



Data Sources for Understanding the Social Determinants of Health: Examples from Two Middle-Income Countries: the 3-D Commission

Irene Torres · Bishnu Thapa · Grace Robbins · Shaffi Fazaludeen Koya · Salma M Abdalla · Onyebuchi A. Arah · William B Weeks · Luxia Zhang · Samira Asma · Jeanette Vega Morales · Sandro Galea · Heidi J. Larson · Kyu Rhee

Accepted: 7 May 2021 / Published online: 1 September 2021
© The New York Academy of Medicine 2021

Abstract The expansion in the scope, scale, and sources of data on the wider social determinants of health (SDH) in the last decades could bridge gaps in information available for decision-making. However, challenges remain in making data widely available, accessible, and useful towards improving population health. While traditional, government-supported data sources and comparable data are most often used to characterize social determinants, there are still capacity

and management constraints on data availability and use. Conversely, privately held data may not be shared. This study reviews and discusses the nature, sources, and uses of data on SDH, with illustrations from two middle-income countries: Kenya and the Philippines. The review highlights opportunities presented by new data sources, including the use of big data technologies, to capture data on social determinants that can be useful to inform population health. We conducted a search

I. Torres
Fundacion Octaedro, Quito, Ecuador
e-mail: irene.torres@octaedro.edu.ec

I. Torres · B. Thapa · G. Robbins · S. F. Koya · S. M. Abdalla · O. A. Arah · W. B. Weeks · L. Zhang · S. Asma · J. V. Morales · S. Galea · H. J. Larson · K. Rhee
Rockefeller Foundation-Boston University 3-D Commission on Determinants, Data, and Decision-making, Boston, USA

B. Thapa
e-mail: bishnu_thapa@brown.edu
G. Robbins
e-mail: grobbins@bu.edu
S. F. Koya
e-mail: fmshaffi@bu.edu
O. A. Arah
e-mail: arah@ucla.edu
W. B. Weeks
e-mail: wiweeks@microsoft.com
L. Zhang
e-mail: zhanglx@bjmu.edu.cn
S. Asma
e-mail: asmas@who.int
J. V. Morales
e-mail: jeanvegamos@ gmail.com

S. Galea
e-mail: sgalea@bu.edu
H. J. Larson
e-mail: Heidi.Larson@lshtm.ac.uk
K. Rhee
e-mail: kyu.rhee@cvshealth.com

B. Thapa
Department of Health Services, Policy and Practice, Brown University, Providence, USA

G. Robbins · S. F. Koya · S. M. Abdalla (✉) · S. Galea
Department of Epidemiology, Boston University School of Public Health, Boston, USA
e-mail: abdallas@bu.edu

O. A. Arah
UCLA Fielding School of Public Health, Los Angeles, USA

W. B. Weeks
Microsoft Research, Redmond, USA

between October 2010 and September 2020 for grey and scientific publications on social determinants using a search strategy in PubMed and a manual snowball search. We assessed data sources and the data environment in both Kenya and the Philippines. We found limited evidence of the use of new sources of data to study the wider SDH, as most of the studies available used traditional sources. There was also no evidence of qualitative big data being used. Kenya has more publications using new data sources, except on the labor determinant, than the Philippines. The Philippines has a more consistent distribution of the use of new data sources across the HEALTHY determinants than Kenya, where there is greater variation of the number of publications across determinants. The results suggest that both countries use limited SDH data from new data sources. This limited use could be due to a number of factors including the absence of standardized indicators of SDH, inadequate trust and acceptability of data collection methods, and limited infrastructure to pool, analyze, and translate data.

Keywords Social determinants of health · Data sources · Big data · Kenya · Philippines

L. Zhang
National Institute of Health Data Science, Peking University,
Beijing, China

S. Asma
World Health Organization, Geneva, Switzerland

J. V. Morales
National Research and Development Agency (ANID Chile),
Santiago, Chile

H. J. Larson
Department of Health Metrics Sciences, University of
Washington, Seattle, USA

H. J. Larson
London School of Hygiene & Tropical Medicine, London, UK

K. Rhee
CVS Health, Woonsocket, USA

Introduction

A broad range of non-biological factors—known as the social determinants of health (SDH)—shape the health of individuals and populations. However, SDH are often overlooked by decision-makers, who predominantly focus on healthcare delivery as the primary determinant of health [1]. Lack of actionable and timely data on SDH may impede efforts to bring attention to these determinants and incorporate them in decision-making about how to improve population health.

The world has seen a dramatic expansion in the scope, scale, and sources of data [2], including on health outcomes and the wider SDH in the past couple of decades [3]. This “data revolution” [4] is expected to help close data gaps towards increased health equity, but requires further development of data collection. It has given rise to a number of global efforts designed to improve the availability of these data. The African Data Consensus 2015, for example, encouraged governments in the region to introduce data initiatives in their home countries and provided guidance on the use of technology, the production of disaggregated data, and making data open and accessible. [5]

Despite the rise in data availability, there remain challenges [6] and limitations to data availability, accessibility, and usefulness for decision-making geared towards improving population health.

First, global health has continued to rely largely on traditional, government-supported data sources while large amounts of data collected by newer sources are controlled by the private sector [7]. The World Health Organization’s (WHO) SCORE initiative to consolidate essential population data and vital statistics is limited to globally comparable data that is reported by member states and has gaps [8]. Therefore, more granular or contextual data may not be available in a number of countries. Additionally, challenges in the use and analysis of complex data exist even in high-income countries (HICs), such as Canada [9].

Second, many countries, particularly low- and middle-income countries (LMICs), face capacity constraints to handle big data. There is the issue of availability; frequently, data are collected on paper, are in unintelligible formats, or are stored without direct or open access [10]. Furthermore, data sharing may not be valued or prioritized across sectors, national or international organizations, and academia. Fragile countries may have only minimal data collection resources, including infrastructure, which means that

there are still a number of LMICs that are not following international standards for data collection and management and that not all have quality controls in place [11, 12]. This may be due to outdated laws or lack of national oversight agencies in those countries.

Third, data on SDH are often collected or acquired privately, including by companies, non-governmental organizations (NGOs), think tanks, private charities, and philanthropies. Data from these sources may not be publicly or readily available; concurrently, data collection and use in HICs by large corporations are being scrutinized due to privacy concerns [5]. Also, capacity for health data collection in the private sector, even in HICs, may become strained, as evidenced during the COVID-19 pandemic [13].

Fourth, data are often being collected and stored by different sectors in isolated ways, often by a multiplicity of governmental agencies. When data are neither harmonized nor interoperable across sectors, integrating data for informed decision-making to address SDH becomes more difficult [14, 15]. Capacity and infrastructure also play an important role: countries with smaller populations and robust public health systems, such as Sweden and Denmark, have more inter-linked registries [16]. In contrast, LMICs have limitations in data infrastructure to collect, store, process, translate, and communicate data.

Fifth, while non-conventional data sources can help bridge data gaps in regions facing deficits in data collection, few systems have the capacity to put data in context, or the capacity to protect users, consumers, or patients [5, 17, 18]. Large information technology companies like Google and Facebook are often suspect for use by decision-makers because of their surveillance-like models [19] and there is the threat of data breaches, as shown by a recent case in Finland [20].

This paper aimed to document the nature, sources, and uses of data on SDH in two middle-income countries—Kenya and the Philippines—to better understand the implications of this shifting data landscape for research, decision-making, and policy on the wider SDH.

Methods

We chose to focus on Kenya and the Philippines because (1) in these countries, there is evidence of recent developments in data collection using new sources, including initiatives and legislation; (2) they are comparable in basic characteristics such as World Bank

country income classification and type of government; and (3) they represent different geographic locations and sociocultural contexts.

Kenya and the Philippines are two of the nine roadmap countries of the Global Partnership for Sustainable Development Data Roadmap Countries/Territories [21]. These countries have open data portals available online: <https://www.opendata.go.ke/> for Kenya and <https://data.gov.ph/> for the Philippines. In 2016, they were the top-ranked LMICs for open data; the Philippines was in the 36th place and Kenya was in the 42nd place in the global rankings [22]. Kenya was the first country in Africa to establish a fully online health information system by 2011 [23]. The Philippines has both a Statistical Development Program (2018–2023) and a National Mapping and Resource Information Authority (NAMRIA) in charge of geospatial data [24]. Since 2013, data-producing government agencies are consolidated in the Philippines Statistics Authority (PSA) [25]. According to the World Bank, the statistical capacity overall score for Kenya is 52.22 and for the Philippines is 81.11 out of 100 [26].

Concepts and Definitions

To select a specific set of SDH, as social and economic factors that may influence health [27], we set three main conditions: (1) Determinants that are generally accepted to directly affect health, to ensure that data will be available across countries; (2) determinants on which data are commonly found across countries, to ensure saturation of findings; (3) and well-documented determinants, to ensure that ample literature examining these determinants using various data sources exist.

We focused on seven of the most acknowledged determinants of health: healthcare (H), education (E), access to healthy choices (A), labor/employment (L), transportation (T), housing (H), and income (Y), collectively termed “HEALTHY” determinants. Access to healthy choices includes food security and physical activity.

Data Focus

We initially focused our search on new sources of data but, because results were limited, we also included traditional sources of data for comparison. These new sources included (1) electronic medical records (EMRs), electronic health records (EHRs); (2) social network data; (3) mobile phone data (e.g., call data records); (4)

GIS data; (5) satellite imagery; (6) economic/market/commerce/consumer data (e.g., retail scanner data, consumer purchase data, patient payments); (7) remote tracking and sensing (e.g., sensor data from digital devices, wearable technology); (8) internet/media content (e.g., search engine, web scraping, text mining); and (9) crowd-sourced and citizen-generated data. The traditional data sources included (i) survey data (e.g., household, facility); (ii) census data; (iii) administrative data (e.g., claims files); (iv) medical records; (v) vital records (e.g., simplified birth records, complete birth records); (vi) community health assessments.

Literature Search Strategy

We limited our search to English language studies between October 1, 2010, and September 30, 2020. English is an official language in Kenya and the Philippines. For our initial search, we broke down the research question into the following component concepts (Supplemental Table 3): Data collection/Data sources, Social determinants of health, Employment/labor, Income, Transportation, Education, Health care, and Access to healthy choices. We created a search strategy with PubMed MeSH terms that included these concepts and synonymous key words/phrases, including colloquialisms and alternative spellings. We used a similar search string at the Boston University Library, Google Scholar, and Google, and we identified additional literature or specialized informants through snowball sampling based on authorship of publications, references of publications, and references by authors of previously identified literature. For literature recommendations, we also contacted subject matter experts at WHO, institutes of public health, and within the private sector, who specialize in data science, public health, and/or SDH. The grey literature included publications from United Nations agencies, the World Bank, the Asian Development Bank, WHO, national statistical agencies and ministries of health, and NGOs.

Analytical Strategy

We compiled publications that used new data sources on any HEALTHY determinant in Kenya or the Philippines. We conducted thematic analyses by first organizing findings by country, publication name, data collection tool, type of data source, description of the data, and research notes. We subsequently identified and documented more specific data sources, their use, information captured, and

additional literature on the data environment of the two countries. The discussion centers on comparing and contrasting the countries and the needs and opportunities that exist to use data better or differently and in new ways. The conclusion focuses on implications for decision-making.

Results

Data Sources in Kenya

Data Context

Kenya is the ninth largest economy in Africa and the highest ranked country in the continent for open data [22]. In 2019, Kenya had 62 active digital platforms, 50% of them being “homegrown” that serve 49.6 out of 52.6 million people [28]. It was the first country in sub-Saharan Africa to establish a fully online health information system using a free and open software by 2011 [23]. The Africa Data Consensus of 2015 triggered Kenya to improve data collection and use with new data technology and sources [29]. However, the country’s health management information system has yet to collect robust data on social determinants of health [30].

Some reforms have been forward looking, including: the National Information, Communications and Technology (ICT) Policy (2019) [31], the Data Protection Act, and ensuing regulations that allowed the use of mobile phone to make payments, which gave way to the “health wallet” app that collects health claims data with government support [12].

Kenya is also one of the five countries in the Africa Regional Data Cube that is harnessing Earth observation data and satellite technology through a public-private partnership [32]. As an example, the country was part of the Urban ARK partnerships between researchers, practitioners, and city- and community-level activists in eight countries in sub-Saharan Africa [33].

One of the limits in the use of new data sources in Kenya is the apparent lack of support from the government. The Kenya Open Data Initiative (KODI) only shares official, not crowd-sourced or other type of data, while initiatives from different private organizations using other new data sources are not consolidated under a single state institution or program [34]. Limited governmental support also affects implementation and continuity of independent initiatives. Two such examples are given. First, Uwezo’s annual, citizen-led

assessments of the education system was originally supported by the Ministry of Education [35], but these assessments ended in 2015 and were replaced by surveys [36]. Second, the Datashift study [35] claimed data were citizen-generated, but data collection was actually based on a scorecard filled out by parents. All this may explain why even a recent study [37] on fast internet in relation to employment does not use any new data sources and relies strictly on surveys [37].

Nevertheless, traditional sources have their own sets of issues in Kenya, such as the “discrepancies between administrative data and independent household surveys [which] suggest official statistics systematically exaggerate development progress.” These discrepancies may be due to an intent on the part of the government to mislead donors or because the government itself is misled by frontline service providers reporting the data [38]. In corroboration of this finding, for apparently similar reasons, the country’s growth figures are also not considered trustworthy [39].

Main Findings

Figure 1 summarizes the traditional and news sources of data that have been used in Kenya in the context of health and its wider SDH (see Supplemental Table 1 for details on more specific types of data sources and information captured, and Supplemental List 1 for references).

Data Sources in the Philippines

Data Context

Philippines’ latest National Strategy on Statistics, which covers the 2018–2023 period, calls for the enhancement of administrative-based data and the exploration of opportunities in the use of big data and citizen-generated data [25]. With the creation of the Philippines Statistics Authority (PSA), the Philippines has also taken steps towards streamlining data. The creation of PSA in 2013, which entailed consolidating four other data-producing government agencies, has (a) increased the timeliness of data updates at the national and regional levels; (b) made national data more transparent; and (c) enhanced innovation in the conduct of government-led household surveys by making geotagging an integral part of such surveys [40].

Through the Smarter Philippines Data Analytics Research and Development, Training and Adoption project (Project Sparta), the government has committed to training 30,000 personnel in data analytics. The project, a collaboration between the Department of Science and Technology (DOST) and the Development Academy of the Philippines (DAP), is aimed at, among other things, establishing the essential infrastructure on data science and analytics [41]. The Nationwide Operational Assessment of Hazards (NOAH) program, focusing on disaster risk management in the Philippines, has been using light detection and ranging (LiDAR)-based topographic maps to help identify vulnerability to natural hazards [42].

Main Findings

Figure 2 below summarizes the traditional and news sources of data that have been used in the Philippines in the context of health and its wider social determinants (see Supplemental Table 2 for details on more specific types of data sources and information captured, and Supplemental List 2 for references).

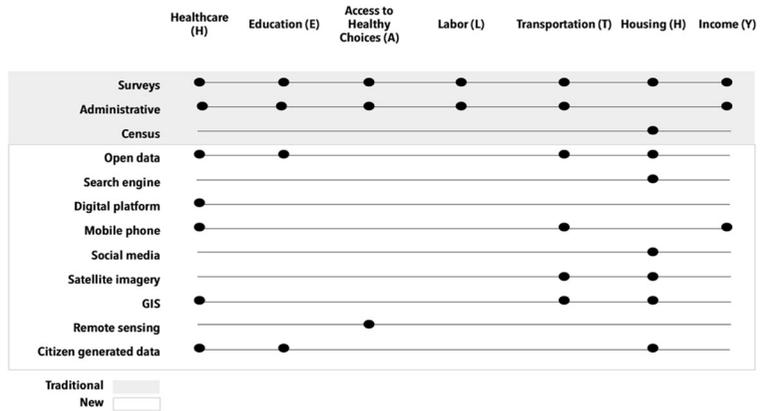
Cross-Country Comparison

Across the two countries, there are a variety of new sources being used for data on SDH, with some determinants having fewer types of data sources than others (Fig. 1). GIS/GPS data appear to be used quite frequently. More often than not, GIS/GPS-based data are used in conjunction with traditional sources of data to undertake predictive analyses. There is relatively greater use of GIS/satellite-based data in transportation. In Kenya, transportation was the determinant with the most varied data sources and, therefore, with the most combinations with other determinants (healthcare, housing, income, and education). Remote sensing is used in the two countries and Google Trends are used in the Philippines but not in Kenya. Education is often viewed as one of the major determinants of health, yet the use of new data sources in education is still very limited (Fig. 3).

Discussion

We conducted an inventory of the nature, sources, and uses of data on HEALTHY SDH, in two LMICs, Kenya and the Philippines, with a focus on new data sources.

Fig. 1 Variety of data sources on SDH in Kenya



We found limited evidence on the use of new sources of data to study the wider SDH, as most of the studies available used traditional sources. HEALTHY determinants were not often combined in the publications. There was also no evidence of qualitative big data being used. Regarding the number of publications and distribution across the HEALTHY determinants, the study found Kenya has more publications using new data sources than the Philippines, with the exception of the labor determinant. The Philippines has a more consistent distribution of the use of new data sources across the HEALTHY determinants compared to Kenya, where there is greater variation in the number of publications across determinants.

In the Philippines, surveys provided data on all HEALTHY determinants except transportation, while administrative data were available only for labor and transportation. No census data were found within the study period (2010–2020). The most commonly used resources in the Philippines include Demographic and Health Survey (DHS); Family Health Survey, Maternal and Child Health Survey (MCHS); Functional Literacy,

Education and Mass Media Survey (FLEMMS); and Family Income and Expenditure Survey (FIES). The DHS is used not just to investigate health behaviors and outcomes but also to understand HEALTHY determinants in relation to health. DHS-based HEALTHY determinants that were examined in the context of the Philippines include education, distance to health facility, income (as proxied by a wealth index), and physical condition of a house.

In terms of the new sources of data, two transportation-based platforms in particular have received national and international attention in the Philippines. They are *Open Roads* and *Open Traffic*. By making it possible for the public to keep track of publicly funded road projects, the *Open Roads* initiative promotes transparency and accountability. *Open Traffic* allows people to gather/analyze information on traffic speed by collecting GPS-based data from the mobile phones of taxi drivers.

Citizen-generated data (CGD), which is produced by Civil Society Organizations (CSOs) and NGOs that

Fig. 2 Variety of data sources on SDH in the Philippines

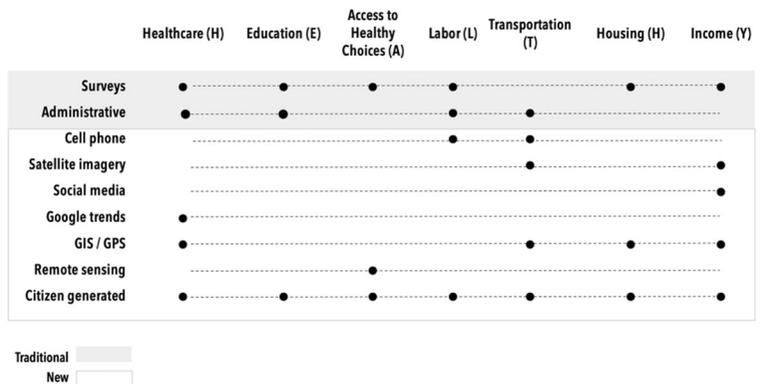
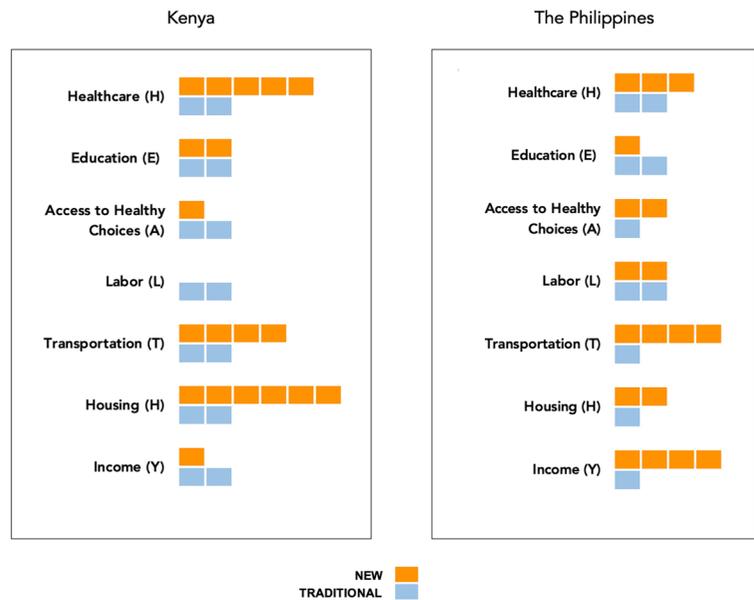


Fig. 3 Publications by determinant and type of source (each square represents a publication included in the review)



compile citizen or beneficiary information for project monitoring and other purposes, holds a great promise for the Philippines. CGD data includes data on health as well as all HEALTHY determinants. A recent publication by the Partnership in Statistics for Development in the 21st Century (PARIS21) in collaboration with the PSA noted that as many as 81 SDG indicators can be based on the CGD [43].

In Kenya, surveys provided data on all HEALTHY determinants, administrative data provided information on all determinants but housing, and the census was used only for housing. The most commonly used sources in Kenya are the Demographic and Health Survey (DHS) and Kenya Integrated Household Budget Survey (KIHBS); two World Bank surveys on service delivery and enterprise are also used.

Regarding new data sources, some innovations can be highlighted, such as the combination of different types of data to locate informal settlements, including open data (for places of worship as an indicator); search engines; social media data (Flickr API); GIS data files, Majidata, OpenStreetMap, Google Map Maker, Google Earth Engine, and LandScan (for information on housing clusters, population and road density, street intersections, pit latrines, water kiosks, and travel

patterns). With an estimated 46.5% of Kenya's inhabitants living in informal settlements, this is an example of how more/better data on SDH may be captured through non-traditional means. Additionally, the use of remote sensing of photosynthetic activity data to gauge vegetation cover as a drought indicator/an indicator of food availability in relation to child malnutrition has been proposed [44]. Such data have the potential to then influence decision-making [45].

Active legislation, needed technology, and investment in health and SDH data collection are largely insufficient for filling data gaps in the two countries. Trust in and acceptability of data collection methods are not government and researcher priorities in Kenya and the Philippines. Furthermore, big qualitative data may not be a focus of research and governments. In the Philippines in particular, despite various efforts, significant data gaps remain. This was evident in the SDG data gap assessment undertaken by the PSA, which found that data are not available for nearly 50% of the SDG indicators that are relevant to the Philippines [43].

As one of the studies in Kenya emphasizes, for data to be truly open, there must be mechanisms in place to guarantee they are available [46]. Giving strong support

to open data can incentivize more open data through the creation of collaborative programs and projects and especially through having this shared goal.

Limitations

This study has several limitations. First, the data mapping exercise is not exhaustive. Although we attempted to gather as many data sources as possible on HEALTHY determinants for each country, private sector data are not publicly available. In addition, the search concentrated on the application and use of data sources, not on conceptual models or proposals. Second, our study does not articulate the causal pathways and relationships that exist between the HEALTHY determinants and health. The relationship between these determinants and health is explored in the existing literature base that examines the association between a specific determinant and health. Therefore, for the purposes of this study, we assume as a given that the HEALTHY determinants do matter for health. Third, while for the most part the data sources are mutually exclusive, it is not always the case. For instance, citizen-generated data comes from a mix of several different sources. We have included it as a “new” data source because it is non-traditional in its involvement of CSOs and NGOs in data collection. Fourth, we did not assess the quality of the studies. Fifth, since the use of new data sources is relatively recent, the field is not yet developed to describe the full extent of their variety or to properly index them to allow for easy discovery. Finally, the lack of global SDH indicators and indexing descriptors may influence the amount of research in the field, and thus the search findings. WHO’s SCORE initiative to standardize country health data does not include SDH measures and a proposal on indicators of government action on SDH is yet to be adopted. Similarly, only one region—Latin American—has expanded the variety of SDH descriptors to index publications [47, 48]. Therefore, our findings may not necessarily reflect the full amount of relevant studies that have been conducted.

Conclusions

Difficulties in finding literature on SDH in Kenya and the Philippines point towards two major problems. On the one hand, there is a limited number of standardized indicators of new data sources and

SDH, and their corresponding search descriptors, which do not truly encapsulate their variety. On the other, the use of new sources on SDH is still in an incipient stage in the exemplar countries of this study.

The fact that proposals to create indicators of government action on SDH [47] or search terms on SDH [48] are recent and also independent from each other points to the novelty and diversity of the field, as well as a shared interest in the matter. However, a lack for a global initiative towards reaching consensus also suggests we may expect delays in new sources of data on SDH being more adequately indexed. In addition, publications in journals from data informatics and related fields that are not indexed as SDH literature may hold a wealth of information on SDH that are not being leveraged. It is crucial that attention is directed toward making studies and their insights more readily accessible and thus applicable to decision-making and further research.

Concurrently, in terms of the need for data-based decision-making (i.e., having the “right” data for the “right” decisions), there is potential for creative intersectoral work. While some capacity and infrastructure building is required for data collection on SDH, some data are already available. For example, estimating food security through remote sensing can help identify malnutrition risk in a timely manner and in a way that may not be equally captured in a survey.

The planning stage of data collection should consider requirements for interoperable data exchange across sectors and compiling data that are not directly managed by a country, so that studies combine multiple SDH. It is relevant to note that Kenya’s participation in different intergovernmental and regional data platforms, data sharing mechanisms, and benchmarking exercises has allowed the country to learn and make reforms.

Some of the data sources discussed here will become conventional as newer data sources continue to emerge, together with innovations in the tools and models that will help to streamline their use. The combination of new with traditional sources of data support the notion that the two types of sources complement each other in the SDH data landscape. It is important to stress that standardizing global measures for cross-country comparisons does not preclude the importance of contextualizing the use of new data sources to

address information gaps according to the needs and conditions of a country.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11524-021-00558-7>.

Acknowledgements We thank Zahra Zeinali for reading earlier iterations of this manuscript. We thank Leona Ofèi for her support with formatting. The Rockefeller Foundation–Boston University 3-D Commission (Grant number: 2019 HTH 024).

References

1. Espelt A, Contente X, Domingo-Salvany A, Domínguez-Berjón MF, Fernández-Villa T, Monge S, et al. La vigilancia de los determinantes sociales de la salud [Monitoring social determinants of health]. *Gac Sanit*. 2016;30:38–44. <https://doi.org/10.1016/j.gaceta.2016.05.011>.
2. Economic Commission for Latin America and the Caribbean (ECLAC). *Data, algorithms and policies: redefining the digital world (LC/CMSI.6/4)*. Santiago: United Nations; 2018.
3. Center for Open Data Enterprise (CODE). *Leveraging data on the social determinants of health roundtable report*. Washington: The Center for Open Data Enterprise (CODE); 2019.
4. United Nations Organization. *A world that counts.*; 2014. <https://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf>. Accessed May 16 2021.
5. Davies T, Walker SB, Rubinstein M, Perini F, editors. *The state of open data: histories and horizons*. Cape Town and Ottawa: African Minds and International Development Research Centre; 2019.
6. Ud Din I, Guizani M, Hassan S, Kim BS, Khurram Khan M, Atiqzaman M, et al. The Internet of Things: a review of enabled technologies and future challenges. *IEEE Access*. 2019;7:7606–40. <https://doi.org/10.1109/ACCESS.2018.2886601>.
7. Pisa M, Dixon P, Ndulu B, Nwankwo U. *Governing data for development: trends, challenges, and opportunities. CGD Policy Paper 190*. Washington: Center for Global Development; 2020.
8. World Health Organization (WHO). *SCORE for health data technical package: tools and standards for SCORE essential interventions.*; 2020. <https://apps.who.int/iris/handle/10665/334006>. Accessed May 16 2021.
9. Lucyk K, Tang K, Quan H. Barriers to data quality resulting from the process of coding health information to administrative data: a qualitative study. *BMC Health Serv Res*. 2017;17(1):766. <https://doi.org/10.1186/s12913-017-2697-y>.
10. Siddiqi S, Aftab W, Siddiqui FJ, Huicho L, Mogilevskii R, Friberg P, et al. Global strategies and local implementation of health and health-related SDGs: lessons from consultation in countries across five regions. *BMJ Glob Health*. 2020;5:e002859. <https://doi.org/10.1136/bmjgh-2020-002859>.
11. Organisation for Economic Cooperation and Development (OECD). *States of fragility 2020*. OECD Publishing; 2020. doi:<https://doi.org/10.1787/ba7c22e7-en>
12. Stuart E, Samman E, Avis W, Berliner T. *The data revolution. Finding the missing millions*. London: Overseas Development Institute; 2015.
13. Prasad V, Sri BS, Gaitonde R. Bridging a false dichotomy in the COVID-19 response: a public health approach to the ‘lockdown’ debate. *BMJ Glob Health*. 2020;5(6) <https://doi.org/10.1136/bmjgh-2020-002909>.
14. Venkataramani AS, O’Brien R, Whitehorn GL, Tsai AC. Economic influences on population health in the United States: toward policymaking driven by data and evidence. *PLoS Med*. 2020;17(9) <https://doi.org/10.1371/journal.pmed.1003319>.
15. Artiga S, Hinton E. *Beyond health care: the role of social determinants in promoting health and health equity.*; 2018. <https://www.kff.org/disparities-policy/issue-brief/beyond-health-care-the-role-of-social-determinants-in-promoting-health-and-health-equity/>. Accessed May 16 2021.
16. Schmidt M, Schmidt SAJ, Adelborg K, Sundbøll J, Laugesen K, Ehrenstein V, et al. The Danish health care system and epidemiological research: from health care contacts to database records. *Clin Epidemiol*. 2019;11:563–91. <https://doi.org/10.2147/CLEP.S179083>.
17. Dodman D, Adelekan I, Brown D, et al. A spectrum of methods for a spectrum of risk: generating evidence to understand and reduce urban risk in sub-Saharan Africa. *Area*. 2019;51(3) <https://doi.org/10.1111/area.12510>.
18. Osuteye E, Johnson C, Brown D. The data gap: an analysis of data availability on disaster losses in sub-Saharan African Cities. *Published online*. 2016; <https://www.urbanark.org/data-gap-analysis-data-availability-disaster-losses-sub-saharan-african-cities>. Accessed May 16 2021.
19. Amnesty International. *Surveillance giants: how the business model of Google and Facebook threatens human rights*. London: Amnesty International Ltd.; 2019.
20. AFP. “Shocking” hack of psychotherapy records in Finland affects thousands. *The Guardian*. Published October 26, 2020. <https://www.theguardian.com/world/2020/oct/26/tens-of-thousands-psychotherapy-records-hacked-in-finland>. Accessed May 16 2021.
21. Global Partnership for Sustainable Development Data. Data for development activities in countries. Accessed February 5, 2021. <https://www.data4sdgs.org/index.php/country-listing>
22. Global Open Data Index (GODI). Tracking the state of open government data. Published 2020. <https://index.okfn.org/place/>. Accessed May 16 2021.
23. Farnham A, Utzinger J, Kulinkina AV, Winkler MS. Using district health information to monitor sustainable development. *Bull World Health Organ*. 2020;98(1) <https://doi.org/10.2471/BLT.19.239970>.
24. Thematic Research Network on Data and Statistics (TRENDS). *Improved data governance leads to better economic outcomes for Philippine citizens.*; 2018. <https://www.sdsntrends.org/research/2018/11/5/case-study-philippines-statistics-authority>. Accessed May 16 2021.

25. Philippines Statistical Authority (PSA). *Philippines statistical development program 2018-2023*; 2018.
26. The World Bank Group. Statistical capacity indicator dashboard. Published 2019. <https://datatopics.worldbank.org/statisticalcapacity/scidashboard.aspx>. Accessed May 16 2021.
27. Commission on Social Determinants of Health (CSDH). *Closing the gap in a generation: health equity through action on the social determinants of health*; 2008. https://www.who.int/social_determinants/final_report/csdh_finalreport_2008.pdf. Accessed May 16 2021.
28. Smit H, Johnson C, Hunter R, Dunn M, Janse van Vuuren PF. *Africa's digital platforms and financial services: an eight-country overview*. Cape Town: insight2impact; 2019. <https://lu.linkedin.com/company/insight2impact>.
29. Van Belle J-P, Lämmerhirt D, Iglesias C, et al. *Africa data revolution report The status and emerging impact of open data in Africa*: Addis Ababa: United Nations Economic Commission for Africa (ECA), New York: United Nations Development Programme (UNDP), Washington: World Wide Web Foundation. Open Data for Development Network; 2018. <https://www.od4d.net/regionalhubs.html>.
30. Mulaki A, Muchiri S. *Kenya health system assessment*. Washington: Palladium, Health Policy Plus; 2019.
31. Government of Kenya. *National Information, Communications and Technology (ICT) policy*; 2019. <https://www.ict.go.ke/wp-content/uploads/2019/12/NATIONAL-ICT-POLICY-2019.pdf>. Accessed May 16 2021.
32. Global Partnership for Sustainable Development Data. Africa regional data cube. Accessed February 5, 2021. <https://www.data4sdgs.org/initiatives/africa-regional-data-cube>
33. Leck H, Pelling M, Adelekan I, Dodman D, Issaka H, Johnson C, et al. Towards risk-sensitive and transformative urban development in Sub Saharan Africa. *Sustainability*. 2018;10(8):2645. <https://doi.org/10.3390/su10082645>.
34. Simche LA, Migisha C, Minjire D, Juma L, Adan S, Alemu W. *Drivers of data for development. Exploring the factors that enable availability and utility of open data for development in Africa*. Nairobi: Local Development Research Institute; 2020. <https://www.developlocal.org/contacts/>.
35. Datashift. Using citizen-generated data to monitor the SDGs. A Tool for the GPSDD Data Revolution Roadmaps Toolkit. Published 2017. <https://www.data4sdgs.org/sites/default/files/2017-09/Making>. Accessed May 16 2021. Use of Citizen-Generated Data - Data4SDGs Toolbox Module.pdf
36. Uwezo. *Are our children learning? Uwezo Uganda Eighth Learning Assessment Report. Twaweza East Africa*. Kampala: Twaweza East Africa; 2019.
37. Hjort J, Poulsen J. The arrival of fast internet and employment in Africa. *Am Econ Rev*. 2019;109(3) <https://doi.org/10.1257/aer.20161385>.
38. Sandefur J, Glassman A. The political economy of bad data: evidence from African survey and administrative statistics. *J Dev Stud*. 2015;51(2):116–32. <https://doi.org/10.1080/00220388.2014.968138>.
39. Jerven M. Random growth in Africa? Lessons from an evaluation of the growth evidence on Botswana, Kenya, Tanzania and Zambia, 1965–1995. *J Dev Stud*. 2010;46(2) <https://doi.org/10.1080/00220380903370161>.
40. Global Partnership for Sustainable Development Data. *Improved data governance leads to better economic outcomes for Philippine citizens a case study of reform of the Philippine Statistics Authority*; 2018. <https://www.data4sdgs.org/sites/default/files/2018-10/PSA%20Case%20Study.pdf>. Accessed May 16 2021.
41. Development Academy of the Philippines. SPARTA smarter Philippines through data analytics R&D, training and adoption. Accessed February 5, 2021. <https://sparta.dap.edu.ph/#about-sparta>
42. Novero AU, Pasaporte MS, Aurelio RM, Madanguit CJG, Tinoy MRM, Luayon MS, et al. The use of light detection and ranging (LiDAR) technology and GIS in the assessment and mapping of bioresources in Davao Region, Mindanao Island, Philippines. *Remote Sens Appl Soc Environ*. 2019;13:1–11. <https://doi.org/10.1016/j.rsase.2018.10.011>
43. Philippines Statistical Authority (PSA). *Use of citizen generated data for SDG reporting in the Philippines: a case study*. Quezon City: Philippine Statistics Authority (PSA); 2020. <https://psa.gov.ph/directory>.
44. Bauer JM, Mburu S. Effects of drought on child health in Marsabit District. *Northern Kenya Econ Hum Biol*. 2017;24: 74–9. <https://doi.org/10.1016/j.ehb.2016.10.010>.
45. The World Bank Group. Population living in slums (% of urban population) – Kenya. Published 2020. <https://data.worldbank.org/indicator/EN.POP.SLUM.UR.ZS?Locations=KE>. Accessed May 16 2021.
46. Williams S, Marcello E, Klopp JM. Toward open source Kenya: creating and sharing a GIS database of Nairobi. *Ann Assoc Am Geogr*. 2014;104(1) <https://doi.org/10.1080/00045608.2013.846157>.
47. Barros A, Boutayeb A, Brown C, et al. Towards a global monitoring system for implementing the Rio Political Declaration on Social Determinants of Health: developing a core set of indicators for government action on the social determinants of health to improve health equity. *Int J Equity Health*. 2018;17:136. <https://doi.org/10.1186/s12939-018-0836-7>.
48. Campos ACE, Treuherz A, Murasaki RT, Gonzalez D, Mújica OJ. Nuevos Descriptores en Ciencias de la Salud para clasificar y recuperar información sobre equidad. *Rev Panam Salud Pública*. 2020;44:e98. <https://doi.org/10.26633/RPSP.2020.98>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.