

## Use of big data on the social determinants of TB to find the “missing millions”

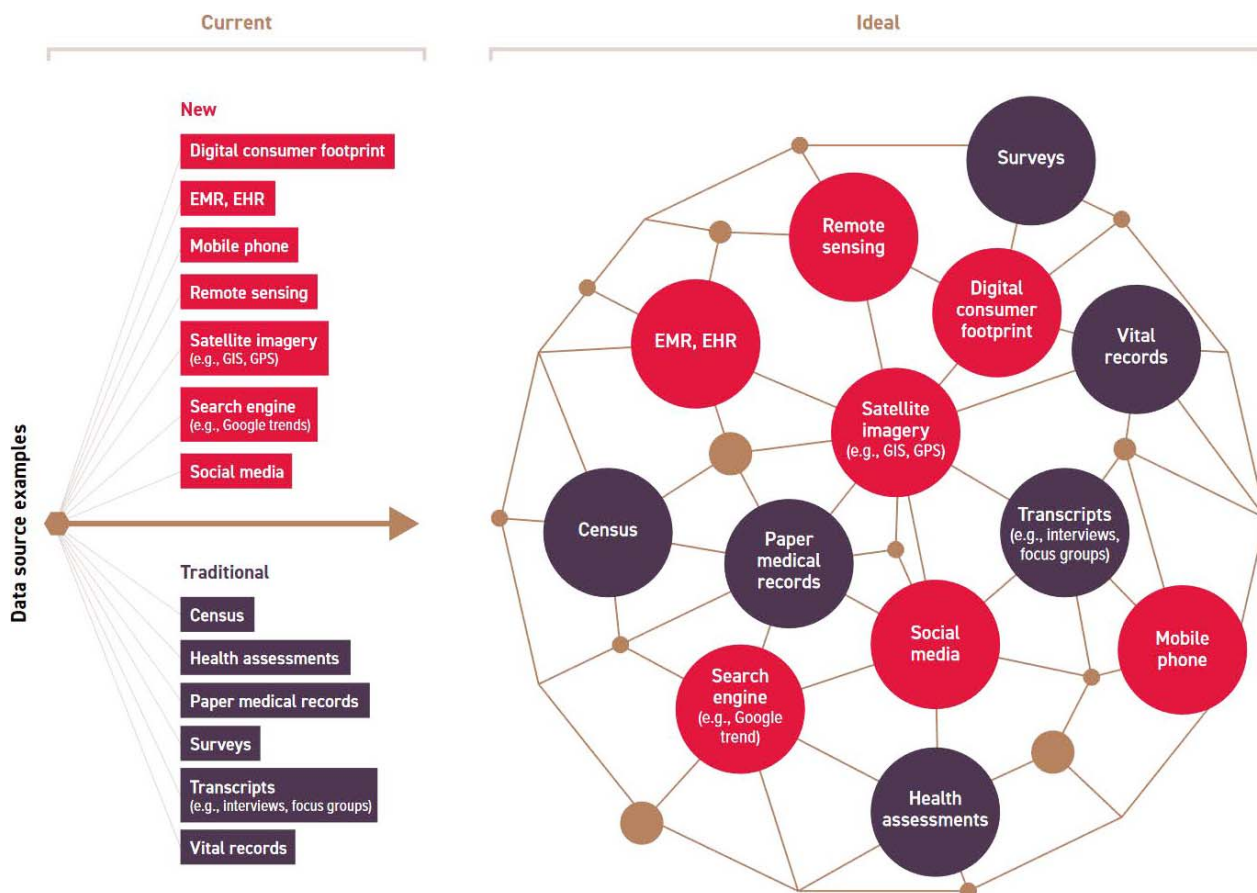
Dear Editor,

Globally, TB remains a leading cause of death by an infectious disease.<sup>1</sup> TB is driven by social determinants such as poverty, food insecurity, and poor living and working conditions.<sup>2,3</sup> Annually, 10 million people are estimated to become ill with TB, and many are missed by health services. Because of the COVID-19 pandemic, these “missing millions” have grown from 2.9 million people in 2019 to 4.2 million in 2020.<sup>1</sup> The WHO and the United Nations Sustainable Development Goals (SDGs) aim to end TB disease by 2030, highlighting the need for active case-finding as part of an integrated approach to finding the “missing millions”. The data sources that are typically used to identify people with TB (e.g., survey data from demographic and health surveys) are limited in their capacity to identify areas of high risk, and therefore in their ability to have precisely targeted interventions. Understanding the social determinants of TB through the use of new data sources, including “big data” (characterised by high volume, velocity and variety), could help to reveal uncounted vulnerable populations and allow better targeted TB interventions.<sup>4</sup> For example, better targeting could enhance the accuracy and yield of active case-finding strategies, help identify communities where standard TB health service delivery is of low quality, educate people about TB and reduce stigma. It could also inform efforts for TB prevention (e.g., the provision of TB preventive therapy), or the delivery of holistic support packages with broader health and socio-economic benefits.

Mining surveillance data,<sup>5</sup> or poverty maps that show dimensions of poverty and food insecurity,<sup>6</sup> could support the geographical tailoring of interventions to the poorest, or most food insecure populations with the highest risk of TB. Further examples of big data that could be useful are environmental data (e.g., related to economic activity, accessibility to facilities and climate change impact), and mobile phone and social media data (e.g., capturing individual and temporal aspects of people, groups and communities). As such, big data could complement and triangulate existing types of data to provide more in-depth, high-quality insights to inform decision-making (Figure). In geographical areas with limited sources of big data (e.g., a lack of smart phones),

governments and other stakeholders could involve communities to provide data on social determinants (e.g., education and housing), making the data cycle complete, representative, and responsive to changing needs.

Technological infrastructure continues to grow rapidly and has the potential to enhance real-time big data generation, storage and analysis. Such infrastructure includes information and communication technology with more speed, flexibility and reliability than traditional data systems (e.g., rural broadband and mobile phones).<sup>7</sup> Where available, big data collected through mobile phones may help to shape the collection and use of data by communities, with substantial potential to contextualise and thus improve TB interventions. For example, self-reported data from communities has been used to inform both the local and global response during the COVID-19 pandemic, providing lessons for improvements in TB prevention and control (e.g., the *Zoe App* in the United Kingdom [[covid.joinzoe.com/data](https://covid.joinzoe.com/data)] and the Johns Hopkins University [Baltimore, MD, USA] dashboard for COVID-19 [[coronavirus.jhu.edu/map](https://coronavirus.jhu.edu/map)]), or the use of digital tools for communication and surveillance in Kerala, India.<sup>8</sup> This latter example showed that digital tools were able to help meet citizens’ needs due to the high levels of collaboration and intersectoral response that connected different levels of government and multiple state departments in Kerala. These involved the private sector, and harnessed input from civil society organisations and community volunteers.<sup>8</sup> In contrast to the improvements in technological infrastructure, research and data infrastructure is often lacking where this information is needed most (i.e., low-income settings with a high burden of TB). Research and data infrastructure needs to include data collection applications and approaches, and the capacity to interpret big data to address contemporary, real-world issues.<sup>9</sup> However, even with infrastructure in place, disaggregated data may be unavailable, and silos of data sources often persist. This may lead to disparate databases (e.g., linked to specific goals and targets of the End TB Strategy or the SDGs), rather than a unified, collective approach. Other challenges include climate-friendly ways of storing data; legal infrastructure to regulate data privacy, security and sharing;<sup>10</sup> protecting



**Figure** Social determinants of health data sources universe: the current state and potential for an integrated system (this figure has been developed and published by the 3-D Commission). EMR = electronic medical record; EHR = electronic health record; GIS = geographic information system; GPS = global positioning system.

vulnerable communities, especially in authoritarian countries where data protection is problematic; and frameworks for engagement with large private and commercial firms such as Facebook/Meta (Menlo Park, CA, USA) or Google (Mountain View, CA, USA).

The use of big data to provide more in-depth, high-quality insights on the social determinants of TB, as part of an integrated approach to improve healthcare delivery, could help in the identification of the “missing millions” with TB. Despite its promise, to date, the use of big data in the field of TB has been limited to describing the global burden,<sup>11,12</sup> predicting the flow of TB patients,<sup>13</sup> optimising algorithms,<sup>14</sup> and improving screening through computer-assisted diagnosis.<sup>15</sup> Decision makers and international organisations (such as WHO) have an opportunity to embrace the power of big data. Dedicated international funding support could help to resolve today’s challenges related to infrastructure, research and data. Countries and institutions with strong capacity could provide secure data storage, access, and support for data analysis. In all this, safeguarding equity through equal partnerships with communities in collecting, analysing and interpreting

big data (as done in the above examples for the COVID-19 response), will be critical for meaningful interventions and sustainable impact.

O. BIERMANN,<sup>1</sup> T. WINGFIELD,<sup>1,2</sup> B. THAPA,<sup>3</sup>  
O. BABAJIDE,<sup>4</sup> Z. ZEINALLI,<sup>5</sup> I. TORRES,<sup>6</sup> S M ABDALLA,<sup>7</sup>  
SANDRO GALEA<sup>7</sup>

<sup>1</sup>Department of Global Public Health, Karolinska Institutet, Solna, Sweden; <sup>2</sup>Departments of International Public Health and Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK; <sup>3</sup>Department of Health Services, Policy and Practice, Brown University, Providence, RI, <sup>4</sup>Urban Health Collaborative, Drexel University, Philadelphia, PA, <sup>5</sup>Rockefeller Foundation, Boston University 3-D Commission on Determinants, Data, and Decision-making, Boston, MA, USA; <sup>6</sup>Fundación Octaedro, Quito, Ecuador; <sup>7</sup>Boston University School of Public Health, Boston, MA, USA  
Correspondence to: Olivia Biermann, Department of Global Public Health, Karolinska Institutet, Tomtebodavägen 18a, 17177 Solna, Sweden. E-mail: [olivia.biermann@ki.se](mailto:olivia.biermann@ki.se)

This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted

use, distribution, and reproduction in any medium, provided the original author and source are credited.

### References

- 1 World Health Organization. Global tuberculosis report, 2021. Geneva, Switzerland: WHO, 2021.
- 2 Lönnroth K, et al. Drivers of tuberculosis epidemics: the role of risk factors and social determinants. *Soc Sci Med* 2009;68(12):2240–2246.
- 3 Wingfield T, et al. Addressing social determinants to end tuberculosis. *Lancet* 2018;391(10126):1129–1132.
- 4 3-D Commission. Data, social determinants, and better decision-making for health. Boston, MA, USA: Rockefeller Foundation-Boston University Commission on Health Determinants, Data and Decision-making, 2021.
- 5 Källestål C, et al. Assessing the multiple dimensions of poverty. Data mining approaches to the 2004–2014 health and demographic surveillance system in Cuatro. *Front Public Health* 2020;7(409):1–12.
- 6 Pokhriyal N, Christophe D. Combining disparate data sources for improved poverty prediction and mapping. *Proc Natl Acad Sci USA* 2017;114(46):E9783–E9792.
- 7 World Economic Forum, Institut Européen d'Administration des Affaires. The Global Information Technology Report, 2015: ICTs for inclusive growth. Geneva, Switzerland: World Economic Forum, 2015.
- 8 Ummer O, et al. Connecting the dots: Kerala's use of digital technology during the COVID-19 response. *BMJ Glob Health* 2021;6(e005355):1–14.
- 9 Bogaardt M, et al. Designing a research infrastructure on dietary intake and its determinants. *Nutr Bull* 2018;43:301–309.
- 10 Protopop I, Shanoyan A. Big data and smallholder farmers: big data applications in the agri-food supply chain in developing countries. *Int Food Agribus Manag Rev* 2016;19(A):173–190.
- 11 GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018;392(10159):1789–1858.
- 12 Martins-Melo FR, et al. The burden of tuberculosis and attributable risk factors in Brazil, 1990–2017: results from the Global Burden of Disease Study 2017. *Popul Health Metr* 2020;18(Suppl 1):10.
- 13 Shi Q, Ma J. Big data analysis of flow of tuberculosis cases in China, 2014. *Zhonghua Liu Xing Bing Xue Za Zhi* 2016;37(5):668–672.
- 14 Peng J, et al. Harnessing big data to optimize an algorithm for rapid diagnosis of pulmonary tuberculosis in a real-world setting. *Front Cell Infect Microbiol* 2021;11:650163.
- 15 Morris M, et al. Computer-aided reporting of chest radiographs: efficient and effective screening in the value-based imaging era. *J Digit Imaging* 2017;30(5):589–594.