



Harnessing real-world evidence to reduce the burden of noncommunicable disease: health information technology and innovation to generate insights

Kelly H. Zou¹ · Jim Z. Li² · Lobna A. Salem¹ · Joseph Imperato¹ · Jon Edwards³ · Amrit Ray⁴

Received: 15 May 2020 / Revised: 1 September 2020 / Accepted: 9 October 2020 /
Published online: 6 November 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Noncommunicable diseases (NCDs) are the leading causes of mortality and morbidity across the world and factors influencing global poverty and slowing economic development. We summarize how the potential power of real-world data (RWD) and real-world evidence (RWE) can be harnessed to help address the disease burden of NCDs at global, national, regional and local levels. RWE is essential to understand the epidemiology of NCDs, quantify NCD burdens, assist with the early detection of vulnerable populations at high risk of NCDs by identifying the most influential risk factors, and evaluate the effectiveness and cost-benefits of treatments, programs, and public policies for NCDs. To realize the potential power of RWD and RWE, challenges related to data integration, access, interoperability, standardization of analytical methods, quality control, security, privacy protection, and ethical standards for data use must be addressed. Finally, partnerships between academic centers, governments, pharmaceutical companies, and other stakeholders aimed at improving the utilization of RWE can have a substantial beneficial impact in preventing and managing NCDs.

Keywords Real-world data · Real-world evidence · Population health · Health information technology · Noncommunicable disease · Disease burden · Risk factors · Data science

Disclaimer This paper was funded by Upjohn, a Division of Pfizer Inc. The authors, Kelly H. Zou, Jim Z. Li, Lobna A. Salem, Joseph Imperato, and Amrit Ray are employees of the Upjohn Division of Pfizer Inc, and the author Jon Edwards is an employee of Envision Pharma Ltd. The views expressed are their own and do not necessarily represent those of their employers. Envision Pharma Ltd has a consultancy agreement with Pfizer Inc.

✉ Kelly H. Zou
Kelly.Zou@pfizer.com

Extended author information available on the last page of the article

1 Introduction

Noncommunicable diseases (NCDs) are the leading causes of mortality globally and a major global health challenge that affects people in all countries, regardless of their socio-economic status (World Health Organization 2018; GBD 2017 Disease and Injury Incidence and Prevalence Collaborators 2018).

Approximately 30% of NCD-related deaths are considered premature, occurring before the age of 70 years, and approximately 85% of premature deaths occur in low- and middle-income countries (World Health Organization 2018). However, the prevalence of many NCDs increases with age (GBD 2017 Disease and Injury Incidence and Prevalence Collaborators 2018; Global Burden of Disease Study Collaborators 2015). The continued growth of the aging population, which is often pronounced in developed economies with relatively advanced healthcare systems, will increase the prevalence of NCDs and accentuate disease burden.

The World Health Organization (WHO) has identified 4 main types of NCDs that contribute the greatest burden. They are cardiovascular diseases (CVDs), cancers, chronic respiratory diseases, and diabetes (Fig. 1). In 2016, NCDs accounted for 71% (41 million) of deaths worldwide, with 44% of these deaths attributable to CVD, 22% to cancer, 9% to chronic respiratory disease, and 4% to diabetes (World Health Organization 2018). In

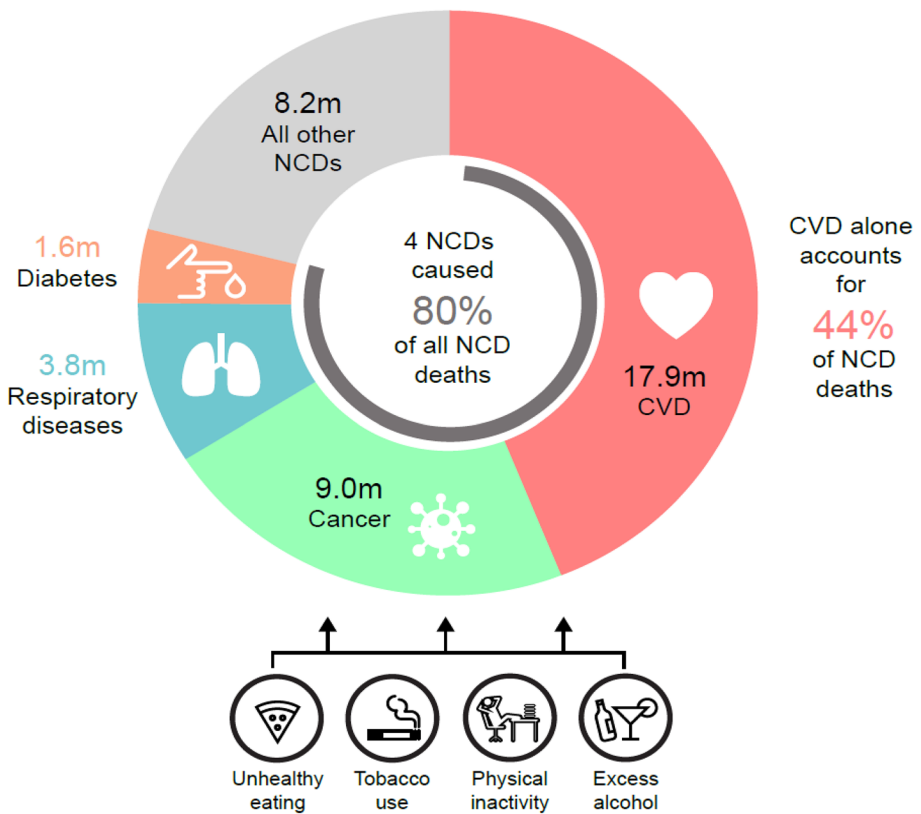


Fig. 1 Global annual deaths by key NCDs; source: World Health Organizations (World Health Organization 2018)

the US, CVD accounted for approximately 1 of every 3 deaths that year (Benjamin et al. 2019). In addition, neurological conditions and mental health disorders, such as anxiety disorders, migraine, major depressive disorder, bipolar disorder, and Alzheimer's disease, have emerged as major causes of disability (Global Burden of Disease Study Collaborators 2015). Indeed, mental health and well-being have been highlighted by the United Nations as important components of the goal to reduce premature mortality from NCDs by 33% over the next 10 years (United Nations 2020).

NCDs are characterized by their long duration or continual recurrence (i.e., chronic) and slow progression, and many NCDs have high prevalence. For example, hypertension alone is diagnosed in over 20% of adults in a vast majority of countries across North America, Latin America, Europe, Asia, Africa, and Oceania (Clarivate Analytics 2020). NCDs are also major causes of morbidity and disability, including anxiety, depression, pain, and mobility impairment (Lisy et al. 2018). Data from the Global Burden of Disease study showed that the most common chronic sequelae, as consequences of disease, are largely attributable to NCDs (Global Burden of Disease Study Collaborators 2015).

Those who are most socioeconomically disadvantaged are often at a high risk of developing NCDs (Murray et al. 2005; Marmot and Bell 2019; Nulu 2017). Furthermore, in the era of COVID-19, NCDs pose an even greater threat. According to the Centers for Disease Control and Prevention, "people of any age with the following conditions are at increased risk for COVID-19: cancer; chronic kidney disease; chronic obstructive pulmonary disease (COPD); immunocompromised state from solid organ transplant; obesity (body mass index [BMI] of 30 or higher); CVD; sickle cell disease; and Type 2 diabetes mellitus" (Centers for Disease Control and Prevention 2020).

The economic costs of NCDs are burdensome to countries around the world (Bloom et al. 2011). Besides NCD treatment and control, the negative impact also includes the reduced productivity at work, more days absent from work (absenteeism), and early retirement, both for individuals with NCDs and due to premature death. These costs can lower family economic status, as well as national economic output. Furthermore, national funds that need to be deployed toward treatment of NCDs take away the funds that might otherwise be invested in infrastructure, research, and education (Chen et al. 2018). Studies also suggest that the medical and economic burdens due to NCDs, which are already high, continues to increase, especially in less-developed economies, as well as in middle- and high-income countries (Global Burden of Disease Risk Factors Collaborators 2016; Timmis et al. 2020). Therefore, how the social determinants of health (SDOH) affect NCDs is an increasingly important area of focus (Marmot and Bell 2019).

Governments around the world and international organizations, such as the World Bank and WHO, are increasingly committed to NCD prevention and control (World Health Organization 2013). Efforts to reduce the burden of NCDs are increasingly shifted from treatment to prevention. This is because the majority of NCDs occur as a result of modifiable risk factors. Therefore, reducing and controlling these risk factors is an effective means of reducing the burden of NCDs.

The WHO has called for NCD prevention efforts to focus on the following modifiable behavioral risk factors (tobacco use, physical inactivity, the harmful use of alcohol and unhealthy diets; Fig. 1) and metabolic risk factors (raised blood pressure, overweight/obesity, hyperglycemia, and hyperlipidemia). Prevention not only reduces NCD-related suffering to patients but is also more cost-effective to societies. For example, programs targeted at preventing or treating NCDs can have a significantly beneficial effect, with recent estimates suggesting that every US \$1 spent on tackling NCDs will have a return of at least US \$7 in the following 10 years (World Health Organization 2014).

A global comprehensive approach is needed to reduce the burden of NCDs, which requires the collaboration across various sectors, including academia, industry and governments (Upjohn 2020). This approach should reduce the risk factors for NCD and promote interventions to prevent and control NCDs and is especially important during the COVID-19 pandemic (Hassan et al. 2020). Organizations such as the WHO have realized the importance of innovative data visualization to help educate on both the importance of reducing the global burden of NCDs and the stopping the spread of the COVID-19 during 2020. Real-world data (RWD) and real-world evidence (RWE) play an important role for international organizations, governments, and societies, helping them to make informed decisions regarding NCD prevention and control.

2 Harness RWE to prevent and control NCDs

2.1 RWD and RWE

According to the US Food and Drug Administration (FDA), RWD means data collected outside of the framework of randomized, controlled trials (RCTs) (US Food and Drug Administration 2019). Furthermore, RWE is generated through the analysis of RWD and is used to answer specific clinical and research questions. The lack of RCTs designed to assess the burden and degree of the comorbid conditions occurring with NCDs means that RWE is the greatest resource available to study this important area. Global advances in information technologies and telecommunication infrastructures have enabled a massive amount of RWD to be generated from diverse data sources. The wide variety of data sources include pharmacovigilance databases, electronic medical records (EMRs) including medical images/imaging data and free-text notes from healthcare providers, electronic health records (EHRs), administrative insurance claims, patient registries, population health surveys, medical researches including genomics studies, data collected from digital apps and digital recording devices including wearable devices, and various other sources (Fig. 2).

RWD can be claims and transactions for healthcare resource utilization, electronic health records, surveys, linked datasets, and other digital data collected outside a traditional clinical trial. Because RWD can be generated at a low cost (relative to RCTs) and rapidly (e.g., streamed from wearable devices), it is often stored and processed in considerable quantity. Thus, it can be so-called big data due to its “volume” (there’s a lot of it), “variety” (the data takes many different forms), “velocity” (the data changes or are updated frequently) and “veracity” (the data may be of poor/unknown quality) (IBM 2020; Seth 2014). The immense potential value of data in the modern world has led to it being described as the “the New Oil” (The Economist 2017). Pharmaceutical companies are now investing in their RWE programs to increase their capabilities in this arena, across all aspects of the drug development and approval process (Deloitte 2017; Davis et al. 2018; Morgan et al. 2020).

2.2 Opportunities

RWE is the foundation for the understanding of disease epidemiology, including rates of disease incidence and prevalence, awareness, diagnosis, treatment, and control. RWE plays

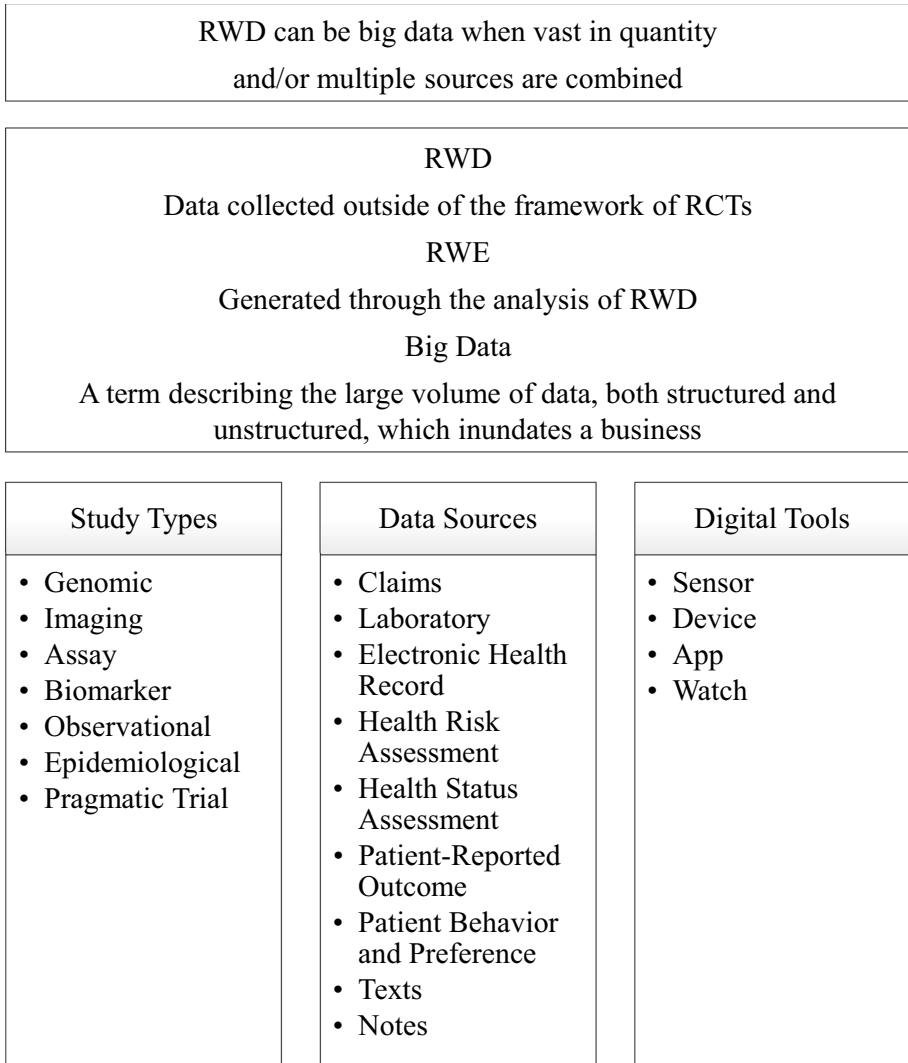


Fig. 2 RWD, RWE, and big data

a critical role in quantifying disease burden, which can be measured according to the following aspects: patients’ lives saved/lost, gain/loss of daily function, work productivity, and income; quality of life; healthcare resource usage. RWE can also be used to detect vulnerable populations (e.g., elderly and persons at high risk of NCDs) and identify the most influential risk factors, which may lead to new and non-traditional solutions to clinical problems (Batra and Cheung 2019). Recently, RWE has been increasingly used in the regulatory arena for gaining label expansion, as well as accelerating drug approvals due to regulatory authorities now being more receptive to reviewing RWE (Katkade et al. 2018; Zou et al. 2020).

There is an opportunity to use RWE to study diverse populations that are frequently underrepresented in both clinical and observational studies. For example, patients who

have certain diseases or clinical/demographic features (e.g. advanced age) are often excluded from RCTs (Kennedy-Martin et al. 2015). Nevertheless, how patients respond to treatment is important especially in areas such as NCDs, which are likely to impact many patients. For example, several observational RWD studies have identified commonly occurring NCDs as frequent comorbidities for COVID-19 (Yang and Jin 2020; Docherty et al. 2020; Richardson et al. 2020; Onder et al. 2020; Hassan et al. 2020). These comorbid NCDs such as CVD, chronic pulmonary disease and diabetes have also been demonstrated to worsen the clinical outcomes and increase the risk of death in those infected with COVID-19 (Docherty et al. 2020; Bergman et al. 2020). This understanding can help deploy preventive strategies to identify people at most risk of contracting severe COVID-19. This will in turn avoid overburdening the healthcare system. Finally, RWE can be used to measure and evaluate the effectiveness and cost-effectiveness of treatments, as well as programs targeting SDOHs (Li et al. 2020), and public policies for NCDs.

Global and societal efforts to reduce the populations' risk of NCDs (World Health Organization 2013, 2018) can be supported by using RWE to monitor incidence and prevalence trends of NCDs and risk factors, and to target prevention measures at populations that are vulnerable to NCDs. On a larger scale, risk reduction and increased access to interventions can be attained through collaboration among public health policymakers, payors, healthcare providers, and patient groups. Disease management can improve, as primary care providers adjust their treatment approaches, access to screening, detection, and treatment services, and more people have access to and can benefit from palliative care. The rationales for these changes will rely heavily on the availability and intelligent use of RWE. Only when stakeholders clearly understand the context and implications of NCDs can they begin to effect and direct changes. Another important aspect to consider is regional and racial differences and variations of the prevalence of NCDs and the various risk factors for these diseases (World Health Organization 2013, 2016). This aspect can provide key information as to the effectiveness of local policy initiatives targeted at NCDs as well as the influence of various factors and behaviors that could influence NCDs, e.g., smoking and obesity (Kontis et al. 2015; Office of Disease Prevention and Health Promotion 2020).

Although RCTs remain critical in determining treatment safety, efficacy, and mechanisms of action (Collins et al. 2020), their focus of patient selection and controlled clinical trial setting make it difficult to generalize the findings of RCTs to real-world clinical practice, which often has many confounding factors affecting the effectiveness of the treatment (Sherman et al. 2016). In addition, RCTs often focus on the effects of a single disease on outcomes, rather than multiple related conditions, which given the high levels of co-occurrence of NCDs, can be a limitation. Furthermore, RCTs are time-consuming and are becoming increasingly costly to conduct (Baumfeld Andre et al. 2020). This is especially true due to the chronic nature of the diseases being discussed. For example, patient follow-up in RCTs is often insufficient to have a clear understanding of patient safety. In contrast, RWE allows a treatment's effectiveness and tolerability to be evaluated in real-world practice, and therefore, RWE is important to assess the long-term safety of medications and can help identify rare adverse events (Collins et al. 2020). Furthermore, RWE provides tremendous opportunities to develop a more holistic understanding of patients and more effective approaches for comprehensive disease management. This is key to modifying patients' behaviors such as improving adherence with treatment, which in turn helps optimize outcomes (see below). While it is premature to suggest that RWE will replace RCTs, RWE provides an important complementary mechanism to RCTs for healthcare professionals seeking to find novel solutions to address the burden of NCDs.

For example, RWE has been used for many years to study adherence with medications to treat NCDs and evaluate methods of improving adherence with these medications. Poor adherence to medications to treat NCDs is a major global issue. Low adherence with treatment increases the morbidity and mortality burden of NCDs, even in high-income countries (Khan and Socha-Dietrich 2018; Brown and Bussell 2011; Cutler et al. 2018) where effective therapies are available, disease burden can only be reduced if patients adhere to the treatment for the prescribed duration (World Health Organization 2003; Shau et al. 2019). Improving patient adherence to existing interventions increases treatment effectiveness, resulting in significant overall cost-savings associated with disease burden. Therefore, an additional key use of RWE in NCD prevention and control is to better understand how we can improve the complex issues surrounding treatment adherence and persistence (Cramer et al. 2008). To address this challenge, RWE can be used to estimate adherence and persistence rates and assess the factors associated with these rates (Chen et al. 2019). Such information can be translated into realistic plans based on authentic insights to improve the proportion of patients adherent with their therapeutic regimens.

2.3 Challenges

The availability of RWD may depend on technology, digitization, data capture systems, and data flow regulations (US Food and Drug Administration 2013). RWD is usually not collected for research purposes and thus can be messy and in many different formats or forms. Data accuracy, reliability, and quality must be taken into account when using RWD for research (Sherman et al. 2016). To make the most of these research opportunities, innovative digital/analytic capabilities and technologies must be enabled and elevated globally for use by medical researchers serving in industry, government, and academia. Given the variety of data sources for RWD, important issues arise around data integrity, integration, access, interoperability, standardization, quality control, security, privacy protection, and ethical standards for data use.

Informed consent is a key consideration in RWD. The US Food and Drug Administration in its guidance for Institutional Review Boards and Clinical Investigators states that “no investigator may involve a human being as a subject in research covered by these regulations unless the investigator has obtained the legally effective informed consent of the subject or the subject’s legally authorized representative. An investigator seeks such consent only under circumstances that provide the prospective subject or the representative sufficient opportunity to consider whether or not to participate and that minimize the possibility of coercion or undue influence. The information that is given to the subject or the representative shall be in language understandable to the subject or the representative. No informed consent, whether oral or written, may include any exculpatory language through which the subject or the representative is made to waive or appear to waive any of the subject’s rights, or releases or appears to release the investigator, the sponsor, the institution, or its agents from liability for negligence” (US Food and Drug Administration 1998). However, informed consent needs to be carefully examined and discussed. Nevertheless, informed consent can limit the access to certain types of data or information from some regions. Advances in the blockchain technology can also enable dynamic informed consent (Mamo et al. 2020).

When data are available and in aggregate, cutting-edge innovations might be introduced to address such challenges. This is because, in order to be of any use, structured data (i.e., encoded in a standardized data format) and unstructured data (e.g., free-text physician

notes) from various sources may be combined and curated into databases using interoperable systems, which can be complicated. This is further complicated when databases obtained from various divergent sources are housed in separate repositories and data warehouses, which creates a major barrier to accessing and collating the relevant data.

Language differences between countries also present a common challenge in this regard. Furthermore, in many countries, data are collected and stored in siloed patient registries in non-standardized ways. These registries are often not connected with each other and contain diverse sets of information, which complicates analysis further.

Even within the same country, such as the United States, linking patient-level data across various sources can be problematic. Sometimes this can be accomplished using tokens that are generated from certain subsets of patient protected health information under the Health Insurance Portability and Accountability Act of 1996 (HIPAA) rule (US Department of Health & Human Services 2013, 2020; Centers for Disease Control and Prevention 2018). Other times, linkage can occur according to dates or locations of the treatment (Curtis et al. 2014). In many instances, some level of probabilistic matching, such as propensity score matching, may be required across multiple data sources or datasets (Desai et al. 2017). Sensitivity analyses based on different probabilistic matches can be considered to examine the robustness of the matching algorithm.

Databases can also be difficult to standardize; different diagnostic and treatment codes need to be harmonized (Observational Health Data Sciences and Informatics 2020). Similarly, it is important to establish and utilize standard and specific analytical methods and algorithms. For example, the Observational Health Data Sciences and Informatics (OHDSI) group is at the forefront of harmonizing and standardizing the analysis of observational data and has created a common-data model (Hripcsak et al. 2015). This, along with similar initiatives, will improve data reproducibility and the potential for groups to collaborate while working towards a common aim.

An additional complication in the analysis of RWD can be data and privacy protection laws, such as the California Consumer Privacy Act of 2018 (CCPA) (State of California Department of Justice 2018) or the European Union General Data Protection Regulation (GDPR) (European Commission 2019), which will need to be considered carefully when utilizing data from the European Union, even without containing personal identifiable information.

Moreover, when utilizing data from digital devices, it is necessary to consider how best to accurately combine data. For example, glucose monitors may not calculate summary measures in exactly the same way, which necessitates getting the raw data and using software package such as “cgmanalysis” in R to compute values in a consistent manner (Vigers 2020).

Once these initial challenges are addressed, robust data curation and analytic algorithms must be derived that are tailored toward specific patient populations. These algorithms will aid the identification of persons with, or at risk of, NCDs and help to address the challenge of NCDs throughout a patient’s NCD journey, including detection, screening, diagnosis, treatment, and monitoring. To achieve this, the current shortage of data translators (who act as shepherds with subject-matter expert knowledge) and data scientists will need to be addressed.

The medical and scientific insights gained through RWE-generation can potentially impact global policies to reduce the burden of NCDs. In addition, RWE can help provide opportunities to analyze both cross-sectional and longitudinal data to assess and monitor the effectiveness of various interventions designed to target NCDs and risk factors for NCDs.

3 Conclusions

There is a critical need to address the substantial burden associated with NCDs. This is even more urgent in the COVID-19 era given the increased risk of poor clinical outcomes in patients with NCDs who become infected with COVID-19. Some of the limitations associated with RCTs, notably their often narrow focus, can be addressed by using RWD. Therefore, over the coming years, we anticipate seeing more of the power of RWD and RWE being harnessed to address the immense healthcare burden associated with NCDs. Innovative techniques for both capturing and analyzing data will be utilized. For example, a distributed research network is useful for generating RWE. To enhance the effectiveness and efficiency of healthcare delivery, it is important to understand the risk factors for disease progression, treatment patterns, and utilization. Educational initiatives regarding the potential for RWD and RWE to deliver patient-centric, value-based healthcare will lead to further improvements in the management of NCDs. However, it is also important to strike the right balance of understanding the challenges and limitations of RWD and RWE to ensure that they are utilized correctly.

Finally, partnerships can play a key role in improving the utilization of RWE. Fruitful collaborative research opportunities exist across different healthcare stakeholders, including academia, industry and government, based on health information technology and innovation for gaining valuable insights. This will, in turn, have a substantial beneficial impact on the prevention and management of NCDs.

References

- Batra, A., Cheung, W.Y.: Role of real-world evidence in informing cancer care: lessons from colorectal cancer. *Curr. Oncol.* **26**(Suppl 1), S53–S56 (2019)
- Baumfeld Andre, E., Reynolds, R., Caubel, P., Azoulay, L., Dreyer, N.A.: Trial designs using real-world data: the changing landscape of the regulatory approval process. *Pharmacoepidemiol. Drug Saf.* **29**(10), 1201–1212 (2020)
- Benjamin, E.J., Muntner, P., Alonso, A., Bittencourt, M.S., Callaway, C.W., Carson, A.P., Chamberlain, A.M., Chang, A.R., Cheng, S., Das, S.R., Delling, F.N., Djousse, L., Elkind, M.S.V., Ferguson, J.F., Fornage, M., Jordan, L.C., Khan, S.S., Kissela, B.M., Knutson, K.L., Kwan, T.W., Lackland, D.T., Lewis, T.T., Lichtman, J.H., Longenecker, C.T., Loop, M.S., Lutsey, P.L., Martin, S.S., Matsushita, K., Moran, A.E., Mussolino, M.E., O'Flaherty, M., Pandey, A., Perak, A.M., Rosamond, W.D., Roth, G.A., Sampson, U.K.A., Satou, G.M., Schroeder, E.B., Shah, S.H., Spartano, N.L., Stokes, A., Tirschwell, D.L., Tsao, C.W., Turakhia, M.P., VanWagner, L.B., Wilkins, J.T., Wong, S.S., Virani, S.S., American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee: Heart disease and stroke statistics-2019 update: a report from the American Heart Association. *Circulation* **139**(10), e56–e528 (2019)
- Bergman, M., Jagannathan, R., Narayan, K.M.V.: Nexus of COVID-19 and diabetes pandemics: global public health lessons. *Diabetes Res. Clin. Pract.* **164**, 108215 (2020)
- Bloom, D.E., Cafiero, E.T., Jané-Llopis, E., Abrahams-Gessel, S., Bloom, L.R., Fathima, S., Feigl, A.B., Gaziano, T., Mowafi, M., Pandya, A., Prettner, K., Rosenberg, L., Seligman, B., Stein, A.Z., & Weinstein, C. The global economic burden of non-communicable diseases. In: World Economic Forum, Geneva (2011)
- Brown, M.T., Bussell, J.K.: Medication adherence: WHO cares? *Mayo Clin. Proc.* **86**(4), 304–314 (2011)
- Centers for Disease Control and Prevention: Health Insurance Portability and Accountability Act of 1996 (HIPAA) (2018). <https://www.cdc.gov/php/publications/topic/hipaa.html>. Accessed 7 Aug 2020
- Centers for Disease Control and Prevention: People with Certain Medical Conditions (2020). <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html>. Accessed 7 Aug 2020

- Chen, S., Kuhn, M., Prettner, K., Bloom, D.E.: The macroeconomic burden of noncommunicable diseases in the United States: estimates and projections. *PLoS One* **13**(11), e0206702 (2018)
- Chen, S.T., Huang, S.T., Shau, W.Y., Lai, C.L., Li, J.Z., Fung, S., Tse, V.C., Lai, M.S.: Long-term statin adherence in patients after hospital discharge for new onset of atherosclerotic cardiovascular disease: a population-based study of real world prescriptions in Taiwan. *BMC Cardiovasc. Disord.* **19**(1), 62 (2019)
- Clarivate Analytics: Incidence and Prevalence Database (IDP) (2020). <http://www.tdrdata.com>. Accessed 2 January, 2020
- Collins, R., Bowman, L., Landray, M., Peto, R.: The magic of randomization versus the myth of real-world evidence. *N. Engl. J. Med.* **382**(7), 674–678 (2020)
- Cramer, J.A., Roy, A., Burrell, A., Fairchild, C.J., Fuldeore, M.J., Ollendorf, D.A., Wong, P.K.: Medication compliance and persistence: terminology and definitions. *Value Health* **11**(1), 44–47 (2008)
- Curtis, J.R., Chen, L., Bharat, A., Delzell, E., Greenberg, J.D., Harrold, L., Kremer, J., Setoguchi, S., Solomon, D.H., Xie, F., Yun, H.: Linkage of a de-identified United States rheumatoid arthritis registry with administrative data to facilitate comparative effectiveness research. *Arthritis Care Res. Hoboken* **66**(12), 1790–1798 (2014)
- Cutler, R.L., Fernandez-Llimos, F., Frommer, M., Benrimoj, C., Garcia-Cardenas, V.: Economic impact of medication non-adherence by disease groups: a systematic review. *BMJ Open* **8**(1), e016982 (2018)
- Davis, B., Morgan, J., Shah, S.: The Future of Real-World Evidence: Biopharma Companies Focus on End-to-End, AI-Driven, Internally Developed Solutions (2018). <https://www2.deloitte.com/be/en/pages/life-sciences-and-healthcare/articles/the-future-of-real-world-evidence.html?>. Accessed 7 Aug 2020
- Deloitte: Getting Real With Real-World Evidence (RWE): 2017 RWE benchmark survey (2017). <https://www2.deloitte.com/us/en/pages/life-sciences-and-health-care/articles/real-world-evidence-benchmarking-survey.html>. Accessed 7 Aug 2020
- Desai, J.R., Hyde, C.L., Kabadi, S., St Louis, M., Bonato, V., Katrina Loomis, A., Galaznik, A., Berger, M.L.: Utilization of positive and negative controls to examine comorbid associations in observational database studies. *Med. Care* **55**(3), 244–251 (2017)
- Docherty, A.B., Harrison, E.M., Green, C.A., Hardwick, H.E., Pius, R., Norman, L., Holden, K.A., Read, J.M., Dondelinger, F., Carson, G., Merson, L., Lee, J., Plotkin, D., Sigfrid, L., Halpin, S., Jackson, C., Gamble, C., Horby, P.W., Nguyen-Van-Tam, J.S., Ho, A., Russell, C.D., Dunning, J., Openshaw, P.J., Baillie, J.K., Semple, M.G.: ISARIC4C investigators.: Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ* **369**, m1985 (2020)
- European Commission: Communication from the commission to the European Parliament and the Council: data protection rules as a trust-enabler in the EU and beyond—taking stock. In. Brussels, Belgium (2019)
- GBD 2017 Disease and Injury Incidence and Prevalence Collaborators: Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **392**(10159), 1789–1858 (2018)
- Global Burden of Disease Risk Factors Collaborators: Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**(10053), 1659–1724 (2016)
- Global Burden of Disease Study Collaborators: Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **386**(9995), 743–800 (2015)
- Hassan, T.A., Saenz, J.E., Li, J.Z., Ducinskiene, D., Imperato, J., Zou, K.H.: A confluence of acute and chronic diseases: risk factors among Covid-19 patients. *Signif. Mag.* (2020). <https://www.significancemagazine.com/science/671-a-confluence-of-acute-and-chronic-diseases-risk-factors-among-covid-19-patients>. Accessed 7 Aug 2020
- Hripscak, G., Duke, J.D., Shah, N.H., Reich, C.G., Huser, V., Schuemie, M.J., Suchard, M.A., Park, R.W., Wong, I.C., Rijnbeek, P.R., van der Lei, J., Pratt, N., Noren, G.N., Li, Y.C., Stang, P.E., Madigan, D., Ryan, P.B.: Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud. Health Technol. Inform.* **216**, 574–578 (2015)
- IBM: The Four V's of Big Data (2020). <https://www.ibmbigdatahub.com/infographic/four-vs-big-data>. Accessed 7 Aug 2020
- Katkade, V.B., Sanders, K.N., Zou, K.H.: 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 (2018)

- Kennedy-Martin, T., Curtis, S., Faries, D., Robinson, S., Johnston, J.: A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials* **16**, 495 (2015)
- Khan, R., Socha-Dietrich, K.: Investigating in Medication Adherence Improves Health Outcomes and Health System Efficiency: Adherence to Medicines for Diabetes, Hypertension, and Hyperlipidemia. OECD Health Working Papers No. 105 (2018)
- Kontis, V., Mathers, C.D., Bonita, R., Stevens, G.A., Rehm, J., Shield, K.D., Riley, L.M., Poznyak, V., Jabbour, S., Garg, R.M., Hennis, A., Fouad, H.M., Beaglehole, R., Ezzati, M.: Regional contributions of six preventable risk factors to achieving the 25 × 25 non-communicable disease mortality reduction target: a modelling study. *Lancet Glob. Health* **3**(12), e746–e757 (2015)
- Li, J.Z., Zou, K.H., Woodruff, S., Imperato, J., Salem, L.A.: Medical analytics on social determinants of health for non-communicable disease prevention and control. Paper Presented at the Joint Statistical Meeting, Philadelphia, Pennsylvania (2020)
- Lisy, K., Campbell, J.M., Tufanaru, C., Moola, S., Lockwood, C.: The prevalence of disability among people with cancer, cardiovascular disease, chronic respiratory disease and/or diabetes: a systematic review. *Int J Evid Based Healthc* **16**(3), 154–166 (2018)
- Mamo, N., Martin, G.M., Desira, M., Ellul, B., Ebejer, J.P.: Dwarna: a blockchain solution for dynamic consent in biobanking. *Eur. J. Hum. Genet.* **28**(5), 609–626 (2020)
- Marmot, M., Bell, R.: Social determinants and non-communicable diseases: time for integrated action. *BMJ* **364**, 1251 (2019)
- Morgan, J., Feghali, K., Shah, S., Miranda, W.: RWE Focus is Shifting to R&D, Early Investments Begin to Pay Off (2020). https://www2.deloitte.com/us/en/insights/industry/health-care/real-world-evidence-study.html?id=us:2sm:3li:4di_gl:5eng:6di. Accessed 7 Aug 2020
- Murray, C.J., Kulkarni, S., Ezzati, M.: Eight Americas: new perspectives on U.S. health disparities. *Am. J. Prev. Med.* **29**(5 Suppl 1), 4–10 (2005)
- Nulu, S.: Neglected chronic disease: the WHO framework on non-communicable diseases and implications for the global poor. *Glob Public Health* **12**(4), 396–415 (2017)
- Observational Health Data Sciences and Informatics: OMOP Common Data Model (2020). <https://www.ohdsi.org/data-standardization/the-common-data-model/>. Accessed 7 Aug 2020
- Office of Disease Prevention and Health Promotion: Social Determinants of Health (2020). <https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-of-health>. Accessed 7 Aug 2020
- Onder, G., Rezza, G., Brusaferro, S.: Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *JAMA* **323**(18), 1775–1776 (2020)
- Richardson, S., Hirsch, J.S., Narasimhan, M., Crawford, J.M., McGinn, T., Davidson, K.W., the Northwell, C.-R.C., Barnaby, D.P., Becker, L.B., Chelico, J.D., Cohen, S.L., Cookingham, J., Coppa, K., Dieffenbach, M.A., Dominello, A.J., Duer-Hefele, J., Falzon, L., Gitlin, J., Hajizadeh, N., Harvin, T.G., Hirschwerk, D.A., Kim, E.J., Kozel, Z.M., Marrast, L.M., Mogavero, J.N., Osorio, G.A., Qiu, M., Zanos, T.P.: Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City Area. *JAMA* **323**(20), 2052–2059 (2020)
- Shau, W.Y., Lai, C.L., Huang, S.T., Chen, S.T., Li, J.Z., Fung, S., Tse, V.C., Lai, M.S.: Statin adherence and persistence on secondary prevention of cardiovascular disease in Taiwan. *Heart Asia* **11**(2), e011176 (2019)
- Sheth, A.: Transforming Big Data into Smart Data: Deriving Value via Harnessing Volume, Variety, and Velocity using Semantic Techniques and Technologies. In: 2014 IEEE 30th International Conference on Data Engineering (ICDE), pp. 2–2. Chicago, IL, USA (2014). <https://doi.org/10.1109/ICDE.2014.6816634>, <https://ieeexplore.ieee.org/document/6816634>. Accessed 7 Aug 2020
- Sherman, R.E., Anderson, S.A., Dal Pan, G.J., Gray, G.W., Gross, T., Hunter, N.L., LaVange, L., Marinac-Dabic, D., Marks, P.W., Robb, M.A., Shuren, J., Temple, R., Woodcock, J., Yue, L.Q., Calif, R.M.: Real-world evidence—What is it and what can it tell us? *N. Engl. J. Med.* **375**(23), 2293–2297 (2016)
- State of California Department of Justice: California Consumer Privacy Act (CCPA) (2018). <https://oag.ca.gov/privacy/ccpa>. Accessed 7 Aug 2020
- The Economist: Regulating the Internet Giants: the World’s Most Valuable Resource is no Longer Oil, But Data (2017). <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>. Accessed 7 Aug 2020
- Timmis, A., Townsend, N., Gale, C.P., Torbica, A., Lettino, M., Petersen, S.E., Mossialos, E.A., Maggioni, A.P., Kazakiewicz, D., May, H.T., De Smedt, D., Flather, M., Zuhlke, L., Beltrame, J.F., Huculeci, R., Tavazzi, L., Hindricks, G., Bax, J., Casadei, B., Achenbach, S., Wright, L., Vardas, P.: European Society of Cardiology.: European Society of Cardiology: cardiovascular disease statistics 2019. *Eur. Heart J.* **41**(1), 12–85 (2020)

- United Nations: Sustainable Development Goal 3: Ensure Healthy Lives and Promote Well-Being for all at all ages (2020). <https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf>. Accessed 7 Aug 2020
- Upjohn: Noncommunicable Diseases Worldwide: An Evidence-Based Review of Key Research and Strategies to Develop Sustainable Solutions - a White Paper (2020). https://pfe-pfizercom-d8-prod.s3.amazonaws.com/Upjohn_NCD_White_Paper-24JAN2020.pdf. Accessed 7 Aug 2020
- US Department of Health & Human Services: What is PHI? (2013). <https://www.hhs.gov/answers/hipaa/what-is-phi/index.html>. Accessed 7 Aug 2020
- US Department of Health & Human Services: Health Information Privacy (2020). <https://www.hhs.gov/hipaa/index.html>. Accessed 7 Aug 2020
- US Food and Drug Administration: A Guide to Informed Consent, Guidance for Institutional Review Boards and Clinical Investigators (1998). <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guide-informed-consent#general>. Accessed 7 Aug 2020
- US Food and Drug Administration: Guidance for Industry Electronic Source Data in Clinical Investigations (2013). <https://www.fda.gov/media/85183/download>. Accessed 7 Aug 2020
- US Food and Drug Administration: Real-World Evidence (2019). <https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence>. Accessed 7 Aug 2020
- Vigers, T.: Cgmanalysis: Clean and Analyze Continuous Glucose Monitor Data (2020). <https://cran.r-project.org/web/packages/cgmanalysis/index.html>. Accessed 7 Aug 2020
- World Health Organization: Health Statistics and Information Systems: Disease Burden and Mortality Estimates 2000–2016 (2016). https://www.who.int/healthinfo/global_burden_disease/estimates/en/. Accessed 7 Aug 2020
- World Health Organization: World Health Organization. Noncommunicable Diseases (NCDs) and Mental Health: Challenges and Solutions (2014). <https://www.who.int/nmh/publications/ncd-infographic-2014.pdf?ua=1>. Accessed 7 Aug 2020
- World Health Organization: Adherence to Long-Term Therapies (2003). https://www.who.int/chp/knowledge/publications/adherence_introduction.pdf. Accessed 7 Aug 2020
- World Health Organization: Global Action Plan for the Prevention and Control of Noncommunicable Diseases 2013–2020 (2013). https://apps.who.int/iris/bitstream/handle/10665/94384/9789241506236_eng.pdf?sequence=1. Accessed 7 Aug 2020
- World Health Organization: Noncommunicable Diseases: Key Facts (2018). <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases>. Accessed 7 Aug 2020
- Yang, C., Jin, Z.: An acute respiratory infection runs into the most common noncommunicable epidemic—COVID-19 and cardiovascular diseases. *JAMA Cardiol.* **5**(7), 743–744 (2020)
- Zou, K.H., Imperato, J., Potkar, C.N., Sethi, N., Edwards, J., Ray, A.: Harnessing real-world data for regulatory use and applying innovative applications. *J. Multidiscip. Healthc.* **13**, 671–679 (2020)

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

Affiliations

Kelly H. Zou¹  · Jim Z. Li²  · Lobna A. Salem¹ · Joseph Imperato¹ · Jon Edwards³ · Amrit Ray⁴ 

Jim Z. Li
jim.li@pfizer.com

Lobna A. Salem
Lobna.A.Salem@Pfizer.com

Joseph Imperato
Joseph.Imperato@pfizer.com

Jon Edwards
Jon.Edwards@envisionpharmagroup.com

Amrit Ray
Amrit.Ray@pfizer.com

- ¹ Research, Development and Medical, Upjohn Division, Pfizer Inc, 235 East 42nd Street, MS 235-9-1, New York, NY 10017, USA
- ² Research, Development and Medical, Upjohn Division, Pfizer Inc, 10777 Science Center Drive, San Diego, CA 92121, USA
- ³ Envision Pharma Group, Envision House, 5 North Street, Horsham RH12 1XQ, UK
- ⁴ Research, Development and Medical, Upjohn Division, Pfizer Inc, 500 Arcola Road, Collegeville, PA 19426, USA