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,

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-administered, 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