INFO 648: Healthcare Informatics

# Assignment 1: Healthcare Datasets Review

For my healthcare dataset, I selected the Big Cities Health Inventory (https://bchi.bigcitieshealth.org/). This dataset was "initially developed by the Chicago Department of Public Health to present epidemiologic data specific to large cities." (Big Cities Health Data Inventory, 2016) and is now produced by the Big Cities Health Coalition. The purpose of this publicly-available dataset is to provide researchers access to a variety of health indicators over time in order to better understand issues which affect a large portion of Americans.

## Part A: What data elements are available?

This dataset features more than 12,000 health data points from 26 large cities, recorded across 39 health indicators and patient demographics ranging from years 2003 to 2014. The indicators selected for this dataset represent prevalent causes of death in the US population, as well as health issues currently impacting many large American cities. There are a total of nine indicator categories: HIV/AIDS, cancer, nutrition/physical activity/obesity, food safety, infectious disease, maternal and child health, tobacco, injury/violence, and behavioral health/substance abuse (Big Cities Health Data Inventory, 2016). The values are shown per 100,000 people. In addition, there are columns for Source, Methods and Notes to place the data in context.

## Part B: Requirements to access the datasets

These datasets are freely available to the public. They are already aggregated and contain no patient information, so there is no need to have a special login to access the information. In addition, the data can be fully accessed at: <https://bchi.bigcitieshealth.org/Government/Big-Cities-Health-Data-Inventory/rwyf-6cby> where it can be filtered and visualized, or exported into a variety of standard formats for further visualization, analysis or aggregation.

## Part C: How do data owners manage privacy?

Privacy is not a concern with aggregated data such as these as there is no identifying information about patients, only values per 100,000 people.

## Part D: Important issues and uses of this dataset

I believe that this data is especially useful and relevant as, according to the US Census Bureau 62% of Americans live in cities (US Census Bureau, 2015), and in addition, urban centers are susceptible to unique health problems that suburban and rural areas are not. Urban centers often have high concentrations of poor people, who are themselves at greater risk for many conditions. In addition, there are increased risks from physical injury from traffic accidents and violence, as well as higher risks from physical inactivity and unhealthy diets due to poverty. These can often lead to greater risk of heart disease, cancer, diabetes and chronic lung diseases (WHO, 2010). Additionally, high concentrations of people living in unhealthy conditions can cause infectious diseases to spread more rapidly. Being able to track trends in demographic groups to observe what diseases are causing illness and death would potentially save many lives.

This dataset is not without its flaws, however. The data is not complete across all time units, indicators or locations, sometimes making it difficult to do comparisons because data won't be available for the same parameters across demographics or locations. Additionally, having the data pre-aggregated limits the analytical options of researchers.

While having more complete data and being able to break down the data into a more granular level would be helpful, this kind of data can still be extremely useful to health policy-makers to set priorities based on trends in morbidity. It offers an easy way to analyze overall trends in health across large population centers in order to spot issues which could use further analysis and research before they turn into crises. It can be a very valuable tool for researchers looking into health patterns for a substantial portion of the American public.

# References

Big Cities Health Data Inventory. (2016). *Search and Browse*. https://bchi.bigcitieshealth.org/browse

US Census Bureau. (2015). *U.S. Cities are Home to 62.7 Percent of the U.S. Population, but Comprise Just 3.5 Percent of Land Area.* Retrieved from <http://www.census.gov/newsroom/press-releases/2015/cb15-33.html>

World Health Organization. (2010). *World Health Day 2010- Frequently Asked Questions.* Retrieved from http://www.who.int/world-health-day/2010/media/whd2010faq.pdf?ua=1