Big Data Training in Global Health Education

Olaniyi Olayinka

Manasi Sheth-Chandra

Muge Akpinar-Elci

Old Dominion University

Author Note

Olaniyi Olayinka, Center for Global Health, Old Dominion University

Manasi Sheth-Chandra, Center for Global Health, Old Dominion University

Muge Akpinar-Elci, Center for Global Health, Old Dominion University

Corresponding Author: Olaniyi Olayinka

The manuscript provided has been read by all the authors. The requirements for authorship have been met, and each author believes the manuscript attached represents honest work.

Correspondence concerning this article should be addressed to Olaniyi Olayinka, Center for Global Health, College of Health Sciences, Old Dominion University

Health Science Bldg, Room 3130, Norfolk VA 23529. Contact: oolayink@odu.edu

**Abstract**

The ability to synthesize and analyze massive amounts of data is critical to the success of organizations including global health. As countries become highly interconnected, increasing the risk for pandemics and outbreaks, the demand for big data is likely to increase. This requires a global health workforce that is trained in the effective use of big data. To assess implementation of big data training in global health, we conducted a survey of members of Consortium of Universities of Global Health. Over half of the respondents did not have big data training program at their institution. Additionally, the majority agreed that big data training program will improve global health deliverables among others. Given the observed gap and benefits, global health educators may consider investing in big data training for students seeking a career in global health.

*Keywords*: big data, global health, education, public health

# 

# **Big Data Training in Global Health Education**

# **Introduction**

# The ability to collect, synthesize, and analyze massive amounts of data obtained from multiple sources, is critical to the success of any organizational endeavor (Brown, Chui, & Manyika, 2011).Over the past decade, interest in the potential use of big data has increased (Hay, George, Moyes, & Brownstein, 2013; Landefeld, 2014; Michael & Miller, 2013; The-White-House, 2014; Wyber et al., 2015; Zavazava, 2015).For instance, the United Nations’ International Telecommunication Union (ITU) supported Ebola response activities in 2014 by using mobile telecommunication data to track the outbreak (Zavazava, 2015). From national security experts to those working in healthcare and global health, there is the realization that big data will potentially transform various aspects of human lives (Bates, Saria, Ohno-Machado, Shah, & Escobar, 2014; Chen, Chiang, & Storey, 2012; Kayyali, Knott, & Van Kuiken, 2013; Michael & Miller, 2013; Wyber et al., 2015).

# More precisely, big data are “large, diverse, complex, longitudinal, and/or distributed datasets generated from instruments, sensors, internet transactions, email, video, click streams, and/or all other digital sources available today and in the future” (The-White-House, 2014). Given that the repository of potentially useful data is rapidly expanding, in addition to efforts to improve transparency in analysis, reporting, and use of global health data (Bollyky, 2013), public health professionals may benefit by leveraging big data for the global good.

# As countries become highly interconnected because of globalization, infectious diseases that were considered local or regional are now emerging and reemerging worldwide (Fineberg & Hunter, 2013). Additionally, the global demographic and epidemiological transitions are projected to increase morbidity and death from injuries and noncommunicable diseases in the future (Wyber et al., 2015). In order to address these and other global health issues, education in data science and especially use of big data is important (Chen et al., 2012; Hay et al., 2013; Savin-Baden, 2015; Wyber et al., 2015; Zavazava, 2015). The aim of this article is to draw the attention of global health educators to the need to introduce the concept of big data to students early in their global health career. The objective was to assess the opinion of educators in global health on (1) whether students need knowledge and training for big data concepts, and if yes, (2) whether this training program would establish collaborative platforms between various disciplines as well as between private and public institutions, and (3) whether training in big data will improve future global health workforce.

# **Methods**

# In this cross sectional study, we collected data via an online survey instrument. A review of the extant peer-reviewed literature indicated that no prototype surveys for investigating big data training needs in global health education exist. In order to assess whether big data training was essential for implementing in global health, a qualtrics survey was established, for which the results would determine the global health education needs and its intervention of the knowledge gap.

# After developing the survey, the study team convened an expert panel consisting of global health educators, biostatisticians, and “big data” scientists from different institutes to evaluate the survey instrument. Subsequently, the study team piloted the survey instrument locally to test its feasibility. The study was approved by the University’s Institutional Review Board. Our target population were members of Consortium of Universities of Global Health (CUGH), a Washington, DC based organization of academic institutions and other organizations from around the world engaged in addressing global health challenges. CUGH was established in 2008 with generous funding from the Bill & Melinda Gates Foundation and The Rockefeller Foundation. No other inclusion or exclusion criteria was used. An invitation to participate in the study was sent to all members of CUGH (n: 115). No other inclusion or exclusion criteria was used.

# From October 2014 through February 2015, the survey was administered. To minimize non-response, participants were contacted, via electronic mail, four times, and the study period extended by one month. In order to ensure privacy protection of participants, all data obtained from the survey were de-identified, so that the researchers never knew members’ identities. Additionally, each participant was assigned a random number using a combination of random number generating algorithm and their current identification number. Data were collected via the survey instrument and analyzed using Microsoft Excel.

## **Results**

## Fourteen respondents (participation rate = 12.2%) completed the survey (Figure 1.). Of the 14 respondents, 69.2% worked in university, 7.7% in public and private institution each, and 23.1% worked in non-profit organization (Table 1.). Nine respondents (64.3%) indicated that their institution did not have formal training for students in the use of big data; five respondents (35.7%) were unsure (Figure 2.). With regards to training in the use of big data for global health, however, 13 (92.9%) of the 14 respondents agreed that training students in big data will enhance global health management and policies governing the use of big data, improve efficiency of global health deliverables, and expand the scope of the prospective global health workforce in the future. Twelve respondents (85.7%) agreed that such training will allow for collaborative platforms between industry and academic partners in global health as well as from different disciplines (Table 1). Three of the respondents provided additional feedback regarding other interest areas where training in the use of big data would beneficial. These areas include global health management, public health professionals in government, education and health, and for infectious disease specialists.

## **Discussion**

## Although the response-rate for this study was low, the respondents’ perspective underscores the need for big data training for students in global health. Overall, the respondents recognize the useful of big data and its potential application in global health. Big data knowledge is predicted to arm global health professionals with the skills required to work with large data sets and design health systems critical to global health (Chen et al., 2012; Hay et al., 2013). For instance, innovative disease surveillance systems that integrate massive, open source data to inform timely, global-public health actions have emerged over the last two decades. Popular platforms that use “big data” for surveilling diseases and health-related events include Ushahidi, HealthMap, and more recently, EpiCore (Brownstein, Freifeld, Reis, & Mandl, 2008; Freifeld et al., 2010; Haddad et al., 2016).

## In health care, big data will play a major role in improving the quality of patient care and reducing the cost of care (Bates et al., 2014). However, only a few understand what big data is, and even fewer leverages big data for public-health decision making. One way to bridge the gap between knowledge of big data and practice is to encourage formalized education that highlights the potentials of big data use in global health (Najafabadi et al., 2015; Peter, Basel, David, & Steven, 2013; Picciano, 2012; Wyber et al., 2015).The essence is for students to learn how to accurately and efficiently integrate data obtained from different traditional and non-traditional sources.

## Competency in data analytics and the use of big data would also foster collaborative relationships between public health professionals in the formal and informal sectors. This is critical for the global health workforce to handle complex or unfamiliar data formats. In countries where electronic health records are available and well-developed, the huge data can be combined with other informal sources of health data to monitor risk factor and non-communicable disease trends. For instance, the ability to search and analyze health-related data obtained via web-based surveillance of diseases have been shown to complete formal disease surveillance as well expedite the timely detection of an outbreak (Carneiro & Mylonakis, 2009).

## Big data, however, pose some challenges including the resources required to track, store, analyze, interpret, disseminate, and maintain confidentiality and privacy of health-related data (Heitmueller, Henderson, Warburton, Elmagarmid, & Darzi, 2014; Wyber et al., 2015). Of note, the inextricable link between lifestyle and health makes it difficult to separate personal data from health-related raising some privacy questions. While these challenges exist, they serve as an opportunity for global health experts to develop strategies that address them. For instance, ITU focuses on individual privacy in its innovative use of big data to track disease outbreaks globally (Zavazava, 2015). Similar efforts can be replicated to reduce suspicion and increase the acceptability of big data for decision-making in global health. A limitation of this study is the small number of respondents highlighting the need for a similar but larger study in the future. Additionally, using a conceptual framework for this study would be valuable and will be considered in the future. To the knowledge of the authors, however, this is the first study to investigate the need for big data training in global health education.

## **Conclusions**

## Although the participation rate was low, our result underscores the need for education and training in big data for global health practice. The importance of the potential of big data in health care and education has just begun. There are challenges that need to be considered when developing global health policy about the use of big data to enhance the health of populations. However, the benefits of big data such as to eradicate health conditions and to take preventive measures against epidemics should not be underestimated. We believe that this study would help to draw the attention of global health educators to the need to introduce the concept of big data to students early in their global health career

**References**

Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Affairs, 33*(7), 1123-1131.

Bollyky, T. (2013). Big Data, Better Global Health. *Expert Brief-Noncommunicable Diseases. Council on Foreign Relations*.

Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of ‘big data’. *McKinsey Quarterly, 4*(2011), 24-35.

Brownstein, J. S., Freifeld, C. C., Reis, B. Y., & Mandl, K. D. (2008). Surveillance Sans Frontieres: Internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Med, 5*(7), e151.

Carneiro, H. A., & Mylonakis, E. (2009). Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clinical infectious diseases, 49*(10), 1557-1564.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS quarterly, 36*(4), 1165-1188.

Fineberg, H. V., & Hunter, D. J. (2013). A global view of health—an unfolding series. *New England Journal of Medicine, 368*(1), 78-79.

Freifeld, C. C., Chunara, R., Mekaru, S. R., Chan, E. H., Kass-Hout, T., Iacucci, A. A., & Brownstein, J. S. (2010). Participatory epidemiology: use of mobile phones for community-based health reporting. *PLoS Med, 7*(12), e1000376.

Haddad, Z., Madoff, L., Cohn, E., Olsen, J., Crawley, A., Brownstein, J., . . . Herrera-Guibert, D. (2016). The EpiCore Project: Using innovative surveillance methods to verify outbreaks of emerging infectious diseases. *International Journal of Infectious Diseases, 45*, 19.

Hay, S. I., George, D. B., Moyes, C. L., & Brownstein, J. S. (2013). Big data opportunities for global infectious disease surveillance. *PLoS Med, 10*(4), e1001413.

Heitmueller, A., Henderson, S., Warburton, W., Elmagarmid, A., & Darzi, A. (2014). Developing public policy to advance the use of big data in health care. *Health Affairs, 33*(9), 1523-1530.

Kayyali, B., Knott, D., & Van Kuiken, S. (2013). The big-data revolution in US health care: Accelerating value and innovation. *Mc Kinsey & Company*, 1-13.

Landefeld, S. (2014). *Uses of Big Data for Official Statistics: Privacy, Incentives, Statistical Challenges, and Other Issues.* Paper presented at the International Conference on Big Data for Official Statistics, United Nations Statistics Division and National Bureau of Statistics of China.

Michael, K., & Miller, K. (2013). Big data: New opportunities and new challenges [guest editors' introduction]. *Computer, 46*(6), 22-24.

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data, 2*(1), 1-21.

Peter, G., Basel, K., David, K., & Steven, V. K. (2013). The ‘big data’ revolution in healthcare: Accelerating value and innovation. Retrieved June 2, 2016, from file:///F:/Big%20Data/The\_big\_data\_revolution\_in\_healthcare.pdf

Picciano, A. G. (2012). The Evolution of Big Data and Learning Analytics in American Higher Education. *Journal of Asynchronous Learning Networks, 16*(3), 9-20.

Savin-Baden, M. (2015). Education and Big Data.

The-White-House. (2014). Big Data: Seizing Opportunities, Preserving Values. Retrieved June 2, 2016, from <https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf>

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*(3), 203-208.

Zavazava, C. (2015). How Big Data Will Help Fight Global Epidemics. Retrieved June 2, 2016, from <https://itu4u.wordpress.com/2015/10/13/how-big-data-will-help-fight-global-epidemics/>

Table 1: Participant’s job title and type of institution (n= 14)

|  |  |
| --- | --- |
| **Job Titles** | **n (%)** |
| Program Directors | 4 (29) |
| Scientist | 4 (29) |
| President/Vice President | 3 (21) |
| Faculty | 2 (14) |
| Administrator | 2 (14) |
| Supervisor/Program Managers | 1 (7) |
| **Institutions** |  |
| University | 9 (69) |
| Public Institution | 1 (8) |
| Private Institution | 1 (8) |
| Non-profit organization | 3 (23) |

Table 2: Participants’ response to key questions on the benefits for big data training program

|  |  |  |  |
| --- | --- | --- | --- |
| **Questions** | **Strongly Agree or Agree**  **n** | **Neutral**  **n** | **Strongly Disagree or Disagree**  **n** |
| Big Data Training Program will help enhance the global health management and policies governing big data | 13 | 0 | 0 |
| Big Data Training program will improve efficiency of global health deliverables | 13 | 0 | 0 |
| Big Data Training program will improve student retention by providing them necessary skills of their respective employer | 11 | 2 | 0 |
| Big Data Training program will offer an ease of transition for graduate students into their future employment | 11 | 3 | 0 |
| Big Data Training program will expand the scope of the prospective global health workforce in the future | 13 | 0 | 0 |
| Big Data Training program will allow for collaborative platforms between industry and academic partners in global health as well as from different disciplines | 12 | 1 | 0 |

Figure 1: Geographic distribution of participants in the big data training survey



Figure 2. Participants' response to two key questions on big data training