Annals of Global Health
VOL. 83, NO. 3–4, 2017

ISSN 2214-9996/$36.00
https://doi.org/10.1016/j.aogh.2017.09.005

VIEWPOINT

Big Data Knowledge in Global Health Education

Olaniyi Olayinka, MBChB, MPH, Michele Kekeh, PhD, Manasi Sheth-Chandra, PhD, Muge Akpinar-Elci, MD, MPH
Norfolk, VA

Abstract

The ability to synthesize and analyze massive amounts of data is critical to the success of organizations, including those that involve 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 pilot survey of members of the Consortium of Universities of Global Health. More than half the respondents did not have a big data training program at their institution. Additionally, the majority agreed that big data training programs will improve global health deliverables, among other favorable outcomes. Given the observed gap and benefits, global health educators may consider investing in big data training for students seeking a career in global health.

KEY WORDS big data, education, global health, pilot study, public health.

INTRODUCTION

Big data refers to "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."[1p3] Wakim[2] refers to big data as "large, complex data that are difficult to understand using traditional data analysis techniques." 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,[3] public health professionals may benefit by leveraging big data for the global good. For example, public health institutions in the United States and across Europe are using big data to augment traditional public health surveillance systems and to inform the treatment of cancers.[4] On the world stage, large datasets are also used to determine the global burden of diseases and estimate global health.[5] However, analyzing and interpreting large datasets for use in global health require professionals with training in big data management and population health.

The ability to collect, synthesize, and analyze massive amounts of data obtained from multiple sources is critical to the success of any organizational endeavor.[6] Over the past decade, interest in the potential use of big data has increased.[1,7-11] For instance, the United Nations' International Telecommunication Union supported Ebola response activities in 2014 by using mobile telecommunication data to track the outbreak.[11] From national security experts to those working in health care and global health, there is the realization that big data will potentially transform various aspects of human lives.[9,10,12-14]

As countries become highly interconnected because of globalization, infectious diseases that were considered local or regional are now emerging and reemerging worldwide.[15] Additionally, the global demographic and epidemiological transitions are projected to increase morbidity and death from injuries and noncommunicable diseases in the future.[10] To address these and other global health issues,

Conflict of Interest: The authors state no conflict of Interests. The results of this study have not been fully or partially published in or submitted to any other printed or electronic publication, in any language.
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.
From the Center for Global Health, Old Dominion University, Norfolk, VA. Address correspondence to M. Kekeh (mkekeh@odu.edu).

education in data science and especially use of big data is important.[7,10,11,13] A review of the extant peer-reviewed literature indicated that no prototype surveys for investigating big data training needs in global health education exist. This study was conducted to collect preliminary data about 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. This study is an initial step in exploring big data training in global health.

METHODS

The study was approved by the Old Dominion University's Institutional Review Board. The target population was members of the 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. The method used in this study is a combination of expert reviews and pilot testing.

Expert Panel Reviews. To assess whether big data training was essential for implementing in global health, a survey instrument was developed. Subsequently, 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. A total of 15 experts from different academic institutions in and outside the United States were invited to participate in the instrument review. Additionally, experts were selected according to their publications on global health education in the last 5 years. Nine experts responded to the invitation and constituted the panel for the instrument review. These experts only reviewed the instrument to determine whether the survey questions essentially captured the topic under exploration; they were not among the study participants in this pilot. All the experts performed their review individually, and the identity of each reviewer was kept confidential by the search team.

Description of the Instrument. During the development of the survey instrument, the research team made sure the questionnaire included items that would accurately address the research questions. Because the survey would be self-administrated, the research team ensured that questions were clear and presented in a manner that was consistent. Each question was designed to measure 1 concept at a time. The response options were chosen to align with the question items. Response categories were mutually exclusive and comprehensive.

The instrument was divided into 4 categories: (1) the characteristics of participants and their institutions (questions 1-3), (2) the opinion of educators in global health as it relates to students' training in big data (questions 4-7), (3) big data training as collaborative platforms between various disciplines (questions 8 and 9), and (4) big data and the global health workforce (questions 10-12).

Selection of Participants. An invitation to participate in the study was sent to all the members of CUGH (n = 115). No other inclusions or exclusions criteria were used. A consent form was sent to all members; participants indicated their agreement by signing the consent form.

From October 2014 through February 2015, the survey was administered. To minimize nonresponse, participants were contacted, via electronic mail, 4 times, and the study period extended by 1 month. To ensure privacy protection of participants, all data obtained from the survey were deidentified, 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

The experts indicated that the instrument was relevant and included items that were essential to big data education in global health. Experts also specified that the questions were clearly worded, short, and precise. They also indicated that the instrument was pertinent to the study population.

Among the 115 CUGH members invited to participate in the pilot study, 14 members (participation rate = 12.2%) completed the survey (Table 1, Fig. 1). Of the 14 respondents, 69.2% worked in universities, 7.7% in public and private institutions, and 23.1% in nonprofit organizations. Nine respondents (64.3%) indicated that their institution did not have formal training for students in the use of big data; 5 respondents (35.7%) were unsure (Fig. 2). With regard to training in the use of big data for global health, however, 13 (92.9%) of the 14 respondents agreed

Annals of Global Health, VOL. 83, NO. 3–4, 2017
May–August 2017: 676–681

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)
Nonprofit organization	3 (23)

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 2). Three of the respondents suggested the following as aspects of global health that would benefit from education in big data: *(1) research in global health management; (2) policies guiding the use of government data, education data and health data and how to use the data to create applications for commercial and public use; (3) how to best understand and learn the rules of the government and how to operate within them.*

DISCUSSION

Big data knowledge is predicted to arm global health professionals with the skills required to work with large datasets and design health systems critical to global health.[7,13] For instance, innovative disease surveillance systems that integrate massive, open source data to inform timely, global health actions have emerged over the last 2 decades. Popular platforms that use "big data" for surveilling diseases and health-related events include Ushahidi, HealthMap, and, more recently, EpiCore (Table 3).[16-18]

In health care, big data will play a major role in improving the quality of patient care and reducing the cost of care.[12] However, only a few understand

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

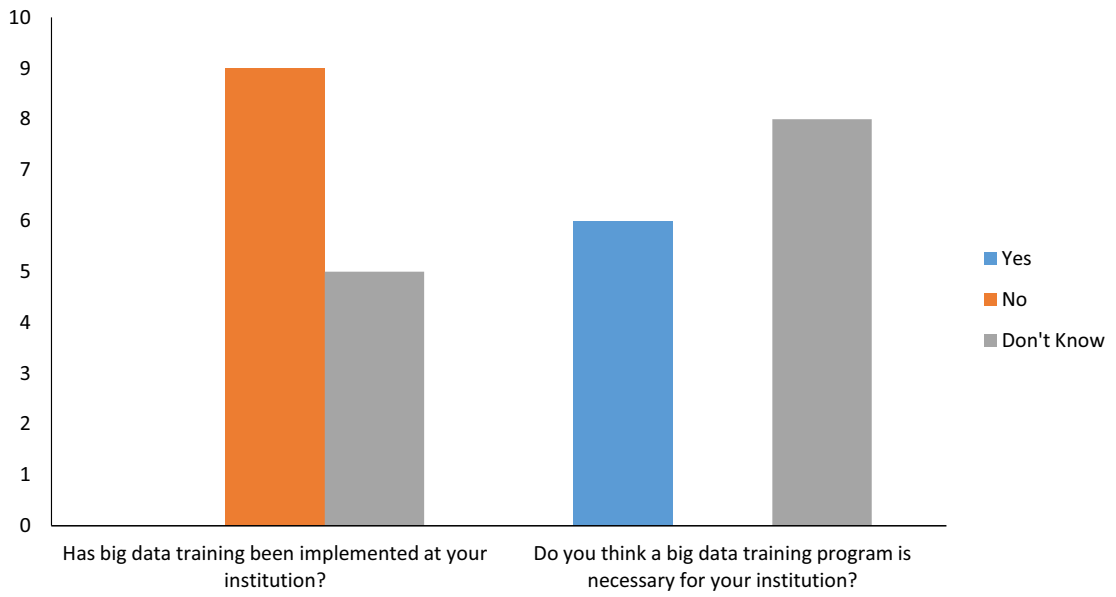

Figure 2. **Participants' responses to 2 key questions on big data training.**

what big data are, and even fewer leverage 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.[10,19-21] The essence is for students to learn how to accurately and efficiently integrate data obtained from different traditional and nontraditional sources.

Table 2. **Participants' Responses to Key Questions on the Benefits of 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

Table 3. **Example of Surveillance Systems Using Big Data to Improve Global Health**

Surveillance Systems Using Big Data	Description
EpiCore https://epicore.org/#/about	EpiCore is an innovative global disease surveillance system run by a global network of health professionals. EpiCore aims to eliminate the time lag in outbreak detection and reporting by augmenting traditional public health surveillance systems.
HealthMap http://www.healthmap.org/site/about	Established in 2006 by Boston Children's Hospital, HealthMap uses information obtained from the internet to track outbreak of diseases outbreak in real time.
Ushahidi https://www.ushahidi.com/about	Initially developed to crowd-map reports of violence related to the 2007-2008 presidential election in Kenya, Ushahidi has leveraged its global data collection, analysis, and information distribution platform for disaster-related surveillance.

Annals of Global Health, VOL. 83, NO. 3–4, 2017
May–August 2017: 676–681

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 noncommunicable disease trends. For instance, the ability to search and analyze health-related data obtained via Internet-based surveillance of diseases have been found to complete formal disease surveillance as well expedite the timely detection of an outbreak.[22]

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; this is because most of the data are obtained via informal sources, including the Internet and biosensors.[10,23] Of note, the inextricable link between life-style and health makes it difficult to separate personal data from health-related data, raising some privacy questions. Although these challenges exist, they serve as an opportunity for global health experts to develop strategies that address them. For instance, the International Telecommunication Union focuses on individual privacy in its innovative use of big data to track disease outbreaks globally.[11] Similar efforts can be replicated to reduce suspicion and increase the acceptability of big data for decision making in global health. 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

This study was designed to examine the feasibility of including big data training in global health education. Although there is anecdotal evidence from this pilot study suggesting that education in big data will be beneficial in global health, the results of this study cannot be generalized beyond the scope of this pilot design. Also, the small size of the sample offers incomplete insight into the validity and reliability of the instrument. Therefore, the authors call for a similar study involving a larger sample size. The potential of big data to transform health care and education is likely to be huge, particularly for global disease surveillance. However, leaders in global health are likely to face some challenges including the development of policy governing the use of big data to enhance the health of populations. These are some areas of big data that may be the focus of educators in global health. Ultimately, this pilot is a call for global health educators to consider the usefulness of "big data education" for the 21st-century students in global health.

REFERENCES

1. The White House. Big data: seizing opportunities, preserving values. Washington, DC: The White House; 2014. Available at: https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf. Accessed September 4, 2015.
2. Wakim N. Big data and graduate education in the United Arab Emirates: challenges, trends and perspectives. In: Proceedings of the Ninth Annual Strategic Global Leaders Summit. Singapore: National University of Singapore; 2015:28-30. Available at: http://cgsnet.org/ckfinder/userfiles/files/book_2015_Global_Summit_booklet_Final.pdf. Accessed May 25, 2016.
3. Bollyky T. Big Data, Better Global Health. Expert Brief—Noncommunicable Diseases. New York, NY: Council on Foreign Relations; 2013.
4. Salas-Vega S, Haimann A, Mossialos E. Big data and health care: challenges and opportunities for coordinated policy development in the EU. Health Syst Reform 2015;1:285–300.
5. World Health Organization. A commitment to improve global health information. Geneva, Switzerland: World Health Organization; 2015. Available at: http://www.who.int/mediacentre/news/releases/2015/health-metrics-evaluation/en/. Accessed June 8, 2016.
6. Brown B, Chui M, Manyika J. Are you ready for the era of "big data". McKinsey Q 2011;4:24–35.
7. Hay SI, George DB, Moyes CL, Brownstein JS. Big data opportunities for global infectious disease surveillance. PLoS Med 2013;10:e1001413.
8. Landefeld S. 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. October 28-30, 2014; Beijing, China.
9. Michael K, Miller K. Big data: new opportunities and new challenges [guest editors' introduction]. Computer 2013;46:22–4.
10. Wyber R, Vaillancourt S, Perry W, Mannava P, Folaranmi T, Celi LA. Big data in global health: improving health in low-and middle-income countries. Bull World Health Org 2015;93:203–8.
11. Zavazava C. How big data will help fight global epidemics. Geneva, Switzerland: itublog; 2015. Available at:

https://itu4u.wordpress.com/2015/10/13/how-big-data-will-help-fight-global-epidemics/. Accessed June 7, 2016.
12. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Aff 2014;33:1123–31.
13. Chen H, Chiang RH, Storey VC. Business intelligence and analytics: from big data to big impact. MIS Q 2012; 36:1165–88.
14. Kayyali B, Knott D, Van Kuiken S. The Big-Data Revolution in US Health Care: Accelerating Value and Innovation. New York, NY: McKinsey & Company; 2013:1–13.
15. Fineberg HV, Hunter DJ. A global view of health—an unfolding series. N Engl J Med 2013;368:78–9.
16. Brownstein JS, Freifeld CC, Reis BY, Mandl KD. Surveillance Sans Frontieres: internet-based emerging infectious disease intelligence and the HealthMap project. PLoS Med 2008;5:e151.
17. Freifeld CC, Chunara R, Mekaru SR, et al. Participatory epidemiology: use of mobile phones for community-based health reporting. PLoS Med 2010;7: e1000376.
18. Haddad Z, Madoff L, Cohn E, et al. The EpiCore project: using innovative surveillance methods to verify outbreaks of emerging infectious diseases. Int J Infect Dis 2016;45:19.
19. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. J Big Data 2015;2: 1–21.
20. Peter G, Basel K, David K, Steven VK. The 'big data' revolution in healthcare: acceleration value and innovation, McKinsey & Company. New York, NY: McKinsey & Company; 2013. Available at: http://200.25.59.71:8081/jspui/bitstream/11146/465/1/1661-The_big_data_revolution_in_healthcare.pdf
21. Picciano AG. The evolution of big data and learning analytics in american higher education. J Asynchronous Learn Netw 2012;16:9–20.
22. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 2009;49:1557–64.
23. Heitmueller A, Henderson S, Warburton W, Elmagarmid A, Darzi A. Developing public policy to advance the use of big data in health care. Health Aff 2014;33:1523–30.