Reliability of a Store Observation Tool in Measuring Availability of Alcohol and Selected Foods

Deborah A. Cohen, Diane Schoeff, Thomas A. Farley,
Ricky Bluthenthal, Richard Scribner, and Adrian Overton

ABSTRACT *Alcohol and food items can compromise or contribute to health, depending on the quantity and frequency with which they are consumed. How much people consume may be influenced by product availability and promotion in local retail stores. We developed and tested an observational tool to objectively measure in-store availability and promotion of alcoholic beverages and selected food items that have an impact on health. Trained observers visited 51 alcohol outlets in Los Angeles and southeastern Louisiana. Using a standardized instrument, two independent observations were conducted documenting the type of outlet, the availability and shelf space for alcoholic beverages and selected food items, the purchase price of standard brands, the placement of beer and malt liquor, and the amount of in-store alcohol advertising. Reliability of the instrument was excellent for measures of item availability, shelf space, and placement of malt liquor. Reliability was lower for alcohol advertising, beer placement, and items that measured the “least price” of apples and oranges. The average kappa was 0.87 for categorical items and the average intraclass correlation coefficient was 0.83 for continuous items. Overall, systematic observation of the availability and promotion of alcoholic beverages and food items was feasible, acceptable, and reliable. Measurement tools such as the one we evaluated should be useful in studies of the impact of availability of food and beverages on consumption and on health outcomes.*

KEYWORDS *Alcohol availability, Alcohol outlets, Price, Shelf space*

INTRODUCTION

In the United States, the primary means of modifying individual health status is through changes in behaviors, particularly diet, physical activity, and use of alcohol, tobacco, and other substances.¹ The consumption of healthy foods, unhealthy foods, and addictive substances is dependent in part on the products that are available and how they are promoted, displayed, and priced.² When such consumer products have a health impact, their relative availability may affect not only an individual's health,

Cohen, Schoeff, Bluthenthal, and Overton are with the RAND Corporation, Santa Monica, CA, USA; Farley is with the Department of Community Health Sciences, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA; Bluthenthal is with the Urban Community Research Center, Department of Sociology, Carson, CA, USA; Scribner is with the LSU School of Public Health, Department of Epidemiology, New Orleans, LA, USA.

Correspondence: Deborah A. Cohen, MD, MPH, Health, RAND Corporation, Santa Monica, CA, USA. (E-mail: dcohen@rand.org)

but also the health of populations. Understanding the relationship between the availability and the consumption of key consumer products is therefore important to developing population-based health promotion strategies.

Sales of consumer products are associated with several in-store variables, including price, placement of product in a salient position, and length of shelf space. Price is well-known to affect sales, with a lower price generally associated with greater frequency of purchase.³ Placement of products near the cash registers, on special floor displays or at eye level has also been shown to be associated with increased purchases, because people are more likely to notice products prominently displayed.⁴ The relative availability of products measured in the length of shelf space of items has also been shown to strongly influence the amount of sales.^{5,6}

To begin to explore whether the general availability of items in particular neighborhoods may influence the group consumption patterns of local residents, it is first necessary to determine whether product availability and salience can reliably be measured. We developed an instrument to measure the availability of items critical to health, specifically alcoholic beverages, fruits, vegetables, and snack foods. Although our initial goal was to measure the availability only of alcohol, we included measurement of food for two reasons: 1) to avoid stigmatization or negative attitude toward the field staff if they were only to measure alcohol and 2) to examine the importance of the relative availability of alcohol to other food and beverage items, as relative availability may influence purchase decisions.^{5,6}

This report details the reliability of the observational instrument and describes our experiences with measurement of the store environment.

METHODS

Study Sites and Selection of Census Tracts

This evaluation of 51 alcohol outlets was part of a larger study of the availability of alcoholic beverages and foods in 891 outlets in 228 census tracts in southeastern Louisiana and in Los Angeles County, California. The larger study was limited to urban residential census tracts, with urban defined as having more than 2,000 residents per square mile in the 2000 U.S. census. We randomly selected 114 of these census tracts in southeastern Louisiana and 114 census tracts in Los Angeles County, for a total of 228 census tracts. In Louisiana, data collection was suspended when Hurricane Katrina struck, after measurements were collected in 103 census tracts.

Store Observation Instrument

The observation instrument was a four-page paper/pencil form adapted from the "ImpacTeen" assessment.⁷ We assessed interobserver reliability on the following items:

- Type of outlet. Alcohol outlets were classified as liquor stores, grocery stores, and convenience stores or "mom and pop" stores, supermarkets, drug store, or other type of store. We defined convenience stores as chain stores selling primarily snack foods and small, privately owned stores as "mom and pop" stores.
- Extent of alcohol advertising inside the store and number of alcohol ads on exterior of store (a scale from "none" to "covering almost all available space")
- Number of cash registers
- Purchase price of standard brands and containers of beer, malt liquor, and cola

- Placement within the store of beer and malt liquor (on shelves, in refrigerated cases, in open ice buckets, in floor displays, and within 1 m of cash register) and whether items were offered self-service or clerk-assisted
- Length of shelf space in feet for alcoholic beverages, including both self-service (refrigerated and non-refrigerated) and clerk-assisted shelf space
- Length of shelf space in feet for all fruits and vegetables
- Availability (presence or absence) and least price of selected fresh fruits and vegetables

When we pilot-tested the form, we found that some store managers would not allow our staff to conduct measures of price and shelf space. Therefore, for these stores we developed a short data form that collected information on only selected key items that we could observe unobtrusively.

Observer Teams and Training

The observations were conducted by a two-person team in each site. Each team followed a standard protocol and used the standard data collection instrument. A quality control supervisor conducted separate independent observations of selected items on the same day.

Field staff took part in a 3-day training session in Los Angeles, which included lectures, brief trials of coding using samples of products and a practice measurement in local alcohol outlets, and concurrent assessment conducted by the co-principal investigators. Six weeks later a 2-day retraining and assessment session for both teams and the quality control supervisor was conducted in New Orleans. Feedback from assessments was provided to counter observer-drift and reduce any inter-observer disagreement.

Observation Procedures

Approximately 2 weeks before the observers surveyed each tract, letters were mailed to outlet store managers describing the study and informing them of the visit, and observers obtained detailed street maps of the tract. On the day observations began in each tract, the observer team drove along every street in the tract, beginning by following the perimeter of the tract and then proceeding through each street from north to south and from east to west, to locate any stores not in the database from the State Alcohol Beverage Control Agency. Upon arrival at each store, observers asked to speak to the store manager, showed him/her a copy of the advance letter, and described the purpose of the study and the information they wished to collect. If permission was not granted to do the measurements, staff returned to the store and conducted a limited number of the observations on a shorter form, recording the information after they left the store.

Length measurements were made using a wheeled measuring device (Measure Master, Rolotape Corporation). The number of shelves in which items were displayed was not recorded, as items only occupied a small portion of a shelf. The total store floor space was estimated by taking measurements of the store length and width. Purchase price was recorded from advertised or labeled prices; if no price was listed staff asked store personnel. Least price reflected the lowest amount that could be paid to purchase a single item (i.e., one can or bottle of beer, one apple, one orange).

Interobserver reliability was assessed in 51 stores (26 in Los Angeles and 25 in Louisiana) by comparing measurements of field staff with those of a single quality control supervisor. Kappa scores were used to assess reliability for categorical items and intraclass correlation coefficients for continuous items.

RESULTS

Observations

Of the 51 stores in which two people observed independently, we used the short form in six and the full form in 45. Stores that did not permit measurement were primarily “mom and pop” grocery stores. The most significant barrier to using the instrument was caused by language differences. Some shopkeepers were suspicious, but when a team member was the same race and spoke the same language, they were more likely to be cooperative.

Measurements in small stores were also more problematic than in large stores. Not only were owners more often suspicious, but the small aisles made it difficult to maneuver when customers were present. Smaller stores were also less likely to have prices posted, so staff had to ask store personnel for prices. Cooperation in larger grocery stores and supermarkets varied. After practice, field staff were able to complete the long form within about 30 min for small stores and 60 min for supermarkets. Supermarket measurements took longer mainly because in these stores snack foods were placed in many locations.

Interobserver Reliability

Table 1 summarizes reliability of a variety of store characteristics and specific items. Agreement was very high on store type. Liquor stores generally had the words “liquor” in the store name. Agreement was poor on the extent of alcohol advertising.

As far as the locations of placement of beer and malt liquor there was perfect agreement for five out of eight items. Low scores on the shelf and floor display items occurred because of disagreements on the definition of the type of display. One observer recorded excess inventory lined across the floor as a floor display rather than as more shelf space. If items were stacked on temporary shelves at the ends of aisles without special advertising or promotion, observers also disagreed whether those should be considered special floor displays or shelf space. The length of shelf space for alcoholic beverages and for fruits and vegetables had excellent agreement.

Agreement was good on the brand and prices of cola, beer, and malt liquor except for the price of the 2-l bottle of cola. This was caused by the fact that several stores carried 1.5-l bottles instead of 2-l bottles of cola. Another reason the value is low is because occasionally both Pepsi and Coca-cola brands were available and the field staff selected different brands from the quality control supervisor.

There was almost perfect agreement on availability of fruits and vegetables, but varying levels of agreement on least price. Least price was meant to refer to what could be purchased by spending the smallest amount of money, rather than what was the best value per unit purchase. Price varied because observers were not necessarily documenting the same item and the same amount. For example, many outlets had several varieties of apples and oranges. In some places they were sold by the pound, making it possible to select one fruit rather than a bag. In other outlets, some fruits were prepackaged, forcing the purchase of more than one fruit; field staff did not always notice the same items. When there were many choices, it was more likely for errors to occur.

Overall, reliability of our methods was good, with average Kappa=0.87 for dichotomous or categorical values and ICC=0.83 for continuous variables.

TABLE 1. Interobserver reliability of store characteristics, placement, shelf space, item availability, and price

Store characteristics	Reliability	Reliability
Type of store	.98 (Kappa)	—
Interior alcohol advertising	.25	—
No. of cash registers	—	.98 (ICC)
No. of exterior alcohol ads	—	.90
Placement of beer and malt liquor	Beer (Kappa)	Malt liquor (Kappa)
On the shelf (non-refrigerated)	.48	.94*
In refrigerator/cooler	1.0	.68
Singles in open ice bucket	.88	.96*
Floor display	.55	1.0
Within 1 m of cash register	1.0	1.0
Over or behind counter	1.0	1.0
In locked or closed cabinet	1.0	1.0
Other placement	1.0	1.0
Length of shelf space		
Refrigerated alcohol	—	.98
Non-refrigerated alcohol	—	.99
Clerk assisted alcohol	—	.86
Fruits	—	.99
Vegetables	—	.96
Item	Brand (Kappa)	Price (ICC)
Cola (Coca-cola or Pepsi-2 liter)	.52	.64
Beer (Budweiser or MGD-6 pack 12 oz cans)	.79	.86
Malt liquor (Old English, Colt 45, Other 40 oz. bottle)	.76	.91
Item	Availability (Kappa)	Least price (ICC)
Apples	1.0	.41
Bananas	.95	.81
Oranges	1.0	.03
Lettuce	.96	.99
Spinach	.94	.79
Tomatoes	.96	.79
Broccoli	1.0	.98

MGD Miller Genuine Draft; *simple agreement (Kappa could not be calculated).

DISCUSSION

We found that measurement of the availability and price of health-impacting consumer goods was feasible and that our instrument and protocol demonstrated high levels of inter-observer reliability for availability and shelf-space measurements, with somewhat lower levels of reliability for price measures. Although overall our instrument should be considered reliable, the errors discovered highlight two principles that may influence the impact of consumer goods and display on consumer purchasing decisions. First, the more beverage and food choices the greater chance of coding error. Second, selling the same products in different units also led to more errors. When items were bundled in multi-unit packs or when fruits were sold by the piece versus by the pound, it was difficult to judge the least price.

Other researchers have developed instruments to assess the availability and placement of alcohol and food in stores.⁷⁻⁹ “ImpacTeen,” a policy research partner-

ship addressing youth substance use, assessed alcohol advertisements, alcohol beverage signage, beer placement, and the presence of alcohol branded functional objects, but not price or shelf space.⁷ They did not report reliability. Another study assessed the grocery store displays of healthy products and measured the relative space and reported reliability of between 0.73 and 0.78 for the proportion of the total display that was considered “healthy,” rather than total shelf space.⁸ Another study compared the agreement between the store managers’ reports of shelf space and direct observation and found poor agreement (0.14–1.0).¹⁰ Excellent reliability was found in the study of Horowitz et al.⁹ that inventoried items and reported the presence or absence of items.

The lower reliability scores on specific items in our study were mainly caused by two situations: 1) categories of products displayed in several locations in the store rather than clustered together and/or 2) judgment differences among coders as to what constituted floor displays and shelf space or unit of purchase. In the first case, one coder might overlook a display. In the latter, new designs, arrangements, and different store practices resulted in new situations for which the interpretation had not been previously agreed upon. In both these cases, additional training would be helpful to attain better reliability. In addition, to be certain of prices when they are not posted and to clearly document which items are priced, it may be beneficial for field staff to actually buy the items. In some cases, staff had the impression that store clerks made up a price if they asked when the price was not labeled. We had poor reliability on the advertising measure, which relied on a general impression of the extent of advertising and use of logos in the store by the field staff, rather than a focused assessment of a particular characteristic of an item. Objective assessments should be more specific to achieve better reliability.

Given that consumption patterns are influenced by in-store marketing, understanding the factors that are most likely to influence our responses is necessary to develop interventions to encourage greater use of healthy products and lesser use of unhealthy products. Documenting the in-store environment is the first step. Our assessment tool is likely to be helpful in studies characterizing the relationship between the availability of key consumer items and health.

ACKNOWLEDGMENTS

We would like to thank Heather Guentzel, Kamau Williams, Michael Murrley, Paul Robinson, and Kelli Trombacco in Los Angeles, and Erica Alarcon and Catherine Haywood in New Orleans. This study was supported by NIAAA # R01AA013749.

REFERENCES

1. McGinnis JM, Foege WH. Actual causes of death in the United States. *JAMA*. 1993;270(18):2207–2212.
2. Paine-Andrews A, Francisco VT, Fawcett SB, Johnston J, Coen S. Health marketing in the supermarket: Using prompting, product sampling, and price reduction to increase customer purchases of lower-fat items. *Health Market Q*. 1996;14(2):85–99.
3. Raynor HA, Polley BA, Wing RR, Jeffery RW. Is dietary fat intake related to liking or household availability of high- and low-fat foods? *Obes Res*. 2004;12(5):816–823.
4. Hausman A. A multi-method investigation of consumer motivations in impulse buying behavior. *J Consum Mark*. 2000;17:403–419.

5. Curhan R. The relationship between shelf space and unit sales in supermarkets. The relationship between shelf space and unit sales in supermarkets. *J Mark Res.* 1972;9:406–412.
6. Wilkinson J, Mason J, Paksoy C. Assessing the impact of short-term supermarket strategy variables. *J Mark Res.* 1982;19:72–86.
7. Point-of-purchase alcohol marketing and promotion by store type—United States, 2000–2001. *MMWR Morb Mortal Wkly Rep.* 2003;52(14):310–313.
8. Cheadle A, Psaty B, Wagner E, et al. Evaluating community-based nutrition programs: assessing the reliability of a survey of grocery store product displays. *Am J Public Health.* 1990;80(6):709–711.
9. Horowitz CR, Colson KA, Hebert PL, Lancaster K. Barriers to buying healthy foods for people with diabetes: evidence of environmental disparities. *Am J Public Health.* 2004;94(9):1549–1554.
10. Housemann R, Orenstein D, Mayer J. Inside the community: A validity study examining availability of heart-healthy foods at urban grocers. *Prev Med.* 2001;33:S27.