Jack Cone

Professor Odhiambo

Methods in Health Development

March 22, 2022

How Data Science Can Improve the Public Health Field

**ABSTRACT**

The biggest issue with any research project or study is when there is not enough data to make an educated final judgment about the subject matter. It is the responsibility of the researcher to gather all the necessary information and to make sure that they will not extrapolate their findings. Extrapolation is when someone makes an assumption about a subject or community based off of data that may not be associated with them at all. For example, if we gathered 1000 data points on individuals from the ages of 25-45 about heart disease, we could not accurately use that data model to make assumptions about a 65-year-old. This is vitally important when looking at public health issues because what is true for one area or country may be totally different for a different location in the world. This literature review will discuss the issues with data collection for areas that aren’t fortunate enough to have advanced technology. This will cause challenges because many of the easy ways to survey a population involve some sort of technology to make researchers lives so much easier.

**INTRODUCTION**

As we were discussing in class our own individual backgrounds on public health, I started to think about how the Data Science field could improve health around the globe. It is fairly easy to gather information from data in high income countries with all of the technology and cell phones. One of the best ways to gain data is by looking at CDR or call data records. These can track human movement across the country accurately to almost 10 meters. It uses GPS (Global positioning systems) to place coordinates of individuals on maps. This helps so much with disease tracking and finding out how things spread or where it may have originated. Natural disasters are also another huge public health issue. CDR data can find misplaced families and bring them back together even with infrastructure totally demolished. This will ensure that rescue teams know exactly where to go in order to save everyone possible.

Another use of data science in public health involves big data. People use the phrase big data to just mean a large sample or collection of data points. Data way to big for any single person or group of people to analyze by hand. Researchers will use programs such as R or other programming languages like Python and Java in order to sort through all the data and make sense of everything. An example of big data could be census records. There will be loads of data on every individual in a country or smaller county. The census data will be vital to help try and draw conclusions about the people who live in that area. That way, government officials or leading public health officials can make the right decisions that will best fit the community.

This is where the biggest gap in research comes for data science. Since most of the data collection relies on technology, not every country has access to all of these wonderful things so they will not have the same information that other high-income countries do. It could be as simple as sending out a mass survey to computers asking people their vaccination status. So, if a scientists needed this information on a certain country, they would either need to contact doctors themselves in hope that they have records or ask citizens to complete paper surveys and analyze the results by hand. The problem with this is that it can takes days or months to look at hundreds of thousands of written surveys. One thing that scientists can do to mitigate this issue is only polling a portion of the country and then using that sample to make assumptions about the rest of the population. This may lead to some margin of error in the results but, without the proper technology this is I believe the best way to gather accurate information on low-income countries.

**METHODS**

For this literature review the method used for research was gathering articles from many different places. Not only this but trying to find articles that were written at different times. This was important so the research could be complete and contain results from all time periods where opinions in the medical field may vary. Even though data science is a fairly new idea, the way people view it has changed incredibly since its birth. Even within the past few years so much has changed. The data science major did not even exist here when I started as a freshman four years ago. Now we have a full major/minor with loads of people applying to be in the program every year.

I wanted to see how exactly data science was being used in the public health industry or how it can be further used to improve methods that may already be in place. That is why I used the keyword “improve” while scouring databases over the internet. The time spent searching for articles was quite interesting because the data science field can be so broad. There were articles talking about medicine, human migration patterns, viral diseases, and so many more. The point is that there will always be more that needs to be done when looking at how data science can improve public health.

|  |  |  |  |
| --- | --- | --- | --- |
| Database | Date of search | Keyword(s) | Number of publications identified |
| **Google Scholar** | **3/20/2022** | **Data Science, Public Health, Improve** | **5** |
| **PubMed** | **3/20/2022** | **Medicine, big data,**  **Data, machine learning** | **9** |
| **Scopus** | **3/21/2022** | **Epidemiology, Ethics, Databases, Data Science** | **7** |

**DISCUSSION**

In Jeff Goldsmith’s article discussing the future for data science in public health states that, “Public health data science is the study of formulating and rigorously answering questions in order to advance health and well-being using a data-centric process that emphasizes clarity, reproducibility, effective communication, and ethical practices.” This quote embodies what data science can do for a field perfectly. He says that many people will see the data science field in different ways but in reality, it all comes down to who is utilizing the field to form their own opinions. In public health, the biggest problems that data scientists can help with almost strictly involve data. That may seem obvious but data itself can mean so many things. It is the scientist’s job to change that data into usable information to then draw accurate and ethical conclusions. Ethics can be a topic that a lot of people either forget about or turn a blind eye to. The power held by data scientists could be used for alternative motives if people are not practicing in an ethical matter. For example, graphs or charts can easily be skewed by numbers on axis in order to show a false finding. The onus should not fall on to the public to sniff out these false statistics, the responsibility is all on the scientists. They need to have the public in their minds at all times and must never mislead government officials or civilians. The other part of the quote that I think is the most important is the part about effective communication. If there is a study that involves many different types of math or models to produce results, the average person will not understand the results at all unless the findings are communicated well. People will also have trouble listening to things they may not fully understand as we learned these past two years with coronavirus.

Another big reason why data science can help improve public health around the world is how it can make health care so much more efficient. In these countries that do not have the most up to date operating systems in their hospitals, their citizens are suffering because of how long some things will take to do. With the right systems in place, someone can go to a hospital or health clinic and consult with a physician regarding their symptoms and find out what they need to do in order to get healthy within minutes. There would be decreased wait times for patients and increased patients that recover from illnesses because they will be treated earlier before their symptoms reach a dangerous level. There will also be more electronic medical records so that doctors will not lose any records and can update them way easier. Data science will also aid in other methods like disaster surveillance and response. This is a big issue in low-middle income countries because it can be almost impossible to find people or landmarks. But with satellite imaging and machine learning, scientist will be able to aid the countries in need so much quicker.

One thing that people may be worried about is data privacy, especially with their own medical records. This was a real issue in the early 2000’s with the rise of the internet and computers. Starting in 1995 countries like the UK and Australia implemented data privacy policies for large hospitals, but even in 2004 hospitals were having trouble complying with the policies. Nowadays, the internet security has improved so much that it should not be a worry for anyone that is just getting into technology. That is not to say that it can’t happen at all, but the chances are incredibly low. So, state of the art operating systems and databases will be put in place in countries that desperately need them. I think as long as the citizens are told the right information about how helpful the new technology is they will be willing to comply with any surveying or data collection methods. That way researchers can analyze all of their data and find patterns that will lead to health care improvements which is one of the main goals of public health in the first place.

**Works Cited**

DeRenzi, B., Borriello, G., Jackson, J., Kumar, V.S., Parikh, T.S., Virk, P. and Lesh, N. (2011), Mobile Phone Tools for Field-Based Health care Workers in Low-Income Countries. Mt Sinai J Med, 78: 406-418. https://doi.org/10.1002/msj.20256

Goldsmith, Jeff, et al. “The Emergence and Future of Public Health Data Science.” *Frontiers*, Frontiers, 1 Jan. 1AD, www.ssph-journal.org/articles/10.3389/phrs.2021.1604023/full.

Lederman, Reeva. “Managing Hospital Databases: Can Large Hospitals Really Protect Patient Data?” Health Informatics Journal, Sept. 2005, pp. 201–210, doi:10.1177/1460458205055685.

Subrahmanya, S.V.G., Shetty, D.K., Patil, V. *et al.* The role of data science in healthcare advancements: applications, benefits, and future prospects. *Ir J Med Sci* (2021). https://doi.org/10.1007/s11845-021-02730-z

Wang, Weiqi, et al. “Big Data and Clinicians: A Review on the State of the Science.” *JMIR Medical Informatics*, JMIR Publications Inc., Toronto, Canada, medinform.jmir.org/2014/1/e1.