Real-World Data in Global Health

60

Jeffrey S. Barrett

Contents

Abstract

Global health (GH) seeks broadly to promote health research and practice with the intention of improving health and achieving equity in health for all people worldwide. The use of real-world data (RWD), principally electronic health records (HER), patient registries, and survey-based outcomes from census and other data sources that describe the global burden of disease, has the potential to accelerate GH initiatives. As this potential becomes realized through the collaborative efforts of relevant stakeholders and the GH ecosystem, greater investment in the infrastructure of low- and middle-income countries (LMICs) will be required. Differences in the developmental status of various economies create a diversity in "readiness" for RWD use both with respect to access and the quality of the underlying source data. In order to fully realize the potential of RWD to inform GH research and practice, education regarding the value and mechanism of sharing, governance, and adoption of appropriate standards (data and policies) must continue to the point of adoption. Current examples of RWD application to GH initiatives are encouraging and illustrate the diversity in scope as well as the broad applicability and dependence on collaboration. Continued investment in

J. S. Barrett (✉)

Quantitative Medicine, Critical Path Institute, Tucson, AZ, USA
e-mail: jbarrett@c-path.org

1861

RWD may ultimately provide more patient-centric deliverables as patients in all parts of the world become more invested in their own health outcomes.

Introduction

Given the broad intention of global health (GH) research and practice to place a priority on improving health and achieving equity in health for all people worldwide, RWD has multiple opportunities to facilitate GH initiatives. The benefit of real-world data (RWD) to inform various aspects of global health research is well supported (Chan et al. 2010; Murray et al. 1996, 2015) with great expectation for expanded utilization (Nord 2013; Salomon et al. 2003; Salomon 2008, 2013). A great deal of recent attention has been given to its benefit in drug development with the potential to improve efficiency and lower costs (Breckenridge et al. 2019). Beyond the drug development setting, the use of RWD has other customers beyond big PhRMA. Provider and payer organizations utilize RWD to increase their knowledge about the effectiveness, safety, and costs associated with a treatment option, while healthcare practitioners benefit from RWD to help inform the real-world implications of their treatment decisions. Financial RWD is obviously also critically important in the assessment of needs and allocation of resources as administered by the WHO. Patients themselves can exploit RWD to help inform discussions with their healthcare professional about treatment options. The positive perspective for RWD and early evidence of improved decision-making is largely realized by development strategies focused for the developed world, i.e., advanced economies with advanced technological infrastructure or high-income countries (HICs). The path for RWD utilization in the global health space (low-/middle-income countries, LMICs) is not straightforward, and additional challenges exist.

Though the use of RWD to bridge populations to gauge various GH solutions works well in some instances, this bridging exercise is often not appropriate in a more global context. Reasons for this can be due to a variety of factors including differences in the standards of care, heterogeneous populations, societal structure/network, migration, and adherence. Some of these issues could be addressed by increasing the availability and utilization of RWD in the different regions of the world; however, the assumption that such data already exists or is accessible is often invalid. The trajectory of RWD utilization in and for developing or low-income countries (LICs) has been very different than in developed countries. Healthcare for the affluent world and the data streams that define strategies and indices of health are both more extensive and varied than those that map health in low-income countries with often little or no data in these populations. Likewise, when we look at global health indices as assessed by RWD, these are often weighted to HICs, and so the

analyst must take care to adjust appropriately if the intention is to truly reflect a global health perspective.

An added complexity is the diversity of the stakeholders involved in various aspects of global health, not the least of which are the governmental interests in the management and monitoring of the health of the populations and economies in scope. Geopolitical organizations such as NATO and the World Bank have a vested interest as well as global health advisors and policy makers such as the WHO, various centers for disease control and philanthropic organizations such as the Bill & Melinda Gates Foundation, the Wellcome Trust, Gavi, IAVI, PATH, The Rockefeller Foundation, and others. All have had and continue to invest in generating and collecting RWD sources for various purposes related to GH. In this chapter we will explore the current landscape and opportunities for RWD to inform various aspects of global health, assess the data available in these situations, and examine the context for which the current applications can be sustained and extended to gain the full value of RWD for the benefit of global health.

Context for RWD to Support Global Health

If one considers the most common forms of RWD to include electronic medical records, electronic health records, claims databases, health surveys, patient registries, data from health-related applications and mobile devices, and data from social media, it is easy to conclude at first glance that most of these data sources simply do not exist in LICs. In reality, of course, many of them do exist, but the access and/or infrastructure supporting their organization/management is not in place, and often the quality of data is suspect. With respect to actual RWD generated in global health target populations, this is accomplished with limited capacity to date given the difficulty gathering the data for the most part. The expectation is that such RWD will increase dramatically as LICs' economies improve, and governments have greater focus on healthcare and healthcare costs. Although not always optimal, most countries have some form of civil registration and vital statistics systems to record births, deaths, and causes of death. As these improve, they may represent another source of RWD that could be utilized for global health research. When we refer to GH research, we speak very broadly about initiatives that place a priority on improving health and achieving equity in health for all people worldwide, both LMIC and HIC economies.

The fact that a disparity in both GH research and practice exists around the world is a surprise to no one. The dynamic nature of the disparity in the face of political and economic uncertainty makes the goal of "leveling the playing field" especially difficult however. Part of the challenge has been the generation of credible data documenting disease prevalence in LMICs where the global burden of disease (GBD) is highest. One needs only to recall the recent past to see the challenge and the effort it took to make progress. The GBD enterprise dates to the early 1990s, when the World Bank commissioned the original GBD study and featured it in the landmark World Development Report 1993: Investing in Health. Co-authored by

Dr. Christopher Murray, who went on to become Director of the Institute for Health Metrics and Evaluation (IHME), the GBD study served as the most comprehensive effort up to that point to systematically measure the world's health problems, generating estimates for 107 diseases and 483 sequelae. It covered eight regions and five age groups with estimates through 1990. The GBD 1990 impacted health policy and agenda-setting throughout the world; it brought global attention to otherwise hidden or neglected health challenges, such as mental illness and the burden of road injuries. GBD work was institutionalized at the World Health Organization (WHO), and the organization continued to update GBD findings. In 1998, the WHO created a Disease Burden Unit, which generated GBD estimates for 2000, 2001, and 2002, publishing the estimates in WHO's annual World Health Reports. In 2008, WHO updated the GBD estimates for 2004. It took over two decades to expand the scope of the effort to its current form and establish a process for the accurate collection and distribution of this information. The WHO continues to recognize the strong interest in the data it collects in LMICs and is one of the few organizations in this area to provide access to such data in an open and transparent manner (see https://www.who.int/gho/publications/world_health_statistics/en/). However, this is often provided without much context, and the metadata that would facilitate the assembly of meaningful, analysis-ready data is simply not there. Likewise, much of the data are flat files (ASCII or Excel-based) and require significant effort to transform to be useful. Nonetheless, the collection and provision of vital sign data, patient comorbidities, and risk factors in LMICs is a significant contribution by the WHO and highlights an important facet of the GH ecosystem – the necessity of sharing in an informative and expedient manner is neither guaranteed nor always recognized. An important aspect of future progress will be the effort to organize/qualify data sources and provide governance over shared, accessible RWD data sources.

As globalization continues to evolve and health issues affecting countries and parts of the world become increasingly interconnected, so does the interest in global health and welfare. Not surprisingly many academic centers have also responded to the interest in and need for global health initiatives by creating innovative educational programs to address global health and welfare concerns. The interest is worldwide and diverse based on affiliation, focus, and research interests (Stone 2014; Sienkiewicz 2019). In addition, the Consortium of Universities for Global Health (CUGH) supports academic institutions and partners to improve the well-being of people and the planet through education, research, service, and advocacy (https://www.cugh.org/members). These programs offer multiple benefits, not the least of which is to give students the opportunity to study, research, and perform clinical rotations within domestic health as well as international health domains. Beyond the clinical/medical incentives, some programs offer specialization in the analysis and interpretation of RDW for global health issues (Withers et al. 2016) which train young scientists in the relevant approaches and methodologies but also give them a context for global health application. In addition, many academic medical centers not only provide teaching and research programs in various aspects of GH but also have arranged country-specific agreements with target LMICs where

the GBD is high-yielding unique RWD source. In some cases, they are directly involved with WHO to propose labelling changes and make GH recommendations.

As global interest in RWE continues to grow, the databases and research methodologies used to collect and analyze these data have become more sophisticated as healthcare researchers are gaining access to new, previously unavailable data. With robust RWD sources, more insight-generating analyses can be conducted to help better inform healthcare decision-making based on everyday patient outcomes.

> Definition *Real-world data (RWD) refers to data related to patient health status and/or the delivery of healthcare that are routinely collected from a variety of sources, such as EHRs, claims and billing data, medical product and disease registries, patient-generated data including in home-use settings, and data gathered from other sources that can inform on health status (e.g., mobile devices).*
>
> Definition *Real-world evidence (RWE) is the clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of RWD.*

The global health ecosystem includes a variety of valuable stakeholders and partners including, but not limited to, the following: the WHO, various decision-makers (e.g., ministries of health, government agencies, other government departments at the national level), health partnerships, foundations, intragovernmental and nongovernmental organizations, civil society, media, professional associations, and various academic collaborating centers. The United Nations at the global, regional, and country level is also a major player. In addition, there are industry partners seeking to market their goods to GH economies, not the least of which are the pharmaceutical and related life science industries. Table 1 describes some of the specific stakeholder agencies and organizations that collect, promote, provide, or analyze RWD for various GH initiatives.

Within the context of drug and vaccine development, regulatory authorities have weighed in and provided both opinions regarding the value and use of RWD to augment a sponsor's submission and the conditions upon which they will review. In December 2018, the FDA published a framework for Real-World Evidence program. In 2018, the EMA published a discussion paper on the use of patient disease registries for regulatory purposes (methodological and operational considerations). A major driver for the integration of RWE into the clinical decision-making process is understanding that RWD may provide valuable information on both short- and long-term medication safety and effectiveness in patient populations that are not well represented in randomized controlled trials (RCTs). RWD may also be used to provide insights into other aspects of medication use in clinical practice, including how a medication compares with other therapies in terms of surrogate outcomes, biomarkers, clinical uptake, safety, and cost. Through this application, RWD plays a critical role in bridging the gaps in information that cannot be met by RCTs alone. In resource-scarce healthcare environments, RWD may increasingly be used to

Table 1 Stakeholder agencies investing in RWD as part of their global health initiatives

GH stakeholder	Primary focus	RWD interest and effort
Bill & Melinda Gates Foundation	Funding basic and clinical research to effect cures where the global burden of disease (GBD) is greatest	Open data sharing policy to promote expanded access to funded research. The Foundation supports both the WHO and IHME for the collection and sharing of GBD data
Gavi Global Vaccine Alliance	Brings together public and private sectors to create equal access to new and underused vaccines for children living in the world's poorest countries	Using data to map immunization in the developing world; partnering with Royal Philips[a]
IAVI International AIDS Vaccine Initiative	Develops vaccines and other biomedical innovations that prevent HIV infection	Using historical RWD to design and conduct prospective vaccine trials (Abaasa et al. 2019)
The Global Fund to Fight AIDS, Tuberculosis and Malaria	Raises, manages, and invests the world's money to respond to three of the deadliest infectious diseases the world has ever known	Management and coordination of financial data – The Data Explorer visualizes where investments come from, where they are, and what they achieve by providing pledge and contribution data, grant financial data, and results data at global, regional, and country levels[b]
UNAIDS	Leads and supports an expanded response to HIV and AIDS including preventing transmission of HIV, providing care and support to those living with the virus, reducing vulnerability of individuals and communities to HIV, and alleviating the impact of the epidemic	Estimates the state of the worldwide HIV epidemic. Data used for decisions on access to services, for treatment and prevention, and to decide how to spend billions of dollars a year; works with individual countries to collect survey data; summary data published yearly[c]
UNFPA United Nations sexual and reproductive health agency	Builds knowledge and capacity to respond to needs in population and family planning; promotes awareness in both developed and developing countries of population problems and possible strategies; leads United Nations system in promoting population programs; and coordinates projects supported by the Fund	Manages a global evaluation report database and makes publicly available evaluation reports conducted by or for UNFPA, available to the broad public. The database is regularly updated with reports from country offices, regional offices, or headquarters units[d]
UNICEF	Mandated by the United Nations General Assembly to advocate for the protection of children's rights, to help meet their basic needs, and to expand their opportunities to reach their full potential	Provides data to the Humanitarian Data Exchange (HDX, https:// data.humdata.org/) an open platform for sharing data across crises and organizations. The goal of HDX is to make humanitarian data easy to find and used for

(continued)

Table 1 (continued)

GH stakeholder	Primary focus	RWD interest and effort
		analysis. It contains a growing collection of data sets that has been accessed by users in over 200 countries and territories. HDX is managed by OCHA's Centre for Humanitarian Data. OCHA is part of the United Nations Secretariat
World Bank	Provides low-interest loans, interest-free credit, and grants; focuses on improving education, health, and infrastructure. It also uses funds to modernize a country's financial sector, agriculture, and natural resources management	The World Bank's Open Data initiative provides all users with open access to World Bank data. Financial indicators, sample surveys, censuses, and administrative system data as well as geospatial data provided – see catalogue at https://datacatalog.worldbank.org/
World Health Organization	Acts as a directing and coordinating authority on international health work, to ensure valid and productive technical cooperation and to promote research	WHO's annual World Health Statistics reports present the most recent health statistics for the WHO Member States

[a]https://www.gavi.org/library/news/press-releases/2016/gavi-and-philips-team-up-to-improve-immunisation-data-quality-in-developing-countries/
[b]https://data.theglobalfund.org/investments/home
[c]https://www.unaids.org/en/resources/fact-sheet
[d]https://web2.unfpa.org/public/about/oversight/evaluations/documentList.unfpa

generate more cost- and time-efficient post-marketing evidence than RCTs. At the moment the submission of RWE derived from RWD is viewed by regulatory agencies as supportive only as opposed to being a substitute for the two well-controlled Phase 3 trials currently required to gain market access. As confidence in the data quality and the approach improves, this may change as many are hoping. Independent of pharmaceutical sponsors, the payer community (health insurance providers, etc.) also look to RWD to verify effectiveness of approved agents and to justify formulary listing and reimbursement policies (Katkade et al. 2018; Amirthalingam et al. 2014).

One of the frontier areas for RWD application to GH initiatives is the extent to which it may influence future policy decisions. To some extent this has already been accomplished though it is based on limited scenarios where data has been highly curated, and the decision-making risk was not extreme. Certainly, the use of RWD to estimate disease prevalence rates and GBD in LMIC is a milestone for RWD impact and provides guidance on funding initiatives likely to improve these situations (GBD Collaborators 2015). At what point however are we comfortable with making policy recommendations for LMIC based on RWE? As with the desire to use such data for regulatory approval in HICs, the issue quite often revolves around uncertainty in the

quality of the underlying source data and the methodologies and approaches used to compare RWE derived from RWD with data obtained in a more rigorously controlled setting. There is also the issue of education of both regulators and policy makers. While legitimate concerns exist regarding the quality of the underlying source data, the interpretation of such data with those collected under more stringent conditions (e.g., RCTs), and the evolving methodologies which have not been wholly validated, the necessity to look beyond the short-term limitations to provide an appropriately critical assessment must be embraced. Before we can engage in how such data can be used to inform policy decisions, greater comfort with the afore-mentioned issues must be addressed. Even though PhRMA looks on with interest and support, most of their efforts are based on late-stage, commercial, or surveillance efforts. The lobby for use of RWE as part of regulatory requirements begins with those that cannot afford the cost of Phase 3 trials and/or patient populations where financial incentives are limited as well as industry interest.

With respect to actual RWD data sources that might support various aspects of GH research, these would include patient registries, healthcare databases including electronic health records, pharmacy and health insurance databases, social media, and patient-powered research networks. Some of these are available publicly, and some are sold typically in a fee-for-service model but also as a subscription service. The vast majority of RWD available for purchase is, of course, predominantly based on HIC patients given the infrastructure, that is, type required to collect, store, and manage such data. Most vendors (that sell RWD) will claim they have LICs represented in their RWD sources, but typically upon investigation, such data comes from affluent regions of these countries and potentially represents foreign influence as opposed to reflecting the true population characteristics of the repre-sentative LIC. While even remote LIC hospitals are beginning to collect HER data, there is often reluctance to provide these data to those interested in GH, especially when the request comes from parties outside the host country. When permission is granted, it is usually with the stipulation that the data does not leave the data repository, and either must be analyzed on site or with significant security assurances regarding access only privilege (Srivastava 2016). Many LICs struggle with the adoption of HER in general, and the mixed healthcare system in India (government, state and private sector influence, and ownership) has created a diversity in readiness and strategy. Likewise, the common issues of standards adoption, limited current use for clinical decision-making, lack of legal safeguards (laws in place) to ensure the appropriate sharing and data privacy, and liability concerns create what seems like an insurmountable obstacle for expedited HER access in the near future though the interest level is high.

Beyond HER there are also disease registries that provide longitudinal data in certain patient populations of interest. In the global health space, the availability of such registries coincides with disease prevalence, but here again that does not mean that rare diseases don't have outreach into LICs. The underlying problem, as always, is the lack of infrastructure to identify and care for these patients. Table 2 provides a landscape view of available RWD sources that serve the global health community. It should be viewed as representative and not exhaustive though it attempts to

Table 2 Common RWD sources (registries, HER, etc.) available for access and/or purchase and the extent to which they represent data from LMICs and usefulness for GH research

Data source	RWD data types	Dimension relevant to GH?	LMIC?	Comments
Known hospitals in India with HER systems (not a complete list): AIIMS, Jai Prakash Narayan Trauma Center (New Delhi), Rajiv Gandhi Cancer Institute & Research Centre (New Delhi), Super Speciality Paediatric Hospital & Post Graduate Teaching Hospital (Noida), India Cancer Institute (Noida), India Cancer Society Cancer Treatment Center (New Delhi), Lok Nayak Hospital Government Hospital (Delhi), Paras Hospitals (Gurgaon), TATA Healthcare IT Consultancy Services, Nepal Cancer Hospital (Nepal), Max Super Speciality Hospital (New Delhi)	HER data; unknown number of patient experiences and duration for all sources	Yes, longitudinal data reflecting risk factors, comorbidities, and SOC	Yes	Unstructured data; no clear agreements in place for access and sharing; concerns about standards and liability (Srivastava 2016)
IQVIA*, https://www.iqvia.com/solutions/real-world-evidence/real-world-data-and-insights	600 million, non-identified patient records	HIC comparator	Yes, global	
Truven Healthcare Analytics https://www.ibm.com/watson-health/learn/truven-health-analytics	225 million unique patients' records		No, US-based	
OM1 https://www.om1.com/solutions/real-world-evidence/clinical-registries/	Multiple clinical registries: rheumatology, cardiometabolic, gastrointestinal, neurology, respiratory, pain management, behavioral health	HIC comparator	No, US-based	

(continued)

Table 2 (continued)

Data source	RWD data types	Dimension relevant to GH?	LMIC?	Comments
AIIMS https://www.aiims.edu/en/hostel_accomodation_contact/102-microbiology/2310-dr-u-b-singh.html https://www.aiims.edu/images/pdf/notice/Digital%20AIIMS/Patient%20Display%20System-%20White%20Paper.pdf	Multi-specialty, tuberculosis	Yes, longitudinal data in TB	India	Hospital-based HER
Sir Ganga RAM Hospital; InterSystems TrakCare: system includes clinical, administrative, laboratory, and community care capabilities, unified by a single data repository. Patient's consolidated history can be shared securely across all care settings	Multi-specialty	Yes	India	InterSystems TrakCare
Hinduja National Hospital and Medical Research Centre https://www.teamhgs.com/news-resources/news/hgs-partners-pd-hinduja-hospital-medical-research-centre-treatment-mdr-tb	Multi-specialty	Yes	India	A multi-specialty, tertiary care hospital in Mumbai, India. HGS partners with P.D. Hinduja National Hospital and Medical Research Centre for treatment of MDR-TB patients
http://www.nicd.ac.za/centres/national-cancer-registry/	National Cancer Registry	Yes	South Africa	

SickleInAfrica registry; https://www.sickleinafrica.org/registries-list	Sickle Cell Disease Registry	Yes	South Africa	Multi-country registry involving different African countries, with different legal or institutional approaches to recording, storing, and sharing and secondary use
National Centre for Disease Informatics and Research (NCDIR) with the following subspecialties: National Cancer Registry Program (NCRP), National Stroke Registry Programme http://stroke.ncdirindia.org/, diabetes and CVD	Disease registries: cancer, stroke, diabetes, CVD		India	Disease registry
The Pan African Clinical Trials Registry (PACTR) is a regional register of clinical trials conducted in Africa. The registry is an African initiative serving the needs of Africans. It provides an open-access platform where clinical trials can be registered free of charge https://pactr.samrc.ac.za/	Database of planned trials and trials currently in progress	Yes	Africa	The PACTR aims to increase clinical trial registration in Africa by developing awareness of the need to register trials and supporting trialists during registration
National Centre for Disease Informatics and Research (NCDIR): National Cancer Registry Program (NCRP); National Stroke Registry Programme (http://stroke.ncdirindia.org/); diabetes; CVD	Disease registry: cancer, stroke, diabetes, CVD	Yes	India	Disease registry

Table 2 (continued)

Data source	RWD data types	Dimension relevant to GH?	LMIC?	Comments
Integrated Disease Surveillance Program (IDSP) https://idsp.nic.in/index4.php?lang=1&level=0&linkid=412&lid=3695	Disease surveillance data medical colleges, health centers, hospitals, labs, etc.	Yes	Global	Under the Ministry of Health and Family Affairs in India; assisted by the World Bank; maintained from 2004 to 2010
RNTCP Government TB program; Nikshay portal: https://Nikshay.in	Tuberculosis	Yes	India	TB program patient portal managed by government
PGIMER: Post Graduate Institute of Medical Education & Research, Chandigarh	Multi-specialty	Yes	India	Hospital using EMR
Artemis Hospital http://nhspl.in/hospitals_details.php?hospitals_details=15&&hospitals_details_id=NHSPL-482964	Multi-specialty	Yes	India	CT scan from Philips Medical Mammography from Siemens Medical
Hospital Information System (HIS) software by Wipro https://www.wipro.com/healthcare/hospital-information-system%2D%2Dhis-/		Yes	India	EMR software
MEDITECH South Africa Ltd https://ehr.meditech.com/global/meditech-south-africa MEDITECH, a commercial system, is also entering African markets		Yes	South Africa, Botswana, and Namibia	MEDITECH

Ministry of Science and Technology iCare piloted at Tikur Anbessa Hospital http://p2pbridge.org/2019/01/22/digitizing-patient-data/ OpenMRS is the most commonly-used EHR open-source system in SSA MEDITECH, a commercial system, is also entering African markets			Ethiopia, Addis Ababa	iCare – newer EMR system
Leishmaniasis Research and Treatment Centre, University of Gondar http://p2pbridge.org/2019/01/22/digitizing-patient-data/ OpenMRS is the most commonly used EHR open-source system in SSA MEDITECH, a commercial system, is also entering African markets			Ethiopia	OpenMRS laboratory management system (OpenELIS)
Global Health Observatory (GHO) data from WHO – World Health Statistics, https://www.who.int/gho/publications/world_health_statistics/en/	Vital signs, health metrics, alcohol and tobacco consumption	Yes	Global	Structured data files in repository; data downloadable

*GE Quintiles and IMS are part of IQVIA

provide a diverse view of varied data sources and data types as well as the extent to which it represents LMICs which are often the subject of global health research and outreach.

Patient disease registries are organized systems that use observational methods to collect uniform data on a population defined by a particular disease, condition, or exposure, and that is followed over time. They are also a critical RWD source utilized in various aspects of research including GH initiatives. Many registries are focused on providing longitudinal data and clinical experience in specific disease populations around the world; many of these are rare diseases. The vast majority of registry data exists in HICs that have established infrastructure for identifying and connecting patients to resources for treatment, data collection, and support. Likewise, the organizing group responsible for managing and maintaining registry data is often specific foundations representing these patient populations in HICs and more likely to be organized and managed by individual government organizations in LMICs (see Table 2).

Most would conclude that registries remain an underutilized resource with some obvious shortcomings including heterogeneity in registry design and in the data collected, unreliable data quality, and data sharing impediments (McGettigan et al. 2019). In an effort to address some of these deficiencies, the Patient Registries Initiative was established by the European Medicines Agency in 2015 to support registries in collecting data suitable to contribute to regulatory assessments, especially post-authorization safety and effectiveness studies. Table 2 contains several registries (e.g., sickle cell, cancer, and stroke) in LMICs targeted to diseases specific to their populations but also many in which there are HIC equivalent sources.

Other Data Sources

In the USA, the Agency for Healthcare Research and Quality (AHRQ) offers practical, research-based tools and other resources (Kronick 2016) to help a variety of healthcare organizations, providers, and others make care safer in all healthcare settings (https://www.ahrq.gov/data/resources/index.html). Other similar organizations around the world including the National Institute for Health and Clinical Excellence (NICE) in the UK, the Institute for Quality and Efficiency in Health Care (IQWiG) in Germany, the Haute Autorité de Santé (HAS) in France, and the Canadian Agency for Drugs and Technologies in Health (CADTH) provide similar resources for the GH community. Additionally, Health Technology Assessment International (HTAi) is an organization with global membership that promotes evidence-based technology assessments.

The contract research organization (CRO) industry has also started to make significant investments in the acquisition of HER data with the intention of providing it to pharmaceutical sponsors mostly for the purpose of aiding patient and site selection but also for Phase 4/ post-marketing surveillance studies and market access facilitation.

Data are for the USA:

- Percent of office-based physicians using any EMR/EHR system: 85.9%
- Percent of office-based physicians with a certified EMR/EHR system: 79.7%

These efforts are at various stages, and in most cases the vendors are transparent in the recognition that the underlying data quality in many of the resident systems is extremely variable. Given this fact every consumer of this data must recognize that they use the information at risk. Likewise, pooling this data across sites is not straightforward given the variation in HER systems and underlying data structures and standards (or lack of). Table 3 provides an assessment of the current CRO landscape with respect to their HER data provision coverage across key LMICs where mature HER experience does not exist but is growing. Some of the individual hospitals, particularly in Africa and India, are listed in Table 2 as well.

An important capability for HER providers is the ability to ensure patient privacy and maintain some level of data standards to confirm that the data is usable and interpretable. Clinerion and other vendors use proprietary technology to accomplish this and confirm that only de-identified patient data unlinked from identifiers is used, ensuring that identifiable patient data does not leave the hospital. Data security is enforced by multiple firewalls, one-way data connections, and user authentication. Compliance with data privacy regulations in the USA (HIPAA) and the European General Data Protection Regulation (GDPR) (EU 2016/679) is a must along with the express permission of participating hospitals and is done in accordance with relevant local/federal legislation.

Table 3 Contract research organizations (CROs) engaged in HER access and provision in key GH target geographic areas (e.g., LMICs)

CRO/vendor	Geographic area (hospitals)	Nature of HER data available	Access and interface details
Clinerion (https://www.clinerion.com/index/OverviewNews/News)	India, Africa, South America HIC as well	Clinerion's Patient Network Explorer currently has 25 million patient records across 20 countries	GUI for data review and analysis; data abstraction limited
IQVIA (https://www.iqvia.com/)	HIC mostly; partnership for LMIC data	>600 million non-identified patients accessible	Service for data provision
ICON (http://www.iconplc.com/)	Works with sponsors to exploit RWD they purchase	Analytics solutions only; collaboration with Intel on platform (agnostic to source)	Service for RWD analytics
TriNetX (https://www.trinetx.com/our-network/)	HIC mostly; partnership for LMIC data	>300 million patient records representing 17 countries	Service for data provision and analytics

Table 4 Professional societies offering RWD programming and educational/networking opportunities

Professional society	Primary focus	Global health application?
ISPOR: Professional Society for Health Economics and Outcomes Research (ispor.org)	Health economics	Yes
ISPE: International Society for Pharmacoepidemiology (pharmacoepi.org/)	Epidemiology	Yes
ISACA: Information Systems Audit and Control Association (iaaca.org) Industry (abpi.org.uk/)	IT/methodology	Not directly
NEHI: Network for Excellence in Health Innovation (nehi. net)	Health sciences	Yes
IMI-GetReal consortium (imi-getreal.eu/)	Pharmaceutical R&D	Yes
ismpp: International Society for Medical Publication Professionals (ismpp.org/)	Methodology/ data access	Yes
ICC: International Chamber of Commerce (iccwbo.org/)	Financial services	Yes
ABPI: Association of the British Pharmaceutical Industry	Pharmaceutical R&D	Yes

An important resource for those getting trained to work in this field is the access to mentors and other resources that would benefit young data scientists. For this purpose, professional societies have recognized the potential of RWD to influence healthcare from a variety of vantage points and provided programming, networking, and educational services. A short list of professional societies along with their primary focus is provided in Table 4 below.

Importance of Data Sharing

Global health research and policy recommendations face increasing demands for transparency at every stage, including an emphasis on improvements in data sharing (Taichman et al. 2016; Longo and Drazen 2016). Data sharing is not without challenge and controversy, particularly as it incurs additional costs, in both time and money, for researchers with limited budgets and timelines. Nevertheless, the likelihood that most or nearly all data pertinent to clinical trials and RWE analysis may be made public may reinforce the commitment of stakeholders to adhere to the highest standards of RWE analysis. More diverse RWD serving the GH ecosystem suffers similarly. Transparency and data sharing may also deter those tempted to release studies that are inadequate, misleading, or even falsified.

Important: The potential for RWD to inform GH initiatives will only be realized with strong and transparent sharing policies with expectation on data quality and recognized governance procedures.

Stakeholders need to commit themselves to achieving a standard in which researchers will post study designs, assumptions, and data sources prior to beginning the research process, at least in cases where analysis is likely to have major implications for public health. Some sources of RWD are likely to enjoy a higher level of trust than others. Data sets of proven quality will typically include well-established and long-running, or longitudinal, data sets that encompass data elements that are well accepted in clinical practice and clinical research, allowing for comparative analysis and for generalizability of results. Newer, less familiar, or novel data sets should be accessible for review and for comparison to similar data sets that meet the test of reliability and comparability.

Patients must not be left behind either. The last decade has witnessed increased health-related research in resource-poor settings mainly in Africa, Asia, and South America. While this increase is a positive outcome for addressing neglected diseases, it has also raised concerns over potential for exploitation through unfair distribution of risks and benefits among the parties involved. Several strategies to address these concerns have been promoted in the ethics literature including calls for universal standards of care, reasonable availability of proven interventions, and, more recently, promotion of the overall social value of research (Lairumbi et al. 2012). Closely related to the idea of promoting the social value of research is the determination of fair benefits through the consideration of what is owed to those participating in research and their communities, a process that has come to be referred to as benefit sharing. Although there now seems to be an international consensus on the need to share benefits arising from global health research, there has been continued debate over what constitutes a fair benefit, whether those that address the micro level issues of justice (those relating to individual circumstances of those participating in research) or those focusing on the macro level (the broader issues that might predispose participants to exploitation). The benefit of recent cross-sectional sharing on stimulating rapid innovation is indisputable so one would hope that the ethical considerations can be appropriately dealt with without stagnating the progress made thus far.

RWD Use Cases

The challenges of generating, analyzing, and applying RWD are particularly problematic in low- and middle-income countries. Given the size of the populations in question and the complexity of current healthcare delivery in these geographic areas (many diseases and thousands of medications and interventions), the reconciliation of data-driven improvements in clinical strategies with good population health is complex. The recent development of new methods to collect, analyze, and apply data has narrowed the gap between healthcare delivery and population health and created a vision of how these two settings can be bridged and with the potential for improved health outcomes. These new methods of collecting, curating, and conceptualizing data offer advantages to populations to see into their health trajectories with greater precision (Wyber et al. 2015).

Challenges notwithstanding, there are good examples which illustrate the diversity of RWD exploitation to the benefit of global health initiatives. Table 4 provides representative, recent examples which illustrate the growing confidence in the approach as well as the diversity of the applications. What is apparent in the few case studies shown is not only the diversity of the application of RWD but the variation in RWD sources and the necessity of merging and interpreting disparate data types, not to mention the differences in stakeholders requesting the analyses.

The health economics example (Campioni et al. 2019) provides an excellent example of the recent trend by pharmaceutical sponsors to appeal to healthcare decision-makers in targeted geographic areas of interest. The study is sponsored by the pharmaceutical industry but conducted and analyzed in collaboration with academic investigators treating the target population. RWD sources represent a combination of RCT data with observational, cost-effectiveness databases. While the focus of the analysis and models developed are based on the payers perspective, the cost-effectiveness and budget constraint recognition is clearly targeted to position the sponsor's contribution to combination therapy in this case as superior to alternative choices. Still, as the world economy gravitates to a more value-based healthcare system overall, these studies and analyses conducted in the target patient populations (multiple myeloma patients in the Czech Republic in this case) using relevant real-world metrics (quality-adjusted life year (QALY) gained) are essential to generated informed healthcare policy decisions. One hopes that these types of studies are also directed to LICs in the future (Table 5).

Similar in intention, the regulatory example (Wedam et al. 2019) illustrates the merger of RWD from electronic health records and insurance claims with safety data (from global safety databases that sponsors are obligated to collect as part of their surveillance programs), RCT data, and early Phase 1 safety data. This analysis provides an early example of what is likely to be an increasing trend in the pharmaceutical industry and allows regulators an opportunity to judge the value of the information in the approval process without lowering the regulatory requirements (thus far). Most importantly, from the GH perspective, this allows a baseline assessment from which one can judge the impact of data quality, reliability, and methodologies from well-defined and curated sources, all of which will be of a concern for data coming from LICs.

The social determinants example (Douthit and Seema 2018) is a somewhat unconventional choice for inclusion in the table to the extent that the source data is not the "big data" RWD type that we typically associate with these use cases. Rather, it refers to 70 *British Medical Journal* (BMJ) case reports from five continents, written by doctors, nurses, students, and allied health professionals. These cases, a burgeoning repository of evidence of how real patients are affected by disease, trauma, violence, sexual assault, conflict, migration, adverse living and working conditions, and poor access to healthcare, discuss, in addition to clinicopathological findings, need to be provided to the patient populations that would stand to benefit from the shared experience. The example highlights the necessity of patient and provider engagement in GH solutions and illustrates the current problems with poor dissemination of critical knowledge that would inform decision-making

Table 5 RWD use case examples with global health implications

Use case and reference	Category	RWD sources	Impact
Health economics (Campioni et al. 2019)	Cost-effectiveness: (carfilzomib in combination with lenalidomide and dexamethasone versus lenalidomide and dexamethasone in relapsed multiple myeloma patients)	HER and registry data of monoclonal gammopathies from Czech Republic cancer patients	The reimbursement of carfilzomib in combination with lenalidomide and dexamethasone represents an efficient allocation of resources within the healthcare system
Support for NDA approval Supplemental support for RCT data (Wedam et al. 2019)	Regulatory decision	HERs of male patients with breast cancer treated with palbociclib, complimentary insurance claim data, information from the global safety database and from two studies in male patients with other solid tumors, post-marketing reports for palbociclib from the FDA Adverse Event Reporting System (FAERS)	Supportive evidence for approval of palbociclib based on RWD analysis
Social determinants of health (Douthit and Seema 2018)	Education and advocacy	BMJ case reports (from five continents, written by doctors, nurses, students, and allied health professionals)	Repository documents the determinants of health as they affect real patients, to use this evidence base in advocating for better health and better healthcare for individuals and societies and to take on global health problems at the level of the individual
Information campaign strategies in country-wide epidemics (Lima et al. 2013)	Disease containment	Cellular network data	Model describes the spreading of disease in a population where individuals move between geographic areas; evolution of the disease described and countermeasures (the quarantine of central geographic areas and a collaborative "viral" information campaign) evaluated

for those that often lack advocates. It also provides a roadmap of how culturally and medically appropriate care was supplemented by a strong universal healthcare system which included ethnic minorities, regular visits from culturally competent nurses and physicians, and strong social support.

Finally, the disease containment and information campaign strategy example (Lima et al. 2013) provides a means of establishing patterns in cellular data to examine the impact of habits linked to healthcare outcomes. Specifically, the example illustrates how human mobility is linked with the spread of disease. Four different data sets containing information about user mobility and call patterns at various levels of granularity and time duration were used to build a model which projects mobility and information spreading. The spreading model was then used in a simulation context to describe how the process progresses in the presence of contagion and mobility. Such examples also provide a mechanism of how such approaches can be used in diverse parts of the world where the mechanisms for the spread of disease are not the same and where mobility is an important factor in disease containment.

Conclusion

There are many stakeholders who stand to benefit from continued investment in RWD to advance GH solutions. The pharmaceutical industry faces challenges to find ways to make its innovative medicines available to patients. The interests of the global healthcare system including the interests of patients, prescribers, payers, and regulators, and the need to measure disease burden, create a complex environment for quantifying clinical value in diverse socioeconomic settings with varying support on the infrastructure side. Likewise, the healthcare industry will continue to be both a generator and exploiter of RWD. The delivery of healthcare is a complex endeavor at both individual and population levels. At the clinical level, the provision of care to individuals is guided, in part, by medical history, examination, vital signs, and evidence. More recently, these traditional tenets have been supplemented by a focus on learning, metrics, and quality improvement. The collection and analysis of data of good quality are critical to improvements in the effectiveness and efficiency of healthcare delivery.

There are also new frontiers for RWD. Specifically, there is a compelling case for the value of greater use of person-generated health data (PGHD) for regulatory and other policy decisions, and especially for integrating these data with other real-world data sources (Taylor et al. 2016; WHO 2015).

> Definition *Patient-generated health data (PGHD) are health-related data created, recorded, or gathered by or from patients (or family members or other caregivers) to help address a health concern.*

Key areas considered where PGHD could add value to clinical research included defining patient subtypes, allowing more frequent measurements of outcomes, and creating more meaningful outcome metrics that measure patient preferences and improvement in their lived experiences. As we move toward a patient-centered health system in the age of personalized medicine, data from patients themselves will be needed to establish more robust patient profiles on how medications with the complexity of various comorbidities are influencing specific outcomes. Having these data available allows researchers and caregivers to ensure the right patient populations are being studied and treated appropriately for specific outcomes.

When PGHD are linked to other RWD sources such as curated, disease-specific registries, they can provide important insight into the patient's clinical care journey and help accelerate biomedical breakthroughs. Linked with multiple RWD sources, PGHD could provide value when developing and validating patient-centered outcomes, particularly those outcomes that may not be easily identified through algorithms that rely on a single data source. These opportunities could be highlighted in a future set of "use cases" – potential applications to show their feasibility and impact on clinical evidence.

Data curation processes will be critical for improving the reliability and characterizing the validity of these data to ensure fitness for use to answer study questions. While there is a growing diversity of PGHD being collected by different systems, many questions remain about what data curation initiatives could improve the quality and assess the validity of PGHD. As the FDA and other regulatory agencies consider how to leverage RWD to inform key decisions, standardized approaches for evaluating these data and curation processes to ensure their fitness for the intended use are necessary. This will require consistent and scalable approaches that can be systematically applied to track the life cycle of PGHD.

Ultimately, policy makers and data users will need to answer the question, "Does this data meet expectations for regulatory use?." This will require well-documented processes related to data generation and provenance, as well as understanding what data quality checks were used to assess the data set's fitness. In addition, an appropriate governance framework must be developed and enforced to protect individuals and ensure that healthcare delivery is tailored to the characteristics and values of the target communities. The big data approach (Wyber et al. 2015) could be associated with health data that are owned by patients; robust governance processes that have been developed to ensure respect of values and principles in the use of data, with an emphasis on risk minimization; data that are aggregated automatically, with little effort and decreasing cost; interoperability standards that allow data to be seamlessly pooled and connected; laws that, while establishing adequate safeguards, allow the sharing and pooling of anonymized data in real time; and data that are presented in a usable format to patients, healthcare providers, entrepreneurs, and policy makers.

Clearly the path to ensuring the sustainability of RWD usage to address challenging GH problems is not straightforward. Nonetheless, there should be great optimism that this effort can be moved forward with confidence. It will require leadership from key GH stakeholders and a willingness from relevant governments

to invest in infrastructure improvements while providing shared access to RWD sources. Data standards and broad transparency consideration are essential along with good communication. Some centralization of governance of such data with an "honest broker" model would also be helpful. The commercialization of RWD is well on the way to providing high-quality sources for commercial interests. While much of these efforts are currently dominated with data from HICs, one would hope that LMIC data will eventually represent a meaningful portion of these sources in the future. It should also be appreciated that these companies (CROs mostly) should move beyond the "use at your own risk" moniker and facilitate realistic solutions to data quality concerns. These are all solvable issues, and there is every expectation that future review of RWD analyses for GH solutions represents a significant advance to current efforts.

References

Abaasa A, Nash S, Mayanja Y, Price M, Fast PE, Kamali A, Kaleebu P, Todd J (2019) Simulated vaccine efficacy trials to estimate HIV incidence for actual vaccine clinical trials in key populations in Uganda. Vaccine 37:2065–2072

Amirthalingam G et al (2014) Effectiveness of maternal pertusis vaccination in England: an observational study. Lancet 384(9953):1521–1528. https://doi.org/10.1016/s0140-6736(14)60686-3

Breckenridge AM, Breckenridge RA, Peck CC (2019) Report on the current status of the use of real-world data (RWD) and real-world evidence (RWE) in drug development and regulation. Br J Clin Pharmacol 85:1874–1877. https://doi.org/10.1111/bcp.14026

Campioni M, Agirrezabal I, Hajek R, Minarik J, Pour L, Spicka I, Gonzalez-McQuire S, Jandova P, Maisnar V (2019) Methodology and results of real-world cost-effectiveness of carfilzomib in combination with lenalidomide and dexamethasone in relapsed multiple myeloma using registry data. Eur J Health Econ. https://doi.org/10.1007/s10198-019-01122-6

Chan M, Kazatchkine M, Lob-Levyt J, Obaid T, Schweizer J et al (2010) Meeting the demand for results and accountability: a call for action on health data from eight global health agencies. PLoS Med 7(1):e1000223. https://doi.org/10.1371/journal.pmed.1000223

Douthit NT, Seema B (2018) Global Health Education and Advocacy: Using BMJ Case Reports to Tackle the Social Determinants of Health. Frontiers in Public Health 6:2296–2565

GBD (2015) Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease study 2015. Lancet. 2016; 388(10053):1545–1602. https://doi.org/10.1016/S0140-6736(16)31678-6. Erratum in: Lancet. 2017. 389(10064):e1. PMID: 27733282; PMCID: PMC5055577

Katkade VB, Sanders KN, Zou KH (2018) Real world data: an opportunity to supplement existing evidence for the use of long-established medicines in health care decision making. J Multidiscip Healthc 11:295–304. https://doi.org/10.2147/JMDH.S160029. PMID: 29997436; PMCID: PMC6033114

Kronick R (2016) AHRQ'S role in improving quality, safety, and health system performance. Public Health Rep 131:229–232

Lairumbi GM, Parker M, Fitzpatrick R, English MC (2012) Forms of benefit sharing in global health research undertaken in resource poor settings: a qualitative study of stakeholders' views in Kenya. Philos Ethics Humanit Med 7:7

Lima A, De Domenico M, Pejovic V, Musolesi M (2013) Exploiting cellular data for disease containment and information campaign strategies in country-wide epidemics. Cornell University Library, Ithaca. http://arxiv.org/pdf/1306.4534v1.pdf

Longo DL, Drazen JM (2016) Data sharing. N Engl J Med 374: 276–277. http://www.nejm.org/doi/full/10.1056/NEJMe1516564#t=article

McGettigan P, Alonso Olmo C, Plueschke K, et al (2019) Patient Registries: An Underused Resource for Medicines Evaluation: Operational proposals for increasing the use of patient registries in regulatory assessments. Drug Saf 42(11):1343–1351. https://doi.org/10.1007/s40264-019-00848-9

Murray CJL, Lopez AD, World Health Organization, World Bank, Harvard School of Public Health (1996) The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020: summary (eds: Murray CJL, Lopez AD). World Health Organization, Geneva. http://www.who.int/iris/handle/10665/41864

Murray CJL et al (2015) Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990–2013: quantifying the epidemiological transition. Lancet 386:2145–2191

Nord E (2013). Disability weights in the global burden of disease 2010: unclear meaning and overstatement of international agreement. Health Policy 111(1):99–104.

Salomon J (2008). Measurement of disability weights in the global burden of disease 2005. In: GBD disability weights expert consultation, Seattle, 4–5 September 2008. Institute for Health Metrics and Evaluation, University of Washington, Seattle

Salomon J (2013) Disability weights measurement in the Global Burden of Disease study 2010 [slides]. In: Global health metrics and evaluation conference, Seattle, 18 June 2013. Institute for Health Metrics and Evaluation, University of Washington, Seattle. http://www.slideshare.net/IHME/disability-weightsmeasurement-in-the-global-burden-of-disease-study-2010

Salomon JA, Murray CJL, Ustun TB, Chatterji S (2003) Health state valuations in to meet the population measures of population health. In: Murray CJL, Evans D (eds) Health systems performance assessment: debate, methods and empiricism. World Health Organisation, Geneva

Sienkiewicz T (2019) Global health and welfare: 20 leading medical school programs, May 15, 2019. https://www.petersons.com/blog/global-health-and-welfare-20-leading-medical-school-programs/

Srivastava SK (2016) Adoption of electronic health records: a roadmap for India. Healthc Inform Res 22(4):261–269

Stone I (2014) 30 great European universities for studying healthcare abroad, Posted November, 2014. https://www.topmastersinhealthcare.com/best/european-universities-studying-healthcare-abroad/

Taichman DB, Backus J, Baethge C, Bauchner H, de Leeuw PW, Drazen JM, Fletcher J, Frizelle FA, Groves T, Haileamlak A, James A, Laine C, Peiperl L, Pinborg A, Sahni P, Wu S (2016) Sharing clinical trial data: a proposal from the inter- national committee of medical journal editors. Annof Intern Med 164(7):505–506. http:// www.icmje.org/news-and-editorials/M15-2928-PAP.pdf

Taylor LA, Tan AX, Coyle CE, Ndumele C, Rogan E, Canavan M, Curry LA, Bradley EH (2016) Leveraging the social determinants of health: what works? PLoS ONE 11(8):1–20. https://pdfs.semanticscholar.org/c2b0/db1a55fddf5cbc685725d919f7bfbc869ed5.pdf

Wedam S, Fashoyin-Aje L, Bloomquist E, Tang S, Sridhara R, Goldberg KB, Theoret MR, Amiri-Kordestani L, Pazdur R, Beaver JA (2019) FDA approval summary: palbociclib for male patients with metastatic breast cancer. Clin Cancer Res. pii: clincanres.2580.2019. https://doi.org/10.1158/1078-0432.CCR-19-2580

Withers M, Press D, Wipfli H, McCool J, Chan C-C, Jimba M, Tremewan C, Samet J (2016) Training the next generation of global health experts: experiences and recommendations from Pacific Rim universities. Glob Health 12:34

World Health Organization (2015) WHO global strategy on people-centred and integrated health services: interim report. WHO, Geneva

Wyber R, Vaillancourt S, Perry W, Mannava P, Folaranmi T, Celi LA (2015) Big data in global health: improving health in low- and middle-income countries. Bull World Health Organ 93:203–208. https://doi.org/10.2471/BLT.14.139022