Big Data in the Assessment of Pediatric Medication Safety

Ann W. McMahon, MD, MS, FISPE,[a] William O. Cooper, MD, MPH,[b] Jeffrey S. Brown, PhD,[d] Bruce Carleton, PharmD,[f]
Finale Doshi-Velez, PhD,[e] Isaac Kohane, MD, PhD,[c] Jennifer L. Goldman, MD,[g] Mark A. Hoffman, PhD,[h]
Rishikesan Kamaleswaran, PhD,[i] Michiyo Sakiyama, MD,[j,k] Shohko Sekine, MS,[k] Miriam C.J.M. Sturkenboom, PhD,[l]
Mark A. Turner, MBChB, PhD,[m] Robert M. Califf, MD, MACC[n]

Big data (BD) in pediatric medication safety research provides many opportunities to improve the safety and health of children. The number of pediatric medication and device trials has increased in part because of the past 20 years of US legislation requiring and incentivizing study of the effects of medical products in children (Food and Drug Administration Modernization Act of 1997, Pediatric Rule in 1998, Best Pharmaceuticals for Children Act of 2002, and Pediatric Research Equity Act of 2003). There are some limitations of traditional approaches to studying medication safety in children. Randomized clinical trials within the regulatory context may not enroll patients who are representative of the general pediatric population, provide the power to detect rare safety signals, or provide long-term safety data. BD sources may have these capabilities. In recent years, medical records have become digitized, and cell phones and personal devices have proliferated. In this process, the field of biomedical science has progressively used BD from those records coupled with other data sources, both digital and traditional. Additionally, large distributed databases that include pediatric-specific outcome variables are available. A workshop entitled "Advancing the Development of Pediatric Therapeutics: Application of 'Big Data' to Pediatric Safety Studies" held September 18 to 19, 2017, in Silver Spring, Maryland, formed the basis of many of the ideas outlined in this article, which are intended to identify key examples, critical issues, and future directions in this early phase of an anticipated dramatic change in the availability and use of BD.

abstract

[a]Office of Pediatric Therapeutics, US Food and Drug Administration, Rockville, Maryland; [b]Departments of Pediatrics and Health Policy, Vanderbilt University Medical Center, Nashville, Tennessee; [c]Departments of Biomedical Informatics, Pediatrics, and [d]Population Medicine, Harvard Medical School and Harvard Pilgrim Healthcare Insititute, Boston, Massachusetts; [e]Paulson School of Engineering and Applied Sciences, Harvard University, Boston, Massachusetts; [f]Division of Translational Therapeutics, Department of Pediatrics, Faculty of Medicine, University of British Columbia, Vancouver, Canada; [g]Divisions of Pediatric Infectious Diseases and Clinical Parmacology, Department of Pediatrics, and [h]Children's Research Institute, Children's Mercy Hospital, Kansas City, Missouri; [i]Departments of Biomedical Informatics, Pediatrics, and Emergency Medicine, School of Medicine, Emory University, Atlanta, Georgia; [j]Office of Vaccines and Blood Products and [k]Office of New Drug IV, Pharmaceuticals and Medical Devices Agency, Tokyo, Japan; [l]Department of Epidemiology, Julius Center Research Program Cardiovascular Edpidemiology, Utrecht University Medical Center, Utrecht, Netherlands; [m]Institute of Translational Medicine, University of Liverpool, Liverpool, United Kingdom; and [n]Division of Cardiology, Department of Internal Medicine, School of Medicine, Center for Health Science, Duke Clinical Research Institute, Duke University, Durham, North Carolina

To cite: McMahon AW, Cooper WO, Brown JS, et al. Big Data in the Assessment of Pediatric Medication Safety. *Pediatrics.* 2020;145(2): e20190562

Big data (BD) in pediatric medication safety research provides many opportunities to improve the overall safety and health of children. In this article, we provide the unique perspective of a group of experts in this field convened at a workshop designed to give a picture of some of the major current pediatric BD systems and to give an overview of gaps in pediatric BD going forward. Topics covered include sample size, generalizability, and duration of observation, all of which can be improved when traditional randomized clinical trials in children are supplemented with data from routine clinical practice and other settings outside of research facilities and trials. In this article, our objectives are to outline our functional definition of BD for medication safety, describe why we need BD to help evaluate pediatric medication safety, give examples to describe the current state of using BD to

SPECIAL ARTICLE

OUR DEFINITION OF BD FOR MEDICATION SAFETY

In 2005, Roger Magoulas[1] coined the term BD to describe data that were too large and complex for traditional data-processing software to manage and analyze. But, given the ever-expanding capabilities of computer processing and data storage capacity, this concept no longer applies. A definition that better reflects our understanding of current data streams in the health care arena might read as follows: BD in health care refers to real-world data (RWD) as well as genomics and other "-omics" data (eg, proteomics, transcriptomics). RWD includes digitized health data generated outside of traditional randomized clinical trial settings that can help inform health status. Examples include data collected from biological samples, physiologic monitoring, electronic health records (EHRs), claims data and billing activities, and product and disease registries; data generated by patients; and data gathered from other sources such as mobile devices. Real-world evidence (RWE) is the clinical evidence derived from the application of a proper study design and analysis to RWD to inform the potential benefits or risks of a medical product. RWE includes not only observational analysis but also various types of randomization, such as cluster, stepped-wedge, and individual-person randomization.[2]

The expansion of measurement in multiple dimensions is at the core of both the BD and the RWD concepts. For example, the National Institutes of Health *All of Us* Research Program (Fig 1)[3] is an effort to gather data over many years from ≥1 million people living in the United States, with the ultimate goal of accelerating public health and medical research. Unlike research studies that are focused on a specific disease or population, *All of Us* will serve as a national research resource to inform thousands of studies. Pilot studies under development in the *All of Us* program use rich analysis of EHR and claims data, health applications, and fitness wearables. Computational power no longer seems to be a limiting factor, and social media have dramatically changed our ability to assess the time dimension because data collection is no longer limited to periodic physical visits to a clinic or intermittent recall of past events.

WHY WE NEED BD TO HELP EVALUATE PEDIATRIC MEDICATION SAFETY

Studies of medication safety in children present special challenges often not faced in studies done in adults. Pediatric development spans from life in utero, potentially affected by maternal medication exposure, to at least 18 years of age. Although the stages of pediatric development involve rapid alterations in somatic and neurologic growth, neurocognitive development, endocrine functions, and medication metabolism, the most widely used measurements in pediatrics are still weight, height, and BMI. There is a paucity of age-dependent or developmental stage–specific standards against which thousands of developmentally influenced changes could be measured. As a result, although the limited standard parameters have been important, traditional approaches to pediatric study designs often lack the statistical power or phenotypic detail needed to detect differences in important subcategories of children. Children are also less likely than adults to take prescription medications and have fewer chronic diseases, limiting the numbers of individual medication exposures. As a result, pediatric clinical trials are typically relatively small compared with adult trials. Although enormous data streams are now available in inpatient and outpatient settings in an appropriate time frame to support medical decisions, availability alone is not sufficient. For example, in the current routine NICU workflow, much of the waveform data (even those captured at lower sampling frequencies) are not captured and retained for reanalysis but are reduced to a handful of vital statistics for entry into the EHR, thus missing potential safety signals. In the future, in addition to capturing these electronic signals, pediatric medication safety studies must necessarily draw from large populations and encompass many exposures and outcomes.

Today, more sophisticated infrastructure is available so that BD can be captured and analyzed in real time to derive meaningful insight from significant volumes of complex sensor data,[4] not only enhancing the detection of safety signals but also providing insight into associated risk factors. And because of both the enormous changes in storage capacity and in the ability to align processors for powerful depth of analysis, modern computational systems can process data in multiple dimensions (number of patients, depth of biological measurement, digital phenotyping) that are orders of magnitude beyond the size of data that could be handled just a few years ago. By processing many more variables than in the recent past, it is possible to facilitate the discovery of patterns that reach beyond traditional boundaries of our understanding of human biology, clinical outcomes, and medical practice.[5–7] These patterns provide a data-driven way to define cohorts of patients with similar biological, clinical, or behavioral and social characteristics and to evaluate health by using sensors and devices that enable measurement of a more continuous time dimension.

Nevertheless, computational methods pose challenges themselves, beyond the technical complexity of the algorithms, especially in specific
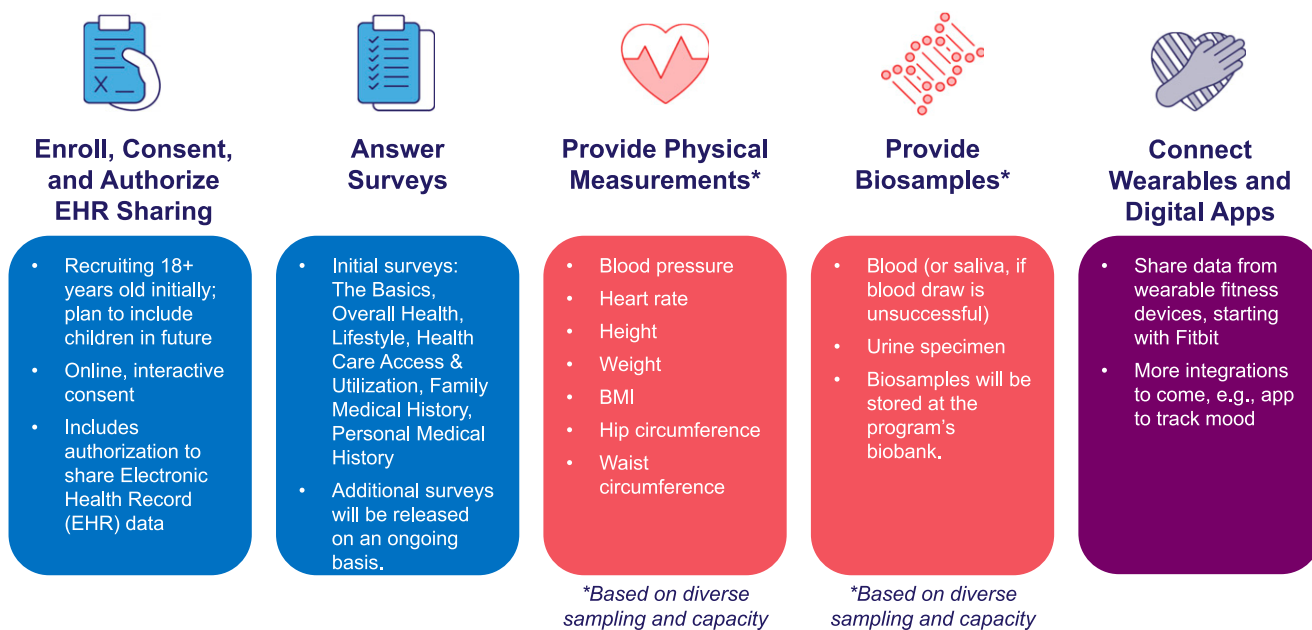
FIGURE 1

All of Us Research Program protocol: actions and data types that participants in the *All of Us* Research Program may be asked to complete or donate based on the program's protocol as of September 2018. The program may expand to collect additional data types in future iterations of the protocol. (Reprinted with permission from National Institutes of Health. *All of Us* Research Program. This figure has been shared publicly, but has not previously been published.)

demographic populations such as pediatrics. Outcomes such as chest pain have 1 meaning in adults and a different meaning in children and adolescents. Decisions about how to represent data and the populations being studied strongly affect the kinds of causal claims that can be made, and predictions from models can be biased by inclusion or exclusion of some populations.

THE CURRENT STATE OF USING BD TO EVALUATE PEDIATRIC MEDICATION SAFETY

Across organized health systems, infrastructure is under development in support of analyzing BD by using tools such as traditional biostatistics, Bayesian methods, machine learning, and artificial intelligence, although the importance of these tools in research involving pediatric patients may be underrecognized.[8] As a deeper understanding is gained of systems biology, much more intensive study will be possible of the impact of genetic polymorphisms on clinical

outcomes and of the interaction of medications with complex systems such as the microbiome. Additionally, until recent advances in computing,[9] the complex biological relationship between mother and fetus was beyond the scope of deep study. Thus, we can expect a new era of understanding of the biological effects of medications and the relationships between these effects and clinical outcomes. In addition, study linkages can be performed to enhance study power (eg, school performance, vital statistics, EHRs).

During the last 10 years, as medical records and pediatric clinical practice have become digitized, the fields of biomedical science and clinical practice have increasingly included EHRs. This trend holds promise that age-specific, developmental stage–specific, and ethnic group–specific standards can be developed from EHR and other BD repositories. One of the most rapidly growing areas of digital interactions is the study of the impact of behavior

and environment on health outcomes.[10] Applications that can guide parents in detecting childhood illnesses, such as autism, or in managing diseases, such as diabetes, will also generate growing volumes of data that can potentially be harvested for research. In addition, especially in the United States, linkages among disparate databases, such as school performance and vital statistics and EHRs, hold significant promise, but the process of linkage is complex and can be labor intensive because of the lack of common identifiers and important privacy considerations when multiple sensitive data are joined together from their primary repositories of consented data management.[11] However, in a US project linking the Cystic Fibrosis Foundation Registry with EHR data on those same patients, ~10 000 patients were successfully linked,[12] and technical "workarounds" are being developed by using federated analysis methods (Sentinel Coordinating Center, Cambridge,

MA).[13,14] These linkage studies are more commonly performed in parts of Europe,[15] where national patient identifiers are available. Whereas unique identifiers to enable linkage are often available in the United States, governance and privacy policies hinder efficient linkage across US health systems.

Some examples follow that depict national and international networks and data warehouses to highlight aspects of BD medication safety research in children. The particular examples were chosen to be some of the prominent and successful systems in pediatric BD safety research, and some representatives of these systems were invited to the workshop "Advancing the Development of Pediatric Therapeutics: Application of 'Big Data' to Pediatric Safety Studies," which was held on September 18 to 19, 2017, in Silver Spring, Maryland. The systems are grouped by geographic area. A summary of highlights and current strengths and gaps going forward is described at the end of each section.

US PEDIATRIC NETWORKS

PEDSnet, Part of PCORnet

PCORnet is a program of the Patient-Centered Outcomes Research Institute, a public and private entity funded through US taxes and governed by a board representing government, industry, academia, and patients and their families. PCORnet combines a base of curated EHRs and claims data from 34 large health systems that were initially organized into 13 clinical data research networks. Dozens of patient-led groups are funded to organize disease-specific information. The network has now evolved into 9 clinical research networks and 2 payer networks. The combination of a disease-specific clinical focus, curated data on >120 million Americans, and interactions among patient groups and clinicians is

intended to produce accelerated, lower-cost, patient-centered research.

One important component of the PCORnet is its sizable pediatric network, PEDSnet, which is producing results after 3 years of infrastructure building.[16–18] The network contains EHR data on >6 million children and has accumulated longitudinal information on 362 550 children. These data can be used to perform association studies such as a recent one on the relationship between antibiotic use and change in weight over a 4-year period.[19] PEDSnet also offers promise for supporting long-term follow-up.[19]

Undiagnosed Disease Network

The Undiagnosed Disease Network,[20] sponsored by the National Institutes of Health Common Fund, is a national network of 12 clinical sites that addresses the problem of undiagnosed diseases that can cause suffering, even death, while incurring significant health costs. Approximately half of the patients seen in this network are children. Because a child's genomic sequence is more informative when interpreted in the context of the parental genomes, the network sequences these "trios," providing superior information, especially when combined with the curated clinical information. Clinical and genomic data are stored in a third party–hosted cloud that meets all relevant security and confidentiality standards. Only authorized users can access these data and collaborate on both the clinical diagnosis and downstream research using the data, a multidisciplinary approach that underscores the importance of both BD and human expertise in diagnosing often-difficult cases.[21]

Medicaid

Much pediatric pharmacoepidemiology has been performed in the Medicaid system.[22,23] The sample size and long-term follow-up in some

populations[23] provided by Medicaid is an advantage. In 1 example, the cumulative incidence of first psychiatric diagnosis and psychotropic medication use (monotherapy or concomitant use of psychotropic medications) was obtained from Medicaid from birth through age 7 years.[22] A challenge to using Medicaid data is that this program is state run, and therefore it is difficult to combine data from >1 state because of the absence of uniform identifiers. Therefore, even when 1 state provides a substantial sample size, child missingness or duplication can be a problem.

Sentinel System

The Sentinel System, initiated by the US Food and Drug Administration (FDA), uses electronic health care data to monitor the performance of FDA-regulated medical products.[24] The Sentinel database contains information for >300 million health plan members, including 13% individuals <19 years of age and Medicare participants who are 65 years and older.[25] Sentinel maximizes local control of data in a federated model in which health systems and insurers retain their data and curate them so that questions can be asked of the data and results can be summarized across the network. In an example of a pediatric medication safety project launched in Sentinel, Raebel et al[26] found that, despite FDA guidelines recommending glucose monitoring in persons starting second-generation antipsychotics, only 11% of children initiating therapy had baseline glucose measured.

Postlicensure Rapid Immunization Safety Monitoring

The Postlicensure Rapid Immunization Safety Monitoring (PRISM) program is a subcomponent of the Sentinel System. PRISM is focused on vaccine safety surveillance. The data in PRISM can support assessments of vaccine use in

pediatric populations and longitudinal follow-up of patients after vaccine exposure.[27] An example of the use of this technology was the finding of Yih et al[28] that RotaTeq was associated with ~1.5 (95% confidence interval [CI] 0.2–3.2) excess cases of intussusception per 100 000 recipients of the first dose.[29] The most practical and easily used data in Sentinel are the billing data, which do not always reflect clinical detail. Some clinical detail may only be accessed through manual medical record searches.[30]

FREESTANDING CANADIAN RESEARCH NETWORK: PAN-CANADIAN NETWORK

The Canadian Pharmacogenomics Network for Drug Safety includes clinical surveillance, largely with EHRs and site-specific funded personnel, located at 13 pediatric and 13 adult academic health centers across Canada. The Canadian Pharmacogenomics Network for Drug Safety collects detailed information on adverse drug reactions (ADRs) identified from electronic and nonelectronic medical records, patients and families, and other sources.[31] As of June 2018, the network database contained 9910 ADR cases and 89 263 medication-matched controls with detailed clinical data and DNA collected from each patient. The goal is to find high-association pharmacogenomic biomarkers of clinical ADR phenotypes (odds ratios of ≥3), create innovative tools (eg, pharmacogenomic tests) to predict the likelihood of ADR risk, and implement medication safety solution strategies. A success from the network of these strategies was the study of anthracycline-induced cardiotoxicity,[32] in which the most important risk factor in children was high cumulative dose, although no dose was absolutely safe. The Canadian, multisite, primary data, collection model collects detailed clinical data and obtains specimens

for genetic analysis, requiring manual support for data collection. This may limit the ultimate sample size because of cost.

EUROPEAN LONG-TERM CARE SYSTEMS WITH PEDIATRIC DATA

Swedish Cancer Register

Several Nordic countries, including Sweden, have longitudinal health data on their entire populations of adults and children. These networks provide an excellent opportunity to follow children into adulthood who may have been exposed to a medical product in childhood. In 1 illustrative study, the Swedish National Patient Register 1964–2014 and the Swedish Cancer Register were used to study the risk of cancer among children and adults who had inflammatory bowel disease as children (median age at the end of follow-up: 27 years). Data from administrative and clinical national registers on demographics, medications, morbidity, and mortality were linked by using the Swedish personal identifier. Sample sizes included thousands of patients with inflammatory bowel disease and comparators from the general population. The hazard ratio (HR) for first all-cancer was 2.2 (95% CI 2.0–2.5).[15] Combining these 2 networks enabled demonstration of the feasibility of long-term follow-up of children into adulthood in the context of RWD. Personal identifiers also enable analysis of multidimensional data.[15]

Clinical Practice Research Datalink

Another example of a long-term and large database containing children is the Clinical Practice Research Datalink, formerly the General Practice Research Database, which is based on the health care delivery system and has collected primary care data in the United Kingdom since 1987.[33] In a pediatric cohort study, conducted by using the Clinical Practice Research Datalink

(1987–2009), a total of 11 934 people with epilepsy, aged 1 to 24 years at diagnosis, and 46 598 people without epilepsy were followed for a median (interquartile range) of 2.6 (0.8–5.9) years. The risk of fractures, thermal injuries, and poisonings was estimated. The authors found that children and young adults with epilepsy were at significantly greater risk than those without epilepsy, and the greatest risk was from medicinal poisonings.[33]

EUROPEAN INITIATIVE EVALUATING PEDIATRIC PUBLIC HEALTH: GLOBAL RESEARCH IN PEDIATRICS

Global Research in Pediatrics is a European Commission–funded Network of Excellence that has completed a 6-year project that included building a pediatric pharmacoepidemiology platform for collaborative studies on medication use, safety, and effectiveness. As part of a gap analysis, Global Research in Pediatrics investigators reviewed the literature and observed that the number of pediatric pharmacoepidemiological safety studies is low, but has increased steadily, and that these studies were concentrated in the United States and European Union. Unfortunately, although they have the majority of the world's pediatric population, low- and middle-income countries have little representation in available pharmacoepidemiological data.[34]

JAPANESE MEDICAL RECORD NETWORK WITH PEDIATRIC DATA: MEDICAL INFORMATION DATABASE NETWORK

Japan's Pharmaceuticals and Medical Devices Agency has developed a new database system that is used to analyze diverse electronic health care data from multiple Japanese medical institutions. Through collaboration with 10 cooperating medical institutions nationwide, the Medical Information Database Network (MID-NET) is able to collect and analyze medical information (eg, EHRs, claims

data) on a scale exceeding 4 million people. Recently, a project was designed to use MID-NET to identify the risk of respiratory depression associated with codeine-containing products in children. MID-NET will be useful as a tool for assessing other adverse reactions in children.[35] The federation of Japanese health care institutions allows searching by billing codes and EHRs to assess the frequency with which an adverse reaction occurs in children.

ISSUES RELATED TO INTERPRETING PEDIATRIC BD

As pediatric RWE has evolved, questions arise about the characteristics of the underlying RWD and how the methods of analysis can affect the interpretation of results. For example, what is the size of the database? The incidence and prevalence of the disease studied and the frequency of adverse events dictate the adequacy of the sample size. Also, it is often difficult to ensure that individual patient follow-up is complete. Does the database have personal identifiers for each patient allowing linkage of information to >1 data source? These linkages will be more commonly performed in parts of Europe,[17] where, unlike in the United States, patient identifiers are available across systems.

Does the data capture specific clinical and laboratory data that might be needed for conclusions related to effectiveness or safety? Characterization of complex clinical outcomes can be difficult and/or may be inconsistent from patient to patient because of variability in how clinicians record unstructured data in an EHR. Also, what level of granularity is sufficient? Adult studies suggest that each incremental increase in the temporal granularity of the data improves predictive performance,[36] but little empirical evidence exists about the level of granularity needed for predictions at different stages of childhood.

Some issues are specific to the situation in which an attempt is made to make a causal inference from nonrandomized data. Confounding due to differences in demographics, underlying illness, or other variables between comparison groups and difficulty establishing an inception time for the cohort make such analyses risky unless effect sizes are large. Follow-up data are also variable, with some European and Korean systems having longer-term follow-up and many US systems having shorter-term follow-up.[37] A system that collects data in a predefined manner may have more easily interpretable data, but such a mechanism may be costly, thereby limiting sample size. Today, in the United States and elsewhere, industry, investigators, and regulators are grappling with the need beyond data collected separately by a professional research team for traditional randomized trials for more generalizable, less expensive, and easily accessed data that are scientifically valid and useable for regulatory decisions.[38]

Finally, 1 of the major barriers to participation in BD consortia is the concern about data privacy and security. These issues are being addressed in many sectors,[39,40] but the solutions will vary depending on the cultural beliefs and health systems in different countries. Within the United States, concerns about privacy seem to be increasing just when the benefits of identifiable data are becoming clearer: benefits such as identifying drug toxicities, as in Sentinel; best clinical practice, as in PCORnet; or even making individual diagnoses, as in the Undiagnosed Disease Network. Even when linkage is possible, depending on the purpose, consent may not be in place to allow for identification of an individual within a database.

Strengths to date in pediatric BD medication safety research include the existence and continuing expansion of large databases with the capacity for long-term follow-up in some populations. Gaps include the need to develop more pediatric-specific networks in almost all countries, but especially in low- and middle-income countries, and the inability to easily identify the duration of individual patient observation in many systems. Also, the personal identifier, present in health care databases in some countries, is not universally present.

WHERE ARE WE HEADED IN USE OF BD TO STUDY PEDIATRIC MEDICATION SAFETY?

Large pediatric-specific databases with long-term follow-up, using pediatric-specific outcome variables, and containing high-quality data are needed to continue to improve pediatric medication safety assessments. These networks are currently being built, and some of the results are impressive. Long-term follow-up of patients with childhood cancer is increasingly possible. In a follow-up study, Lega et al[41] studied 10 438 1-year survivors of childhood cancer who were diagnosed before age 21. The mean follow-up was 11.2 years. Cancer survivors had a 55% increased rate of developing diabetes compared with matched controls (HR 1.51; 95% CI 1.28–1.78). Individuals treated for cancer between the ages 6 and 10 had the highest increased rates of diabetes (HR 3.89; CI 2.26–6.68).

Another critical issue in drug safety is pregnancy exposure. Using RWD through chart review and clinical assessment, the authors of 1 observational study demonstrated that children exposed to valproate in utero experienced decreased IQ at 6 years of age. High doses of valproate were negatively associated with IQ ($r = -0.56$; $P < .001$), verbal ability ($r = -0.40$; $P = .005$), nonverbal ability ($r = -0.42$; $P = .003$), memory ($r = -0.30$; $P = .04$), and executive function ($r = -0.42$; $P < .001$), whereas in utero exposure

to carbamazepine, lamotrigine, or phenytoin was not negatively associated.[42]

Despite the notable progress, several areas need more thought on the part of researchers before large-scale conversion to pediatric BD safety research. One area includes overcoming cultural and political barriers to collaboration across health systems, countries, and continents. The vast majority of the published literature on the use of BD in pediatric research comes from the United States and Europe, and barriers to those countries sharing information are formidable at every level.[15,26] In addition, to be able to realize the potential of BD in the area of pediatric research, given our evolving technological and computing environments, a much more coordinated approach is needed. For example, an established reference set of pediatric-specific outcomes for BD health care research will be critical. Moreover, for both rare and more common pediatric diseases and adverse events, genomic considerations are often critical. Privacy issues remain a serious

barrier to the use of BD regarding pediatric health research as well as to our daily health care lives.

CONCLUSIONS

Tremendous progress has been made during the last 20 years in the ability to develop drugs and devices that can be safely used in children. We have also seen key advances in computer processing and data storage capacity along with the ubiquitous digitization of health care–related information. Many national and international networks and data warehouses already are making use of these advances to apply BD in medication safety research in children. However, now that a comprehensive understanding of children, their biology, their systems pharmacology, their interactions, and their behaviors is possible, acceleration of the development of interoperable, large-scale networks is needed. To be able to expand our capabilities in this field and address the many important pediatric-specific safety issues of the future, researchers will have to be able to easily access multiple types of

BD, including biological, clinical, behavioral, and social data, collected over the long-term and shared in a manner that enables sophisticated analysis without compromising privacy or security.

ABBREVIATIONS

ADR: adverse drug reaction
BD: big data
CI: confidence interval
EHR: electronic health record
FDA: US Food and Drug Administration
HR: hazard ratio
MID-NET: Medical Information Database Network
PRISM: Postlicensure Rapid Immunization Safety Monitoring
RWD: real-world data
RWE: real-world evidence

REFERENCES

1. Ross JM. Roger Magoulas on big data. 2010. Available at: http://radar.oreilly.com/2010/01/roger-magoulas-on-big-data.html. Accessed September 6, 2018

2. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence - what is it and what can it tell us? N Engl J Med. 2016;375(23):2293–2297

3. National Institutes of Health. All of Us Research Program overview. Available at: https://allofus.nih.gov/about/all-us-research-program-overview. Accessed December 3, 2019

4. Sutton JR, Mahajan R, Akbilgic O, Kamaleswaran R. PhysOnline: an open source machine learning pipeline for real-time analysis of streaming physiological waveform. IEEE J Biomed Health Inform. 2019;23(1):59–65

5. Doshi-Velez F, Ge Y, Kohane I. Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis. *Pediatrics*. 2014;133(1). Available at: www.pediatrics.org/cgi/content/full/133/1/e54

6. Kale DC, Che Z, Bahadori MT, Li W, Liu Y, Wetzel R. Causal phenotype discovery via deep networks. *AMIA Annu Symp Proc*. 2015;2015:677–689

7. Saria S, Rajani AK, Gould J, Koller D, Penn AA. Integration of early physiological responses predicts later illness severity in preterm infants. *Sci Transl Med*. 2010;2(48):48ra65

8. Sadler NC, Nandhikonda P, Webb-Robertson BJ, et al. Hepatic cytochrome P450 activity, abundance, and expression throughout human development. *Drug Metab Dispos*. 2016;44(7):984–991

9. Quinney SK, Gullapelli R, Haas DM. Translational systems pharmacology studies in pregnant women. *CPT Pharmacometrics Syst Pharmacol*. 2018;7(2):69–81

10. Insel TR. Digital phenotyping: technology for a new science of behavior. *JAMA*. 2017;318(13):1215–1216

11. Ball R, Robb M, Anderson SA, Dal Pan G. The FDA's sentinel initiative—a comprehensive approach to medical product surveillance. *Clin Pharmacol Ther*. 2016;99(3):265–268

12. Cogen JD, Hall M, Loeffler DR, et al. Linkage of the CF foundation patient registry with the pediatric health information system database. *Pediatr Pulmonol*. 2019;54(6):721–728

13. Hale C. Datavant, PCORnet to build real-world data network spanning 60M patients. 2019. Available at: https://www.fiercebiotech.com/medtech/datavant-pcornet-to-build-real-world-data-network-spanning-60m-patients. Accessed October 16, 2019

14. Sentinel Coordinating Center. Exploration of potential for sentinel and PCORnet data linkage. Available at: https://www.sentinelinitiative.org/sentinel/data/complementary-data-sources/exploration-potential-sentinel-and-pcornet-data-linkage. Accessed October 16, 2019

15. Olén O, Askling J, Sachs MC, et al. Childhood onset inflammatory bowel disease and risk of cancer: a Swedish nationwide cohort study 1964-2014. *BMJ*. 2017;358:j3951

16. Kuhlthau KA, Bailey LC, Baer BL, et al. Large databases for pediatric research on children with autism spectrum disorder. *J Dev Behav Pediatr*. 2018;39(2):168–176

17. Khare R, Utidjian L, Ruth BJ, et al. A longitudinal analysis of data quality in a large pediatric data research network. *J Am Med Inform Assoc*. 2017;24(6):1072–1079

18. Forrest CB, Margolis P, Seid M, Colletti RB. PEDSnet: how a prototype pediatric learning health system is being expanded into a national network. *Health Aff (Millwood)*. 2014;33(7):1171–1177

19. Block JP, Bailey LC, Gillman MW, et al; PCORnet Antibiotics Childhood Growth Study Group. PCORnet antibiotics and childhood growth study: process for cohort creation and cohort description. *Acad Pediatr*. 2018;18(5):569–576

20. Undiagnosed Diseases Network. About the Undiagnosed Diseases Network. Available at: https://undiagnosed.hms.harvard.edu/. Accessed December 3, 2019

21. Macnamara EF, Schoch K, Kelley EG, et al; Undiagnosed Diseases Network. Cases from the Undiagnosed Diseases Network: the continued value of counseling skills in a new genomic era. *J Genet Couns*. 2019;28(2):194–201

22. Pennap D, Zito JM, Santosh PJ, Tom SE, Onukwugha E, Magder LS. Patterns of early mental health diagnosis and medication treatment in a Medicaid-insured birth cohort. *JAMA Pediatr*. 2018;172(6):576–584

23. Cooper WO, Habel LA, Sox CM, et al. ADHD drugs and serious cardiovascular events in children and young adults. *N Engl J Med*. 2011;365(20):1896–1904

24. Platt R, Brown JS, Robb M, et al. The FDA Sentinel Initiative - an evolving national resource. *N Engl J Med*. 2018;379(22):2091–2093

25. Sentinel Coordinating Center. Snapshot of database statistics. Available at: https://www.sentinelinitiative.org/sentinel/data/snapshot-database-statistics. Accessed April 11, 2019

26. Raebel MA, Penfold R, McMahon AW, et al. Adherence to guidelines for glucose assessment in starting second-generation antipsychotics. *Pediatrics*. 2014;134(5). Available at: www.pediatrics.org/cgi/content/full/134/5/e1308

27. Baker MA, Nguyen M, Cole DV, Lee GM, Lieu TA. Post-licensure rapid immunization safety monitoring program (PRISM) data characterization. *Vaccine*. 2013;31(suppl 10):K98–K112

28. Yih WK, Lieu TA, Kulldorff M, et al. Intussusception risk after rotavirus vaccination in U.S. infants. *N Engl J Med*. 2014;370(6):503–512

29. Nguyen M, Ball R, Midthun K, Lieu TA. The Food and Drug Administration's Post-Licensure Rapid Immunization Safety Monitoring program: strengthening the federal vaccine safety enterprise. *Pharmacoepidemiol Drug Saf*. 2012;21(suppl 1):291–297

30. Duffy J, Weintraub E, Hambidge SJ, et al; Vaccine Safety Datalink. Febrile seizure risk after vaccination in children 6 to 23 months. *Pediatrics*. 2016;138(1):e20160320

31. Carleton B, Poole R, Smith M, et al. Adverse drug reaction active surveillance: developing a national network in Canada's children's hospitals. *Pharmacoepidemiol Drug Saf*. 2009;18(8):713–721

32. Visscher H, Rassekh SR, Sandor GS, et al; CPNDS consortium. Genetic variants in SLC22A17 and SLC22A7 are associated with anthracycline-induced cardiotoxicity in children. *Pharmacogenomics*. 2015;16(10):1065–1076

33. Prasad V, Kendrick D, Sayal K, Thomas SL, West J. Injury among children and young adults with epilepsy. *Pediatrics*. 2014;133(5):827–835

34. Osokogu OU, Dukanovic J, Ferrajolo C, et al. Pharmacoepidemiological safety studies in children: a systematic review. *Pharmacoepidemiol Drug Saf*. 2016;25(8):861–870

35. Pharmaceuticals and Medical Devices Agency. *Pharmaceuticals and Medical Devices Safety Information. No. 351*.

MCMAHON et al

Tokyo, Japan: Pharmaceuticals and Medical Devices Agency; 2018

36. van Wyk F, Khojandi A, Mohammed A, Begoli E, Davis RL, Kamaleswaran R. A minimal set of physiomarkers in continuous high frequency data streams predict adult sepsis onset earlier. *Int J Med Inform.* 2019;122: 55–62

37. Gerhard T, Moride Y, Pottegård A, Pratt N. Encounter Databases. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology,* 6th ed. Chichester, United Kingdom: John Wiley & Sons; 2019:211–240

38. US Food and Drug Administration. *Submitting Documents Using Real-World Data and Real-World Evidence to FDA for Drugs and Biologics: Guidance for Industry: Draft Guidance.* Rockville, MD: US Food and Drug Administration; 2019

39. Vaidya J, Shafiq B, Asani M, Adam N, Jiang X, Ohno-Machado L. A scalable privacy-preserving data generation methodology for exploratory analysis. *AMIA Annu Symp Proc.* 2018;2017: 1695–1704

40. Zhu H, Tan YA, Zhu L, Wang X, Zhang Q, Li Y. An identity-based anti-quantum privacy-preserving blind authentication in wireless sensor networks. *Sensors (Basel).* 2018;18(5): E1663

41. Lega IC, Pole JD, Austin PC, Lau C, Nathan PC, Baxter NN. Diabetes risk in childhood cancer survivors: a population-based study. *Can J Diabetes.* 2018;42(5):533–539

42. Meador KJ, Baker GA, Browning N, et al; NEAD Study Group. Fetal antiepileptic drug exposure and cognitive outcomes at age 6 years (NEAD study): a prospective observational study. *Lancet Neurol.* 2013;12(3):244–252

Updated Information & Services	including high resolution figures, can be found at: http://pediatrics.aappublications.org/content/145/2/e20190562
References	This article cites 31 articles, 6 of which you can access for free at: http://pediatrics.aappublications.org/content/145/2/e20190562#BIBL
Subspecialty Collections	This article, along with others on similar topics, appears in the following collection(s): **Administration/Practice Management** http://www.aappublications.org/cgi/collection/administration:practice_management_sub **Safety** http://www.aappublications.org/cgi/collection/safety_sub **Health Information Technology** http://www.aappublications.org/cgi/collection/health_information_technology_sub **Electronic Health Records** http://www.aappublications.org/cgi/collection/electronic_health_records_sub
Permissions & Licensing	Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: http://www.aappublications.org/site/misc/Permissions.xhtml
Reprints	Information about ordering reprints can be found online: http://www.aappublications.org/site/misc/reprints.xhtml

PEDIATRICS®

OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

Big Data in the Assessment of Pediatric Medication Safety

Ann W. McMahon, William O. Cooper, Jeffrey S. Brown, Bruce Carleton, Finale
Doshi-Velez, Isaac Kohane, Jennifer L. Goldman, Mark A. Hoffman, Rishikesan
Kamaleswaran, Michiyo Sakiyama, Shohko Sekine, Miriam C.J.M. Sturkenboom,
Mark A. Turner and Robert M. Califf

The online version of this article, along with updated information and services, is
located on the World Wide Web at:
http://pediatrics.aappublications.org/content/145/2/e20190562

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™