available: http://quickfacts.census.gov/qfd/states/37000.html

U.S. Census Bureau, Economic Census. [Accessed November 2010] American Factfinder. 2010. http://factfinder2.census.gov

U.S. Census Bureau. 2007 NAICS definitions. 2007. http://www.census.gov/eos/www/naics/

Widome R, Brock B, Noble P, Forster JL. The relationship of point-of-sale tobacco advertising and neighborhood characteristics to underage sales of tobacco. Eval Health Prof. 2012; 35:331–345. [PubMed: 22615496]

Highlights

- We conducted primary data collection to identify all tobacco retail outlets in three counties.

- Two commercial data sources were validated against primary data collection.

- Nearly 90% of tobacco outlets were identified by combining secondary data sources.

- Over 90% of outlets were geocoded to the correct census tract by ReferenceUSA.

Table 1

Primary NAICS[a] codes used to search for probable tobacco outlets in ReferenceUSA and Dun & Bradstreet, Inc. in three North Carolina counties

NAICS code	Industry	Description
445110	Supermarkets and other grocery (except convenience) stores	Establishments generally known as supermarkets and grocery stores primarily engaged in retailing a general line of food, such as canned and frozen foods; fresh fruits and vegetables; and fresh and prepared meats, fish, and poultry. Included in this industry are delicatessen-type establishments primarily engaged in retailing a general line of food.
445120	Convenience stores	Establishments known as convenience stores or food marts (except those with fuel pumps) primarily engaged in retailing a limited line of goods that generally includes milk, bread, soda, and snacks.
453991	Tobacco stores	Establishments primarily engaged in retailing cigarettes, cigars, tobacco, pipes, and other smokers' supplies.
447110	Gasoline stations with convenience stores	Establishments engaged in retailing automotive fuels (e.g., diesel fuel, gasohol, gasoline) in combination with convenience store or food mart items. These establishments can either be in a convenience store (i.e., food mart) setting or a gasoline station setting. These establishments may also provide automotive repair services.
452910	Warehouse clubs and supercenters	Establishments known as warehouse clubs, superstores or supercenters primarily engaged in retailing a general line of groceries in combination with general lines of new merchandise, such as apparel, furniture, and appliances.
451212	News dealers and newsstands	Establishments primarily engaged in retailing current newspapers, magazines, and other periodicals.
445310	Beer, wine, and liquor stores	Establishments primarily engaged in retailing packaged alcoholic beverages, such as ale, beer, wine, and liquor
446110	Pharmacies and drug stores	Establishments known as pharmacies and drug stores engaged in retailing prescription or nonprescription drugs and medicines.
452112	Discount department stores	Establishments known as department stores that have central customer checkout areas, generally in the front of the store, and that may have additional cash registers located in one or more individual departments. Department stores in this industry sell a wide range of general merchandise (except fresh, perishable foods).
447190	Other gasoline stations	Establishments known as gasoline stations (except those with convenience stores) primarily engaged in one of the following: (1) retailing automotive fuels (e.g., diesel fuel, gasohol, gasoline) or (2) retailing these fuels in combination with activities, such as providing repair services; selling automotive oils, replacement parts, and accessories; and/or providing food services.

[a]North American Industry Classification System

Table 2

Tobacco outlets identified through primary and secondary tobacco retail data sources for three North Carolina counties

Outlet Type	Primary Data	ReferenceUSA				Dun & Bradstreet, Inc.			
	Total No. Tobacco Outlets	Total No. Outlets Identified	Open Probable Tobacco Outlets (%)	Tobacco Outlets (%)	Closed or Not Found (%)	Total No. Outlets Identified	Open Probable Tobacco Outlets (%)	Tobacco Outlets (%)	Closed or Not Found (%)
All outlets	662	761	86.5	71.2	13.5	553	90.6	82.3	9.4
Convenience Stores	102	122	77.0	68.9	23.0	88	83.0	76.1	17.0
Convenience Stores with Gas Stations	297	247	100.0	100.0	0.0	197	100.0	100.0	0.0
Other Gas Stations	28	54	70.4	46.3	29.6	33	75.8	51.5	24.2
Supermarkets & Grocery Stores	100	148	78.4	61.5	21.6	115	83.5	72.2	16.5
Pharmacies & Drug Stores	79	91	86.8	62.6	13.2	76	92.1	85.5	7.9
Tobacco Stores	38	31	74.2	74.2	25.8	16	87.5	87.5	12.5
Warehouse clubs & Supercenters	12	12	100.0	91.7	0.0	11	100.0	90.9	0.0
Discount Department Stores	2	4	100.0	0.0	0.0	0	NA	NA	NA
News Dealers & News Stands	0	3	33.3	0.0	66.7	1	0.0	0.0	0
Beer, Wine & Liquor Stores	4	31	83.9	12.9	16.1	9	88.9	22.2	11.1
Other Store Types	0	18	100.0	0.0	0.0	7	100.0	0.0	0.0

Note: Percentages are of total number of outlets on each list.

[a] Tobacco outlets were added if identified in the field.

Table 3

Evidence for Validity of Secondary Data Sources for Probable Tobacco Outlets Compared to a Field Census, North Carolina, 2011

Outlet Type	ReferenceUSA		Dun & Bradstreet		ReferenceUSA and Dun & Bradstreet	
	Probable Tobacco Outlet Sensitivity	Probable Tobacco Outlet PPV	Probable Tobacco Outlet Sensitivity	Probable Tobacco Outlet PPV	Probable Tobacco Outlet Sensitivity	Probable Tobacco Outlet PPV
All outlets	0.8374 (0.0132)	0.8647 (0.0124)	0.6366 (0.0171)	0.9060 (0.0124)	0.8971 (0.0108)	0.8537 (0.0123)
Convenience stores	0.8246 (0.0356)	0.7705 (0.0381)	0.6404 (0.0449)	0.8295 (0.0401)	0.8509 (0.0334)	0.7462 (0.0382)
Gas/convenience stores	0.8316 (0.0217)	1.0000 (0.0000)	0.6633 (0.0274)	1.0000 (0.0000)	0.8889 (0.0182)	1.0000 (0.0000)
Other gas stations	0.9048 (0.0453)	0.7037 (0.0621)	0.5952 (0.0757)	0.7576 (0.0746)	0.9524 (0.0329)	0.6780 (0.0608)
Supermarkets & grocery stores	0.9063 (0.0258)	0.7838 (0.0338)	0.7500 (0.0383)	0.8348 (0.0346)	0.9531 (0.0187)	0.7578 (0.0338)
Pharmacies & drug stores	0.7822 (0.0411)	0.8681 (0.0355)	0.6931 (0.0459)	0.9211 (0.03090)	0.9307 (0.0253)	0.8624 (0.0330)
Tobacco stores	0.6053 (0.0793)	0.7419 (0.0786)	0.3684 (0.0783)	0.8750 (0.0827)	0.7368 (0.0714)	0.7778 (0.0693)
Warehouse clubs & Supercenters	0.9231 (0.0739)	1.0000 (0.0000)	0.8462 (0.1001)	1.0000 (0.0000)	0.9231 (0.0739)	1.0000 (0.0000)
Discount department stores	0.5714 (0.1870)	1.0000 (0.0000)	0.0000 (0.0000)	_a_	0.5714 (0.1870)	1.0000 (0.0000)
News Dealers & News Stands	1.0000 (0.0000)	0.3333 (0.2722)	0.0000 (0.0000)	0.0000 (0.0000)	1.0000 (0.0000)	0.3333 (0.2722)
Beer, wine & liquor stores	1.0000 (0.0000)	0.8387 (0.0661)	0.3077 (0.0905)	0.8889 (0.1048)	1.0000 (0.0000)	0.8387 (0.0661)
Other store types	0.9500 (0.0487)	1.0000 (0.0000)	0.3500 (0.1067)	1.0000 (0.0000)	0.9000 (0.0671)	1.0000 (0.0000)

PPV= Positive predictive value; standard errors are reported in parentheses.

_a_There were no discount stores on the Dun & Bradstreet list.

Table 4

Evidence for Validity of Secondary Data Sources for Actual Tobacco Outlets Compared to Primary Data for three North Carolina counties, 2011

Outlet Type	ReferenceUSA		Dun & Bradstreet		ReferenceUSA and Dun & Bradstreet	
	Tobacco Outlet Sensitivity	Tobacco Outlet PPV	Tobacco Outlet Sensitivity	Tobacco Outlet PPV	Tobacco Outlet Sensitivity	Tobacco Outlet PPV
All Outlets	0.8187 (0.0150)	0.7122 (0.0164)	0.6873 (0.0180)	0.8228 (0.0162)	0.8897 (0.0122)	0.7122 (0.0157)
Convenience Stores	0.8235 (0.0377)	0.6885 (0.0419)	0.6569 (0.0470)	0.7614 (0.0454)	0.8529 (0.0351)	0.6692 (0.0413)
Convenience Stores with Gas Stations	0.8316 (0.0217)	1.0000 (0.0000)	0.6633 (0.0274)	1.0000 (0.0000)	0.8889 (0.0182)	1.0000 (0.0000)
Other Gas Stations	0.8929 (0.0585)	0.4630 (0.0679)	0.6071 (0.0923)	0.5152 (0.0870)	0.9286 (0.0487)	0.4407 (0.0646)
Supermarkets & Grocery Stores	0.9100 (0.0286)	0.6149 (0.0400)	0.8300 (0.0376)	0.7217 (0.0418)	0.9700 (0.0171)	0.6025 (0.0386)
Pharmacies & Drug Stores	0.7215 (0.0504)	0.6264 (0.0507)	0.8228 (0.0430)	0.8553 (0.0404)	0.9114 (0.0320)	0.6606 (0.0454)
Tobacco Stores	0.6053 (0.0793)	0.7419 (0.0786)	0.3684 (0.0783)	0.8750 (0.0827)	0.7368 (0.0714)	0.7778 (0.0693)
Warehouse Clubs & Supercenters	0.9167 (0.0798)	0.9167 (0.0798)	0.8333 (0.1076)	0.9091 (0.0867)	0.9167 (0.0798)	0.9167 (0.0798)
Discount Department Stores	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	[a]	0.0000 (0.0000)	0.0000 (0.0000)
Beer, Wine & Liquor stores	1.0000 (0.0000)	0.1290 (0.0602)	0.5000 (0.2500)	0.2222 (0.1386)	1.0000 (0.0000)	0.1290 (0.0602)

PPV= Positive predictive value; standard errors are reported in parentheses. Newsstands and other store types are not presented because there were no tobacco outlets found in these categories.

[a]There were no discount stores on the Dun & Bradstreet list.

Table 5

Comparison of tobacco retailer density calculated from primary data and two secondary retail data sources

Source	Number of retailers per tract, Mean (SD)		Number of retailers per 1000 people, Mean (SD)	
	Field Verified	Cleaned list, not verified	Field Verified	Cleaned list, not verified
Direct field observation	4.58 (2.93)	na	1.02 (0.85)	na
ReferenceUSA	3.80 (2.46)	4.63 (3.53)	0.94 (0.78)	1.35 (1.56)
Dun & Bradstreet, Inc.	3.41 (2.16)	3.68 (2.44)	0.83 (0.69)	0.95 (1.09)
Combined ReferenceUSA and Dun & Bradstreet, Inc.	4.10 (2.67)	4.99 (3.75)	1.01 (0.83)	1.40 (1.38)