

Is there a need for graduate-level programmes in health data science? A perspective from Pakistan



Historically, health care has been at the vanguard of developing and adopting evidence-based approaches to addressing complex medical problems.¹ However, despite a large amount of information it produces, the health-care industry has been slow to embrace technological advances and the data revolution.² These vast amounts of data can be used for personalised medicine, improving access, quality of care, and health equity.³ Informed decision making with high-quality data is the foundation of making health-care delivery efficient and cost-effective. With a forecasted revenue growth of 2.3%, health care is the second largest industry in which data analytics can help to reduce health-care costs.⁴

To capitalise on these benefits, low-income and middle-income countries (LMICs) have also been gradually adopting health-care technology, such as developing electronic health records and mobile health applications. For example, novel analytical techniques (eg, machine learning) are being applied to identify high-risk pregnancies through clinical and ultrasound data acquired in underserved communities in Pakistan, highlighting the potential of data tools for task shifting in settings with low resources.⁵ In Africa, although there have been efforts to use data analytic approaches to inform health care since the 1980s, most of the efforts have been small pilot studies that show artificial intelligence to hold a lot of promise for enhancing health care.⁶

Although the potential impact of big data on health care in LMICs could be substantial, its adoption has lagged. Unlike the digital ecosystems in the business community, health-care systems are still largely paper-based with a scarcity of technological integration and analytics to quantify the quality of care. To improve the preparedness, efficiency, and effectiveness of health care provision not only under normal conditions but also in crisis situations, health-care services need to be re-engineered with technical solutions. The National Digital Health Framework of Pakistan published in November, 2021, provides a roadmap for ensuring “availability, integration, and adaption of digital health solutions for an improved and strengthened healthcare

delivery system by 2030”.⁷ This detailed plan highlights the willingness of the Government of Pakistan to invest in the digital health space. However, implementation is hindered by limited infrastructure and human resource capacity for implementing these technologies to advance the digital health-care ecosystem in Pakistan.⁷

McKinsey Global Institute predicted that, by 2018, there would be a dearth of approximately 1.5 million trained individuals in the USA who would be able to drive the data revolution. Although Pakistan produces approximately 20 000 information technology and engineering graduates annually, less than 20% have gained skillset-appropriate employment,⁸ and those who are well trained search for better prospects outside of the country contributing to the so-called brain drain. The World Development Report 2021 reported an out-migration of 70 000 information and communication technology specialists in 2015–19 from LMICs to high-income countries (HICs).⁹ This out-migration might be offset by providing high-quality training and employment opportunities within the region.

Health data science requires a composite skill set, including knowledge of and ability to work with health data, mathematics, statistics, and computer science. Responsive to the needs of the future, HICs are offering comprehensive training programmes; however, to the best of our knowledge, none of these programmes have been offered in LMICs. The existing courses in HICs focus on knowledge and skills for research study designs, health management, and data analytics, including machine learning. Such specialised degree programmes represent the future training paradigm to leverage the existing health workforce through data-driven decision making. The absence of such training programmes in LMICs and an adequately trained health data science workforce worsens the divide between existing data and its use for decision making.

With the current global shortage of approximately 15 million health-care providers, with the largest shortage from southeast Asia,¹⁰ digital tools (eg, artificial intelligence) augment the limited human resources and capacity to provide high-quality, efficient, and essential health-care services to the most marginalised

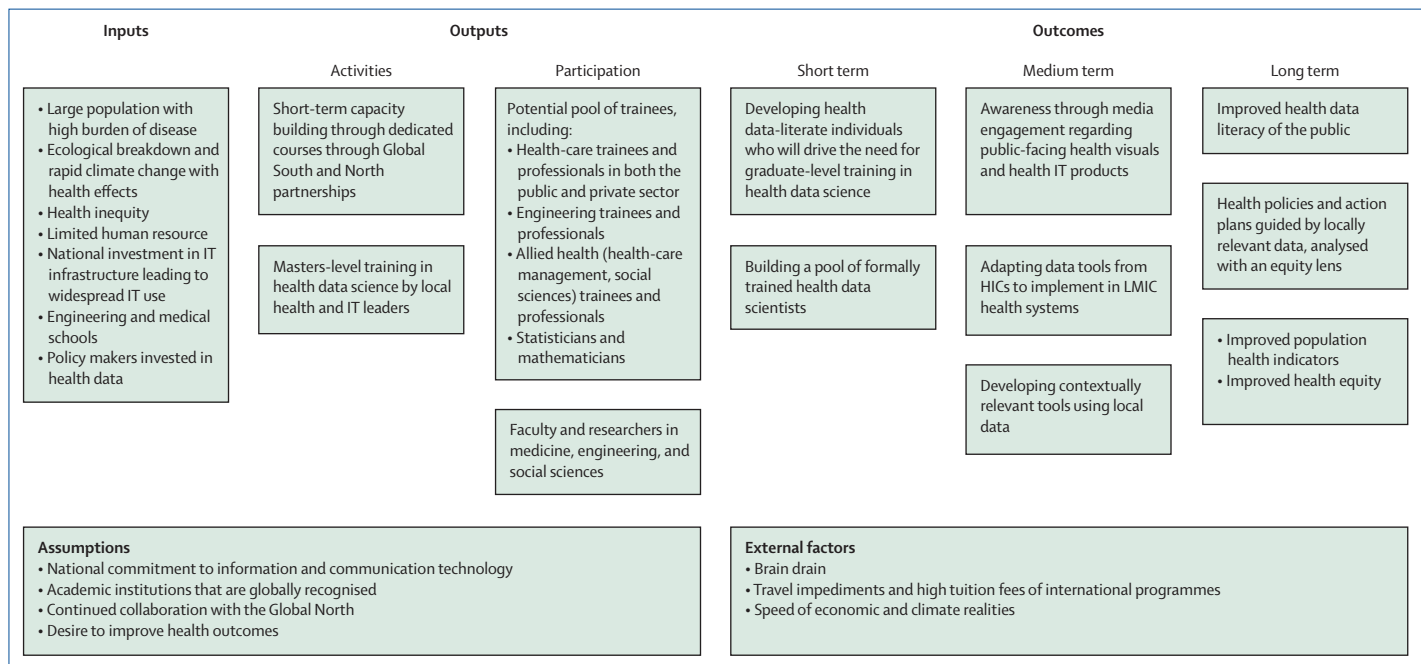


Figure: Theory of change
IT=Information technology.

populations. However, the realisation of this opportunity requires the development of algorithms that utilise local data, are contextually relevant, and thus have to be developed by individuals who can respond to or who have experience with these needs, understand the context and equity issues, and can use both health and data science proficiently.

Building local capacity for health data science in LMICs will require a multi-pronged approach. Efforts for a collaborative programme between engineering and medical and public health schools in the Global North and Global South might act as a gateway to draw domain and contextual knowledge, as well as experience, between the regions. Besides the North-South collaborations, support from HICs for South-South collaborations between relevant institutions is also one approach to create the push for the health data revolution for contextual health problems. The development of local training programmes in health data science would be able to leverage academic medical centres with local expertise to drive contextually relevant yet insightful decisions regarding individual or population health pertinent parameters to LMICs. A degree programme in isolation is not the universal solution for health issues in LMICs. Instead, the technological revolution, if fully availed

through a digitally literate population, would act as a catalyst in achieving sustainable development goals. In our opinion, the inputs, activities, and outcomes of these local training programmes have been illustrated as a theory of change (figure).

The momentum towards creating technological solutions for health data to be curated and analysed and yield insights has been increasing and thus improving health outcomes. In LMICs where disease burden is high and resources are scarce, clinicians and public health specialists need to be empowered with a thorough understanding of data for appropriate decision making. The lack of a specialised graduate-level programme in low-resource settings indicates the dearth of trained individuals to generate these insights. Training and capacity building through structured degree programmes is essential to achieve this target, along with a demand for health data scientists in the public and private health sectors. The balance between the demand and supply of well trained professionals would help to reduce the gap between knowledge and skills for data science and its application to improve health outcomes.

ZH received support from the Bill & Melinda Gates Foundation, VITAL Pakistan Trust, and Islamic Development Bank. RC received support from the National Science Foundation (NSF 1845487), the National Institutes of Health (NIH U2RTW012121), Blue Cross and Blue Shield of Louisiana via a grant awarded to New York University (New York, NY, USA), and from Arizona State

University (Tempe, AZ, USA) to attend the Public Health Technology Meeting in collaboration with New America Foundation. ZS received support from the Bill & Melinda Gates Foundation, Fogarty International Center of the National Institutes of Health (INV-021944, D43TW011625), the UK National Institute for Health and Care Research, Duke University (Durham, NC, USA), Certara, and the Population Health Research Institute. ZS also received support as an advisor for participation in the Heart Failure Publication Committee meeting from Cytokinetics. AW and AAB declare no competing interests. This work is part of a larger grant funded by Bill & Melinda Gates Foundation. The funding agency had no role in the development of this Comment.

Copyright © 2022 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

Zahra Hoodbhoy, Rumi Chunara, Akbar Waljee, Amina AbuBakr, *Zainab Samad samad.zainab@aku.edu

Department of Pediatrics and Child Health, Aga Khan University, Karachi, Pakistan (ZH); Department of Biostatistics, School of Global Public Health (RC), and Department of Computer Science & Engineering, Tandon School of Engineering (RC), New York University, New York, NY, USA; Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, MI, USA (AW); Aga Khan University Institute for Human Development, Aga Khan University Hospital, Nairobi, Kenya (AAB); CITRIC Health Data Science Center, Aga Khan University, Karachi 74800, Pakistan (ZS); Department of Medicine, Aga Khan University, Karachi, Pakistan (ZS)

- 1 Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA* 2013; **309**: 1351–52.
- 2 Safran C, Bloomrosen M, Hammond WE, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. *J Am Med Inform Assoc* 2007; **14**: 1–9.
- 3 Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health information science and systems. Health Inf Sci Syst* 2014; **2**: 3.
- 4 BYTE Academy. Top 3 ways data analytics can reduce costs in healthcare and medicine. 2020. <https://www.byteacademy.co/blog/3-ways-data-analytics-reduce-healthcare-costs#> (accessed Sept 4, 2022).
- 5 Hoodbhoy Z, Hasan B, Jehan F, Bijmens B, Chowdhury D. Machine learning from fetal flow waveforms to predict adverse perinatal outcomes: a study protocol. *Gates Open Res* 2018; **2**: 8.
- 6 Owoyemi A, Owoyemi J, Osiyemi A, Boyd A. Artificial intelligence for healthcare in Africa. *Front Digit Health* 2020; **2**: 6.
- 7 Health Planning System Strengthening & Information Analysis Unit and Ministry of National Health Services, Regulations, and Coordination. National Digital Health Framework of Pakistan 2022–2030. 2021. <https://phkh.nhsrsc.pk/sites/default/files/2022-06/National%20Digital%20Health%20Framework%20Pakistan%202022-30.pdf> (accessed Sept 4, 2022).
- 8 The Express Tribune. Pakistan's IT industry set to double by 2020: Ignite CEO. 2017. <https://tribune.com.pk/story/1586541/pakistans-industry-set-double-2020-ignite-ceo> (accessed Sept 2, 2022).
- 9 The World Bank. World Development Report 2021: Data for Better Lives. 2021. <https://www.worldbank.org/en/publication/wdr2021#:~:text=World%20Development%20Report%202021%3A%20Data%20for%20Better%20Lives%20explores%20the,individuals%2C%20businesses%2C%20and%20societies> (accessed Sept 4, 2022).
- 10 WHO. Health workforce requirements for universal health coverage and the sustainable development goals: human resources for health observer, 17. Geneva: World Health Organization, 2016.