

# Datasets used by policymakers, scientists for public health analyses inconsistent

November 17 2014

---

Commercially available datasets containing a wealth of information about food and alcohol establishments differ significantly, raising concerns about their reliability as sources of information that could be used to set public policy or conduct scientific research, according to a University of Pittsburgh Graduate School of Public Health investigation.

The analysis, funded by the Aetna Foundation, will be presented Monday at the American Public Health Association's (APHA) annual meeting in New Orleans. It examined systematic differences in two commercially available datasets when they were used to determine the relationship between neighborhood socioeconomic characteristics and the density of food and alcohol establishments.

"If we're making decisions about setting [public policy](#) to improve [public health](#) - such as incentives for grocery stores that offer fresh produce in economically depressed areas - then we need to be making these decisions based on accurate data to back up the need for such incentives," said lead investigator Dara Mendez, Ph.D., M.P.H., an epidemiologist at Pitt Public Health. "Our study found that relying on just one of these commercially available datasets likely wouldn't provide robust [information](#)."

There are numerous datasets available for a fee that give detailed information about food and alcohol establishments across the U.S. Typically, these datasets are purchased by companies that use them for marketing purposes.

Dr. Mendez and her team used two different commercially available datasets containing information about food and alcohol establishments in Allegheny County, which includes Pittsburgh. The information was divided into the 416 distinct census tracts in the county as a means to define neighborhoods. Each census tract consists of an average of 4,000 people.

Both of the datasets showed that the density of alcohol outlets increased as neighborhood poverty increased. However, the datasets differed when it came to grocery stores. One showed that as poverty increased, the number of grocery stores increased. The other showed no association.

"This is a perplexing disagreement that likely comes down to the datasets using different classification systems and also not accurately capturing all the information. For example, because we are familiar with Allegheny County, my team was able to determine that some of the key [grocery stores](#) in our area were not included," said Dr. Mendez.

"However, if we were doing a similar analysis for a city we were not familiar with, we likely wouldn't catch the discrepancy and could come to an inaccurate conclusion."

**More information:** APHA Abstract No. 302593, "Examining systematic biases in secondary commercial data sources of food and alcohol environments: Differences across neighborhood racial and socioeconomic characteristics," will be given as an oral presentation at 12:30 p.m. CST on Monday, Nov. 17, 2014.

Provided by University of Pittsburgh Schools of the Health Sciences

Citation: Datasets used by policymakers, scientists for public health analyses inconsistent (2014, November 17) retrieved 18 April 2024 from <https://medicalxpress.com/news/2014-11-datasets->

[policymakers-scientists-health-analyses.html](http://policymakers-scientists-health-analyses.html)

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.