

Leveraging Data Analytics to Advance Personal, Population, and System Health: Moving Beyond Merely Capturing Services Provided

Michael E. Thompson, Michael F. Dulin

The health care industry collects ever-increasing volumes of patient data. Currently, this largely untapped “big data” primarily documents encounters and facilitates billing. This issue of the *North Carolina Medical Journal* explores the promise and the perils of big data as we seek to transform our health care system into one that is more proactive, equitable, and value based.

To make big data useful while protecting patient privacy, we need new governance models that allow organizations to treat data as a fundamental asset. Next, we need to ensure data quality and integrity, and then the timely and seamless integration of clinical data with data describing social determinants and health behaviors. This more holistic patient and community view will reveal key health outcomes drivers and guide focused, evidence-based interventions. As we share data across health systems, we will prevent duplication of services and ensure better coordination of care. Connecting with individuals through social media will speed dissemination of evidence-based information and help overcome limitations in health literacy while also allowing patients to share data with their providers. We need to learn from other industries and to support innovation through entrepreneurship. While big data can help us overcome health disparities, we must implement these new approaches keenly aware of their potential to disenfranchise those with limited digital access and those whose data are not part of the analyses.

Done right, these approaches and insights will help ensure that big data fulfills its promise of helping us achieve our ultimate goal of health for all, individually and collectively.

The Need to Produce Information

The health and health care delivery system is awash in data, yet barely ankle deep in information.

Health care data in the United States is projected to exceed 2,314 exabytes (a trillion megabytes) by 2020 [1] and is growing at 80 megabytes per patient per year [2]. Those values equate to approximately 665 million megabytes (665 terabytes) of data for the average hospital [3], with an annual growth rate of 36%-48% [1, 4]. In addition to

the data generated from clinical encounters, we have untold amounts of data generated from health-related social media messages and the burgeoning area of patient (consumer) generated data from devices like Fitbit activity trackers. Globally, the social media platform Twitter alone generates approximately 280,000 health-related tweets per day [5, 6] and a third of Facebook users post about health experiences [7], while Fitbits, one of many consumer health wearables linked to only one of over 165,000 health and fitness mobile apps that generate health data [7], have already captured over 150 billion hours of heart data [8]. So, we have big data in health care (maybe not in the form we thought we would have), but we are still sadly lacking in *intelligence* in the form of insight or actionable information.

In this issue brief, we explore many of the facets of the promise, perils, trends, and trajectories of health informatics and analytics, preparing the reader for the many thoughtful and revealing papers that comprise this issue of the *North Carolina Medical Journal*. While these issues play out on the national stage, we see—and acutely feel—them here in North Carolina. We conclude with our thoughts on principles and strategies to help us avoid analysis paralysis and move closer to achieving our ultimate goal of health for all, individually and collectively.

Health and health care lag far behind other economic sectors in unlocking value from big data and advancing insights and practices [9]. This laggardly pace—and the general acceptance of this status quo as intransigent—is even more alarming when one remembers that health care accounts for 18% of US gross domestic product, and is projected to surpass 19% by 2027 [10]. As Premier, Inc.’s Leigh Anderson notes, hardly a better case can be made for the near unlimited potential to improve health outcomes, increase efficiency in deploying our finite resources, and advance our knowledge and insights that in turn will further

Electronically published July 8, 2019.

Address correspondence to Michael E. Thompson, Department of Public Health Sciences, UNC Charlotte, 9201 University City Blvd., Charlotte, NC 28223 (methomp1@uncc.edu).

N C Med J. 2019;80(4):214-218. ©2019 by the North Carolina Institute of Medicine and The Duke Endowment. All rights reserved. 0029-2559/2019/80405

improve outcomes and efficiencies [11].

The causes underlying this problem are complex and many. In simplest terms, five intertwined forces perpetuate this logjam. First, the Health Information Technology Act [12] and the Affordable Care Act (ACA) [13] aside, few forces compel the industry to move forward. The payment-for-services side of the process is well developed and functional. The shift to value-based payments has yet to materialize, limiting investment in efforts that ensure robust, precise clinical data are captured in electronic health records (EHRs), even if available [14]. Second, data governance is lacking in consistency and rigor, within and across organizations. As Shannon Fuller notes in his article, without assurances of data quality, integrity, and security, any conclusions drawn from analyses are suspect at best [15]. Worrying then is a 2017 survey of hospital CIOs by business intelligence provider Dimensional Insight where 56% of hospitals reported having incomplete or non-existent governance structures [16]. Third, due to this lack of governance, uniform standards for capturing and reporting the largely unstructured data found in medical records are absent. Aside from a few high-level data elements, the interoperability and easy aggregation and pooling of analytic data promised by the federal Health Information Exchange (HIE) are woefully undeveloped [17]. Fortunately, Christy Revels and Christie Burris report a far more reassuring picture of North Carolina's HIE that gives us cause for optimism moving forward [18]. Furthermore, health care systems and providers are reluctant to share what they consider in many respects to be proprietary, competitive data [19]. In fact, as Leigh Anderson asserts, acquiring and controlling data across the continuum of care is one force driving the many vertical integration mergers we are seeing in health care [11]. Fourth, the sharing, linking, aggregating, and disaggregating of patient, system, and community data, while now technically feasible, are simultaneously required and prohibited by a variety of conflicting laws and policies. As Shannon Fuller notes, organizational and professional cultures and customs block many efforts to advance the state of the art as we move from the 'could we' barriers to the 'should we' barriers [15]. Fifth, the processes needed to link and analyze these disparate data, and make them accessible to patients, providers, payers, and researchers, also create the risks (and liabilities) of breach and abuse [19]. Based on advances at UNC Health Care, Jeff Fuller's article hints at what those promises may be [20], but Doug Hague and Greg Nelson caution about the pitfalls and biases such integration also brings [21, 22]. Collectively, these shortcomings reflect the absence of a coherent national policy strategy and shared set of values in addressing these challenges, exacerbated by inherent lack of trust and concerns over privacy, ethics, security, and the misuse of such information.

Big Data to the Rescue

The exponential growth of health data creates an unprecedented opportunity for advancing health care delivery

at the individual, population, and system levels across the state. Capturing and integrating the data is an essential step in the process. As Revels and Burris report in their article, a key example of our success in this domain is seen in the adoption of new policies in North Carolina that advance our HIE [18]. Their commentary covers our state's 7-year journey to implement NC HealthConnex, which now holds data for 41,000 providers and 6 million patients. This system (powered by SAS and Orion) is on track to become one of the nation's largest in the next three years [18]. It will be a driving force in our state's journey to transform care for patients covered by Medicaid. Using integrated clinical data has widespread potential including: (1) reducing unnecessary or duplicative medical services; (2) better informing providers about needed preventive services; (3) enhanced management of transitions of care; (4) advancing public health surveillance; and (5) better enabling providers to respond in emergencies [18]. Perhaps most importantly, the NC HIE can advance population and community health in novel ways using locally integrated data that can reveal a community's true disease burden and better inform service and workforce needs across the state.

At the county level, the commentary by Gibbie Harris and Jonathan Ong [23] provides insight into how data can be used locally to advance public health. In this example, data obtained during an immunization clinic was used to understand childhood vaccination needs in Mecklenburg County and to tailor a local intervention—in real time—to better serve clients. Moreover, use of tools like the North Carolina Immunization Registry (NCIR) [24] also holds promise for population health initiatives designed to advance uptake of immunizations and avoid preventable infections across the community. This use of data at the individual and community levels is a key component of a new model of public health called Public Health 3.0. As John W. Wallace and colleagues amplify, in the Public Health 3.0 model, public health planners and providers access integrated clinical and community data to assist in understanding public health needs, targeting evidence-based responses, and evaluating the impact of public health interventions [25].

Social media and patient-generated data are also rapidly expanding the data that can inform multiple levels of health care and public health delivery. As outlined by Albert Park and colleagues in this issue [26], these types of data provide a different view of health needs and can be used to detect emerging issues earlier and to better target health-related interventions. In addition, social media provide new channels for delivering information to individuals that can be targeted to an individual's level of health literacy and more rapidly inform a community about a crisis than traditional media. Many North Carolina public health entities could immediately utilize social media to provide timely, valid information to our communities and overcome mis- and disinformation. Furthermore, social media can advance health across the state by empowering advocacy campaigns as well

as directly influencing policy decisions.

Understanding and integrating patient-generated data is another domain that we must address. This type of data, generated from personal devices (fitness trackers and internet-connected scales) as well as online through self-reported information, contributes directly to the exponential growth of data available for health applications. This information is often stored online or on a patient's computer and can inform providers about potential health risk behaviors. In its raw format and without algorithms that can provide insight, however, these data are impossible for individual providers to review. In fact, these data may actually pose risk if shared without context, aside from the pragmatic concern of further drowning time-pressed providers in data that, in its current form and limited by our current analytic capacity, provides no validated clinical insight.

The Systems Context: Health, Not Just Health Care

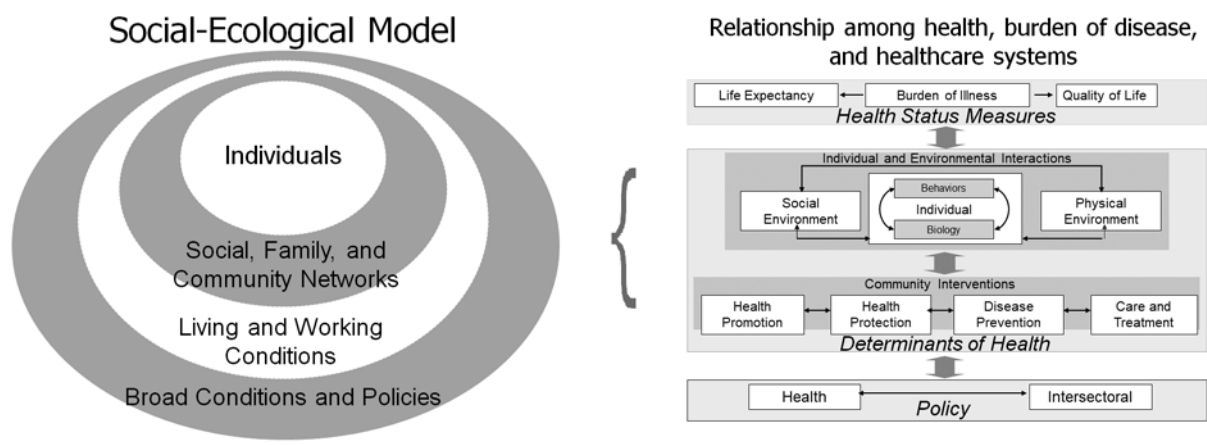
The lack of a coherent, integrated systems approach to health informatics and analytics reflects the many dysfunctions in the larger health and health care delivery system. While public health has long worked within the social-ecological model [27], the highly fragmented US health care delivery system works within multiple silos, a situation reinforced by law, policy, and tradition [28]. These various silos are incentivized to maximize revenue and market share. Consequently, systems-level approaches, which arguably could improve outcomes and reduce societal costs, languish. Figure 1 depicts alternate formulations of the social-ecological model. Clinical services (care and treatment) largely focus on the individual level, meaning the data contained in an EHR are seldom, in and of themselves, useful for affecting communities or targeting primary and secondary prevention efforts that belie the greatest opportunity for cost savings and improved outcomes: the social determinants of health.

Underlying the popularized and oft-cited quote that a person's ZIP code is the most important health predictor is the work of Sir Michael Marmot, who famously demonstrated geographic gradients in lifespan across London neighborhoods [29]. That work resonates today in our ongoing focus on social determinants of health here in North Carolina. In Charlotte, Dr. Alisahah Cole, through the efforts of the One Charlotte Health Alliance [30], brings analytics to the mapping of social determinants [31]. As John Wallace and colleagues note, our newfound data and analytics capacities are enabling the goals of public health and clinical medicine to align around community health and the power of upstream prevention [25]. A case study on the importance of social determinants provided in this issue describes how MedLink of Mecklenburg County, an umbrella safety net provider organization, works with members to coordinate and align their provision of social services through a community-wide referral system that improves coordination of care and reduces administrative burden on agencies and their clients alike [32].

From a systems perspective, the power of data and analytics can be harnessed to reduce total health care spending by enhancing prevention [33]—promoting health and preventing disease, especially through efforts that target early intervention among the most vulnerable—and by improving the efficiency and effectiveness of our treatments, achieving medicine's triple aim [34]. So far, our limited health care big data analytic capacity has focused on improving the efficiency of health services delivery. Shifts to valued-based payments and models that hold integrated delivery systems accountable for the health of defined populations may spur development of prevention-focused analytics [9].

Big data gives us the power to link clinical and ecological data toward identifying priorities and opportunities for promotion and prevention. Clinical medicine's forced shift away

FIGURE 1. The Social Ecological Model (left) and an Alternate Representation Linking Policy Input and Health Outcomes through It (right)



Source: CDC.

from service-volume-based payments to a value-based system will further incentivize collaboration with public health and fuel demand for data-driven holistic insights at patient, facility, and community levels.

Developing the Sociology to Govern our Technology

Key barriers preventing rapid use of data to advance clinical care delivery and public health include poor data quality, inability to integrate data, and the resultant lack of trust in the data by both providers and individuals contributing the data [18]. We can overcome these issues by developing better practices around data governance [15, 20] that lead to better transparency around data use, improved data quality, and enhanced data security. Although relatively nascent in health care, leveraging knowledge from other industries can allow health providers to advance our culture and structures around governance quickly.

In addition to overcoming issues around data quality and oversight, we need to be cognizant of the potential for technology and data to worsen health disparities. For example, the commentaries by Hague [21] and Nelson [22] show how bias in data collection and the use of data from populations that have greater access to services could lead to analytic models and artificial intelligence (AI) algorithms that are effective for only those advantaged populations. In addition, data quality, our ability to integrate data, and the availability of data are compromised for patients with low socioeconomic status who may change jobs, health providers, location of their home residence, and contact information more frequently.

Improving Clinical Analytics

As we look to reduce medical costs and improve outcomes for a diverse and changing population, applying data to drive better care delivery is essential across North Carolina [35]. The approach outlined by J. Fuller [20] provides an exciting glimpse into these opportunities. First, changing providers, hospitals, and health systems into data-driven learning organizations [36] is a major lift that requires high-quality, reliable data and an understanding of how to create and sustain culture change. This approach requires implementation of a data governance strategy and a focus on creating analytic products that can assist providers.

Developing this clinical use capacity is essential: the rapid expansion of data (particularly those describing behaviors and genetics) coupled with advances in the medical evidence mean that no individual alone is capable of serving as the point of integration [37]. Furthermore, J. Fuller rightly asserts that prioritization of work and drawing a clinician's attention to salient measures in the clinical environment must occur quickly: no individual can (or should) track the now over 9,000 clinical measures described in the literature [20]. The analytics capacity that we must build also needs to offer a clear return on investment (ROI). Without that use

case to create incentives for providers, adherence and the needed cultural change to enshrine the value of data and analytics will fade over time.

Finally, in the domains of clinical analytics as well as public health analytics, we must define new roles, followed closely by training and professional development to support clinicians and staff to succeed in these new roles. For example, we need to identify, train, and deploy data stewards into the clinical environments where they can work to advance data quality and procurement [38].

Fostering Innovation

A common theme behind the deployment of data analytics into clinical care delivery and public health is the need to innovate. Again, lessons learned in other industries (eg, data governance and understanding bias in development of models, implementation of data security standards, and use of streaming/real-time data) can inform the transformation of health [39]; however, the health care environment is unique and many new discoveries will need to be made for us to complete this journey. Innovation can be an important driver of culture change [20], governance advancement [15], and creation and deployment of analytic products [20]. Open-source tools and data models foster rapid innovation. Industry disruption is a key consequence of innovation. Entrepreneurs like Charlotte-based Tresata are developing new solutions that leverage next-generation, open-source software and are faster, cheaper, and better than current solutions [40]. As Burriss notes, policy changes that support interoperability of data and EHR systems using FHIR (Fast Healthcare Interoperability Resources) [18] will assist in advancing these EHR systems by making them easier to use and better at providing clinical decision support.

Guiding Thoughts

As amply demonstrated throughout this issue, big data and analytics hold much promise—and present some peril—but are woefully under-developed in health care when compared to other sectors. Given the many pitfalls, security concerns, and risks engendered by our lack of a universal coverage/single risk pool system, perhaps we are fortunate to have the added breathing space to more fully develop our governance structures, ensure data quality and security, and align our policies before charging ahead into the brave new electronically integrated world. Building systems and networks based upon trust and transparency among health care's many stakeholders will take time and effort outside the scope of the technology that makes it possible. The move toward valued-based payments will reinforce the need for health care industries to adapt to a systems-thinking-driven model and will further foment a culture of data-driven learning organizations. In that future, the business case—the ROI for all parties—should be self evident and will in turn further drive sound policy development. Only then will we move on from our obsession with big data to our amazement with the

grand solutions that our having turned those data into information and intelligence will engender. *NCMJ*

Michael E. Thompson, MS, DrPH associate professor; associate chair, Department of Public Health Science; director, Health Informatics and Analytics Programs; director, Master of Health Administration Program, University of North Carolina at Charlotte, Charlotte, North Carolina.

Michael F. Dulin, MD, PhD professor and director, Academy for Population Health Innovation, University of North Carolina at Charlotte, Charlotte, North Carolina.

Acknowledgments

Potential conflicts of interest. M.E.T. and M.F.D. have no relevant conflicts of interest.

References

1. Stanford Medicine. Stanford Medicine 2017 Health Trends Report: Harnessing the Power of Data in Health. Stanford, CA: Stanford Medicine; 2017.
2. Huesch MD, Mosher TJ. Using It or Losing It? The Case for Data Scientists Inside Health Care. *NEJM Catalyst* website. <https://catalyst.nejm.org/case-data-scientists-inside-health-care/>. Published May 4, 2017. Accessed May 30, 2019.
3. Sullivan T. Healthcare is swimming in data, but what to do with it? *HealthCareITNews.com*. <https://www.healthcareitnews.com/news/healthcare-swimming-data-what-do-it>. Published October 23, 2017. Accessed May 30, 2019.
4. Kent J. Big Data to See Explosive Growth, Challenging Healthcare Organizations. Health IT Analytics website. <https://healthitanalytics.com/news/big-data-to-see-explosive-growth-challenging-healthcare-organizations>. Published December 3, 2018. Accessed May 30, 2019.
5. Twitter Usage Statistics. Internet Live Stats website. <https://www.internetlivestats.com/twitter-statistics/>. Accessed May 23, 2019.
6. Tufts C, Polsky D, Volpp KG, et al. Characterizing tweet volume and content about common health conditions across Pennsylvania: retrospective analysis. *JMIR Public Health Surveill*, 2018;4(4):e10834.
7. Health Research Institute, PwC. Social media “likes” healthcare: From marketing to social business. New York, NY: PwC; 2012. <https://www.pwc.com/us/en/health-industries/health-research-institute/publications/pdf/health-care-social-media-report.pdf>. Accessed May 30, 2019.
8. Pogue D. Exclusive: Fitbit's 150 billion hours of heart data reveal secrets about health. *Finance.Yahoo.com*. <https://finance.yahoo.com/news/exclusive-fitbits-150-billion-hours-heart-data-reveals-secrets-human-health-133124215.html>. Published August 27, 2018. Accessed May 22, 2019.
9. Kayyali B, Knott D, Van Kuiken S, Groves P. The 'big data' revolution in healthcare: Accelerating value and innovation. McKinsey & Company website. <https://www.mckinsey.com/industries/health-care-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care>. Published April 2013. Accessed May 30, 2019.
10. Centers for Medicare and Medicaid Services. NHE Fact Sheet. CMS website. <https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpend-data/nhe-fact-sheet.html>. Updated 2019. Accessed May 30, 2019.
11. Anderson L. Realizing the future we want: leveraging technology to achieve a better health care experience for all. *N C Med J*. 2019;80(4):224-228 (in this issue).
12. Health Information Technology Act, 42 USC § 139 (2009).
13. Patient Protection and Affordable Care Act, 42 USC § 18001 (2010).
14. Diaz-Garelli JF, Wells BJ, Yelton C, Stroud R, Topaloglu U. Biopsy records do not reduce diagnosis variability in cancer patient EHRs: are we more uncertain after knowing? *AMIA Jt Summits Transl Sci Proc*. 2018;2017:72-80.
15. Fuller S. Leveraging governance to derive value from advanced analytics. *N C Med J*. 2019;80(4):234-236 (in this issue).
16. Dimensional Insight Dives into the Data Governance Challenges Plaguing U.S. Hospitals: New survey reveals healthcare CIOs struggle with inconsistency and mistrust in data [press release]. Burlington, MA: Dimensional Insight; September 27, 2017.
17. Reisman M. EHRs: the challenge of making electronic data usable and interoperable. *P T*. 2017;42(9):572-575.
18. Revels C. NC HealthConnex and value-based care: statewide health information exchange as a technology tool for all. *N C Med J*. 2019;80(4):229-233 (in this issue).
19. Cohen L. Health data alone doesn't account for much. We need better ways to extract its value. *STATNews.com*. <https://www.statnews.com/2019/03/06/health-data-extract-value/>. Published March 6, 2019. Accessed May 30, 2019.
20. Fuller J. From data to insights: priorities for maturing your health analytics strategy. *N C Med J*. 2019;80(4):234-236 (in this issue).
21. Hague D. Benefits, pitfalls, and potential bias in health care AI. *N C Med J*. 2019;80(4):219-223 (in this issue).
22. Nelson G. Bias in artificial intelligence. *N C Med J*. 2019;80(4):220-222 (in this issue).
23. Harris G, Ong J. Public health 3.0 at Mecklenburg County Public Health. *N C Med J*. 2019;80(4):249-252 (in this issue).
24. North Carolina Department of Health and Human Services, North Carolina Immunization Branch. The North Carolina Immunization Registry (NCIR). NC DHHS website. <https://www.immunize.nc.gov/providers/ncir.htm>. Accessed May 22, 2019.
25. Wallace JW, Decosimo KP, Simon MC. Applying data analytics to address social determinants of health in practice. *N C Med J*. 2019;80(4):244-248 (in this issue).
26. Park A, Bowling J, Shaw G Jr, Li C, Chen S. Adopting social media for improving health: opportunities and challenges. *N C Med J*. 2019;80(4):240-243 (in this issue).
27. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. *Health Educ Q*. 1988;15(4):351-377.
28. Elhauge E. Why we should care about health care fragmentation and how to fix it. In: Elhauge E, ed. *The Fragmentation of U.S. Health Care: Causes and Solutions*. New York, NY: Oxford University Press; 2010:1-20.
29. Marmot M, J Allen. Social determinants of health equity. *Am J Public Health*. 2014;104(suppl 4):S517-S519.
30. ONE Charlotte Health Alliance website. <http://www.oneclhealth.org/>. Accessed May 22, 2019.
31. Cole J. Mapping Social Determinants Proves a Positive Rx for Charlotte's Underserved. *NorthCarolinaHealthNews.com*. <https://www.northcarolinahealthnews.org/2017/02/20/mapping-social-determinants-proves-positive-rx-charlottes-underserved/>. Published February 20, 2018. Accessed May 30, 2019.
32. Tailor S. Spotlight on the Safety Net: The MedLink hub. *N C Med J*. 2019;80(4):253-254 (in this issue).
33. Leavell HR, EG Clark. *Preventive medicine for the doctor in his community: An Epidemiologic Approach*. 3rd ed. New York, NY: McGraw-Hill; 1965.
34. Phillips RL Jr, Bazemore AW, DeVoe JE, et al. A family medicine health technology strategy for achieving the triple aim for US health care. *Fam Med*. 2015;47(8):628-635.
35. Dulin MF, Lovin CA, Wright JA. Bringing big data to the forefront of healthcare delivery: the experience of Carolinas Healthcare System. *Front Health Serv Manage*. 2016;33(1):1-12.
36. Forrest CB, Chesley FD Jr, Tregear ML, Mistry KB. Development of the learning health system researcher core competencies. *Health Serv Res*. 2017;53(4):2615-2632.
37. Greenhalgh TJ, Howick J, Maskrey N. Evidence based medicine: a movement in crisis? *BMJ*. 2014;348:g3725.
38. Noreen N. Data Stewards Play an Important Role in the Future of Healthcare. *IGIQ: A Journal of AHIMA Blog* website. <https://journal.ahima.org/2017/11/17/data-stewards-play-an-important-role-in-the-future-of-healthcare/>. Published November 17, 2017. Accessed May 30, 2019.
39. Wang Y, Kung L, Byrd TA. Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*. 2018;126:3-13.
40. Mehta A. How Data Will Transform Health Care: We Just Need it Better, Faster, and Cheaper. *N C Med J*. 2019;80(4):225-226 (in this issue).