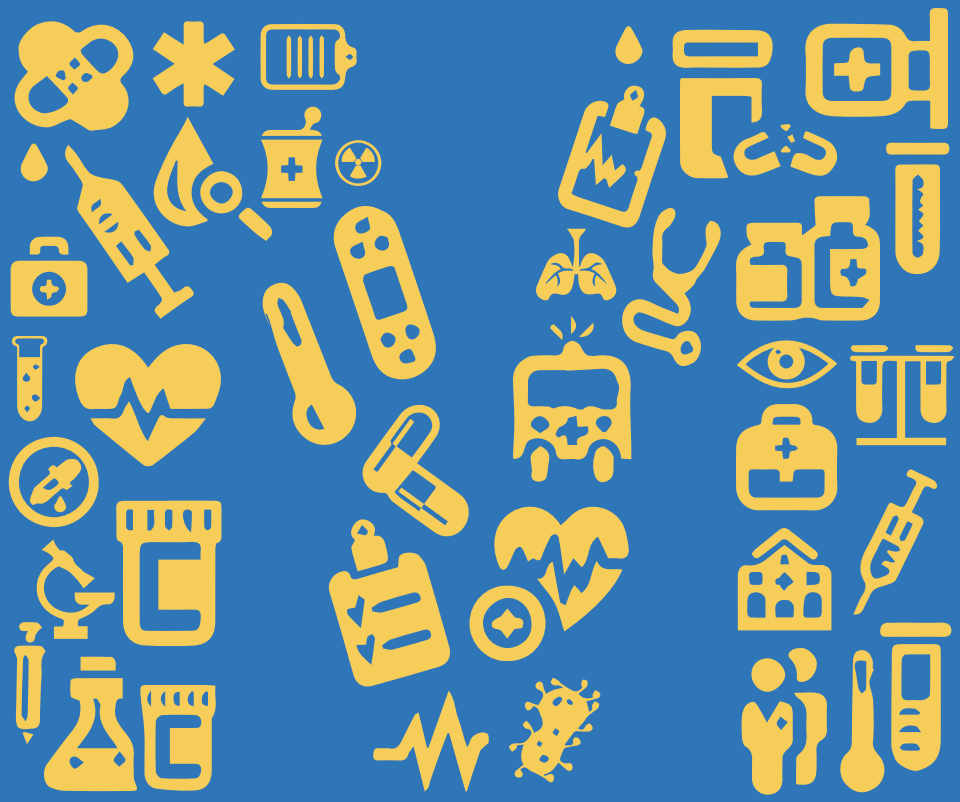
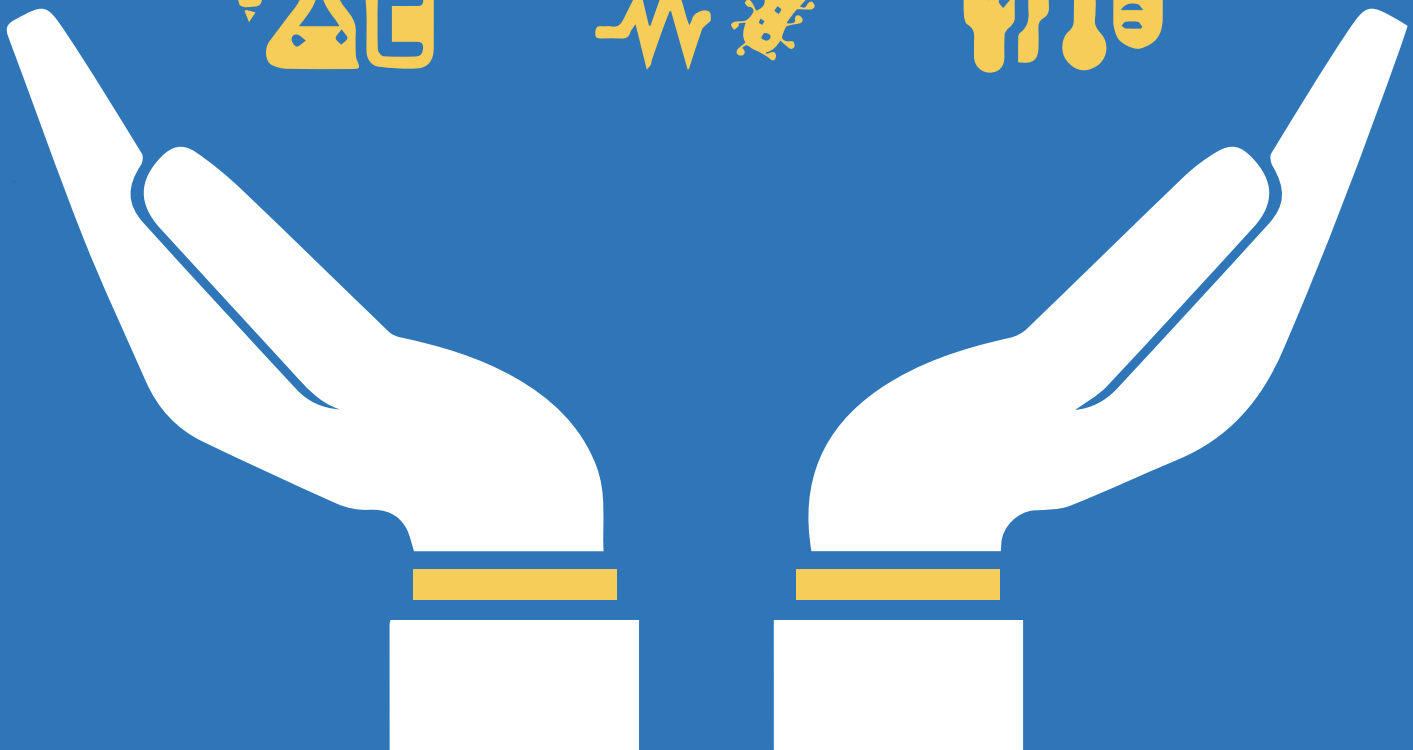
THE UNDERGRADUATE JOURNAL OF PUBLIC HEALTH





The Undergraduate Journal of Public Health

At the University of Michigan

Volume 1—Issue 1

April 2017

The Undergraduate Journal of Public Health

With Support From
University of Michigan School of Public Health
University of Michigan Department of Science, Technology, and Society
Michigan Publishing
University of Michigan Taubman Health Sciences Library

UJPH Editorial Board

Table of Contents

Field Notes

Unintended Consequences of the Bangladesh Tubewell Intervention: A Case Study 1
Neha Tiwari

Independent Research

Indoor Residual Spraying, the AGAMal Program, and Malaria Prevention in Nangodi, Ghana 6
Tara Von Mach, Swati Sudarsan, Anisa Shaw, and Bright Amoore

Literature Review

The Development of Clinical Trials and Safety Determination of HPV Vaccinations for Men 22
Kelly Vermandere

Perspectives

"More of the Same" Not Enough to Reduce U.S. Suicide Rates 29
Matt Dargay

The Effect of Climate Change on Vector-Borne Infectious Diseases 36
Sydney Foote

The Progression, Persistence, and Future of the Opioid Epidemic 43
Merissa Maccani

The Potential of Big Data and Data Science in Public Health Research 53
Nisha Patel

Letter from the Editors

Dear Readers,

It is our pleasure to present to you the inaugural issue of *The Undergraduate Journal of Public Health* at the University of Michigan. Our journey began as a product of passionate conversation in which we both realized there was no forum to independently engage with public health issues. Recognizing this unfulfilled niche on campus, we conceptualized an undergraduate student–run journal, which would create a platform that promotes an active dialogue about relevant issues in public health. We hope to give a voice to a new generation of aspiring physicians, policy makers, and public health professionals.

Although sometimes arduous, this process has been incredibly fulfilling and ultimately humbling. We have been immensely lucky to be surrounded by fellow undergraduates who possess an unparalleled commitment to this publication. The leadership of our executive board, Hannah, Sanjay, and Ajay, has been essential to the establishment of this organization. With their unique perspectives, the five of us were able to refine our vision. Our editorial board is comprised of driven individuals who share our love for public health and are the backbone of the journal. They were vital liaisons between the board and our writers and we couldn't be more grateful for having this particular group of individuals for our inaugural issue. We would also like to thank our PhD student reviewers who took the time to lend their expertise to our editing process. Of course, the journal would not be a reality without the skillful pieces from student writers from Ann Arbor and beyond, which showcase a compelling array of topics.

This process has deepened our understanding of community at the University of Michigan. First and foremost, we would like to thank the School of Public Health for their generous financial support and advice as we crafted this publication. We would also like to thank the Department of Science, Technology, and Society for its contribution to our efforts. Throughout the past year, we formed a close relationship with individuals from Michigan Publishing; the guidance we received from Jason and Allison was integral to the development of this journal. We were also fortunate enough to receive assistance from the informationists at Taubman Health Sciences Library, and resounding support from our undergraduate peers. We hope we can carry this momentum forward for future issues because the creation of this publication is as important as it is timely.

Public health is striking in its pervasiveness; it knows no borders. This allows it to thrive in an increasingly globalized world where societies are an amalgamation of different cultures, races, and identities. Its ability to incorporate perspectives from other fields allows it to transcend conventional ways of understanding health. Therefore, it is increasingly necessary that we, as a community, strengthen our commitment to exploring public health in a global context.

Best Regards,
Krittika Pant and Sonia Ahluwalia
Editors in Chief

PERSPECTIVES

The Potential of Big Data and Data Science in Public Health Research

Nisha Patel

University of Michigan

Abstract:

Although big data already informs and influences the workings of many fields today, it has substantial potential to be applied in the field of public health. Big data refers to large amounts of data that cannot be analyzed with regular data analytics methods. To conduct such analysis, more involved data science methods have come in such as statistical methods, mathematical modeling, and other advanced algorithms, so the large amount of data can be properly mined and analyzed. These new methods of research and investigation in health care can provide new insights in disease management and drug discovery while informing health behaviors and health management. This article advocates for the use of big data when it comes to large data sets in health care such as pharmaceutical data, claims data, and clinical data such as doctors' notes. This also applies to any mobile apps built for public health–related initiatives using which data can be collected from users in real time for analysis. There is immense applicability of big data and analytical methods for handling public health crises on a global scale and for the improvement of health outcomes in low- and middle-income countries, as well as to better inform future public health research.

Keywords: Data Science, Public Health, Research Methods

Introduction

From newspaper articles, scholarly journal articles, blogs, videos, and other sources of information on the Internet and print, there are two buzzwords that have been increasingly featured in many different fields: big data and data science. Even more so, these two words have taken on a larger role in shaping public health research for the future by opening up more possibilities for data collection, data analytics, discovery of treatments for human diseases, and a better understanding of health behavior and management. Furthermore, as with any new innovative method, there are both advantages and disadvantages that characterize the application of big data and data science to research, particularly for public health crises. Despite the few disadvantages that may arise from utilizing big data, its advantages go beyond, making its usage powerful and effective.

Table 1. Sources of Big Data in Health Care.

Type of data	Features of data
Claims and cost data	Reveals cost, type, and amount of care with the potential to help track and identify cost-effective treatments
Clinical data	Consists of doctors' notes and patient medical records, which upon analysis can help identify at-risk individuals and a more comprehensive understanding of diseases
Pharmaceutical data	Consists of data from clinical trials of drugs that can be analyzed to understand participant demographics. It is also possible to use this data to compare drug performance across different trials and studies
Patient behavior and sentiment data	Consists of data from sources such as over-the-counter drug sales and patient behavior as monitored by devices; by understanding patient behavior, this data help gain a better understanding of how devices can be used to understand the population

Big Data, Data Science, and Data Analytics

In simplest of terms, "Big data" refers to large, complex data sets. The data may be "in such unprecedented quantities that terabytes (10^{12} bytes), petabytes (10^{15} bytes), or even zettabytes (10^{21} bytes) of storage may be required" for its storage, beyond the capacity of a single computer by itself (Wyber et al., 2015, p. 203). Not only does this data come in all different sizes but also from a variety of sources, thus supporting the idea that big data analytics can be performed in a variety of different fields. Considering the technology that exists today, such data can also be streamed in real time for analysis. For instance, consider a mobile app developed by a company that tracks a user's weight, diet, and medication. A data analyst at the company can collect in real time (stream) a user's information and analyze it to improve the app's design or better support the user with advice on their diet, and so forth. Thus, big data can be likened to "digital crumbs," which enable humans to have an "unprecedented, ubiquitous and continuous view of our individual lives and behavior: where we live and work, our activity level, travel patterns, shopping habits, what we eat and drink, and which people we interact with" (Pentland, Reid, & Heibeck, 2013, p. 4).

The term "data science" refers to the cleaning, processing, and analysis of data in order to extract insight, patterns, and relationships for variables. This involves a combination of "statistics, mathematics, programming, problem-solving" and more to extract insights from data (Monnappa, 2016). One such method applies machine-learning algorithms, which are predictive in nature and allow for a model that looks at a particular feature in the data to learn from itself. As more data is collected, the model is updated to more accurately reflect the data. Thus, it is natural for big data and data science to be paired together because the complexity of big data sets requires advanced data analytical tools.

According to Bernard Marr, a data scientist who writes for Data Science Central, a community website for individuals involved in big data and data science, there are four types of data from health care that can be analyzed using big data methods. Table 1 details these four types (Marr, 2015).

The data sources described in Table 1 represent the large data sets in public health that have the potential to be analyzed using data science methods for new insights. All of this data regarding patients in hospitals exists across the United States in the form of clinical data, and by tapping into big data and data science, public health research initiatives can be developed to address patient management or bring about new information about health behaviors or diseases. Even the everyday actions of users on a mobile app can be analyzed to track the public's health and give insight into the realm of public health.

Furthermore, not only is examining one data source possible with big data and data science but cross-comparison analysis of multiple, large data sets can also be performed. This is especially valuable for public health. By examining data such as electronic health records, genetic information, and other data sets together, there is greater potential to discover treatments in medicine, uncover patterns in human behavior, search for improvements to reconfigure health systems, or even analyze epidemics to create better prevention measures. Such large amounts of data can be mined using data science methods to uncover relationships and significance that previously would have been undetectable due to statistical constraints and missing information if analyzed by itself. Thus, there are many sources of data in public health that can be powerfully analyzed using big data and data science methods.

Global Efforts in Expanding Use of Big Data in Research Initiative

Using big data has potential in the field of public health, no matter where it is implemented: There is significant potential for their application in low- and middle-income countries. For example, such an approach has changed the epidemic pattern of dengue fever in Lahore, Pakistan, over the past few years. In 2011, the worst outbreak of dengue fever affected the city (Pentland et al., 2013). In 2012 and 2013, however, the combination of smartphones and big data analytics changed this. The development of software created from an open-source repository by the Centers for Disease Control and Prevention (CDC) in the United States enabled early detection of dengue fever once it was modified from its original use to detect outbreaks of flu epidemics (Pentland et al., 2013). The modifications allowed investigators to "identify high-risk areas for infection, and then to aggressively eliminate breeding grounds for the mosquito larvae" (Pentland et al., 2013, p. 25). In this manner, data science has applications that reach beyond the grasp of countries like the United States to help with public health measures instituted in other countries.

On the global scale, there is also a push for implementing data science and using big data for research initiatives, particularly public health–focused ones. The United Nations Global Pulse is a network of innovation labs that work to implement big data practices efficiently across the globe. Set up by the United Nations Secretary General, Global Pulse's mission aims to "accelerate discovery, development and scaled adoption of big data innovation for sustainable development and humanitarian action" (United Nations Global Pulse, n.d.). There are many projects piloted by the United Nations Global Pulse that operate in many different countries around the globe. One such project involves implementing an HIV Mother-to-Child prevention program. This lab is operating in Kampala, Uganda, and aims to use an application to monitor in real time the administration of an antiretroviral treatment that helps protect against HIV in women (United Nations Global Pulse, n.d.). Other projects analyze social media for an increased understanding of public knowledge and their attitudes related to

public health crises such as sanitation. For instance, a lab in Jakarta, Indonesia, is using social media to understand the public's views on immunization, tapping into a search for understanding health behavior in relation to a public health topic of concern (United Nations Global Pulse, n.d.). With such a focus on expanding big data and data science applications in public health on a global scale, there is widespread agreement on the power contained in these methods of data analytics and a push for more application of them in the future.

Disadvantages of Big Data

There are many advantages to using big data and data science methods as outlined above, but there are also a few disadvantages because with more data, there is more potential for errors in analysis, false outcomes, and confounding. False outcomes may surface due to "large-scale examination of putative associations with disease outcomes," where relationships may be discovered between variables that are seemingly significant but in reality insignificant (Khoury & Ioannidis, 2014, p. 1054). Another way to imagine this large-scale data problem is to refer to it as "finding a needle in a haystack" or as "signal and noise" where "separating the true signal from the gigantic amount of noise is neither easy nor straightforward, but it is a challenge that must be tackled if information is ever to be translated into societal well-being" (Khoury & Ioannidis, 2014, p. 1054). However, to combat this, principles of scientific research, statistical design, and refined data analytical methods can be used. Reproducibility of findings is a key factor in determining whether signals are actually signals in fields like genomics (Khoury & Ioannidis, 2014). Moreover, having research teams comprised of specialists in certain areas of statistical analysis, programming methods, and other data science areas can be effective as well. In this manner, the power of big data can be harnessed and false findings avoided. From this, it is clear that the benefits of using big data and data science analytics in public health research outweigh the potential negative consequences if certain procedures are followed and data is analyzed effectively.

Conclusion

Overall, it is clear that using big data and data science methods for data analytics is an emerging, interdisciplinary approach that is increasingly becoming relevant to the field of public health. By applying such an approach toward the handling of public health crises, public health researchers can redefine the way research is conducted. This approach may also result in new discoveries of treatments or provide new insight into health behavior and management. At the global level, the United Nations is pursuing this through innovative labs around the world in all kinds of countries from low income to high. It is clear that the potential of big data and its positive effects on the general public outweigh any negative consequences that may arise from data analysis. Through repeated testing and combination of data sets to confirm new discoveries, big data and data science can lead to more technological innovation and ultimately save lives.

References

Khoury, M. J., & Ioannidis, J. P. A. (2014, November 28). Big data meets public health. *Science*, *346*, 1054–1055. doi:10.1126/science.aaa2709

Marr, B. (2015, January 29). *4 ways big data is transforming healthcare* [Web log post]. Retrieved from http://www.datasciencecentral.com/profiles/blogs/4-ways-big-data-is-transforming-healthcare

Monnappa, A. (2016, April 5). *Data Science vs. Big Data vs. Data Analytics*. Retrieved from https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article

Pentland, A., Reid, T. G., & Heibeck, T. (2013). *Big data and health: Revolutionizing medicine and public health*. Retrieved from https://kit.mit.edu/sites/default/files/documents/WISH_BigData_Report.pdf

United Nations Global Pulse. (n.d.). Available from http://www.unglobalpulse.org/

Wyber, R., Vaillancourt, S., Perry, W., Mannava, P., Folaranmi, T., & Celi, L. A. (2015). Big data in global health: Improving health in low- and middle-income countries. *Bulletin of the World Health Organization*, *93*, 203–208. doi:10.2471/BLT.14.139022