

# **An Investigation of Health Status Indicators: An Analytics Approach to Community Health in the United States**

**Shilpa Balan**  
California State University, Los Angeles

**Kanishka Sunnam**  
California State University, Los Angeles

**Aysha Khan**  
California State University, Los Angeles

*It is vital to improve individual outcomes for overall community health; therefore, this study reviewed some of the major components of community health status indicators—heart disease, lung cancer, and breast cancer—from 1994 to 2003 in the United States. Using Tableau analytics software, we analyzed an open data set made available by the Centers for Disease Control and Prevention to extract insights and represent the findings visually. Our analysis revealed that the most deaths in the United States were due to heart diseases. From 1994–2003, Kentucky had the highest number of deaths due to lung cancer and Ohio had the highest number of deaths due to breast cancer. The current findings can be used to advance community health policy strategies to deliver better healthcare models and improve public health.*

## **INTRODUCTION**

The healthcare industry has historically generated enormous amounts of data (Raghupathi, 2010); however, there is a need to manage this data efficiently. There is also a need to improve the quality of healthcare delivery while reducing healthcare costs. Large amounts of data (i.e., big data) are used to describe large data sets. Traditional data processing applications are insufficient to deal with these large data sets. Big data are thus an asset that enables enhanced decision making (De Mauro et al., 2016). There has been an enormous progress in the healthcare industry in digitizing medical records (Groves et al., 2013). Rapid increase of data has made it essential to extract useful data from large datasets (Johri et al., 2017). In this paper, we analyzed trends and provided insights on various diseases such as heart disease, lung cancer, and breast cancer. The findings from this study can help advance community health policy strategies to deliver better healthcare models and improve public health (CDC, 2015).

Applications of big data are now prevalent in every industry such as banking, finance, healthcare, and so on. Without the efficient use of big data technologies, healthcare organizations cannot draw on increasing volumes of data and medical knowledge in an organized, strategic way. In the healthcare industry, big data helps doctors make informed decisions quickly (Cresswells et al., 2017). Moreover,

without big data technologies, it is difficult to exploit the knowledge gained from big data to improve patient safety and quality (Cresswells et al., 2017). According to community health policies, preventive care is essential as it helps prevent illness and provide initial-stage treatment (CDC, 2017). This leads to superior public health.

Anomalies in patient data can lead to incorrect diagnoses. Analytics can help healthcare organizations improve treatment of patients. Data analysis can be used to make clinical decisions. Data, if analyzed correctly, can be an asset to hospitals in reducing administrative costs, providing clinical decision support, reducing fraud, and improving patient wellness (Brimmer, 2013). Data analytics in healthcare can also be used to discover various associations and patterns within the data, which can improve care, save lives, and lower costs. When big data are analyzed, healthcare providers can develop more insightful diagnoses and treatments, resulting in higher quality care and decreased costs (Raghupathi and Raghupathi, 2014). Some additional benefits of big data in healthcare are to predict epidemics, cure diseases, and avoid preventable deaths (Marr, 2015). This leads to improved quality of community life. The drive to understand patient data is to determine the warning signs of serious illnesses early (Marr, 2015).

The Centers for Disease Control and Prevention (CDC) is one of the major operating components of the Department of Health and Human Services. The three diseases we analyzed in this study are some of the major components of community health strategies (CDC, 2016). We identified their associated trends and risks. Specifically, lung cancer is prevalent worldwide, and it has a high mortality rate. Smoking is the main risk factor for lung cancer. Lung cancer can be prevented by avoiding smoking and air pollution (Mustafa et al., 2016). Early detection results in better outcomes. Further, it was estimated in 2018 that a total of 266,120 women in the United States would be diagnosed with breast cancer in 2018 (American Society of Clinical Oncology, 2018). Lastly, about 610,000 people in the United States die from heart disease each year (Centers for Disease Control and Prevention National Center for Health Statistics, 2015). Consequently, we examined these three diseases in detail to identify factors to improve community health. They are each described in further detail below.

## **RESEARCH BACKGROUND**

According to Sabharwal et al. (2016), analyzing large amounts of data in healthcare has previously been time consuming. Semi-structured or unstructured data in the healthcare industry includes doctor notes, office medical records, radiograph films, paper prescriptions, and so on (Sabharwal et al., 2016). Recent technologies in big data have made it easier to extract relevant insights from enormous amounts of data that can be used to provide better care, including detecting diseases at earlier stages, which will help to reduce mortality rates. The CDC identifies best practices to improve public health and lower healthcare spending (CDC, 2015).

Cancer progresses over time and develops in nearly every part of the body. Rashbass (2016) states that the most common cancers occur in the lung, breast, prostate, and colon and the less common cancers occur in the brain, blood, pancreas, ovary, stomach, and skin. In this study, we only examined lung cancer and breast cancer.

Lung cancer involves a malignant tumor that causes rampant cell growth in the lung tissues (Tsao, 2007). The highest rates of lung cancer occur in North America, Europe, and East Asia, while the lung cancer rates in Africa and South Asia are much lower (Stewart et al., 2014). In 2012, according to the World Cancer Report (2014), 1.8 million people worldwide contracted lung cancer, resulting in 1.6 million deaths. Further, the report states that among all types of cancer, lung cancer results in the most deaths among men and the second most among women (after breast cancer).

Most breast cancer cases occur in women aged older than 50 years. If a woman has some history of breast cancer, then she is at a higher risk of developing breast cancer (Roche, 2017). Breast cancer can be detected before the symptoms have developed or after noticing a lump. Breast cancer can be predicted by the stage of the disease; that is, the extent or spread of the cancer when it is first diagnosed (American Cancer Society, 2017).

Heart disease affects millions of U.S. adults (Centers for Disease Control and Prevention National Center for Health Statistics, 2015). Further, with predictive analytics, healthcare providers can track early signs of heart disease risk before illness strikes. In addition to saving lives, healthcare costs can be reduced without having to impact the quality of the provided care. It is essential to keep track of patients' attributes that could indicate an elevated risk of a cardiac episode. Patients may also have previously been diagnosed with chronic obstructive pulmonary disease (Clarke et al., 2017).

A hospital in Toronto, Canada—North York General Hospital—which has an occupancy of 450 beds, uses real time data analytics to improve patient health (Raghupathi and Raghupathi, 2014). The hospital can obtain a greater insight into the operations of healthcare delivery. The hospital has successfully implemented a real-time analytics application to acquire multiple perspectives, including clinical, administrative, and financial. Further, Columbia University Medical Centers analyze data related to patients with brain injuries (Raghupathi and Raghupathi, 2014). The aim is to provide medical professionals with timely information to treat complications. The hospital uses reports generated by these analyses to forecast and diagnose complications sooner than previously able to for patients who experienced a bleeding stroke from a ruptured brain aneurysm.

Developing solutions for such disease-related risk is extremely challenging in healthcare informatics due to large volumes of the data that is increasingly becoming available within hospital systems. Such insights may guide healthcare providers to improve the quality of care, thus reducing the readmission rate and the cost incurred with these readmissions (Rumsfeld et al., 2016). Therefore, using Tableau analytics software, we explored the trends and indicators of these diseases from 1994 to 2003 in the United States.

## **RESEARCH METHODOLOGY**

We used an open dataset made available to the public by the CDC (2016), which includes key community health status indicators. According to the CDC, heart disease and cancer are chief components of the community health data initiative. We used an analytics software (i.e., Tableau) for data analyses, which is a business intelligence and data visualization software (Tableau, 2018). Tableau transforms data into visuals that are appealing and make it easy to spot trends to drive better decisions.

Descriptive analytics is the study of data, typically achieved by visualization tools or business intelligence tools (Raghupathi and Raghupathi, 2014). The goal is to answer questions such as, “What happened or what is happening?” using several types of charts such as pie charts, bar charts, line graphs, tables, heat maps, bubble charts, and others. An example of descriptive analytics is summary health reports that provide a historic review of healthcare providers' care transactions and cost/quality ratios (Ward et al., 2014). Descriptive analytics helps to describe and present data in a format that can be understood easily (Raghupathi and Raghupathi, 2014, Ward et al., 2014).

The current research domain of community health indicators is characterized by increasing prevalence and incidences of people with heart diseases and cancers, to name a few. In this scenario, visualization can represent relevant information such as association between diseases, death rates, and communities requiring more medical attention, and others, to offer insights for effective community health policy management. The next section on Analysis and Results details the insights of our findings.

## **ANALYSIS AND RESULTS**

Figure 1 shows the positive correlation between stroke and coronary heart disease, which is consistent with previous research that noted that stroke is a risk factor for heart disease (Homocysteine Studies Collaboration, 2002). Stroke and coronary heart disease are also associated with other health issues such as high blood pressure (MacMahon et al., 1990).

**FIGURE 1**  
**CORRELATION BETWEEN CORONARY HEART DISEASE AND STROKE**

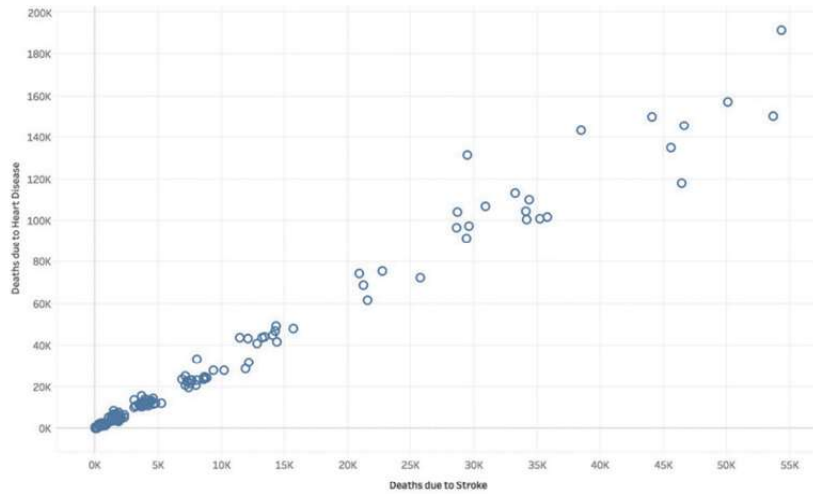


Figure 2 shows the positive correlation between stroke and lung cancer, which is consistent with previous research. When patients with cancer have a stroke, or vice versa, neurological outcomes significantly worsen, resulting in misdiagnosis. Specifically, patients with lung cancer have a 1.5 times higher chance of having a stroke (Chen et al., 2011).

**FIGURE 2**  
**CORRELATION BETWEEN LUNG CANCER AND STROKE**

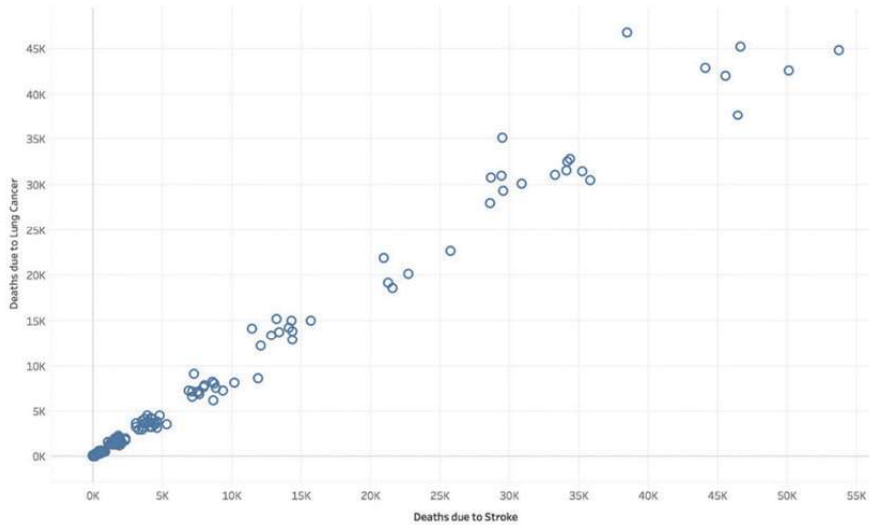


Figure 3 represents the five U.S. states that had the highest number of lung cancer deaths from 1994 to 2003. Kentucky had the highest lung cancer mortality. Of note, lung cancer exceeded most other cancers in the year 2015 (Siegel et al., 2016).

**FIGURE 3**  
**TOP 5 U.S. STATES BY LUNG CANCER DEATHS FROM 1994 TO 2003**

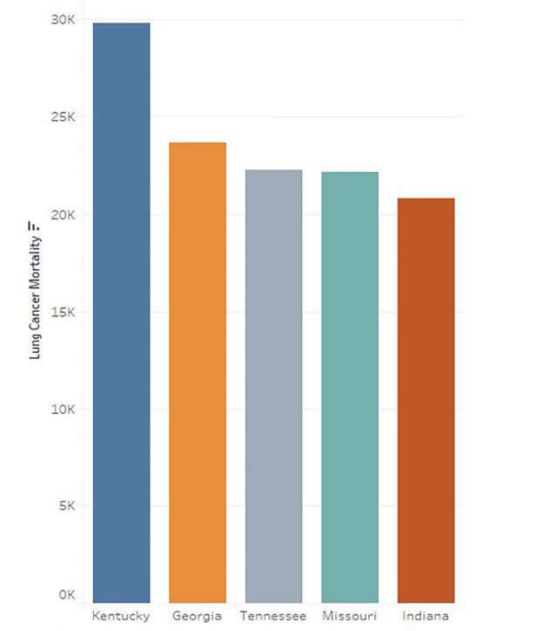
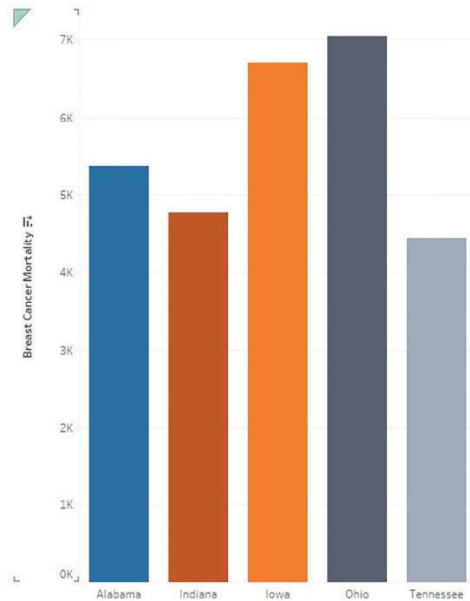


Figure 4 represents the number of breast cancer deaths from 1994 to 2003. Ohio, had the highest breast cancer mortality from 1994 to 2003, followed by Iowa and Alabama. According to Ferlay et al. (2014), breast cancer is the most common cancer in women worldwide and the second most common cancer in general.

Coronary heart disease indicators include factors such as age, blood pressure, cholesterol, diabetes, prediabetes, smoking, obesity, and an unhealthy diet. Table 1 represents indicators caused by coronary heart diseases. In this context, ‘unfavorable to peers’ means that a county requires more effective healthcare compared to the other counties or the entire nation (in this case, the United States), whereas ‘favorable to peers’ means that a county does not need immediate attention as it is doing well regarding overall healthcare progress. As shown, some counties require attention.

**FIGURE 4**  
**TOP 5 U.S. STATES BY BREAST CANCER DEATHS PER 100,000**  
**POPULATION FROM 1994 TO 2003**



**TABLE 1**  
**CORONARY HEART DISEASE INDICATORS FROM 1994 TO 2003**

<b>Time Span</b>	<b>Coronary Heart Disease Indicator</b>	<b>CHSI State Name</b>	<b>CHSI County Name</b>
1994-2003	Favorable to peers	Georgia	Jefferson
			Washington
		Kentucky	Jackson
			Lincoln
	Unfavorable to peers	Texas	Franklin
		Georgia	Franklin
		Missouri	Washington
1999-2003	Favorable to peers	Texas	Jackson
			Washington
	Unfavorable to peers	Virginia	Franklin
		Georgia	Jackson
		Kentucky	Franklin
		Missouri	Franklin
			Lincoln
	Virginia	Washington	
2001-2003	Favorable to peers	Kentucky	Jefferson
		Missouri	Jackson
		Jefferson	
	Unfavorable to peers	Texas	Jefferson

Further, Table 2 represents the indicators caused by lung cancer. Lung cancer indicators include factors such as age, blood pressure, cholesterol, diabetes, prediabetes, smoking, obesity, and an unhealthy diet, which also point to the presence of heart disease.

**TABLE 2  
LUNG CANCER INDICATORS FROM 1994 TO 2003**

<b>Time Span</b>	<b>Lung Cancer Indicator</b>	<b>State</b>	<b>County</b>
1994-2003	Favorable to peers	Georgia	Washington
			Madison
			Franklin
		Kentucky	Washington
		Virginia	Madison
	Unfavorable to peers	Georgia	Lincoln
			Jefferson
		Kentucky	Lincoln
Missouri		Washington Madison	
1999-2003	Favorable to peers	Kentucky	Madison
		Tennessee	Madison
			Jefferson
			Franklin
	Unfavorable to peers	Kentucky	Franklin
		Missouri	Lincoln
			Franklin
		Tennessee	Lincoln
			Franklin
Virginia	Washington		
2001-2003	Favorable to peers	Tennessee	Washington
	Unfavorable to peers	Kentucky	Jefferson
		Missouri	Jefferson

Table 3 shows the mortality rate per disease type in the United States from 2001 to 2003. The mortality rate due to heart disease was much higher than that due to breast or lung cancer. Table 3 also indicates less mortality rate deviation due to breast cancer as compared to other diseases. A higher standard deviation indicates that the deaths caused by the corresponding disease are widely distributed over several disease characteristics including and not limited to the associated diseases and risks.

**TABLE 3**  
**MORTALITY RATE PER DISEASE TYPE IN THE UNITED STATES FROM 2001 TO 2003**

<b>Year</b>	<b>Disease Category</b>	<b>Average Mortality Rate (per 100,000)</b>	<b>Standard Deviation of Mortality Rate (per 100,000)</b>
2001-2003	Heart Disease	171.21	38.07
2001-2003	Breast Cancer	26.32	26.95
2001-2003	Lung Cancer	56.09	56.94

**DISCUSSION**

We analyzed community health status indicators such as heart disease and cancer that were noted by the CDC. Certain correlations were found for lung cancer and heart disease. These diseases were co-occurring with other diseases such as stroke. For the period 1994-2003, Ohio had the highest breast cancer mortality in the United States, which was followed by Iowa and Alabama. The average number of lung cancer deaths from 1994 to 2003 was more than 20,000. In the United States, Kentucky had the most lung cancer deaths during this period. Georgia, Tennessee, Missouri, and Indiana were the other states that also had many cases of lung cancer deaths. These states require more attention to foster health welfare; therefore, healthcare policies should focus on improving community health in these states. Notably, Kentucky has made efforts to expand its medical insurance to improve health outcomes (Reed, 2016), and community health policies in Kentucky have made it possible for its residents to afford medication (Reed, 2016).

The Health Policy Institute of Ohio conducted a review of the hospital community activities (Ohio Department of Health, 2018). Approximately 5 percent of the Ohio hospitals benefit expenditures were allocated toward community health improvement services (Ohio Department of Health, 2018). Thus, Ohio is already making significant investment in their communities. In 2016, a needs assessment was conducted by the Iowa Department of Public Health (Iowa Department of Public Health, 2018). The assessment identified critical issues and indicated Iowa’s health needs. This laid the foundation for the community health in Iowa. In an earlier needs assessment in 2011, some counties in Iowa lacked financial resources (IDPH, 2011). The needs from this assessment were classified as Promoting Healthy Behaviors (IDPH, 2011).

Further, a large population in the rural areas of Alabama do not have good transportation to healthcare providers (Alabama Department of Public Health, 2018), thus hindering access to quality healthcare service. In 2000, about 8.5% of rural households in Alabama did not own a vehicle (Alabama Department of Public Health, 2018); consequently, rural residents face a lack of primary care practitioners in their communities (Alabama Department of Public Health, 2018).

The Health Professional Shortage Area developed by the U.S. Department of Health and Human Services, Health Resources and Services Administration is the most widely recognized method for identifying health professional shortages (Alabama Department of Public Health, 2018). According to this methodology, 54 of Alabama’s 55 rural counties are classified as having a shortage of primary care physicians (Ohio Department of Health, 2018). As of 2017, the CDC’s Healthy Communities Program funded 331 communities and 52 states and territorial health departments (NCCDPHP, 2018). This may address some issues faced by several communities concerning health wellness.



## CONCLUSION AND FUTURE RESEARCH

This study summarized varied community health status indicators in the United States, which illustrates the public health status of the respective counties. In this study, we analyzed three critical diseases that are major components of community health—heart disease, breast cancer, and lung cancer—including disease trends and risks from 1994 to 2003. Our analysis revealed that stroke is highly correlated with lung cancer and heart disease. Of the three diseases, the mortality rate of heart disease was the highest. Kentucky had the highest number of deaths due to lung cancer and Ohio had the highest number of deaths due to breast cancer. CDC's healthy communities program is now working with several states and communities to address such challenges by funding them (NCCDPHP, 2018). This leads to improved community health wellness.

In conclusion, we require more effective approaches to analyze data so that hospitals and healthcare facilities can progress and improve healthcare quality and reduce costs. McKinsey estimates that analytics can aid with savings of more than \$300 billion per year in U.S. healthcare (Kayyali and Van Kuiken, 2013). It is essential to prevent diseases and improve the mortality rate by improving the quality of public health. Using big data analytics, it is possible for healthcare providers to gain radical insights leading to informed clinical decisions. Additional analytic techniques and models can be applied in future research, which will enhance the healthcare delivery model and promote a healthier community.

## REFERENCES

- Alabama Department of Public Health (2018). *The Alabama Community Health Resource Guide*. National Organization of State Offices of Rural Health and the Alabama Rural Health Association. Retrieved February 2018, from <http://www.arhaonline.org/data/publications/community-health-resource-guide/>
- American Cancer Society (2017). *Breast Cancer Facts and Figures 2017–2018*. Atlanta: American Cancer Society, Inc. Retrieved February 2018, from <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017-2018.pdf>
- American Society of Clinical Oncology (2018). *Breast cancer: statistics*. Cancer.Net. Retrieved March 2018, from <https://www.cancer.net/cancer-types/breast-cancer/statistics>
- Brimmer, K. (2013). *5 Ways hospitals can use data analytics*. Healthcare IT News. Retrieved March 2018, from <http://www.healthcareitnews.com/news/5-ways-hospitals-can-use-data-analytics>
- CDC (2017). *CDC Prevention Checklist*. Centers for Disease Control and Prevention. Retrieved May 21, 2018, from <https://www.cdc.gov/prevention/index.html>
- CDC (2015). *CDC Health Policy Series*. Centers for Disease Control and Prevention. Retrieved June 23, 2018, from <https://www.cdc.gov/policy/hst/healthpolicy/index.html>
- CDC (2016). *Community health status indicators (CHSI) to combat obesity, heart disease and cancer*. Centers for Disease Control and Prevention. Retrieved December 2017, from <http://www.healthdata.gov/dataset/community-health-status-indicators-chsi-combat-obesity-heart-disease-and-cancer>
- Centers for Disease Control and Prevention National Center for Health Statistics (2015). *Underlying cause of death 1999–2013 on CDC WONDER Online Database*. Data are from the Multiple Cause of Death Files, 1999–2013, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program.
- Chen, P., Muo, C., Lee, Y., et al (2011). Lung cancer and incidence of stroke. A population-based cohort study. *Stroke*, 42, 3034–3039.
- Clarke, J., Bourn, S., Skoufalos, A., Beck, E., & Castillo, D. (2017). An innovative approach to health care delivery for patients with chronic conditions. *Popul Health Manag*, 20(1), 23–30.

- Cresswells, K., Bates, D., & Sheikh, A. (2017). *Why every health care organizations needs a data science strategy*. NEJM Catalyst. Retrieved April 2018, from <http://catalyst.nejm.org/healthcare-needs-data-science-strategy/>
- De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of big data based on its essential features. *Libr Rev.*, 65(3), 122–135.
- Ferlay, J., Soerjomataram, I., Ervik, M., et al. (2014). *Cancer incidence and mortality worldwide: IARC CancerBase No. 11*. Lyon, France: International Agency for Research on Cancer. Retrieved December 2017, from <http://globocan.iarc.fr>
- Groves, P., Kayyali, B., Knott, D., et al. (2013). *The big data revolution in healthcare*. McKinsey & Company, 1–17.
- Homocysteine Studies Collaboration (2002). Homocysteine and risk of ischemic heart disease and stroke. *JAMA*, 288(16), 2015–2022.
- IDPH (2011). Bureau of Communication and Planning (2018). *Understanding Community Health Needs in Iowa*. Retrieved March 2018, from [https://idph.iowa.gov/Portals/1/Files/CHNAHIP/health\\_needs\\_2011.pdf](https://idph.iowa.gov/Portals/1/Files/CHNAHIP/health_needs_2011.pdf)
- Iowa Department of Public Health (2018). Community Health Needs. Retrieved January 2018, from <https://idph.iowa.gov/chnahip/community-health-needs>
- Johri, P., Singh, T., Das, S., et al. (2017). Vitality of big data analytics in healthcare department. *IEEE*, 1–5.
- Kayyali, B., Knott, D., & Van Kuiken, S. (2013). *The big-data revolution in US health care: accelerating value and innovation*. Retrieved November 2017, from <https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care>
- MacMahon, S., Peto, R., Collins, R., et al. (1990). Blood pressure, stroke, and coronary heart disease: part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet*, 335(8692), 765–774.
- Marr, B. (2015). How big data is changing healthcare. *Forbes*. Retrieved March 2018, from <https://www.forbes.com/sites/bernardmarr/2015/04/21/how-big-data-is-changing-healthcare/#4ee3a4782873>
- Mustafa, M., Azizi, J., & Azizan, N. (2016). Lung cancer: risk factors, management and prognosis. *IOSR J Dent Med Sci.*, 15(10), 94–101.
- NCCDPHP (2018). *Healthy Communities Program (2008–2012)*. Retrieved April 2018, from <https://www.cdc.gov/nccdphp/dch/programs/healthycommunitiesprogram/index.htm>
- Ohio Department of Health (2018). *Improving Population Health Planning in Ohio: Guidance for Aligning State and Local Efforts*. Retrieved February 2018, from <https://www.odh.ohio.gov/-/media/ODH/ASSETS/Files/chss/ship/CommunityPlanningGuidanceV1.pdf>
- Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst.*
- Raghupathi, W. (2010). Data mining in health care. In: *Healthcare informatics: improving efficiency and productivity*, 211–223.
- Rashbass, J. (2016). *Understanding cancer—the importance of patient data*. Public Health Matters. Retrieved February 2018, from <https://publichealthmatters.blog.gov.uk/2016/02/04/understanding-cancer-the-importance-of-patient-data/>
- Reed, P. (2016). *The future of Kentucky's community health care clinics*. Retrieved March 2018, from <https://www.ket.org/health/the-importance-of-family-health-centers-to-public-health/>
- Roche (2017). *Breast Cancer Report*. Retrieved May 2, 2018, from <https://www.roche.com/dam/jcr:5260dc48-ffc1-4991-9f3f-0bdae3e42128/en/med-breast-cancer.pdf>

- Rumsfeld, J., Joynt, K., & Maddox, T. (2016). *Big data analytics to improve cardiovascular care: promise and challenges*. Retrieved April 2018, from <http://www.nature.com/nrcardio/journal/v13/n6/abs/nrcardio.2016.42.html>
- Sabharwal, S., Gupta, S., & Thirunavukkarasu, K. (2016). Insight of big data analytics in healthcare industry. *International Conference on Computing, Communication and Automation (ICCCA)*, 95–100.
- Siegel, R.L., Miller, K.D., & Jemal, A. (n.d.). Cancer statistics. *CA Cancer J Clin.*, 66(1), 7–30.
- Stewart In: Bernard W, Wild CP, editors (2014). World cancer report. Lyon: IARC Press, 350–352.
- Tableau (2018). Data visualization and discovery for better business decisions. Retrieved January 2018, from <https://www.tableau.com/learn/webinars/data-visualization-and-discovery-better-business-decisions>
- Tsao, A.S. (2007). Lung carcinoma: tumors of the lungs. *Merck Manual Professional Edition*, 15.
- Ward, M.J., Marsolo, A., & Froehle, C.M/ (2014). Applications of business analytics in healthcare. *Bus Horiz.*, 57(5), 571–582.
- World Cancer Report (2014). World Health Organization. Chapter 1.1 and Chapter 5.1.