The COMMONWEALTH FUND

Attachment G3

REALIZING THE PROMISE:
APPLYING BIG DATA AND ADVANCED ANALYTICS TO IMPROVE CLINICAL CARE

Ran D. Balicer, Clalit Research Institute
Anural Agrawal, Institute of Genomics & Integrative Biology, CSIR, India
Aziz Sheikh, The University of Edinburgh
Peter Fitzgerald, Center for Cardiovascular Technology, Stanford University
Gabriel Escobar, Kaiser Permanente Northern California

Prepared for:
The Commonwealth Fund
2016 INTERNATIONAL SYMPOSIUM ON HEALTH CARE POLICY

ABSTRACT Big data analytics, its proponents argue, holds promise for major improvements in stretched health care systems worldwide. Hitherto there have been relatively limited data on whether implementing such applications in large populations has measurable impact on patient safety or the quality or efficiency of care. We present a diverse set of case studies that show how applying sophisticated techniques to extract insights from large data sets can be implemented at scale and achieve measurable health improvements in different settings across the globe. These include: enhanced patient safety in high-risk inpatient clinical settings; targeted, proactive, preventive outpatient treatment of noncommunicable illnesses; decision support in treatment selection; and addressing common access and quality challenges in developing countries. The case studies highlight the global challenges associated with realizing the promise of advanced analytics in practice and offer a glimpse into applications that we are likely to see more of in the near future.

INTRODUCTION

Despite mounting health spending, inadequacies in health care persist. Many practitioners believe that if successfully implemented into daily practice, the quality and efficiency of health care can be greatly improved by leveraging large medical data sets via advanced analytical tools to better align individual patient needs and economic efficiency.[1]

Two major categories of health-system shortcomings could benefit from advanced data analytics: structural variations in health care delivery and variable approaches to care. Structural fragmentation in health care delivery systems and differences in health professionals' capabilities often result in variations in the quality of care in different settings.[2] **Standardization** of procedures can help solve these problems by creating systems with the capacity for workflow-integrated guidelines and embedded decision support, such as timely identification of potential errors. The second shortcoming relates to the prevailing practice of reactive, generalized, therapeutic care (rather than proactive, personalized, preventive care), which fails to account for patients' specific characteristics and needs. **Customization and personalization** of health care can help to address this problem by capturing the complexity of patient-specific attributes and needs. As some of the largest health care systems are pursuing major reforms that emphasize patient-centered care impact,[3-6] the stage is set to consider adopting potentially transformative processes toward these aims.

"Big data" is an ill-defined term that refers to massive, noisy data sets that require advanced analytical tools to identify patterns and associations that can be used to gain predictive and prescriptive insight into complex systems.[7] The term is often used to describe both the data and the technology required to capture, store, and analyze it. It is claimed that the multidimensionality of big data sets—i.e., volume, velocity, variety, veracity—could help drive significant improvements in health care by informing meaningful standards of care, while also providing the flexibility (i.e., depth, variety, and real-time dimensions) to deliver personalized care strategies.[8]

Despite recent advances in handling and analyzing big data sets, health care organizations have been slow to adopt or fully exploit their potential.[9] In this paper, we describe case studies in which innovations, which are based on large data streams and advanced techniques to analyze them (whether or not they constitute "big data" by formal definitions), have been implemented at scale. We have not conducted an exhaustive review of global achievements; rather, we aim to highlight relatively mature case studies from different health care systems and clinical settings. These examples span inpatient and outpatient care and in-person and remote telemedicine care provision, in both economically developed and developing countries.

Case studies from economically developed countries presented here focus on increasing personalized patient care and predictive prevention to achieve greater precision and safety through three types of applicative use cases: **inpatient early warning systems**, **chronic disease risk stratification**, and **treatment course selection decision support**. Case studies we present from economically developing countries address their unique challenges, exemplified by the **frugal innovations** in India used to tackle variations in access to care and quality of care due to structural fragmentation and organizational inefficiencies. For both high- and low-income health systems, we set out to outline the challenges that arose and the strategies used to overcome them.

IMPROVING PATIENT SAFETY: PREDICTING AND PREVENTING LIFE-THREATENING EVENTS THROUGH REAL-TIME WARNING SYSTEMS

For hospital patients at high risk for sudden deterioration, care provided outside the intensive care unit can be ineffective at averting adverse consequences, which can often go undetected.[10] Utilizing data from large clinical data sets and predictive modeling allows stratification of the population by risk for the event(s) of interest, which, in turn, allows targeted, proactive, preventive interventions that are otherwise too resource-intensive or clinically inappropriate.

One program that makes use of this type of data is Kaiser Permanente Northern California (KPNC), an integrated health care delivery system serving 3.9 million members at over 200 locations, including 22 hospitals. KPNC is deploying an early warning system (EWS) for early detection of possible deterioration of patients on regular wards in all of its hospitals. The decision to tackle the EWS challenge emerged from the use of predictive analytics to predict and risk-adjust inpatient and 30-day mortality in all hospitalized patients.[11,12] Using predictive modeling, KPNC identified a subgroup of patients—ward patients who deteriorate outside the ICU, whose mortality rates were three to four times higher than expected.[13-15]

In 2011, KPNC finished deploying a comprehensive electronic medical record (EMR). Using EMR data, researchers developed a predictive model as proof of concept for early detection of deterioration outside the ICU.[16] The model now in use is based on a large sample (650,000 inpatient admissions, of whom 20,000 experienced an adverse outcome; the total number of data points for this model was over 250 million).[17]

Once a patient's probability of deterioration within the next 12 hours is 8 percent or more, a rapid-response team is alerted to assess the patient and implement a rescue protocol. This triggers an escalation component that, depending on the case and patient preference, can lead to early transfer to the ICU for a "soft landing" or, if appropriate, referral to palliative care. This pilot program was implemented in two hospitals in November 2013.[17] In two years, the EWS identified over 1,500 at-risk patients, and preliminary data suggest that 90-day mortality has decreased by 20 percent to 25 percent in patients who triggered an alert.[18]

Following the pilot, KPNC has begun deploying the EWS system-wide, a challenging process given both structural and clinical variations. Dedicated implementation teams are a crucial part of this process, including experts on patient safety, clinical specialists, and skilled negotiators who work with local teams to develop workflows. KPNC teams have found that the difficulties involved in developing the predictive model and embedding it into the EMR pale in comparison to the difficulties involved in persuading clinicians to adopt new workflows that incorporate predictive models. In such planning and execution, complex issues require explicit decision-making. For example, some patients with a high predicted probability of deterioration may not desire clinical rescue; consequently, prior to deployment of the EWS, extensive discussions, negotiations, and resource-allocation decisions were required to ensure availability of social workers and palliative care staff (publication in progress).

Moving beyond the use of currently available traditional monitoring devices, experimental, noninvasive sensors—for example, those that are worn, mounted to the patient's bed, or video-based—have been shown to allow continuous remote monitoring, preemptively warning of imminent inpatient decompensation. A surgical unit in a California medical center used one such tool to track over 7,500 patients during a nine-month period and, when compared to a matched control unit, the ICU transfer rate dropped by nearly 50 percent (from 120.1 to 63.5 ICU days/1,000 patients) and length of stay was shortened (from 4.0 to 3.6 days).[19]

TRANSITIONING CARE UPSTREAM: PREDICTIVE AND PROACTIVE OUTPATIENT CARE

In an era when the prevalence of noncommunicable disease multimorbidity is increasing,[20] two major trends emerge for developing more effective, efficient, people-centered care: integrating care across settings and providers, and influencing the incidence trends by targeting early detection and prevention instead of late-stage disease control.

Reducing Readmissions Through Risk Stratification

Preventing 30-day hospital readmissions has been prioritized in many large health systems in recent years through powerful incentive programs.[21] In Israel, a large-scale risk-stratified intervention program was developed and has been implemented since 2012 for 500,000 people 65 year old and older (making up two-thirds of the elderly population). This initiative, called the 3-ARM program, is comprised of automated predictive modeling, readmission-prevention stratified intervention, and ongoing monitoring through e-indicators.[22,23]

This 3-ARM program was formulated in Israel's largest integrated payer/provider health care organization, Clalit Health Services (Clalit), which provides primary, specialty, and hospital care for over half of the Israeli population. It relied on Clalit's integrated data warehouse, which has received extensive clinical data feeds (diagnoses, medication, labs, imaging, etc.) dating back 15 years from EMRs deployed in all of its inpatient and outpatient settings.

The Clalit "preadmission readmissions" prediction model (where risk for readmission is determined immediately upon initial admission) uses a feature selection stage that assesses hundreds of clinical variables with millions of data points. A predictive model comprised of 16 variables allowed adequate accuracy (c-statistics of 0.7). Inpatients who are deemed high-risk upon admission based on the predictive model receive early integrative assessment by a care-transition nurse, early post-discharge telephone outreach, preset appointments to the general practitioner, and home-visit teams, as needed. These new workflows were supported by feedback loops of clinician education and notes in the EMR to facilitate the new clinical pathways.

An improvement in key process measures was observed following the launch of the 3-ARM program. The seven-day post-discharge outpatient clinic contact rates increased from 60 percent to over 85 percent within six months of implementation. Readmission rates among patients identified by the prediction model declined by 5 percent overall (95% CI: 3.1–5.8) between 2012 and 2015. In contrast, readmission rates over the same period among similar members aged 65 and older, who were not targeted for the intervention (flagged as lower risk), increased by 17 percent (95% CI: 11–22). This suggests that increase in readmission rates among patients targeted for intervention (high-risk) were not only minimized, but that an upward trend was reversed. This relative decline in readmission rates amounts to avoiding approximately 6 percent of the annual expected admission days in these high-risk patients.

Preventing Progression of Chronic Disease to End-Stage Disease Through Risk Stratification

Many chronic diseases have an identifiable pre-disease or early stage disease state, in which patients may benefit from less intensive interventions to arrest disease progression or exacerbation.[24,25] Yet, in many cases, the relative high prevalence of such pre-disease states makes it difficult to implement outreach and labor-intensive care.[26] Stratifying the risk of these pre-disease patients may allow focused outreach efforts for the manageable, best-yield subgroup.

In one such example, Clalit identified candidates for renal replacement therapy (through either dialysis or renal transplantation) from a group of approximately 150,000 patients with early chronic kidney disease (stage 3, with an estimated glomerular filtration rate of 30–60 $mL/min/1.73m^2$) by flagging those at high risk for future end-stage renal failure within five years. A high-risk state indicator in the EMRs and computer-generated “action list” of patients for proactive intervention helped primary care clinics identify the patients at highest risk. Since 2012, a multidisciplinary prescriptive intervention program for these high-risk patients has been implemented in all 1,500 Clalit primary care clinics, targeting the 10,000 patients at highest risk among approximately 130,000 stage 3 chronic kidney disease patients. The program is being evaluated for its impact on reducing five-year renal replacement therapy rates. Preliminary ecologic data suggest that the annual increase in renal replacement therapy rates in Clalit witnessed before 2010 is declining (unpublished data).

In another example, a predictive score for cardiovascular-disease deterioration was implemented throughout the United Kingdom to allow earlier, more accurate targeting of at-risk patients. A derivation- and validation-linked cohort of general practice EMR and mortality data was constructed using 500 general practitioners’ clinical records, yielding tens of millions of data points representing over two million patients, to create more reliable predictive scores compared with traditional tools such as the modified Framingham score.[27] The tool and accompanying screening recommendations were endorsed by the U.K. National Institute for Health and Care Excellence and have become a national standard across primary care systems in the United Kingdom. An incentive program for screening based on this tool has contributed to its widespread adoption. The program is now used in thousands of primary care clinics and pharmacies across the United Kingdom and is expected to reduce the risk of cardiovascular events in adults aged 40 to 74 years.[28]

IMPROVING CARE PRECISION: TREATMENT COURSE SELECTION DECISION SUPPORT

Selecting Appropriate Anti-HIV Drug Combinations in View of Potential Resistance Patterns

One well-established project that harnesses large clinical data sets to select the most appropriate treatment choice is EuResist, a drug-interaction modeling tool for human immunodeficiency virus (HIV) treatments.[29] Recommendations for most appropriate chemotherapeutics combinations are based in this prescriptive model on the likelihood that various anti-HIV drug combinations will work given the infecting virus's genome and patient characteristics, such as gender, age, and past treatments. The model, created through a commercial and academic partnership in which over 20 years of data from more than 66,000 HIV patients in 10 countries, combined three analytic approaches: phylogenetic trees that capture virus mutation patterns, a generative-discriminative method based on a Bayesian model that captures drug–drug and drug–mutation associations enhanced with regressors, and a random forest model. An online tool created based on this model allows clinicians to enter the infecting virus's genome and patient data for the case at hand and receive recommendations for the optimal anti-HIV drug combinations to be prescribed. It is estimated that the tool, freely available via a self-explanatory online portal since 2008, has supported treatment decisions in thousands of cases, helping to shorten the trial-and-error process and improve care.[30,31]

Selecting Chronic-Disease Treatment Courses During Telemedicine Disease Management

Another path toward customizing chronic disease treatment decisions is through the use of data streams from data transmission–enabled medical devices that engage patients and physicians in continuous monitoring.[32]

It is widely recognized that blood pressure readings at a primary care physician's office can be potentially misleading (possibly due to "white coat hypertension").[33] In a large, pragmatic, multicenter, randomized controlled trial in Scotland (the HITS trial),[34-36] patients with uncontrolled hypertension received six months of telemonitored support during self-monitored blood pressure readings and were randomized into two groups; one received optional patient decision support and appropriate supervision from primary care clinicians, while the other received usual care. Participants in the intervention group monitored their blood pressure using a validated automated sphygmomanometer, which enabled automatic transmission of blood pressure readings to a server from which data-driven insights could be accessed by patients and physicians via a secure website. A closed-loop feedback of automated short message service texts or emails sent to participants was also available and, based on set abnormality thresholds, physicians were notified so they could alter the treatment regimen, if necessary. This intervention achieved clinically relevant reductions in systolic and diastolic ambulatory blood pressure.

This program is now being implemented through the Scale-up Blood Pressure Project in general practices across NHS Lothian among roughly 2,500 patients with uncontrolled hypertension. If the program is successful, the plan is to work with health boards to embed this telemonitoring approach in primary care practices throughout Scotland over time.[37] Recent work by the same team also demonstrated the potential for this monitoring and treatment approach to improve outcomes in type 2 diabetes mellitus.[38]

Frugal Adaptation to Low-Resource Systems: Tackling Variations in Access and Quality

Economically developing countries are struggling to provide effective health care with severely restricted funding. Having over one billion citizens and a disproportionately high contribution to the global burden of disease, India illustrates an urgent need for major health care reform, particularly as recent economic growth has not translated into better health care for most people.[39]

Digital health information technology and advanced analytics, when used properly, can form cornerstones of systemic transformation efforts if they can be implemented through affordable and scalable platforms. In the absence of sufficient advanced facilities and enough highly trained clinical experts, telehealth augmented with automatic data-driven clinical decision support offers clear medical benefits.

Actionable Insights Using Data Streams from Telemedicine eHealth Centers in India

The electronic Health Center (eHC) is a rapid deployable solution for increasing health care access. It repurposes surplus cargo containers into telemedicine-capable integrated health centers, retrofitted with an electronic workflow and cloud-based data storage. The eHC was jointly developed by the Council of Scientific and Industrial Research, India, and Hewlett Packard Enterprise, India.[40] It was implemented nationwide in 2012, with over 50 sites established and more underway. The output from medical devices such as glucometers, spirometers, and electrocardiogram machines is directly integrated into the electronic workflow, along with an online dashboard summarizing patient visits, diagnoses, and utilization of medical equipment in real time. Tens of thousands of data points are centrally collected daily on numerous health care variables, including symptoms, vital signs, use of medical equipment, procedures, and treatment. This approach illustrates the potential gains in providing health care by packaging innovative technologies into affordable, new platforms.

The data collected through the eHC are used through dashboard visualization and in-depth analytics to influence policy and practice in several key applications. For example, gender gaps in care specific to individuals in their 20s and 30s were detected across the country, consistent with social inhibitions present among young women when discussing health issues with male health care providers. According to specialists involved in this project, this information is expected to lead to greater integration of local female accredited social health activists to address the gender gap in access to care.

In another recent case, unexpectedly high frequencies of skin problems at one site were detected; follow-up inquiry connected the problems to a high burden of undiagnosed fungal infections and scabies.[40] Educational initiatives for healthier practices and awareness appear to have helped reduce the frequency of such complaints over a three-year period (unpublished data).

ACTIONABLE INSIGHTS USING DATA STREAMS FROM LARGE CROSS-SECTIONAL DATA COLLECTION

It is likely that economically developing countries (and underserved regions in some developed countries) have the most to gain from harnessing data-related innovations because of large variance in care quality and access. Implementation of such innovations will necessitate blending contemporary technologies with existing crude methodologies of data collection such as questionnaire-based point-prevalence studies. The largest study of this type from India is the POSEIDON study, conceived and led by the Chest Research Foundation, Pune.[41] The investigators reached 12,000 primary care physicians and pediatricians from 880 Indian cities and towns and asked them to record demographic details, symptoms, and medical conditions for every patient they saw on February 1, 2011. The final data set included responses from 7,400 health care practitioners, representing data for 204,912 patients, comprising 554,146 symptoms, medical conditions, and basic demographic data.

One example of generating insight from advanced analytics came from layering POSEIDON data over geographic information system data, which include regional data on sales of liquefied petroleum gas (LPG), a clean fuel used for cooking in India. This ecologic analysis revealed that low penetration of clean fuels, such as LPG, was an important cause of increasing respiratory ailments in non-metropolitan Indian cities, probably due to use of alternative dirty fuels, which are related to household air pollution.[42] Attempts to increase LPG penetration are underway nationally,[43] which we estimate could reduce incidence of chronic obstructive pulmonary disease by as much as half.[42]

CHALLENGES FOR IMPLEMENTATION AND THEIR REMEDIES: LESSONS LEARNED

Multiple characteristics common to most health systems may be hindering the application of innovations such as those presented on a large scale and in everyday care processes. These include conservative mindsets, embedded routines, complex regulatory requirements, and inappropriate incentives. In the case studies presented, several common practical challenges are evident.

Availability of Sufficiently "Big" Data

The effort to provide more precise patient-tailored therapeutics and prevention has become a strategic goal of governments and global technology giants alike. While the term "precision medicine" is often considered the *sine qua non* of genetic data analysis,[44-47] it is possible to provide considerably more precise personalized treatment without patient genetic data. The case studies presented suggest that advanced analytics of large data sets of purely "phenotypic" (i.e., demographic and clinical characteristics) records, available in an increasing number of health systems, holds much promise for allowing more precise, proactive preventive care.

In the case studies presented, each research and development process required an extensive longitudinal medical record data set. The relative paucity of such data sets in health care settings worldwide may limit the reproducibility of these described model development processes in lower resource countries and in poorer parts of high resource countries, where these data are lacking. Yet, once these and similar models are developed and validated in one of the more affluent and digitized health care system, it is possible for the model itself to be implemented in a much wider group of settings where little historical data is available.

Attempts by global corporate giants[48-50] and health care organizations[28,51,52] are ongoing to create and curate massive, clinically annotated, longitudinal databases that include, among others, genomic data, wearable or cell-phone sensor-based data, and detailed medical history from medical records. Once achieved, such databases coupled with advanced machine learning will create a whole new playing field for innovation.

Workflow Engineering

It is difficult to implement predictive algorithms in clinical practice, because changing established workflows is a labor-intensive, time-consuming, collaborative process that also requires skilled leadership and dedicated resources.

Accurate predictive models, more and more of which have become publicly available, are only the first step toward achieving clinical impact. Embedded alerts in the EMR can improve provider practice, but careful user-experience engineering is necessary to manage "alert fatigue," especially as more and more alerts are introduced into the system.[53]

The challenge in changing workflows sometimes requires explicitly addressing sensitive "unspoken" clinical decision rules, such as addressing the needs of patients who do not desire clinical rescue. In such cases, embedding the decision support required first addressing the complex issue of identifying and accepting a do-not-resuscitate state for some patients, thus excluding them from the early warning process. This seemingly small step was more complex and lengthy than the model development phase.

The KPNC experience with its EWS case study highlights the importance of a seldom-described resource: dedicated implementation teams consisting not only of clinical experts but also skilled negotiators who can work with local clinical teams to develop, agree, and refine these workflows.[11,12] It is important to realize that the work does not end with deployment. It is necessary to plan workflows for both expected (e.g., onset of influenza season) and unexpected situations (e.g., discovery of a gap in the planned workflow).

Another concern has to do with increased liability in view of the new incoming data streams. In the Scotland hypertension telemedicine case study, interviewed participating physicians were ambivalent about automatic integration of the new tele-measured, data-driven inputs into their EMRs due to potential liability associated with abnormal readings being missed or seen too late.[35,54]

Feedback-Loop Implementation

In view of the difficulties anticipated in introducing new workflows, giving staff and their managers timely feedback on their achievements is critical for adherence. In the Israeli readmission-reduction case study, several feedback-loop mechanisms were used to process and present custom indicators (i.e., number of days from hospital discharge to outreach to patient by primary care clinic) to clinicians, as well as administrative staff. In that case, setting readmission rates as a key performance indicator in the balanced scorecard of both hospitals and general practice regional management was shown to be a key facilitating factor in its implementation.

Partnerships for Driving Large-Scale Innovation Implementation

Due to the absence of readily accessible health information technology infrastructure in most economically developing countries and underserved localities, stakeholders must engage in public–private partnerships to generate successful implementation. This was well demonstrated in the case of India’s telemedicine-capable integrated health centers: The coordinated use of a corporate player and public-sector resources permitted the eHCs to transition from concept to practice, providing care and collecting digital data in less than a year at minimal cost.[40]

The deployment of the POSEIDON study required similar partnership from its start. Because most Indian health care coverage is not provided by the government, there was no list of active practicing physicians or practice types available from the government. Such a list was made available from internal databases of a large pharmaceutical company that had a nationwide network for distributing and collecting questionnaires. This contribution allowed the POSEIDON study to be executed on a budget of about US$20,000.[41]

Evidence-Based Practice and External Validity

Because some of the predictive and prescriptive models afforded by big-data analytics may be detached from classic clinical reasoning, there may be too few control mechanisms employed by health care staff in implementation, potentially raising questions about patient safety safeguards required to be installed in these processes in case of recommendation error.

Moreover, there are non-reproducibility risks associated with implementing models developed for one health care setting into another, regardless of similarities between the two. Data-driven models are always, to some extent, fitted to the data set where they were created and may not be as reliable when data entry, storage, and coding differ, limiting model external validity. Careful validation and impact testing of these models across systems must thus be an integral part of their installation and implementation.

Gathering the evidence base to the model efficacy, effectiveness, safety, and external validity across systems is a daunting task that few systems can afford to provide. New insights into the process of approving and quality assurance of these systems can hopefully be anticipated as these innovations become more widely implemented within health care settings.

LOOKING AHEAD: TWO DOMAINS LIKELY TO DEMONSTRATE CLINICAL IMPACT

The case studies presented here used more traditional, health care provider–generated data types (i.e., EMRs) for their advanced analytic applications. But new data sources and tools may soon be available as several trends continue: genetic sequencing services becoming cheaper and more widely used; wearable sensors become more accurate and affordable; and large, international projects produce massive clinically annotated data sets of phenotypic, genetic, and sensor-driven data. Advanced tools created to analyze these data sources will continuously be introduced and are likely to outperform the current precision-care tools. Yet, implementing these tools at scale to improve health outcomes will take time. We briefly mention two domains that are most likely to be profoundly changed in the near future by such analytic innovations, namely those that rely on interpretation of images and decision support in oncology.

Domains Relying on Visual Interpretation of Images

Interpretation of imaging studies is highly dependent on the professional skills of the radiologist, and the interpretation of a multiple-image study (as many as hundreds of images for a single computerized-tomography study) may be time-consuming and error-prone. Many imaging studies are not reviewed by a radiologist in a clinically relevant timeframe (especially at the time of admission to the hospital), and the number of radiology specialists is unlikely to grow sufficiently to meet global demands in the coming years.[55,56]

For over five years, computer vision was shown to surpass human vision in multiple professional benchmark testing.[57] In such settings, even the most qualified radiologist may soon be unable to match the accuracy and scope of insights achieved by well-trained machines. Preliminary data suggest that in some radiological clinical tasks, such as characterizing lung nodules and diagnosing tuberculosis in chest X-rays, supervised learning algorithms currently deployed in Africa achieve accurate diagnoses in rates superior to those of radiologists.[58,59] Due to its success, computer-aided decision support is expected to become a critical component in the radiology domain workflow, and automated screening will become part of triage, suggesting which studies will be further reviewed by a radiologist.

The promise that lies in using computer vision to interpret imaging studies goes well beyond matching the abilities of human radiologists: When massive databases of radiology images coupled with longitudinal clinical data (pre- and post-study) are fed into advanced machine learning algorithms, an opportunity arises for identifying predictive patterns unrecognizable by human eyes. Available algorithms can already identify correlates for bone density, artery calcification, liver density abnormalities, and more, hinting at the potential for profound insights from computer-assisted image analysis.

Decision Support in Oncology

Oncologists face increasing challenges in selecting the ideal treatment choice due to an increasingly complex set of inputs (phenotype and genotype) that must be considered.[60] Treatment guidelines are beneficial in some cases, but in many cases deviation from the classic pathways offers better outcomes.[38] This complexity creates a clear need for better data-backed decision support systems.

In one of the more widely communicated efforts, IBM has been collecting a knowledge base of published scientific papers, as well as clinical data on oncology patients, such as case characteristics, treatment selection, and outcomes. This data has been coupled with IBM's artificial intelligence analytic platform, called Watson. In collaboration with Memorial Sloan Kettering clinicians and analysts, these algorithms were trained to use clinical reasoning to extract and interpret physician notes, lab results, and clinical research. By combining attributes from a single patient's record with the aggregated knowledge graph and data, Watson for Oncology aims to improve the accuracy of care by suggesting optimized treatment plans for every patient individually.[61,62] Despite considerable media coverage, large-scale studies testing system accuracy against expert opinion and current practice are only emerging. The true impact of large-scale implementation of this system and its comparators is yet to be determined.

CONCLUSIONS

The harnessing of advanced analytics to improve health care has barely begun. Although the case studies presented in this paper offer meaningful innovation and impact, they still rely mainly on traditional data sources and care-provision platforms. With ambitious initiatives for the formation of extensive clinical data sets that integrate detailed clinical records with genomic and sensor-based data underway, it is likely that machine learning will allow critical insights for clinical care at an increasing rate. The potential of big data in health care is undoubtedly exciting, but we believe it will take quite a few more years to see its transformative impact on health care quality disseminating around the globe.

References

1. Murdoch TB, Detsky AS, "The Inevitable Application of Big Data to Health Care." *JAMA*. 2013 309(13):1351-52. Epub April 4, 2013.

2. James BC, Savitz LA, "How Intermountain Trimmed Health Care Costs Through Robust Quality Improvement Efforts." *Health Aff (Millwood)*. 2011 30(6):1185-91. Epub May 21, 2011.

3. Magnussen J, Vrangbaek K, Saltman RB, eds, *Nordic Health Care Systems: Recent Reforms and Current Policy Changes*. Berkshire, England: Open University Press, McGraw-Hill Education (England), 2009.

4. Yip WC, Hsiao WC, Chen W, Hu S, Ma J, Maynard A, Early Appraisal of China's Huge and Complex Health-Care Reforms. *Lancet.* 2012 379(9818):833-42. Epub March 6, 2012.

5. Olejaz M, Juul Nielsen A, Rudkjøbing A, Okkels Birk H, Krasnik A, Hernández-Quevedo C, Denmark: Health system Review. *Health Systems in Transition*. 2012 14(2):1-192.

6. VG R.. Six Countries, Six Reform Models: The Healthcare Reform Experience of Israel, The Netherlands, New Zealand, Singapore, Switzerland and Taiwan: Healthcare Reforms "Under the Radar Screen." *JAMA.* 2010 304(18):2067-71.

7. Khosrow-Pour M, ed., *Big Data: Concepts, Methodologies, Tools and Applications*. Information Science Reference, IGI Global. 2016; Information Resources Management Association.

8. Bellazzi R, Big Data and Biomedical Informatics: A Challenging Opportunity. *Yearbook of Medical Informatics*. 2014 9:8-13. Epub May 24, 2014.

9. Cortada JW, Gordon D, Lenihan B, The Value of Analytics in Healthcare: From Insights to Outcomes. IBM Institute for Business Value, 2012.

10. Kang MA, Churpek MM, Zadravecz FJ, Adhikari R, Twu NM, Edelson DP, Real-Time Risk Prediction on the Wards: A Feasibility Study. *Critical Care Medicine*. 2016 44(8):1468-73. Epub April 15, 2016.

11. Escobar GJ, Gardner MN, Greene JD, Draper D, Kipnis P, Risk-Adjusting Hospital Mortality Using a Comprehensive Electronic Record in an Integrated Health Care Delivery System. *Medical Care*. 2013 51(5):446-53. Epub April 13, 2013.

12. Escobar GJ, Greene JD, Scheirer P, Gardner MN, Draper D, Kipnis P, Risk-Adjusting Hospital Inpatient Mortality Using Automated Inpatient, Outpatient, and Laboratory Databases. *Medical Care*. 2008 46(3):232-39. Epub April 5, 2008.

13. Escobar GJ, Greene JD, Gardner MN, Marelich GP, Quick B, Kipnis P, Intra-hospital Transfers to a Higher Level of Care: Contribution to Total Hospital and Intensive Care Unit (ICU) Mortality and Length of Stay (LOS). *Journal of Hospital Medicine*. 2011 6(2):74-80. Epub Feburary 4, 2011.

14. Liu V, Kipnis P, Rizk NW, Escobar GJ, Adverse Outcomes Associated with Delayed Intensive Care Unit Transfers in an Integrated Healthcare System. *Journal of Hospital Medicine*. 2012 7(3):224-30. Epub November 1, 2011.

15. Delgado MK, Liu V, Pines JM, Kipnis P, Gardner MN, Escobar GJ, Risk Factors for Unplanned Transfer to Intensive Care Within 24 Hours of Admission from the Emergency Department in an Integrated Healthcare System. Journal of Hospital Medicine. 2013 8(1):13-9. Epub October 2, 2012.

16. Escobar GJ, LaGuardia JC, Turk BJ, Ragins A, Kipnis P, Draper D, Early Detection of Impending Physiologic Deterioration Among Patients Who Are Not in Intensive Care: Development of Predictive Models Using Data from an Automated Electronic Medical Record. *Journal of Hospital Medicine*. 2012 7(5):388-95. Epub March 27, 2012.

17. Escobar GJ, Critical Care Syndrome Surveillance Using the Electronic Medical Record: Gizmo Idolatry or Practical Solution? Work in Progress at Kaiser Permanente: The EDIP Project. 2014.

18. Escobar GJ, LaGuardia JC, Turk BJ, Ragins A, Kipnis P, Draper D, Early Detection of Impending Physiologic Deterioration Among Patients Who Are Not in Intensive Care: Development of Predictive Models Using Data from an Automated Electronic Medical Record. *Journal of Hospital Medicine*. 2012 7:388-95.

19. Brown H, Terrence J, Vasquez P, Bates DW, Zimlichman E, Continuous Monitoring in an Inpatient Medical-Surgical Unit: A Controlled Clinical Trial. *The American Journal of Medicine*. 2014 127(3):226-32. Epub December 18, 2013.

20. Tinetti ME, Bogardus ST Jr, Agostini JV, Potential Pitfalls of Disease-Specific Guidelines for Patients with Multiple Conditions. *The New England Journal of Medicine*. 2004 351(27):2870-74. Epub December 31, 2004.

21. Leppin AL, Gionfriddo MR, Kessler M, et al., Preventing 30-Day Hospital Readmissions: A Systematic Review and Meta-analysis of Randomized Trials. *JAMA Internal Medicine*. 2014 174(7):1095-107. Epub May 14, 2014.

22. Shadmi E, Flaks-Manov N, Hoshen M, Goldman O, Bitterman H, Balicer RD, Predicting 30-Day Readmissions with Preadmission Electronic Health Record Data. *Medical Care*. 2015 53(3):283-89. Epub January 31, 2015.

23. Balicer RD, Shadmi E, Israeli A, Interventions for Reducing Readmissions – Are We Barking Up the Right Tree? *Israel Journal of Health Policy Research*. 2013 2(1):2. Epub January 25, 2013.

24. Ruggenenti P, Schieppati A, Remuzzi G, Progression, Remission, Regression of Chronic Renal Diseases. *Lancet*. 2001 357(9268):1601-18. Epub May 30, 2001.

25. Levey AS, Coresh J, Chronic Kidney Disease. *Lancet*. 2012 379(9811):165-80. Epub August 16, 2011.

26. Liu R, Li M, Liu ZP, Wu J, Chen L, Aihara K, Identifying Critical Transitions and Their Leading Biomolecular Networks in Complex Diseases. *Scientific Reports*. 2012 2:813. Epub December 12, 2012.

27. Hippisley-Cox J, Coupland C, Vinogradova Y, et al., Predicting Cardiovascular Risk in England and Wales: Prospective Derivation and Validation of QRISK2. *BMJ*. 2008 336(7659):1475-82. Epub June 25, 2008.

28. Robson J, Dostal I, Sheikh A, et al., The NHS Health Check in England: An Evaluation of the First 4 Years. *BMJ Open*. 2016 6(1):e008840. Epub January 15, 2016.

29. Zazzi M, Incardona F, Rosen-Zvi M, et al., Predicting Response to Antiretroviral Treatment by Machine Learning: The EuResist Project. *Intervirology*. 2012 55(2):123-27. Epub January 31, 2012.

30. Rosen-Zvi M, Altmann A, Prosperi M, et al., Selecting Anti-HIV Therapies Based on a Variety of Genomic and Clinical Factors. *Bioinformatics*. 2008 24(13):i399-406. Epub July 1, 2008.

31. Informa s.r.l. - Roma IUdS, Italy; Karolinska Institutet - Stockholm, Sweden; Universitätsklinikum Köln, Germany; IBM Israel; Max Planck Institut für Informatik - Saarbrücken, Germany. EuResist: Integration of Viral Genomics with Clinical Data to Predict Response to Anti-HIV Therapy. 2016 [August 10, 2016]; http://engine.euresist.org/.

32. McLean S, Protti D, Sheikh A. Telehealthcare for Long Term Conditions. *BMJ*. 2011 342:d120. Epub February 5, 2011.

33. Mancia G, Grassi G, Parati G, Zanchetti A, White Coat Hypertension: An Unresolved Diagnostic and Therapeutic Problem. Springer International Publishing. 2015 doi:10.1007/987-3-319-07410-8.

34. McKinstry B, Hanley J, Wild S, et al., Telemonitoring Based Service Redesign for the Management of Uncontrolled Hypertension: Multicentre Randomised Controlled Trial. *BMJ*. 2013 346:f3030. Epub May 28, 2013.

35. Hanley J, Ure J, Pagliari C, Sheikh A, McKinstry B. Experiences of Patients and Professionals Participating in the HITS Home Blood Pressure Telemonitoring Trial: A Qualitative Study. *BMJ Open*. 2013 3(5). Epub June 26, 2013.

36. Stoddart A, Hanley J, Wild S, et al., Telemonitoring-Based Service Redesign for the Management of Uncontrolled Hypertension (HITS): Cost and Cost-Effectiveness Analysis of a Randomised Controlled Trial. *BMJ Open*. 2013 3(5). Epub June 26, 2013.

37. Informatics UIoPHSa, Scale-up BP. Edinburgh [cited August 10, 2016]; www.telescot.org/scale-up-bp.html.

38. O'Connor S, Why Doctors Are Rethinking Breast-Cancer Treatment. *Time Magazine*. October 1, 2015.

39. Subramanian SV, Subramanyam MA, Economic Growth & Health of Poor Children in India. *The Indian Journal of Medical Research*. 2011 133:685-86. Epub July 6, 2011.

40. Agrawal A, Bhattacharya J, Baranwal N, et al., Integrating Health Care Delivery and Data Collection in Rural India Using a Rapidly Deployable eHealth Center. *PLoS Medicine*. 2013 10(6):e1001468. Epub July 5, 2013.

41. Salvi S, Apte K, Madas S, et al., Symptoms and Medical Conditions in 204 912 Patients Visiting Primary Health-Care Practitioners in India: A 1-Day Point Prevalence Study (the POSEIDON Study). *The Lancet Global Health*. 2015 3(12):e776-84. Epub November 15, 2015.

42. T. Sethi AA, Nagori A, Angraal S, Salvi S. Graphical Modeling Driven Policy Insights for Household Air Pollution: Insights from 204,912 Patients' POSEIDON Study. The 2nd Indian Workshop on Machine Learning; IIT Kanpur 2016.

43. PTI. PM Modi Launches Scheme to Provide Free LPG Connections to Poor. *The Economic Times*. 2016; Section: Politics and Nation.

44. Auffray C, Caulfield T, Griffin JL, Khoury MJ, Lupski JR, Schwab M, From Genomic Medicine to Precision Medicine: Highlights of 2015. *Genome Medicine*. 2016 8(1):12. Epub January 31, 2016.

45. Huang BE MW, Rajagopal G, The Path from Big Data to Precision Medicine. Expert Review of Precision Medicine and Drug Development. 2016 1(2):129-43.

46. House TW, The Precision Medicine Initiative, https://wwwwhitehouse/gov/precision-medicine. 2015. Accessed August 4, 2016.

47. Quackenbush J, When the Genes Fit: A Cancer Cure. How Precision Medicine, Genomics, and Big Data Are Treating Disease. Thomson Reuters, Stateofinnovation.thomsonreuters.com/when-the-genes-fit-a-cancer-cure. Accessed August 4, 2016.

48. Broad Institute Genomic Services, Quality at Scale: Engineered Quality Meets High Throughput Genomics, http://www.genomics.broadinstitute.org. Published April 4, 2016. Accessed August 10, 2016.

49. Google Cloud Platform D, DNAstack Tackles Massive, Complex DNA Datasets with Google Genomics, http://cloud.google.com/customers/dnsstack/. Accessed August 10, 2016.

50. Human Longevity I, Health Nucleus, https://www.healthnucleus.com. Accessed August 10, 2016.

51. National Institutes of Health, Precision Medicine Initiative Cohort Program, https://www.nih.gov/precision-medicine-initiative-cohort-program. Accessed August 10, 2015.

52. U.S. Department of Veterans Affairs, Million Veterans Program (MVP), https://www.nih.gov/precision-medicine-initiative-cohort-program. 2016. Accessed August 10, 2016.

53. Patient Safety Network, U.S. Department of Health and Human Services, Alert Fatigue. https://psnet.ahrq.gov/primers/primer/28/alert-fatigue. Updated July 2016. Accessed August 10, 2016.

54. Davidson E, Simpson CR, Demiris G, Sheikh A, McKinstry B, Integrating Telehealth Care-Generated Data with the Family Practice Electronic Medical Record: Qualitative Exploration of the Views of Primary Care Staff. *Interactive Journal of Medical Research*. 2013 2(2):e29. Epub November 28, 2013.

55. Australian Health Ministers Conference, National Health Workforce Strategic Framework. Sydney, Australia. 2004; https://www.health.gov.au/internet/main/publishing.nsf/content/c66647AAB9871405CA257BF000209A71/$File/national_strategic_health_workforce_framework.pdf. Accessed August 10, 2016.

56. Joyce CM, The Medical Workforce in 2025: What's in the Numbers? *The Medical Journal of Australia*. 2013 199(suppl 5):6S-9S. Epub September 2, 2013.

57. He K, Zhang X, Ren S, Sun J, Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. Epub: arXiv:1502.01852. 2015.

58. Breuninger M, van Ginneken B, Philipsen RH, et al., Diagnostic Accuracy of Computer-Aided Detection of Pulmonary Tuberculosis in Chest Radiographs: A Validation Study from Sub-Saharan Africa. *PloS One*. 2014 9(9):e106381. Epub September 6, 2014.

59. Zennaro F, Oliveira Gomes JA, Casalino A, et al., Digital Radiology to Improve the Quality of Care in Countries with Limited Resources: A Feasibility Study from Angola. *PloS One*. 2013 8(9):e73939. Epub October 3, 2013.