Evaluation and Program Planning 53 (2015) 72–79

Contents lists available at ScienceDirect

Evaluation and Program Planning

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

Matching study areas using Google Street View: A new application for an emerging technology

Elyse Levine Less *, Patricia McKee, Traci Toomey, Toben Nelson, Darin Erickson, Serena Xiong, Rhonda Jones-Webb [1]

Alcohol Epidemiology Program, Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, 1300 South Second Street, Suite 300, Minneapolis, MN 55454-1015, USA

ARTICLE INFO

Article history:
Received 17 March 2015
Received in revised form 26 June 2015
Accepted 5 August 2015
Available online 10 August 2015

Keywords:
Google Street View
Audit tools
Matching
Ecological settings
Alcohol-related crime

ABSTRACT

Google Street View (GSV) can be used as an effective tool to conduct virtual neighborhood audits. We expand on this research by exploring the utility of a GSV-based neighborhood audit to measure and match target and comparison study areas. We developed a GSV-based inventory to measure characteristics of retail alcohol stores and their surrounding neighborhoods. We assessed its reliability and assessed the utility of GSV-based audits for matching target and comparison study areas. We found that GSV-based neighborhood audits can be a useful, reliable, and cost-effective tool for matching target and comparison study areas when archival data are insufficient and primary data collection is prohibitive. We suggest that researchers focus on characteristics that are easily visible on GSV and are relatively stable over time when creating future GSV-based measuring and matching tools. Dividing the study area into small segments may also provide more accurate measurements and more precise matching.

1. Introduction

As web-based geographic imaging applications evolve, social scientists have an expanding arsenal of new measurement tools. Evaluating the reliability, utility, and feasibility of these new tools is crucial to understanding how they can be applied most effectively in social science research.

Assessing ecological settings, such as neighborhood characteristics that may impact health and social issues, is an area where this new technology can be applied. Standard methods to assess neighborhoods typically involve direct observation through fieldwork. Researchers walk or drive through small neighborhood areas to directly observe and systematically document neighborhood features (Rundle, Bader, Richards, Neckerman, & Teitler, 2011; Sampson & Raudenbush, 1999). These systematic field audits allow researchers to objectively assess neighborhood conditions and reliably measure constructs such as public disorder (Sampson & Raudenbush, 1999). Audit results are commonly used to evaluate relationships between environmental features and outcomes, such as parenting and childhood development (Caughy, Brodsky, O'Campo, & Aronson, 2001), obesity (Franzini et al., 2010) and crime (Jones, Pebley, & Sastry, 2011). However, this work can often be cost prohibitive and impractical in studies evaluating large geographic regions, multiple study areas, or distant locations (Odgers, Caspi, Bates, Sampson, & Moffitt, 2012; Rundle et al., 2011).

Recent studies have explored whether Google Street View (GSV)[1] can be effectively used to assess ecological settings. GSV, an online virtual imaging application, is a feature of Google Maps/Google Earth that enables users to view and navigate through 360-degree horizontal and 290-degree vertical panoramic street level images (Google, 2007). Google uses many methods to capture GSV images, including specially equipped vehicles that can be driven down streets, and back packs equipped with cameras (Google, 2015). GSV allows the user to virtually observe streets and neighborhoods. One can navigate forward and backward, up and down, and zoom in and out. Image quality allows for quite detailed observation including sidewalk cracks, bars on windows, street signs, and even roofs in need of repair.

* Corresponding author. Tel.: +1 612 625 6200.
E-mail address: levin381@umn.edu (E.L. Less).
[1] PI.

[1] Abbreviations: GSV—Google Street View; SSO—systematic social observation; SB—store block; AB—adjacent blocks.

http://dx.doi.org/10.1016/j.evalprogplan.2015.08.002
0149-7189/

Researchers have conducted virtual neighborhood audits to evaluate whether GSV could be used in lieu of the standard neighborhood observation methods currently in practice by comparing results of in-person and GSV-based observations. For example, in one study, researchers virtually observed urban neighborhoods using a GSV-based audit tool to measure neighborhood conditions, such as physical disorder and decay, in order to study how those conditions impacted children's health (Odgers et al., 2012). Another study found that GSV-based observations offered benefits over traditional environmental audits, but cautioned researchers to be mindful of GSV image capture dates (the dates the photos were taken) which may flip back and forth between years during a virtual observation (Curtis, Curtis, Mapes, Szell, & Cinderich, 2013). The study referred to this capture date issue as "spatio-temporal instability (Curtis et al., 2013)." It presents a particular problem when matching neighborhoods, where ideally target and comparison neighborhoods would be evaluated during the same time frame. Reliably judging certain physical characteristics, such as pavement quality, may also be challenging using GSV (Griew et al., 2013; Vanwolleghem, Van Dyck, Ducheyne, De Bourdeaudhuij, & Cardon, 2014). Despite these minor limitations, findings suggest that GSV-based audits are a reliable and cost-effective method for assessing neighborhood conditions (Bader, Ailshire, Morenoff, & House, 2010; Ben-Joseph, Lee, Cromley, Laden, & Troped, 2013; Curtis et al., 2013; Griew et al., 2013; Kelly, 2013; Odgers et al., 2012; Rundle et al., 2011).

In addition to making large-scale neighborhood observations more practical and cost-efficient, GSV-based audits may help address other methodological challenges. Researchers often rely on census data when matching target and comparison study areas to account for the potential confounding influence of socio-demographic covariates on outcomes. Relying solely on such archival data limits the types of characteristics that can be assessed and may result in suboptimal selections.

1.1. Addressing the methodological challenges presented in our larger study

We wanted to examine the feasibility of conducting GSV-based audits to assess certain neighborhood characteristics, and then evaluate the effectiveness of using these results to inform how we match target and comparison study areas. We found no studies to date that have used data collected by virtual audit to match study areas.

As part of our larger study, we examined the utility of GSV-based audits for matching study neighborhoods more precisely. The overall goal of our larger study was to assess the relationship between policy restrictions on the retail sale of malt liquor (e.g., limits on container size, bans on sales for individual stores) and neighborhood crime rates in 15 U.S. cities. Malt liquor sales and consumption are associated with heavy drinking (Bluthenthal, Browntaylor, Guzman-Becerra, & Robinson, 2005; Collins, Bradizza, & Vincent, 2007) and a spectrum of alcohol-related nuisance crimes in neighborhoods, such as public drinking, disorderly conduct, and loitering, as well as with more serious crimes such as assaults and robberies (Campbell et al., 2009; Gruenewald & Remer, 2006; Paschall, Lipperman-Kreda, & Grube, 2014; Zhu, Gorman, & Horel, 2004).

We compared crime rates in a 0.25-mi buffer around off-premise alcohol outlets (retail stores that sell alcohol to be consumed in another location) that have a malt liquor policy (target stores), before and after policy adoption, while controlling for pre-post crime around similar stores that do not have a malt liquor policy (comparison stores). We chose crime outcomes we expected to be associated with off-premise malt liquor sales based on previous studies (Franklin, Laveist, Webster, & Pan, 2010; Snowden & Pridemore, 2014; Tarnai, 2003).

To achieve more precise matches, we wanted to assess neighborhood characteristics beyond simply matching on typical socio-demographic variables. Characteristics such as land use, neighborhood decay, and social and physical disorder can be important considerations when evaluating neighborhood crime and health outcomes (Milam, Furr-Holden, Harrell, Whitaker, & Leaf, 2012; Skogan, 1989). We wanted to avoid, for example, matching a supermarket in a well-maintained residential neighborhood with a small liquor store in a physically decaying commercial neighborhood. The "broken windows" theory of social disorganization, which suggests that the presence of neighborhood disorder encourages more disorder, (Sampson & Raudenbush, 1999) influenced how we built our neighborhood audit tool.

Building on the GSV validation literature, we created a GSV-based neighborhood audit and "GSV Inventory Form" to systematically observe and measure store and surrounding neighborhood characteristics. The GSV validation studies noted above have demonstrated the reliability of GSV-based neighborhood audits. However, because we created our own audit tool to measure characteristics specifically associated with alcohol-related crime, we needed to test the audit's reliability prior to exploring its effectiveness as a matching tool. In the current study we: (1) describe the GSV Inventory Form development; (2) determine the inter-rater reliability of the GSV Inventory Form (i.e., the degree of concordance between two independent raters); (3) compare the inter-source reliability between GSV-based virtual audits and in-person observations; and (4) determine the effectiveness of using GSV-based audit results for matching target and comparison study areas.

2. Methods

2.1. Defining the neighborhood audit area

We began by identifying the target stores. Targets are retail stores that sell alcoholic beverages for consumption off-premise and have restrictions placed on their ability to sell malt liquor beverages by state statute or local ordinance (e.g., city conditional use permits, state alcohol license operating restrictions). We identified stores with malt liquor restrictions; and these stores with effective dates within our study time frame (2005–2011) became targets.

We entered the store's street address into Google Earth to obtain a map with an aerial view and labeled street names. When the GSV icon is clicked in this mode, streets marked with blue lines indicate GSV availability. We used this map to create the audit observation area, comprised of the block with the store (the "store block") and the immediately adjacent blocks surrounding the store block (the "adjacent blocks"). We divided the observation area into eight block face segments (Fig. 1). A "block face" is one side of a street between two consecutive intersections (Clarke, Ailshire, Melendez, Bader, & Morenoff, 2010). For example, buildings across the street from each other are part of two separate block faces.

2.2. Developing the GSV inventory form

We created the GSV Inventory Form to collect observational data based on a neighborhood's built, physical, and social environmental features that have relative permanence; and are environmental determinants of alcohol-related crime (Furr-Holden et al., 2008; Grubesic, Pridemore, Williams, & Philip-Tabb, 2013; Jones-Webb, McKee, Hannan, Wall, Pham, Erickson, et al., 2008). Our GSV Inventory Form was modeled after Sampson's Systematic Social Observation (SSO) Inventory (Sampson & Raudenbush, 1999)

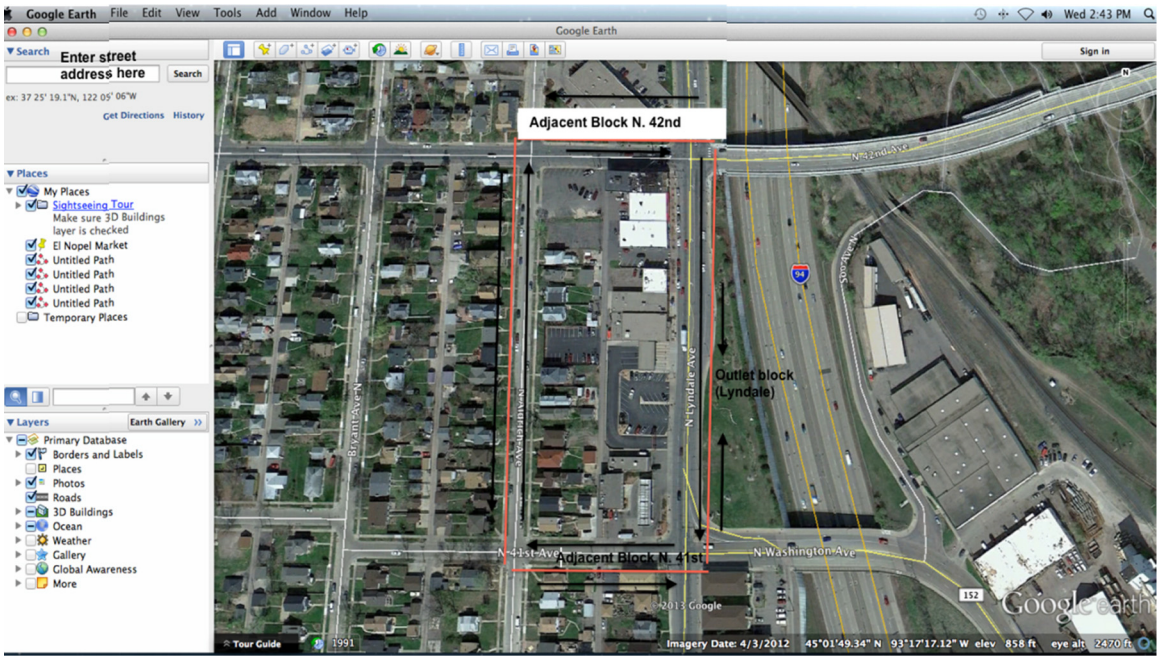

Fig. 1. Sample observation area.

and Odgers' use of the SSO Inventory for virtual observations (Odgers et al., 2012). We chose characteristics that could be easily observed in GSV, and excluded characteristics we could not measure, such as noise and smell; and dynamic features such as broken glass and traffic and pedestrian patterns.

2.3. Measurement development and scoring[2]

We developed seven matching criteria. We prioritized criteria to assess neighborhood features based on their known relationship with crime, our primary study outcome.

(1) *Neighborhood type*: We assessed neighborhood type with an aerial screen shot of the observation area. We created the following categories and definitions informed in part by U.S. government definitions (U.S. Census Bureau 2015; Rural Health, www.hrsa.gov 2015).
   a. Downtown: Commercial and/or geographic center of the city with a large concentration of retail and office buildings (within a two-mile radius of the city's largest city hall).
   b. Urban: Very developed with a highly dense settled core of census tracts or blocks including housing, commercial buildings and public transportation (population over 50,000).
   c. Suburban: Bounded by major cities and characteristically made up of sole family separated residences, strip malls, and shopping centers, with more green space and characterized by lower population density than a city/urban area.
   d. Rural: Non-metropolitan counties characteristically made up of low population density and undeveloped land.

(2) *Store type*: Our study focuses on off-premise alcohol retail stores (e.g., liquor stores, convenience stores) where malt liquor is typically purchased, rather than on-premise outlets (e.g., bars and restaurants). We created store categories informed by North American Industry Classification System definitions (North American Industry Classification System, 2012) and prior research (Jones-Webb et al., 2008). We grouped these store types and ranked them 1–4 based on their relationship with neighborhood crime and the likelihood of malt liquor purchases as indicated in prior research (Jones-Webb et al., 2008).
   1. Large Grocery or Supermarket: Retail sales of a general line of food, such as canned and frozen foods; fresh fruits and vegetables; and fresh and prepared meats, fish, and poultry (e.g., Cub Food (MN), Ralph's (CA)). Excludes Convenience stores.
      General Merchandise: Retail sales of groceries in combination with a general lines of new merchandise, such as apparel, furniture, and appliances (e.g., Wal-Mart, warehouse club).
   2. Mom and Pop Grocery: Retail sales of a limited line of goods that generally include milk, bread, soda, and snacks, but not part of a chain of stores (e.g., "corner store" or "bodega" type of store).
      Convenience Store: Retail sales of a limited line of goods that generally include milk, bread, soda, and snacks, but excludes automotive fuel sales (e.g., 7 Eleven).
      Gas: Retail sales of automotive fuels in combination with a convenience store or food mart (e.g., Super America).
   3. Liquor store—high-end/moderate
   4. Liquor store—low-end/dilapidated
      Liquor store: Retail sales of packaged alcoholic beverages, such as beer, wine, and liquor; high-end versus low-end determined by physical appearance set forth in glossary (e.g., signage quality, percent of windows covered by advertisements, bars on windows).

(3) *Primary land use*: Primary land use was determined to be an important neighborhood feature for matching retail alcohol stores, as a strong correlation exists between land use type and crime (Lockwood, 2007). The primary land use for each audit area was determined by first rating each of the eight block faces separately. A land use was 'primary' if it encompassed at least 50 percent of the block face and appeared in at least two block faces. Using these data, we selected the most frequently-occurring primary land use (or two, if tied). We created categories based on various land use definitions used in prior

[2] A detailed glossary, created and used by the research team, containing photographs, descriptions and definitions of each term, is available upon request from the authors.

research (Odgers, Bates, Caspi, Sampson, & Moffitt, 2009; Stucky, 2009).

a. Residential: single family homes, duplex, multi-unit, apartments, mobile homes.
b. Commercial/Business: retail store, service, professional, office, entertainment, restaurant.
c. Industrial/Warehouse: storage facility, processing, manufacturing, factory, industrial park.
d. Institutional: schools–public or private, places of worship, hospital, government buildings.
e. Airport: commercial or military.
f. Other: park, green space, parking lot (indoor or outdoor), highway.

(4) *Opportunity to loiter*: Physical space that provides an opportunity to loiter and consume alcohol was another matching criterion. Loitering and public alcohol consumption are social disorder indicators (Sampson & Raudenbush, 1999). However, since GSV may not capture these behaviors at the point in time when the pictures were taken, we assessed the number of physical spaces conducive to engaging in these behaviors. Retail alcohol stores and their immediate surroundings, especially in urban areas, can serve as a gathering area for alcohol consumption (Grubesic et al., 2013; Snowden & Pridemore, 2014). These spaces can create the conditions for public safety problems and crimes associated with malt liquor consumption (Franklin et al., 2010; Jones-Webb et al., 2008). For example, vacant lots near retail alcohol stores may provide a space for drinking or serve as a base for gang activity (Bernasco & Block, 2011; Tita, Cohen, & Engberg, 2005).

Scores ranged from 0 to 3. A score of 1 indicated that there was a vacant lot or park within the observation area; a score of 2 indicated a parking lot, vacant lot or park directly adjacent to the store. A score of 3 indicated the presence of both of these features, and a score of 0 indicated the absence of these features.

(5) *Physical decay*: Physical decay indicators are cues of neighborhood disorder, and as such, are associated with violence (Sampson & Raudenbush, 1999; Snowden & Pridemore, 2014). We assessed street and sidewalk conditions (e.g., curbs, pavement, street traffic lines) and building conditions (e.g., roof, windows, age and upkeep) to measure neighborhood physical decay. We scored these conditions "Good," Fair" or "Poor," adopting Odgers' SSO Inventory for measures of decay (Odgers et al., 2009). Both types of physical decay were scored separately on the GSV Inventory Form.

Street and sidewalk conditions:

a. Good: Streets are very well maintained. Few to no sizable cracks, potholes or broken curbs are present. Crosswalk and/street traffic lines are prominently marked.
b. Fair: Streets are moderately maintained. Sizable cracks, potholes or broken curbs, faded crosswalk and/or street traffic lines are present along some of the streets and sidewalks.
c. Poor: Streets are poorly maintained. Sizable cracks, potholes, or broken curbs, faded crosswalk and/or street traffic lines are present along most of the streets and sidewalks.

Building conditions:

a. Good: Majority of the building units are kept in good condition and yards/landscaping are well-maintained. Newer construction or good to high quality building materials used.
b. Fair: Majority of the building units are kept in fair condition. Yards/landscaping are moderately maintained. Average quality building materials used or buildings in need of some cosmetic upkeep like paint or small repairs.
c. Poor: Majority of the units and property are in poor/badly-deteriorated condition. Presence of peeling paint, roofs need repair, broken windows, burned out/boarded up/or appear to be abandoned. Little or no landscaping along buildings and sidewalks.

(6) *Physical disorder*: A neighborhood with high physical disorder appears run-down and dirty and suggests that people living there do not care about their physical surroundings (Ross & Mirowsky, 2009). Disorder indicators such as litter and graffiti are associated with violence and can be reliably measured by direct observation (Sampson & Raudenbush, 1999). We counted the presence or absence of characteristics such as bars on windows, extensive graffiti, garbage, and overgrown lawns and weeds to capture a neighborhood's physical disorder. We gave each item a score of 1 if present and a score of 0 if absent (Sampson & Raudenbush, 1999). We measured physical disorder for the store block (SB) and the adjacent blocks (AB) separately.

(7) *Social disorder*: Like physical disorder, neighborhood social disorder is associated with crime (Sampson & Raudenbush, 1999). Observable signs of social disorder include loitering, public drinking, and panhandling (Ross & Mirowsky, 2009; Sampson & Raudenbush, 1999). These activities are dynamic features that cannot be effectively captured by GSV. To find a proxy for social disorder indicators, we looked at neighborhood crime attractors (i.e., certain types of businesses where these activities often occur) (Bernasco & Block, 2011). We counted the number of these businesses in the observation area, such as bars/taverns, check-cashing services, and pawnshops to assess this characteristic (Bernasco & Block, 2011). We counted 1 for each business present, assessing the store block (SB) and the adjacent blocks (AB) separately (Table 1).

2.4. Procedure and data collection

2.4.1. Training

Four raters, three with master's level research experience or above, and one graduate level research assistant, completed training on using the GSV application (e.g. how to navigate), the GSV Inventory Form, and how to conduct both virtual and in-person neighborhood audits. Raters were provided with a glossary defining all constructs and measures (including screenshots with examples) and step-by-step instructions.

2.4.2. Data collection

Raters independently recorded the street names of the eight block faces comprising the observation area. They then conducted the observation in a clockwise direction starting with the block face containing the alcohol retail store. Raters assessed each characteristic, using the glossary as necessary, and entered the appropriate score in the GSV Inventory Form.

2.4.3. Inter-rater reliability

To evaluate inter-rater reliability, two raters independently conducted virtual audits on 20 Oakland, California alcohol retail stores and their surrounding neighborhoods. We chose Oakland because it is a study city and has a variety of store types.

We compared audit results from the two raters to determine their concordance. Kappa is the statistic most commonly used to determine precision as it relates to agreement between observers (Viera & Garrett, 2005). We measured inter-rater reliability using the kappa statistic for ordinal, categorical measures, a simple correlation coefficient for continuous measures, and percent agreement for multinomial variables (neighborhood type and

Table 1
Measuring store and neighborhood characteristics.

Measure	Score
Neighborhood type	
Downtown, urban, suburban, rural	*
Store type	
Large grocery, general merchandise	1
Mom & pop, convenience, gas	2
Liquor store (high-end or moderate)	3
Liquor store (low-end or dilapidated)	4
Primary land use (*choose up to 2*)	
Residential, commercial, industrial, airport, institutional, other	**
Opportunity to loiter	
None	0
Vacant lot or park in observation radius	1
Adjacent parking lot, vacant lot, or park	2
Both of the above	3
Physical decay: street & sidewalk conditions	
Mostly good	1
Mostly fair	2
Mostly poor	3
Physical decay: building conditions	
Mostly good	1
Mostly fair	2
Mostly poor	3
Social disorder proxies (*Score 1 for each*)	
None	0
Liquor store, bar, tavern	1
Auto body shop	1
Adult entertainment venue	1
Check-cashing service	1
Pawnshop	1
Motel	1
Physical disorder	
None	0
Bars/grates on at least two doors or windows	1
Extensive graffiti	1
Majority of yards overgrown with weeds	1

* Used in initial screening process only.
** At least one primary use must match target.

land use) where kappa would not be appropriate as a measure of reliability.

2.4.4. Inter-source reliability

To evaluate the inter-source reliability of our GSV-based audit tool for both virtual and in-person audits, we used the GSV Inventory Form to conduct both types of audits on 20 Minneapolis and Saint Paul, Minnesota liquor stores and their surrounding neighborhoods. Convenience and cost dictated conducting this test locally. These cities were not appropriate venues for testing inter-rater reliability because in Minnesota, high-alcohol beverages are sold only in liquor stores. This circumstance would have limited our ability to test the store-type measure. As with the inter-rater reliability analysis, we used the Kappa statistic, the correlation coefficient, and percent agreement to evaluate concordance between virtual and in-person neighborhood audits.

2.4.5. Instrument modification

The GSV Inventory Form and glossary evolved during reliability testing based on rater feedback. For example, after conducting eight virtual and in-person audits, raters determined that more precise physical decay measurements were needed. Originally, both "street & sidewalk" and "building conditions" were assessed with one score for the store block and one score for the three adjacent blocks, combined. Raters found it difficult to score these features, since one block could be in "good" condition while the next block was in "poor" condition. We modified the GSV Inventory Form so that each of the eight block faces could be scored individually, and then summed, allowing for more precise scoring. For purposes of reliability testing, measures were considered to be concordant if summed scores were within 0.5 of each other.

2.4.6. Matching

After validation, we used our GSV-based audit results to match target study areas with comparison study areas[3] This paper describes the matching results for Minneapolis. We conducted a full GSV Inventory on all targets and their neighborhoods to create the target area's score and developed a three-step process for choosing comparisons.

First, we chose the 10 best candidates based on census demographics (population density, percent males aged 15–29, and percent female-headed households) and alcohol outlet density (both on and off-premise) in a 0.25-mi-radius created around each target store. We then vetted each candidate store to confirm that it had no malt liquor policy restrictions.

Second, because audits are time-consuming, we conducted a cursory assessment of these candidates to reduce the number of full audits required. Using GSV, we took an aerial screenshot of each observation area to assess neighborhood type. Poor neighborhood type matches (defined solely as a downtown area matched with a rural area) were disqualified from further consideration. We then used GSV to evaluate land use and store type for each remaining candidate, recording the observations in the GSV Inventory Form. We prioritized these two characteristics based on their known relationship with crime, our primary study outcome. A candidate was disqualified at this point if it did not have a primary land use in common with the target area, or if its store type was a poor match (defined solely as a supermarket or general merchandise store matched with any other store type).

Third, we conducted a full GSV-based audit on the five best remaining candidates, using the GSV Inventory Form to assess and record neighborhood characteristics. We chose the three candidates with total GSV scores closest to the target's score as final comparisons.

3. Results

To assess inter-rater reliability and inter-source reliability, we compared scores in 10 measurement categories (as noted earlier, "street and sidewalk conditions" and "building conditions" originally had two scores each).

3.1. Inter-rater reliability results

Inter-rater reliability results for the 20 Oakland neighborhood audits were generally very strong. Scores from the two raters were in perfect to very good agreement on 7 out of 10 measurements. The two physical disorder indicator categories had the lowest concordance. Perception of "substantial" or "excessive" disorder indicators was subjective and inconsistently applied between raters. Raters also reported that it was challenging to see some physical disorder characteristics such as garbage on GSV. These challenges most likely account for these two low reliability scores (Table 2).

[3] A detailed description of the methods used in our larger study to generate potential comparison candidates chosen for the GSV-based audits is set forth in an upcoming paper (Jones-Webb, R., McKee, P., Toomey, T., Nelson, T., Erickson, D., Bestrashniy, J., & Levine Less, E. unpublished results).

Table 2
Inter-rater reliability ($N = 20$).

	Agreement
Neighborhood type	1.0[a]
Primary land use	0.95[a]
Store type	1.0[b]
Loitering opportunity (0–3)	0.87[b]
Street & sidewalk condition (1–3)	0.51[c]
Building condition (1–3)	0.76[c]
Social disorder indicators, OB (0+)	0.74[b]
Social disorder indicators, AB (0+)	0.76[b]
Physical disorder indicators, OB (0–4)	0.29[b]
Physical disorder indicators, AB (0–4)	0.08[b]

[a] Percent agreement.
[b] Kappa.
[c] Correlation (agreement defined as within 0.5).

Table 3
Inter-source reliability ($N = 20$).

	Agreement
Neighborhood type	0.83[b]
Primary land use	0.85[a]
Store type	0.78[b]
Loitering opportunity (0–3)	0.73[b]
Streets & sidewalk conditions (1–3)	0.67[c]
Building conditions (1–3)	0.87[c]
Social disorder indicators, OB (0+)	0.86[b]
Social disorder indicators, AB (0+)	0.64[b]
Physical disorder indicators, OB (0–4)	0.15[b]
Physical disorder indicators, AB (0–4)	0.34[b]

[a] Percent agreement.
[b] Kappa.
[c] Correlation (agreement defined as within 0.5).

3.2. Inter-source reliability results

Substantial agreement was found between ratings for virtual and in-person observations for 8 out of 10 neighborhood characteristics measured. The two physical disorder categories had the lowest concordance among categories, likely due to the temporal nature of items such as garbage. Observations conducted in-person tended to pick up more physical disorder indicators such as extensive graffiti and litter. Also, as noted above, such items were easier to see in-person than in GSV (Table 3).

3.3. Matching

Target area GSV scores ranged from 9 to 16 for the six target areas studied in the pilot city of Minneapolis, Minnesota. The average difference between the target area scores and the scores of the candidates selected as final comparisons ranged from 0.7 to 2.0, while the average difference between the target area scores and the scores of the candidates not selected ranged from 3.0 to 5.0. Thus, final comparisons were more similar to the targets than were the other candidates (Table 4).

4. Discussion

Our results indicate that GSV can be used as a reliable tool for measuring many types of neighborhood characteristics associated with alcohol-related crime. We found that the most reliable characteristics to measure, for purposes of matching target and comparison study areas, were those that are easy to see in both virtual and in-person observations, and unlikely to change over time. This is evidenced by substantial concordance on more stable characteristics such as store type, neighborhood type, primary land use, and opportunities to loiter. Results from inter-rater and inter-source reliability testing in this study are comparable to those in previous GSV validation studies (Odgers et al., 2012; Rundle et al., 2011).

Table 4
GSV scores of target areas, final comparison areas, and other candidates.

	Target score	Range of final comparisons	Range of other candidates	Ave. difference final candidates	Ave. difference other candidates
Group A	**9**	9–10	13–15	0.7	5.0
Group B	**9**	7–10	13–15	1.0	5.0
Group C	**10**	10	12–15	0.0	3.5
Group D	**12**	10–12	16	1.0	4.0
Group E	**14**	13–16	11	1.5	3.0
Group F	**16**	13–16	12	2.0	4.0

Our results confirm that data from GSV-based neighborhood audits can be used to inform matching of target and comparison study areas. Using GSV to conduct virtual neighborhood audits allowed more precise matching of target and comparison areas than would have been possible using census demographics alone. Across 34 targets in our larger study, 18 of the most closely-matched comparisons based on census demographics were disqualified after store and neighborhood characteristics were taken into consideration.

Fig. 2 illustrates how GSV helped us find the best matches. It shows screenshots of a Minneapolis target store and two comparison candidate stores and surrounding areas: one is a good match that was chosen as a final comparison, and the other is a poor match that was not chosen. Had we not evaluated neighborhood characteristics, we would have chosen the poor match as a comparison.

5. Lessons learned

We experienced some of the same challenges noted in previous GSV validation studies. Neighborhood physical disorder was difficult to measure (Rundle et al., 2011). Either items were too small to see on GSV (e.g., garbage), or raters had inconsistent perceptions of "excessive" (e.g., graffiti, overgrown lawns). Parked vehicles sometimes blocked the view, as also noted by Griew et al. (2013), hampering our ability to assess building, street and sidewalk conditions. Also, block length was inconsistent and, in a few instances, GSV did not provide imagery for one of the adjacent blocks. Both issues required us to extend our observation area into the next block. Finally, as with Griew's team, we experienced low inter-rater reliability when assessing certain physical characteristics (e.g., street, sidewalk and building conditions). However, we found that raters could more accurately and reliably measure these characteristics by dividing the study area into smaller segments.

We also experienced spatio-temporal instability. For example, new buildings observed in a 2014 in-person audit did not yet exist when the GSV image was taken. Similarly, a Minneapolis street, shown in disrepair in a 2011 GSV image, had been completely transformed by the time of the 2014 in-person observation. However, spatio-temporal instability may become less problematic as GSV imagery continues to be updated. A new GSV feature called "Time Machine" allows users to view all previous versions of a particular image, with capture dates going back to 2007, but is not yet available for all imagery captured by GSV (Vinay, 2014).

Due to variable image quality, sightline issues, and spatio-temporal instability, we suggest focusing on characteristics that are easy to see in GSV, and are relatively stable over time, when creating GSV-based tools for future research. We also suggest dividing the observation area into small segments for more accurate measurements.

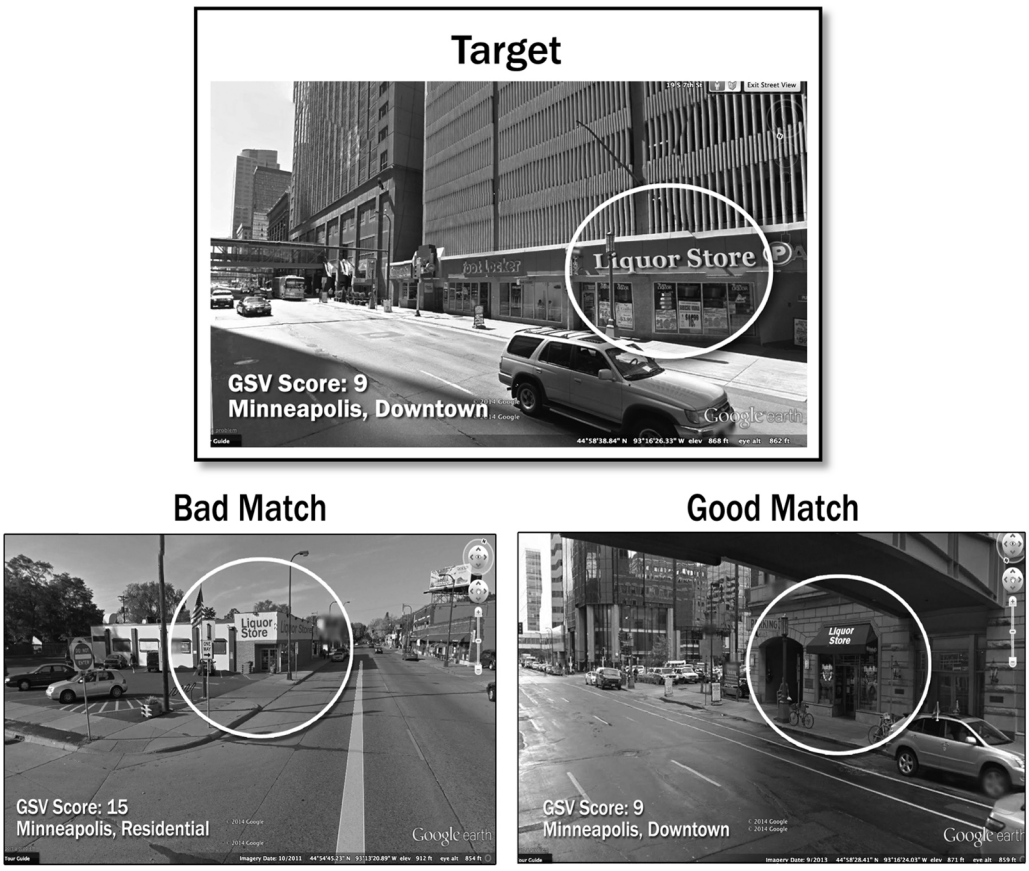

Fig. 2. Example of good match versus poor match.

6. Conclusion

This study evaluated the reliability, utility, and feasibility of a new application for GSV-based research tools. We found that GSV-based neighborhood audits can be a useful, reliable, and cost-effective tool for matching target and comparison study areas when archival data sources are inadequate and in-person audits are not feasible. Use of this technology allowed us to make more precise study area matches. Over half of our top comparison candidates selected, based solely on demographic variables, were disqualified after store and neighborhood characteristics were considered. These results indicate that researchers cannot make assumptions about a neighborhood based on census data alone. Use of GSV saved us significant time and resources, allowing us to focus our efforts on data collection and analysis. Exploring the uses and limitations of GSV as a research tool positions social scientists to more effectively use the next iterations of this emerging technology as it continues to expand.

Acknowledgements

This project was funded by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) NIAAA grant #R01AA020496.

References

Bader, M. D., Ailshire, J. A., Morenoff, J. D., & House, J. S. (2010). Measurement of the local food environment: A comparison of existing data sources. *American Journal of Epidemiology, 171*(5), 609–617. http://dx.doi.org/10.1093/aje/kwp419

Ben-Joseph, E., Lee, J. S., Cromley, E. K., Laden, F., & Troped, P. J. (2013). Virtual and actual: Relative accuracy of on-site and web-based instruments in auditing the environment for physical activity. *Health & Place, 19*, 138–150. http://dx.doi.org/10.1016/j.healthplace.2012.11.001

Bernasco, W., & Block, R. (2011). Robberies in Chicago: A block-level analysis of the influence of crime generators, crime attractors, and offender anchor points. *Journal of Research in Crime and Delinquency, 48*(1), 35–57. http://dx.doi.org/10.1177/0022427810384135

Bluthenthal, R. N., Browntaylor, D., Guzman-Becerra, N., & Robinson, P. L. (2005). Characteristics of malt liquor beer drinkers in a low-income, racial minority community sample. *Alcoholism, Clinical and Experimental Research, 29*(3), 402–409. 00000374-200503000-00015 [pii].

Campbell, C. A., Hahn, R. A., Elder, R., Brewer, R., Chattopadhyay, S., Fielding, J., et al. (2009). The effectiveness of limiting alcohol outlet density as a means of reducing excessive alcohol consumption and alcohol-related harms. *American Journal of Preventive Medicine, 37*(6), 556–569. http://dx.doi.org/10.1016/j.amepre.2009.09.028

Caughy, M. O., Brodsky, A. E., O'Campo, P. J., & Aronson, R. (2001). Perceptions of parenting: Individual differences and the effect of community. *American Journal of Community Psychology, 29*(5), 679–699. http://dx.doi.org/10.1023/A:1010461016939

Clarke, P., Ailshire, J., Melendez, R., Bader, M., & Morenoff, J. (2010). Using Google Earth to conduct a neighborhood audit: Reliability of a virtual audit instrument. *Health & Place, 16*(6), 1224–1229. http://dx.doi.org/10.1016/j.healthplace.2010.08.007

Collins, R. L., Bradizza, C. M., & Vincent, P. C. (2007). Young-adult malt liquor drinkers: Prediction of alcohol problems and marijuana use. *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors, 21*(2), 138–146. 2007-08148-002 [pii].

Curtis, J. W., Curtis, A., Mapes, J., Szell, A. B., & Cinderich, A. (2013). Using Google Street View for systematic observation of the built environment: Analysis of spatio-temporal instability of imagery dates. *International Journal of Health Geographics, 12*. http://dx.doi.org/10.1186/1476-072X-12-53 53-072X-12-53.

Franklin, F. A., Laveist, T. A., Webster, D. W., & Pan, W. K. (2010). Alcohol outlets and violent crime in Washington, DC. *The Western Journal of Emergency Medicine, 11*(3), 283–290.

Franzini, L., Taylor, W., Elliott, M. N., Cuccaro, P., Tortolero, S. R., Janice Gilliland, M., et al. (2010). Neighborhood characteristics favorable to outdoor physical activity: Disparities by socioeconomic and racial/ethnic composition. *Health & Place, 16*(2), 267–274. http://dx.doi.org/10.1016/j.healthplace.2009.10.009

Furr-Holden, C. D., Smart, M. J., Pokorni, J. L., Ialongo, N. S., Leaf, P. J., Holder, H. D., et al. (2008). The NIfETy method for environmental assessment of

neighborhood-level indicators of violence, alcohol, and other drug exposure. *Prevention Science: The Official Journal of the Society for Prevention Research, 9*(4), 245–255. http://dx.doi.org/10.1007/s11121-008-0107-8
Google (2007). *Google Street View [geographic software]*. Available from ⟨http://www.google.com/maps/views/streetview⟩.
Google, 2015. Retrieved May 18, 2015 ⟨https://www.google.com/intl/en-US/maps/about/behind-the-scenes/streetview/#devices⟩.
Griew, P., Hillsdon, M., Foster, C., Coombes, E., Jones, A., & Wilkinson, P. (2013). Developing and testing a street audit tool using Google Street View to measure environmental supportiveness for physical activity. *The International Journal of Behavioral Nutrition and Physical Activity, 10*. http://dx.doi.org/10.1186/1479-5868-10-103 103-5868-10-103.
Grubesic, T. H., Pridemore, W. A., Williams, D. A., & Philip-Tabb, L. (2013). Alcohol outlet density and violence: The role of risky retailers and alcohol-related expenditures. *Alcohol and Alcoholism (Oxford, Oxfordshire), 48*(5), 613–619. http://dx.doi.org/10.1093/alcalc/agt055
Gruenewald, P. J., & Remer, L. (2006). Changes in outlet densities affect violence rates. *Alcoholism, Clinical and Experimental Research, 30*(7), 1184–1193. ACER141 [pii].
Jones, M., Pebley, A. R., & Sastry, N. (2011). Eyes on the block: Measuring urban physical disorder through in-person observation. *Social Science Research, 40*(2), 523–537. http://dx.doi.org/10.1016/j.ssresearch.2010.11.007
Jones-Webb, R., McKee, P., Hannan, P., Wall, M., Pham, L., Erickson, D., et al. (2008). Alcohol and malt liquor availability and promotion and homicide in inner cities. *Substance Use & Misuse, 43*(2), 159–177. http://dx.doi.org/10.1080/10826080701690557
Kelly, C. M. (2013). Using Google Street view to audit the built environment: Inter-rater reliability results. *Annals of Behavioral Medicine, 45*(Suppl. 1), S108–S112. http://dx.doi.org/10.1007/s12160-012-9419-9
Lockwood, D. (2007). Mapping crime in savannah: Social disadvantage, land use, and violent crimes reported to the police. *Social Science Computer Review, 25*, 194–209. http://dx.doi.org/10.1177/0894439307298935
Milam, A. J., Furr-Holden, C. D., Harrell, P. T., Whitaker, D. E., & Leaf, P. J. (2012). Neighborhood disorder and juvenile drug arrests: A preliminary investigation using the NIfETy instrument. *The American Journal of Drug and Alcohol Abuse, 38*(6), 598–602. http://dx.doi.org/10.3109/00952990.2012.701357
North American Industry Classification System, 2012. Retrieved May 18, 2015 ⟨http://www.census.gov/eos/www/naics/⟩.
Odgers, C., Bates, C., Caspi, A., Sampson, R. J., & Moffitt, T. E. (2009). *Systematic social observation inventory—Tally of observations in urban regions (SSO i-tour)*. Retrieved from ⟨https://www.yumpu.com/en/document/view/11331430/systematic-social-observation-inventory-tally-of-observations-in-⟩ [screencast].
Odgers, C. L., Caspi, A., Bates, C. J., Sampson, R. J., & Moffitt, T. E. (2012). Systematic social observation of children's neighborhoods using Google Street View: A reliable and cost-effective method. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 53*(10), 1009–1017. http://dx.doi.org/10.1111/j.1469-7610.2012.02565.x; 10.1111/j.1469-7610.2012.02565.x
Paschall, M. J., Lipperman-Kreda, S., & Grube, J. W. (2014). Effects of the local alcohol environment on adolescents' drinking behaviors and beliefs. *Addiction (Abingdon, England), 109*(3), 407–416. http://dx.doi.org/10.1111/add.12397
Ross, C. E., & Mirowsky, J. (2009). Neighborhood disorder, subjective alienation, and distress. *Journal of Health and Social Behavior, 50*(1), 49–64. Retrieved from ⟨http://www.jstor.org/stable/20617619⟩.
Rundle, A. G., Bader, M. D., Richards, C. A., Neckerman, K. M., & Teitler, J. O. (2011). Using Google Street View to audit neighborhood environments. *American Journal of Preventive Medicine, 40*(1), 94–100. http://dx.doi.org/10.1016/j.amepre.2010.09.034
Rural Health, n.d. Retrieved May 18, 2015 ⟨http://www.hrsa.gov/ruralhealth/policy/definitions-of-rural.html⟩.
Sampson, R., & Raudenbush, S. (1999). Systematic social observation of public spaces: A new look at disorder in urban neighborhoods. *American Journal of Sociology, 105*(3), 603–651. http://dx.doi.org/10.1086/210356
Skogan, W. (1989). Communities, crime, and neighborhood organization. *Crime & Delinquency, 35*(3), 437–457. http://dx.doi.org/10.1177/0011128789035003008
Snowden, A. J., & Pridemore, W. A. (2014). Off-premise alcohol outlet characteristics and violence. *The American Journal of Drug and Alcohol Abuse, 40*(4), 327–335. http://dx.doi.org/10.3109/00952990.2014.918622
Stucky, T. (2009). Land use and violent crime. *Criminology, 47*(4), 1223.
Tarnai, J. (2003). *Evaluation of the Tacoma, Washington, alcohol impact area (AIA)*. Social & Economic Sciences Research Center, Washington State University Retrieved from ⟨http:\\www.sesrc.wsu.edu⟩.
Tita, G., Cohen, J., & Engberg, J. (2005). An ecological study of the location of gang "set space". *Social Problems, 42*, 272–299. http://dx.doi.org/10.1525/sp.2005.52.2.272
U.S. Census Bureau, n.d. Retrieved May 18, 2015 ⟨http://www.census.gov/population/metro/⟩.
Vanwolleghem, G., Van Dyck, D., Ducheyne, F., De Bourdeaudhuij, I., & Cardon, G. (2014). Assessing the environmental characteristics of cycling routes to school: A study on the reliability and validity of a Google Street view-based audit. *International Journal of Health Geographics, 13*. http://dx.doi.org/10.1186/1476-072X-13-19 19-072X-13-19.
Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine, 37*(5), 360–363.
Vinay, S. (April 23, 2014). *Go back in time with street view [weblog]*. Retrieved from ⟨http://googleblog.blogspot.com/2014/04/go-back-in-time-with-street-view.html⟩.
Zhu, L., Gorman, D. M., & Horel, S. (2004). Alcohol outlet density and violence: A geospatial analysis. *Alcohol and Alcoholism (Oxford, Oxfordshire), 39*(4), 369–375. http://dx.doi.org/10.1093/alcalc/agh062

Elyse Levine Less, JD, MPH is a Study Coordinator in the Division of Epidemiology and Community Health at the University of Minnesota, School of Public Health. Her interests include food environment and alcohol policy research.

Patricia McKee, MA is an Assistant Program Director in the Division of Epidemiology and Community Health at the University of Minnesota, School of Public Health. Her interests include alcohol policy research and environmental influences on eating and exercise behaviors.

Traci L. Toomey, PhD is a Professor in the Division of Epidemiology and Community Health at the University of Minnesota, School of Public Health. Her interests include policy research; community organizing; prevention of alcohol and tobacco-related problems; and intentional and unintentional injury prevention.

Toben F. Nelson, ScD is an Associate Professor in the Division of Epidemiology and Community Health at the University of Minnesota, School of Public Health. His interests include health policy, organization change, health behavior during developmental transitions, influence of sports participation on health, social determinants of health, program evaluation, prevention of alcohol—attributable harm, physical activity promotion, obesity prevention, and motor vehicle safety.

Darin J. Erickson, PhD is an Assistant Professor in the Division of Epidemiology and Community Health at the University of Minnesota, School of Public Health. His research interests include prevention of alcohol-related problems, behavioral methodology, analysis of longitudinal and multilevel data, and structural equation modeling.

Serena Xiong, BA is currently an MPH student at the University Of Minnesota, School of Public Health.

Rhonda Jones-Webb, DrPH is a Professor in the Division of Epidemiology and Community Health at the University of Minnesota, School of Public Health. Her interests include alcohol epidemiology and policy with a special emphasis on race and social class issues, policy as a prevention strategy, and minority health issues.