Numeracy
Advancing Education in Quantitative Literacy

Volume 8 | Issue 1 Article 7

2015

Cancer Clusters in Delaware? How One Newspaper Turned Official Statistics into News

Victor W. Perez
University of Delaware, victorp@udel.edu
Joel Best
University of Delaware, joelbest@udel.edu
Rachel J. Bacon
Pennsylvania State University, rjbacon@udel.edu

Recommended Citation
Perez, Victor W., Joel Best, and Rachel J. Bacon. "Cancer Clusters in Delaware? How One Newspaper Turned Official Statistics into News." *Numeracy* 8, Iss. 1 (2015): Article 7. DOI: http://dx.doi.org/10.5038/1936-4660.8.1.7

Cancer Clusters in Delaware? How One Newspaper Turned Official Statistics into News

Abstract

The flagship newspaper for the state of Delaware, the *News Journal*, has been instrumental in disseminating information from state-generated reports of *cancer clusters* to its readers over the past 7 years. The stories provide colorful maps of census tracts designated as clusters, often on the front page, and detail the types of elevated cancers found in these tracts and the purported relationship of elevated cancer rates to local industry pollution. Though the *News Journal* also provided its readers with advice about interpreting these data with caution, it uncritically presented these data. Using the state's unusual definition and measurement of elevated cancer incidence as *cancer clusters*, it transformed questionable statistics into an alarming public issue. This article critically examines these news reports and the state-generated reports they utilized.

Keywords

public health, disease clusters, statistics

Cover Page Footnote

Victor W. Perez is an Assistant Professor in the Department of Sociology and Criminal Justice at the University of Delaware. His research interests include the social construction of health and environmental knowledge. He teaches courses in introductory sociology, research methods and statistics, and environmental and health social movements.

Joel Best is Professor of Sociology & Criminal Justice at the University of Delaware. His most recent books are: *The Student Loan Mess* (with Eric Best); and *Kids Gone Wild* (with Kathleen A. Bogle).

Rachel J. Bacon is a graduate student at Penn State University. She is working towards her M.A. in Sociology and Demography with concentrations in health and spatial methods.

Introduction

The principal headline on the front page of the Sunday, September 12, 2010, issue of Wilmington, Delaware's *News Journal* must have alarmed many readers: "Mapping Out Clues to a Cancer Mystery"; the subhead was "Study Focuses on Clusters to Find Cause of Del.'s High Rates" (Montgomery 2010: A1). Next to the story was a small map of the state and readers were invited to view an "enlarged, detailed map" on page A6. In fact, the larger map covered about half the page and identified 45 census tracts that had unusually high incidence rates (i.e., new cases) of cancer (see Figure 1).

Delaware is a small state—less than 100 miles long, and only 9-35 miles wide— and the *News Journal* is the state's principal daily newspaper. Circulation is highest for the paper's Sunday edition, thus the story was meant to attract maximum attention. Moreover, many readers probably recognized this story as one in a series, originating with the 2007 front-page headline that reported a "confirmed cancer cluster" near the Indian River power plant in Millsboro had been identified (Nathans 2007: A1). Several similar pieces in 2008 examined the cancer cluster issue, and after the 2010 front-page story, another would appear in 2014—each accompanied by its own maps (Barrish 2008; Barrish and Montgomery 2008; Miller 2008; Montgomery 2008; Montgomery and Miller 2014; Shortridge 2008).

These stories made frequent reference to a familiar fact—familiar at least to Delaware residents—that the state has had higher cancer rates than the national average. Many residents assume that this is related to the long-standing presence of the chemical industry in Delaware, that pollution from chemical plants and other types of industry found around the state undoubtedly explains elevated cancer risks. For instance, the *News Journal* quoted one woman (Montgomery and Miller 2014: A8):

> My husband is a cancer patient. I lost my mother to cancer. Everybody I know, truly everybody, has a close connection to cancer in some way. I think my whole family has always assumed it was the chemical companies, and we are in some way paying the price for pollution.

Our purpose is to critically examine these newspaper stories, particularly the ways their maps communicate information about public health, and to further examine the reports from state agencies that provided the basis for these news reports. In a sense, these stories presented data and let the facts speak for themselves, except that the impression conveyed was both alarming and misleading.

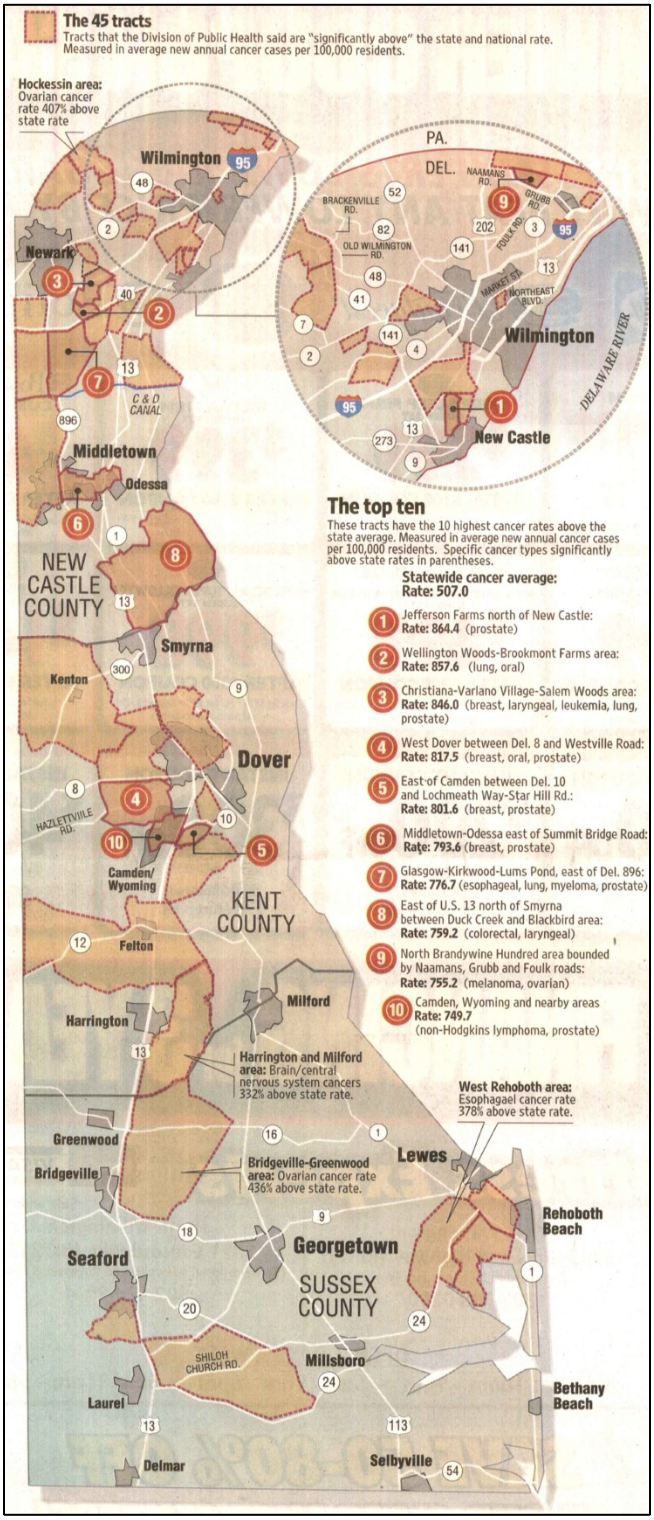

Figure 1. From *The News Journal*, September 12, 2010 © Gannett. All rights reserved. Used by permission and protected by the Copyright Laws of the United States. The printing, copying, redistribution, or retransmission of this Content without express written permission is prohibited. http://www.delawareonline.com/

Thinking about the 2010 Cancer Map

Consider the map in Figure 1. It identifies 45 census tracts with high rates of different types of cancers, and features notations explaining which types have elevated rates in 14 of the tracts. The elevated cancer incidence rates of ten of the census tracts are included in a "Top Ten" list, while 4 census tracts are referred to by way of text pop-outs on the map, describing how much more cancer they have (in percent) relative to the state as a whole. Further, note the types of cancer identified in the census tract areas included in the "Top Ten" list: 1 tract had an elevated rate for one type of cancer, 5 for two types, and the remaining 4 for three to five types (Montgomery 2010). Overall, the map mentions elevated rates for at least 10 different types of cancer (i.e., prostate, ovarian, etc.).

Right away the map presents a statistical issue. The story notes that Delaware has 196 census tracts, and that the state report classified rates for 23 types of cancer. Delaware census tracts average approximately 4,000 residents, but the number of residents in each tract varies greatly across the state. Additionally, for rarer types of cancer a tract might have only "one or two new cases a year to a few dozen" (Montgomery 2010: A6). If the number of residents in a census tract is small, small fluctuations in the number of cancer diagnoses can result in significant fluctuations in the incidence rate. Additionally, if the number of cancer diagnoses in a tract is small, even in tracts with higher numbers of residents, confidence intervals can be wide and denote a significant amount of uncertainty in interpreting the incidence rate. Thus, it is easy for rates to vary in both thinly populated tracts and in tracts with fewer diagnoses, both resulting in apparent "clusters" of cases (Gelman and Nolan 2002). For example, the census tract located west of Dover, Delaware's capital, was number 4 on the top-ten list of census tracts with elevated incidence rates, having a rate of 817.5 per 100,000 for the time period 2002-06 (DHHS 2010a). The average number of cancer diagnoses in this census tract was 19.2 cases per year for the years 2002-06 (DHHS 2010b), and it had a population of approximately 4,000 during this time period (the Census estimate for the time period 2005-2009 is approximately 4,400) (U.S. Census Bureau 2014). Ultimately, this tract had a cancer incidence confidence interval ranging from a lower limit of 647.2 to an upper limit of 1018.9 cases per 100,000, which is a wide range of uncertainty around the rate of 817.5 per 100,000. Upon close inspection, we know that this range of uncertainty is due to the small number of cases diagnosed, however, the lack of confidence intervals around the incidence rates on the "Top Ten" list could easily lead readers to misinterpret the presented rate.

The map also presents us with a geographic puzzle. It doesn't suggest much in the way of clear spatial patterns; the 45 tracts with one or more high cancer rates are sprinkled throughout the state. If we look at the "Top Ten" tracts and the specific types of cancer elevated within them, there is relatively little evidence of contiguous tracts sharing a particular sort of cancer: the tracts designated 2 and 3 share a border, and both have elevated rates of lung cancer, but the one labeled 2 has a second type of cancer, and the tract labeled 3 has four other types for which the other tract does not have elevated rates; similarly the tracts labeled 4, 10 and 5 are contiguous, and all three had high rates of prostate cancer, but each had high rates for one or two other types of cancer that were not shared with its immediate neighbor. If cancer rates were indeed patterned, we might expect more instances of high rates for particular types of cancer being found in adjacent census tracts (i.e., a *spatial* pattern).

In other words, while Delaware residents who looked at the *News Journal*'s 2010 map might have been alarmed to discover that cancer rates were above average in or near the census tracts where they lived, the information conveyed in the map is insufficient to make a convincing case that there are actual cancer clusters in Delaware.

Comparing Maps

The confusion increases when we consider the two maps accompanying a 2014 *News Journal* story. In line with this story's upbeat headline ("Conquering Cancer"), the story featured a smaller map showing "more than 30 census tracts [with] . . . significantly above average cancer rates" for 2002-2006 (Montgomery and Miller 2014: A1). The 2010 story discussed above also had been based on 2002-2006 data, but there is no explanation why the earlier story identified 45 problematic tracts, vs. only 30 in this newer map (the tracts identified in the newer map all seem to have been among those specified in the 2010 story).

Even more interesting is the 2014 story's larger map identifying 14 "Hot Spots"—census tracts that "recorded consistently high overall cancer rates for periods from 2001-2009" (Montgomery and Miller 2014: A1). Comparing this map with the 2010 map, and with the smaller 2014 map is confusing. Of the 14 "Hot Spots," only 2 were among the "Top Ten" tracts identified in the 2010 map, and only 5 are identified on the smaller 2014 map showing 30 tracts with high cancer rates during 2002-2006. Presumably, cancer clusters should display some stability–we might expect that a cluster would reappear on map after map (i.e., a *temporal* pattern). The discovery that, year after year, a particular type of cancer is unusually common in a particular census tract would suggest that the high cancer incidence rate is not merely a product of random variation and could possibly be connected to environmental sources.

The *News Journal* defines a "Hot Spot" as a census tract with an elevated rate of all-site cancer across any two or more of the time periods analyzed in a recent state report examining cancer incidence for the years 2001-05, 02-06, 03-07, 04-08, and 05-09 (DHHS 2013a). Thus, a "Hot Spot" is a census tract that has an elevated cancer incidence rate, regardless of type of cancer, in two or more of the time periods analyzed by the state in one of its 2013 reports. However, this should be interpreted with some degree of caution because it does not mean that any one specific type of cancer is elevated consistently for any specific time period in a tract. It is true that some tracts have elevated cancer diagnoses for some types of cancer over periods of time, but a more precise examination of the types of cancer, and for how long they are elevated (and if consecutively across time periods) is warranted. The fact that the Delaware data do not display clear temporal or spatial patterns–at least as they are portrayed in the *News Journal*'s maps–suggests that these findings should be interpreted with caution.

The *News Journal*'s 2014 coverage seemed more reserved than its earlier articles. It conceded that the reports by state agencies (Montgomery and Miller 2014: A8):

> have yet to reveal a verified, environmentally caused "cluster" of the disease. The tract studies grew out of public concern about a possible pollution-related cancer hot spot in the Millsboro area in 2006. Health officials eventually did find higher than average rates for bladder and male lung cancers in one Millsboro area census tract, between 2001 and 2006, but follow-up studies never confirmed [a] pollution link and rates diminished in subsequent reports.

In sum, the *News Journal* maps sought to summarize a great deal of official data, but it did so in an uncritical fashion. Neither the number of census tracts having higher cancer rates, nor their geographic proximity, nor the temporal patterns in the findings gave a clear sense that the paper had indeed identified health risks. These stories may have commanded readers' attention, but it's not clear that they served the public interest. But, of course, the *News Journal* did not originate the data it mapped; those data came from official agencies. Next, in an attempt to provide some clarity to what the statistics and the maps actually demonstrate, we will consider another puzzling issue: what is a cancer cluster, and how did the state of Delaware identify them?

What is a Cancer Cluster?

To its credit, the *News Journal* did a reasonable job of providing its readers with the general limitations of using "cancer cluster" maps as evidence of industrial or corporate harm on public health. Further, without the luxury of being able to go into detail about the recent state reports re-analyzing cancer incidence data in its news stories, confusion among *News Journal* readers was likely. To boot, there are numerous issues in identifying clusters and drawing conclusions about the environmental impact on public health from these data, such as too few cases to sustain statistical validity (i.e., the "small numbers problem"), resident mobility and migration, resident health behavior, utilization of cancer screening, timing of disease, cancer registries, and statistical chance (Goodman et al. 2014; Margai 2010). Even though the *News Journal* has presented these issues to readers, the chief problem is that the very use of the term "cancer cluster" by the *News Journal* was erroneous.

A cancer cluster falls under the rubric of a non-communicable disease cluster, which the National Cancer Institute (NCI), a division of the National Institutes of Health (NIH), defines as "the occurrence of a greater than expected number of cases of a particular disease within a group of people, a geographic area, or a period of time" (National Cancer Institute 2014a). The *News Journal* used data from reports generated by Delaware's Department of Health and Human Services (DHHS), Division of Public Health (DPH). These reports are unusual, as they did not identify cancer clusters using a standard method known as *relative risk*, where researchers calculate the expected number of cancer diagnoses and then compare that number to what was observed (i.e., the number of diagnoses that actually happened); what they did is different (and will be explained later). To further complicate the use of the term "cancer cluster" in the *News Journal* articles, the NIH reserves the use of *cluster* to a high incidence of one type of rare cancer (not many types lumped together) that affects an age group not usually impacted by that type of cancer or disease (National Cancer Institute 2014b). Delaware officials, on the other hand, began by identifying census tracts in which the combined rate of all cancers exceeded the statewide rate, and labeled those tracts as elevated.

In general, if the public is concerned about cancer in their local area, they can catalyze an investigation by contacting the state's health department, which then seeks to identify the population at risk (e.g., children of a certain age), estimates the expected number of cancer cases in that group in a given time period, and then compares that value to the number of cancer cases that actually occurred in that group. Again, the maps used by the *News Journal* do not reflect this approach; instead, they start with the geographic boundaries of census tracts (not a specific group at risk) and compare cancer *incidence rates* of census tracts to the overall state rate to determine if a tract is a "cluster" (i.e., elevated relative to the state). Using census tracts and incidence rates to identify clusters invites uncertainty in their interpretation because: 1) the incidence rates are made up of numerous cancers lumped together, including the most common types of cancer that are so much more common than other types that they skew the overall rate; and 2) tracts will include people who have the disease but are unrelated to the *real* at-risk group (i.e., the people we suspect have cancer due to exposure to environmental burdens) (National Cancer Institute 2014b).

Thus, what the state reports contain and what the *News Journal* stories reported are age-adjusted cancer incidence rates for 5-year time periods for individual census tracts, which are then compared to the state as a whole. There were two ways that cancer incidence in census tracts were presented using this method. First, the "all-site" incidence rate, which lumped together diagnoses for all 23 types of cancer, including the "big four" of lung, prostate, colorectal, and breast cancer. These four cancer types make up the majority of diagnoses. There are lots of types of each of these cancers (i.e., there are dozens of types of breast cancer), lots of ways people get them, and they have the most opportunity for chance variation. In short, they're so common and varied that including them all in cluster investigations invalidates the method of finding true "clusters."

Second, the state investigators went on to examine incidence rates for 23 specific types of cancer in those tracts that had elevated all-site cancer incidence. Adding to the complexity of including so many types of cancer is the need to calculate an incidence rate per 100,000 people for all of Delaware's census tracts, even though many of the tracts were so small that they averaged only 25 to 35 new cancer cases of all types during the time period being analyzed (DHSS 2010b). Thus, we have a complicated mix of using incidence rates that involve combining five years worth of data at once, extrapolated from relatively small numbers of new cases, for many types of cancer, which are then compared to the incidence rate for the state as a whole. Got it? So, how did these statistics become cancer clusters?

Deconstruction of State Reports

The method of identifying elevated tracts in state reports and the availability of mapping software to present census tracts in color-coded maps, combined with using the term "cluster" in news stories, resulted in the "cancer cluster issue" the readers of the *News Journal* imagine. Arguably the most influential report used by the *News Journal* in recent years was published by the DPH in May 2010. Using year 2000 Census population estimates and census tracts, the report mapped all-site cancer incidence rates for each census tract, for the years 2002-2006 as a whole (i.e., summed together) (DHSS 2010a). We will use that report to describe how clusters came to be identified in the *News Journal*. (It is worth noting again that there have been several DPH reports that used different methodologies for calculating and mapping cancer incidence, resulting in various numbers of elevated tracts across reports – this will be discussed in more detail later.)

The 2010 DPH report contained startling information: Delaware had 45 census tracts that had statistically significantly higher cancer incidence rates for at least one type of cancer than the overall rate for the entire state. These areas quickly became "clusters" in the *News Journal*'s September, 2010 front-page story (Montgomery 2010). Indeed, the article also printed links to interactive online maps that the *News Journal* had created that allowed people to find their census tract and get more detailed information on it.[1]

However, what the Delaware DPH identified in its reports were actually census tracts with higher age-adjusted all-site cancer incidence rates, relative to the state, over the 5-year time period from 2002-06. These reports do not identify clusters in the same way that the NIH does, but instead give spatial and temporal data on cancer *incidence* (Margai 2010). As we noted above, this is not the same as a cancer cluster. So, how did the state identify the census tracts with higher levels of cancer incidence rates, and how did these tracts become "clusters" in the *News Journal*? Furthermore, what are the issues in creating and understanding this type of data? Though the following paragraphs are expository, spelling out precisely how the state generated these data is of vital importance to understanding the issue.

This is where things get a little complicated, so we'll lay out the basics first, drawing on the 2010 report from the DPH. First, the address of a person diagnosed with cancer is assigned to a census tract and this address-to-tract data is validated. For the years 2002-2006, cumulatively (i.e., all diagnoses within that time period added together), all malignant cancer cases that were diagnosed were included (benign tumors and basal and squamous cell cancers were excluded). Because of low census tract validity (i.e., cases where the address of a person could not be assigned to a census tract with confidence), 599 malignant diagnoses were excluded from the analysis at the census tract level (471 of these exclusions came from Sussex—the state's most rural county), but were retained for use in calculating the state's overall cancer incidence rate. In all, 22,161 diagnoses from 2002-2006 were included in the analysis (DHSS 2010a).

Before age-adjusted cancer incidence rates could be calculated, it was necessary to calculate population estimates for census tracts and the state, and to divide them into 5-year age groups so that age could be taken into account (cancer diagnoses occur more frequently among older groups). Let's start with the census tracts, using year 2000 Census guidelines. Of the 197 census tracts that the state of DE had in the year 2000 Census, 1 was not populated, leaving 196 census tracts. Using these data, the DPH calculated 5-year age group proportions by gender for each census tract, providing estimates of annual population for every tract in Delaware. Census tract population estimates for the time period 2002-2006 ranged from 3,132 to 65,136 people across tracts (DHSS 2010a).

[1] These maps have been replaced by a single, all-encompassing interactive map at http://www.delawareonline.com/story/news/health/2014/01/24/map-incidence-cancer-delaware/4834779/ (last accessed Dec 22, 2014)

Given these estimates for the census tracts' populations, how did the report generate age-adjusted cancer incidence rates for them? First, the DPH ran a *cross-tabulation* of age group by census tracts, providing a way to show all the people, according to age group, with a cancer diagnosis in every tract. This gave the number of people in each tract, by age grouping, with a cancer diagnosis. With this information, both crude and age-adjusted incidence rates could be calculated for each 5-year age group and entire census tracts, starting with specific age groups. First, the crude incidence rate was derived by taking the number of cancer diagnoses in a particular age group (e.g., 40-44 year olds) in a census tract, then dividing that number by the population estimate for that age group in the tract; next, multiplying that value by 100,000 to get a cancer incidence rate of that particular age group in that census tract per 100,000 people.

Now, what about a cancer incidence rate for an entire tract and everyone in it together (i.e., not grouped by age)? That's an easy one: just take the number of cancer diagnoses for 2002-2006 in the tract, divide it by the 2002-2006 population estimate for that tract, and then multiply it by 100,000. Want to get, finally, your age-adjusted cancer incidence rates for a census tract? Take the product of the crude incidence rate for each 5-year age group and its corresponding year 2000 Census population weight, then sum all age group incidence rates, and you have an age-adjusted cancer incidence rate per 100,000 for a census tract.

Ultimately, in order for elevated tracts to appear, census tracts were compared to the state as a whole to see how they differed—this is essentially a two-step process. First, we have to calculate the age-adjusted cancer incidence rate for the state of Delaware. Using the 22,161 cases from 2002-2006, the DPH followed similar calculation procedures used to get age-adjusted incidence rates for census tracts in order to calculate the incidence rate for the state as a whole, which turned out to be 507 per 100,000. This rate is an estimate based on the known number of cancer diagnoses and the population estimates used by the DPH. Therefore, the rate is not necessarily exact, but we can situate it within a *confidence interval*, or an interval around this estimate within which we are pretty sure the true cancer incidence rate for the state would fall (i.e., 95% sure). Remember that a confidence interval is calculated using a relatively simple formula involving the age-adjusted rate of a tract, the square root of the number of cancer cases in that tract, and the value 1.96, which reflects a standard deviation value. The 95% confidence interval around the state's all-site cancer incidence rate for the years 2002-2006 ranged from a lower limit of 500.4 to a high of 513.6 per 100,000 (DHSS 2010a). Notice that this confidence interval is relatively narrow around the estimate, indicating an incidence rate that has a good degree of precision because it involves all of the cases (over 20,000 diagnoses for the 5-year time period) and the entire state's population.

In order for census tracts to be identified as elevated, the state compared incidence rate confidence intervals of individual census tracts to the state's. Specifically, a tract's incidence rate for all-site cancer must fall above the state's and the confidence interval for the tract's incidence rate must not overlap with the state's, either. For example, census tract 122, an area located immediately west of the City of Wilmington, had an all-site cancer incidence rate of 670.4 per 100,000, with a 95% confidence interval ranging from 556.9 to 783.9 (DHSS 2010a). Notice how much wider the confidence interval for the census tract is than the state's, highlighting the spread in values necessary in order to be 95% confident that the true number of cancer diagnoses in this tract falls within that range. (This is because the 2002-2006 cancer incidence for that census tract was only an average of 26.8 cases per year for the time period from 2002-2006 (a "small number problem") (DHSS 2010b)) Remember that the confidence interval for the state of Delaware had an upper limit of 513.6 and a lower limit of 500.4 per 100,000. Now, compare the two confidence intervals: the lower limit of the confidence interval for census tract 122 (556.9 per 100,000) does not overlap with the upper limit of the state's confidence interval (513.6 per 100,000), making it "elevated." This is how a statistically significant difference in cancer incidence rate between a census tract and the state came to be identified in official reports, and, subsequently, how "cancer clusters" appeared in the *News Journal*.

In other words, in order to label census tracts as statistically significantly higher, lower, or non-significantly different than the state in terms of cancer incidence over time, the DPH used confidence intervals for census tract incidence rates and compared them to the state's confidence interval. If a census tract's cancer incidence rate was higher than the state's, and the respective confidence intervals did not overlap, the DPH deemed that census tract to have an elevated cancer incidence rate that was not due to chance. Thus, "cancer clusters" were census tracts with confidence intervals that do not overlap with the state's incidence rate for the years 2002-2006 (cumulatively). This approach is not unprecedented, but certainly not the typical definition of a cancer cluster, although it is a very effective way for the *News Journal* to report cancer data according to census tract and incidence.

How the *News Journal* Made Cancer Clusters "Real"

The census tract analyses in the DPH reports are the result of legislation that required such analyses be done, and news reporting on these investigations was inevitable in a state where there is a history of concern about cancer-environment links. It is easy to see how a census tract map, color-coded on the front page of the state's leading newspaper, could be understood to be a "cluster" to readers. Although true cancer clusters are rare, the term is useful for citizens who feel that industry pollution has affected their health in a specific area, and for news media outlets making claims about connections among toxic substances, pollution, and human health. In communities that have an anecdotal history of high cancer rates, as well as communities with long-standing contentious relationships with local industry, the idea of the cancer cluster is a powerful rhetorical tool for making claims about environmental justice. With the availability of mapping software and this type of public health data, organized by incidence rate and already geo-coded by census tract, the *News Journal* was able to provide to the public visually attractive and convincing data on the relationship between cancer and industrial pollution throughout DE, appealing to a variety of communities.

The *News Journal* has been instrumental in championing the cause of citizens who perceive that their area's high cancer rates are the result of industrial pollution since they began publishing a series of stories on the Indian River power plant emissions in the Millsboro area of Delaware, and the disproportionately high number of resident-reported cancers (Nathans 2007). Concern about elevated cancer incidence begins with residents, when one or more concerned citizens ask the state to investigate cancer in their area. In 2007, the *News Journal* portrayed a familiar trajectory for these investigations (Nathans 2007: A1):

> For years, residents in the small towns around the Indian River power plant have noticed friends and relatives falling sick in greater numbers than they thought normal. Years after citizen activists first asked the state for data to establish a pattern, the Division of Public Health has finally confirmed what they suspected: There's a cluster of cancer cases near the coal burning plant – the state's worst polluter.

The 2007 story included only a small map that displayed the Millsboro area and its close surroundings, but soon the *News Journal* was producing statewide maps of cancer incidence by census tract in a series of articles, many of which were based on the release of new and updated state reports. In 2008, for example, the front-page map of the entire state of Delaware showcased 8 areas designated as clusters, and also included a list of high incidence areas and the locations of the top 20 largest polluters in the state (Barrish 2008: A1). The map was made possible by the state's DPH 2008 report that reported cancer incidence rates according to the 27 Census County Division (CCD) areas in Delaware, using methods similar to those described earlier for census tracts, for determining whether a CCD had a statistically significantly higher incidence rate than the state's (DHHS 2008).

These maps are integral to the claimsmaking activities of the *News Journal* and reify the existence of cancer clusters in Delaware, as does the very use of the term. The state's DPH reports, because of the data they included and the way that incidence rates were geo-coded, were easily transferable onto visual displays that made for attractive, easily understood wellsprings of information for the public. Connections to environmental burdens were easy to proclaim, supported by such visual evidence, as one journalist noted in 2008 (Montgomery 2008: A1):

> While state officials cautioned about speculation about potential causes, each of the clusters identified in the state report to be released today was found in an area with clear environmental burdens, from industrial and farm pollution to jammed and smog-laced interstate highways to heavy reliance on shallow, private groundwater supplies that are vulnerable to contamination.

To be sure, the *News Journal* did not misrepresent the data that was provided by the state–it reported them with its own maps and discussed the many limitations of these data. However, it did use the opportunity to call county divisions and census tracts with statistically significantly higher incidence rates *cancer clusters*, and use this information to make claims about environmental burdens and their connection to cancer incidence in Delaware. Only in the early DPH reports did the state actually use the term "cancer cluster," but that term quickly fell out of use in later reports that analyzed census tracts. The idea of a cancer cluster is an effective rhetorical tool for making claims stick to environmental justice issues, suggesting a pattern of excessive cancer in a specific location that needs explanation. The *News Journal* helped to reify cancer clusters in the state of Delaware through the use of state reports, but those early reports were about to get a facelift and fundamentally change the number of "cancer clusters" in Delaware.

From 8 to 59 to 45 to 11: The Problem of Cancer Incidence Statistics and the Frontier of "Hot Spots"

Since 2008, there have been no fewer than 5 comprehensive reports on cancer in Delaware that included incidence-rate-by-census-tract analyses. Recent analyses of cancer incidence by census tract using updated population data, as well as secondary analyses of older reports also using updated population data, have identified significantly fewer "clusters" or elevated incidence areas in the state. In this section, we briefly return to the ways that the state of Delaware has analyzed cancer data and reported on it, how that has changed over time, and how the *News Journal's* reporting on cancer incidence, while remaining as true as possible to these reports, ultimately created a befuddling picture of the issue.

In 2008, the state of Delaware's Division of Public Health used the 27 Census County Divisions (CCD) in Delaware, along with 5-year all-site cancer incidence rates of each CCD to determine if CCDs had elevated cancer incidence rates, using methods similar to those described earlier for census tracts (with confidence intervals for CCDs and comparing them to the state as a whole) (DHSS 2008). The result of that investigation revealed that 8 of the CCDs were elevated. The DPH has always used 5-year interval data to include as many cases as possible to gain some statistical power, but also to "average out" any year-to-year fluctuation of cancer diagnoses. A year later, the DPH started analyzing cancer incidence rates in Delaware using census tracts, using both year 2000 Census and Delaware Population Consortium data, including the 196 tracts that had people living in them, and subsequently revealing *29* census tracts with elevated cancer incidence rates for the years 2001-2005 (DHHS 2009). Then, in 2010, the Delaware DPH released yet another analysis for the years 2002-2006 (described in detail earlier), using year 2000 Census data and 196 census tracts, revealing now *45* tracts with statistically significantly elevated cancer incidence rates (DHSS 2010a). Another report followed, using year 2000 Census data and 196 census tracts for the years 2003-2007 to reveal a startling *59* tracts with elevated rates (DHSS 2012).

However, with the introduction of newer, updated population and census tract data from the Census, things changed dramatically. Armed with year 2010 Census data, which partitions DE into *214* census tracts, the new population estimates used in calculations of incidence rates, and the resulting confidence intervals around them, revealed only *11* elevated tracts in the years 2004-2008 and 9 for the years 2005-2009 (DHSS 2013b). Still another, very recent report analyzing the years 2006-2010 revealed *11* census tracts with elevated cancer incidence rates, relative to the state (DHSS 2014). Further, the availability of 2010 Census data allowed for the re-calculation of population estimates and the re-calculation of cancer incidence rates for the years 2001-2005, 2002-2006, and 2003-2007, revealing *15*, *10*, and *10* elevated census tracts, respectively (DHSS 2013a). In sum, the number of areas officially identified as having high cancer rates has varied widely, from 8 to 59.

The *News Journal* has kept up with drastically changing official statistics on cancer incidence throughout Delaware in reporting on this issue and is now focusing on those tracts with some consistency in elevated incidence rates, designating 14 of them as "hot spots" that serve as areas for increased efforts for screening and prevention. Indeed, one can view these consistently elevated areas using their online, interactive map.[2] These hot spots may be useful in identifying areas that at least have some measurable level of elevated cancer incidence over some time period during the years 2001-2009, but unless these tracts have the *same* type of cancer diagnoses that are elevated over time, and are not merely a tract with *any* cancer diagnosis elevated over time, temporal patterns in elevated tracts provide little value.

Anyone following this series of stories since 2007, however, is confronted with a confusing array of stories, maps, and statistics, and the term "cancer cluster," used often in early *News Journal* stories, has now faded from the headlines.

---

[2] http://www.delawareonline.com/story/news/health/2014/01/24/map-incidence-cancer-delaware/4834779/.

Conclusion

Cancer is a hot button issue for Delaware residents, many of whom live in areas with long histories of controversy concerning industry, pollution, and disease. In the *News Journal* stories, not only can residents read about and connect with others' experiences with cancer, but they are also exposed to public health data that they may view as objective, scientific, and authoritative. The ease with which the *News Journal* could obtain, interpret, and disseminate the state's cancer incidence data is something we need to recognize because it allowed—in relatively short order—the troubling social issue of "cancer clusters" to take on a life of its own through the widespread dissemination of questionable maps and statistics.

Discussions of popular epidemiology often take the form of a morality tale, in which activists expose corporate wrongdoing. Thus, Nathans (2007) cites the Millsboro area cancer cluster investigation as the beginning of the DPH's efforts to look into the issue more broadly for the entire state. A resident of the area—a local doctor—is credited, along with a handful of others, with catalyzing that investigation. As noted earlier, the impetus for cancer cluster investigations is most often one or more community members who believe that cancer in their area is unusually high. To be sure, citizen activists and citizen science alliances can have positive results on community and individual health by shedding light on an issue in a community, providing direction to scientists for asking new questions, and integrating the community's experiences into the scientific paradigm that guides public health investigations (Brown 1992). In order to do this, and in order to raise awareness, the public must have access to good public health data and the media outlets that present it must do so reliably. The *News Journal* did not misrepresent the data it used from Delaware state reports, but the "cancer cluster" issue is one that is already so complex, so statistically tenuous, and so emotionally charged, that the series of stories since the 2006 Millsboro investigation have made for a confusing and misleading saga.

The Delaware case illustrates the problems with statistical "facts." The *News Journal* did not invent numbers; it relayed data first presented in a series of official DPH reports. And those DPH reports were efforts to collate and make sense of reports of medical diagnoses. The officials who wrote the reports explained their methods, although they can be criticized for applying the term cancer cluster (at least early on in their reports) with its alarming implications to phenomena that do not fit the definition of the term. In turn, the *News Journal*'s maps of those "clusters" conveyed a troubling impression to its readers.

What is striking is the failure to ask obvious questions: why were there not clearer geographic patterns; and why did the findings vary so much from report to report, so that the number and locations of high-cancer areas shifted from one report to the next? In a world where numbers are equated with facts, and where maps are increasingly used to display numeric data, there is a risk that colorful images will displace critical thinking (Monmonier 1996).

In sum, what the DPH reports and what the *News Journal* provides its readers are temporal and spatial arrangements of average cancer incidence over time, by census tract, but these are *not* cancer clusters and they should be handled cautiously as evidence of the effect of local industry pollution. In the most recent *News Journal* stories, the tracts with elevated rates are being referred to as "hot spots" of disease, areas to focus cancer prevention efforts. Stopping the use of the term cancer cluster is a step in the right direction, as this undoubtedly will have an effect on how the issue is interpreted by the public. However, the identification of census tracts with elevated cancer incidence - the "hot spots" – should still be interpreted with a healthy degree of caution when trying to determine what the data *really* mean and how they serve as causal evidence of the effect of local industry and pollution. This is not to say that the census tract analysis doesn't have any value, as areas with high levels of cancer incidence (at least relative to the state using confidence intervals) may reveal underlying social conditions that also influence health.

We do not mean to suggest that either the DPH or the *News Journal* were guilty of especially bad practices. They were trying to make data available for public consumption. No doubt other sorts of agencies in other states find themselves trying to produce data to address a wide range of public concerns, and other news media find themselves reporting on those data. But Delaware's small size makes it easier to understand how these data were produced, disseminated, and interpreted. This case reveals many issues in thinking critically about a relatively small data set, which suggests that there are special challenges to promoting numeracy about big data.

Acknowledgments

The authors would like to thank the reviewers for their helpful suggestions and recommendations to improve the paper.

References

Barrish, Cris. 2008. "Eight Cancer Clusters Discovered in Delaware." *Wilmington News Journal*, April 24: A1, A6.

———, and Jeff Montgomery. 2008. "No Money for Cancer Studies, Del. Says." *Wilmington News Journal*, April 27: A1, A8–9.

Brown, Phil. 1992. Popular epidemiology and toxic waste contamination: Lay and professional ways of knowing. *Journal of Health and Social Behavior* 33(3): 267–281. http://dx.doi.org/10.2307/2137356

Delaware Health and Social Services, Division of Public Health. 2008. Average Annual Age-Adjusted Cancer Incidence Rates, 2000-2004, at the Delaware Sub-County Level. Delaware Division of Public Health. (downloaded October 6, 2010)

———. 2009. Delaware Cancer Incidence by Census Tract: Methods. Delaware Division of Public Health. (downloaded February 2, 2014)

———. 2010a. Cancer Incidence and Mortality in Delaware: May 2010. Delaware Division of Public Health. (downloaded August 19, 2011)

———. 2010b. Analysis of Census Tracts with 2002-06 Elevated All-Site Cancer Rates: August 25, 2010. Delaware Division of Public Health. (downloaded August 2, 2011)

———. 2012. Delaware's Cancer Incidence and Mortality Report: 2003-2007. Delaware Division of Public Health. (downloaded February 2, 2014)

———. 2013a. Secondary Analysis of Census Tracts with Consistently-Elevated All-Site Cancer Rates in Delaware, 2001-2005, 2002-2006, 2003-2007, 2004-2008 and 2005-2009. Delaware Division of Public Health. (downloaded February 2, 2014)

———. 2013b. Cancer Incidence and Mortality in Delaware: April 2013. Delaware Division of Public Health. (downloaded February 2, 2014)

———. 2014. Cancer Incidence and Mortality in Delaware, 2006-2010. Delaware Division of Public Health. (downloaded May 2, 2014)

DHSS. See Delaware Health and Social Services, Division of Public Health.

Gelman, Andrew, and Deborah Nolan. 2002. *Teaching Statistics: A Bag of Tricks*. New York: Oxford University Press.

Goodman, Michael, Judy S. LaKind, Jerald A. Fagliano, Timothy L. Lash, Joseph L. Wiemels, Deborah M. Winn, Chirag Patel, Juliet Van Eenwyk, Betsy A. Kohler, Enrique F. Schisterman, Paul Albert, and Donald R. Mattison. 2014. Cancer cluster investigations: Review of the past and proposals for the future. *International Journal of Environmental Research and Public Health* 11:1479–1499. http://dx.doi.org/10.3390/ijerph110201479

Margai, Florence. 2010. *Environmental Health Hazards and Social Justice: Geographical Perspectives on Race and Class Disparities.* Washington, D.C.: Earthscan.

Miller, J.L. 2008. "Del. Cancer Maps Raise Questions about Incidence Rates." *Wilmington News Journal*, December 19: A1–A2.

Monmonier, Mark. 1996. *How to Lie with Maps*, 2nd ed. Chicago: University of Chicago Press.

Montgomery, Jeff. 2008. "High Incidence Seen in Most Polluted Areas." *Wilmington News Journal*, April 24: A1, A6–A7.

———. 2010. "Mapping Out Clues to Cancer Mystery." *Wilmington News Journal*, September 12: A1, A6–A7.

———, and Beth Miller. 2014. “Conquering Cancer.” *Wilmington News Journal*, January 26: A1, A8–9.
Nathans, Aaron. 2007. “Delaware Confirms Cancer Cluster.” *Wilmington News Journal*, August 5: A1, A11–A12.
National Cancer Institute. 2014a. National Cancer Institute Fact Sheet: Cancer Clusters. “Defining Disease Clusters.” (www.cancer.gov/cancertopics/factsheet/Risk/clusters)
———. National Cancer Institute Fact Sheet: Cancer Clusters. “Facts About Cancer Clusters.” (www.cancer.gov/cancertopics/factsheet/Risk/clusters)
Shortridge, Dan. 2008. “State Zeroes in on Millsboro’s Cluster.” *Wilmington News Journal*, May 18: A1, A6.
U.S. Census Bureau. 2014. *American Community Survey, American Fact Finder*. Report generated on October 26, 2014 at http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?fpt=table