Pharmacy Practice, Education, and Research in the Era of Big Data: 2014-15 Argus Commission Report

Jeffrey N. Baldwin, Pharm.D.,[a] J. Lyle Bootman, Ph.D.,[b] Rodney A. Carter, Pharm.D.,[c] Brian L. Crabtree, Pharm.D.,[d] Peggy Piascik, Ph.D.,[e] Jeffrey O. Ekoma, M.S.[f] and Lucinda L. Maine, Ph.D.[f]

[a] University of Nebraska Medical Center College of Pharmacy
[b] University of Arizona College of Pharmacy
[c]Regis University School of Pharmacy
[d]Wayne State University Eugene Applebaum College of Pharmacy and Health Sciences
[e]University of Kentucky College of Pharmacy
[f]American Association of Colleges of Pharmacy

INTRODUCTION

The American Association of Colleges of Pharmacy Argus Commission is comprised of AACP's five immediate past presidents and is annually charged by the current President to examine one or more strategic questions related to pharmacy education. The term "Argus" refers to a character from Greek mythology purported to have 100 eyes and be "all seeing".[1] President Patricia Chase charged the 2014-15 Argus Commission with an analysis of how the use of "big data", or bioinformatics sciences using large and diverse data sources, might affect pharmacy education.

The Commission met in person in October 2014 and divided the analysis into the distinct study of big data and the primary missions of academic pharmacy: education and specifically the assessment of teaching and learning, patient care practice at the individual and population levels, and research and graduate education. The work was framed by the science plenary keynote by Atul Butte,[2] MD, PhD from Stanford University, presented at the 2014 AACP Annual Meeting, and also included key informant interviews and literature searches.

It became clear through these analyses that the availability over the last decade or longer of large databases and super computer computational power has influenced research programs across many of the areas of scholarship in the pharmaceutical and translational sciences (e.g., computational biology, high throughput drug screening, pharmacogenomics and other "–omic" sciences). Education and practice have, for the most part, only begun to be affected by the big data revolution.

BIG DATA: WHAT IS IT?

In the book, Big Data: A Revolution that Will Transform How We Live, Work and Think[3], Mayer-Schonberger and Cukier note that "there is no rigorous definition of big data". They continue to frame how they approach the topic in their acclaimed book noting that: "big data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationships between citizens and governments, and more." They continue that this era will require society "to shed its obsession with causality in exchange for simple correlations: not knowing why but only what." They liken the impact of this era as no less than how the telescope and microscope changed our comprehension of the universe and appreciation of the presence of germs, respectively. Perhaps, "big data" should be referred to as our "datascope".

The authors continue their clarification by noting that the characterization by some of big data being artificial intelligence is inaccurate. Rather than "teaching computers to think like a human", the analysis of huge data sets allow people to infer probabilities and make predictions. They identify the future potential of the use of big data to diagnose illness and identify treatments, certainly activities that resonate in the realms of pharmacy and health care.

The journal *Health Affairs* published a themed issue on big data in July 2014, and Roski, Bo-Linn, and Andrews[4] described three defining features of this phenomenon as the three V's: volume, variety, and velocity. *Volume* relates to the availability of massive amounts of data which requires flexible and easily expanded data storage, retrieval, and management systems. *Variety* refers to the fact that data come in many formats. In health care this is structured and free-text data (e.g., insurance claims, electronic health records (EHRs), diagnostic images, genomic information, social media, personal fitness device data streams.) *Velocity* refers to the characteristic of the big data infrastructure that makes it possible to manage data more flexibly and quickly. The authors distinguish the use of data from a single large dataset such as Medicare claims data, which they would not characterize as big data analysis, from the combination and analysis of data from the Center for Medicare and Medicaid Services, data from electronic health records (EHRs), and sources of additional data such as a population's fitness or nutrition information for which the authors believe there will be significant advances in health care in the near future.

Former Democratic Senate leader Thomas Daschle contributed a timely commentary to the January 2015 issue of *Academic Medicine* reflecting on the role of academic medicine in a transformational time.[5] He identifies five transformational forces in health policy and technology that will impact every aspect of academic medicine. These are big data, greater transparency, new payment methods, emphasis on wellness, and scope of practice. He identifies "big data" as the first and most consequential force among these five, noting that the vast new supply of information is generating thousands of new products and services and will, with time, allow for personalized medicine, real-time access to evidence-based treatment decision support, and new tools for patient management. He states that "academic medicine ought to be the epicenter of this revolution with three goals in mind. First, academic medicine must understand how big data can be utilized and coordinated effectively in all aspects of health care delivery. Second, it should lead the data education and acquisition efforts for both current and future health care providers. Third, it must help create a catalytic environment for research so that these transformative advances can continue."

Daschle's view parallels that of the Argus Commission. This report will examine the implications of big data on patient/population health, pharmacy education, and numerous fields of relevant research and graduate education.

BIG DATA AND HEALTH CARE

"Big data" has been defined as "the linking of disparate large data sets to provide insight at the individual level."[6] In the area of population health, big data is not new. However, the sources are multiplied and often non-traditional, and the volume of data generated is huge compared to traditional methods.

The Institute of Medicine has promoted the concept of a "learning health care system" in which "each patient-care experience naturally reflects the best available evidence from basic, translational, comparative effectiveness, and health service research, and, in turn, adds seamlessly to learning what works best in different circumstances."[7] It is estimated that in 2005, approximately 75,000 deaths could have been prevented if all states delivered the same quality of care as the highest performing state. In addition, inefficiency in the healthcare system due to unneeded services, fraud, administrative costs, and other issues cost in excess of $750 billion in 2009, about 30% of health spending that year.[8]

Healthcare leads many industries in recognizing and utilizing scientific inquiry over expert opinion and testimonials to guide rational decision-making. However, the evidence-based techniques used in healthcare lag behind many industries in utilizing information technology and analytic techniques to optimally utilize big data to guide personal and population health decisions. The torrent of health-related information, from genomics and payer-provider data to web-based social sites and mobile wearable data sources, must be responsibly gathered, analyzed, and utilized. The rich information contained in EHRs has the potential to improve quality and efficiency in healthcare. Elements of the Affordable Care Act that provide incentives to Accountable Care Organizations (ACOs) to collect, analyze, and exchange health care information may accelerate this process. The information provided by EHRs can be combined with other sources to transform medical practice and improve quality of care.

Murdoch and Detsky[9] describe four ways in which big data may improve quality and efficiency in healthcare:

1. Big data increases the capacity to generate new knowledge by using computational techniques to analyze unstructured data within EHRs including natural language processing to extract medical concepts from free-text. These techniques can create an observational evidence base for clinical questions that was not previously possible. Lin et al[10] have identified clinically relevant, accurate symptom-disease-treatment associations for seven diseases from cancers to chronic and infectious diseases from mining of approximately 2.1 million EHRs. Miller[11] has utilized health-related social media sites such as PatientsLikeMe to research healthcare decision support and patient empowerment for chronic diseases.

2. Personalized medicine initiatives may be integrated into clinical practice by analytical methods that integrate genomics information with EHR data. The Electronic Medical Records and Genomics Network uses natural language processing to phenotype patients.

3. Knowledge dissemination can be optimized. Physicians struggle to stay current with the latest evidence guiding clinical practice due to the volume of information. Despite digital access to much relevant evidence and numerous guidelines, sifting through large amounts of information to make clinical decisions is complex and time-consuming. Analyzing existing EHRs to produce a dashboard for guiding clinical decisions is the focus of a collaboration between IBM's Watson supercomputer and Memorial Sloan-Kettering Cancer Center to assist in diagnosis and choice of treatment for cancer patients. This approach offers suggestions from real-time patient data analysis, rather than rule-based algorithms. In a manner similar to messaging from Amazon about purchases you might consider based on shopping habits of others like you, clinicians may receive messages about diagnostic and therapeutic choices made by well-respected colleagues considering similar patient profiles.

4. Big data analytics may allow delivery of information directly to patients, thereby encouraging a more active role in their healthcare. In the future, medical records may reside with patients rather than healthcare providers. Big data could link traditional health-related data found in the EHR to personal data from other sites without the need to gather the information through patient interview. This offers a chance to integrate traditionally-obtained patient information with social determinants of health such as dietary habits, income, education, and merchant data (tobacco and alcohol products, for example) in a patient-directed fashion. Public health initiatives to reduce smoking and obesity could be delivered by targeted messaging to appropriate people based on social media profiles.

Likewise, Shaikh et al[12] discuss "P5" medicine, an approach that combines molecular immunology, advanced computation, and genomics. P5 medicine is described as follows:

1. Predictive profiles of risk, using biologic and phenotypic information.

2. Preventive medicine, based on early detection and prevention, including population-screening from vast data sets to identify potential benefits, harm, and costs of primary prevention.

3. Personalized medicine, based on targeted therapeutics and diagnostics, enhanced by genomics and principles of evidence-based medicine.

4. Participatory medicine, engaging patients, providers, and health-systems (policy, regulatory, health services research).

5. Population medicine, based on broad structural issues around cost and potential harm that could actually result in worsened social, economic, and health disparities.

In addition to big data, we also have an explosion of small data. Small data is the output of individual patient-tracking data using devices such as smart phone applications (apps). The app collects passive data including patient movement, sleep, and communication data as well as gathering patient-reported data via app-supported surveys. Provider dashboards alert a provider to situations requiring intervention via phone or email. Companies that provide these types of data suggest the following patient care benefits: improved care by detecting poor health-related patient behaviors, permitting early intervention, improved outcomes by early detection of at-risk populations, and decreased costs by preventing emergency department visits and readmissions with early detection and quality follow-up using patient generated data.[13] Patient understanding of a personal genome profile can empower an individual to make better-informed health and treatment choices. Along with self-generated quantified data, patients are likely to become more eager to discuss personal information with their providers, including pharmacists. Pharmacists and other providers could compare this individualized data to large data sets to gain valuable insights to help guide care.

Despite potential benefits, there are barriers to widespread adoption of big data analytics in health care:[9]

1. Health service researchers, pharmaceutical companies, and public health and other government organizations are strongly supportive of the use of big data. However, there are few incentives or champions for big data within clinician groups or hospitals at this time.

2. Solutions to privacy concerns are required and may be more extensive than typical protections for financial data. Research indicates that even large-scale data cannot always be completely and reliably de-identified. HIPAA and IRB requirements for protecting privacy and conducting ethical research greatly complicate typical big data analytics.

3. Current systems may be fragmented and have limited interoperability as we have seen in other areas where multiple systems are used for capturing data.

4. Overreliance on electronic systems can be a safety problem. Training and backup plans are required in case of system failure.

5. Amarasingham[14] cites the challenges posed by health care predictive analytics on the doctor-patient relationship and on medical education and training. As available information and data become more massive, clinicians could potentially become overly reliant on algorithms and predictive formulas at the expense of clinical judgment. There is also the risk of focusing on data and ignoring information that could be acquired through patient-provider discussions and other interactions.

It is often stated that data-intensive approaches to medicine will help solve some of the biggest and most difficult problems in health care. However, we have not yet determined how best to utilize this tsunami of information from numerous data sources, including wearable technologies. Neff[15] suggests that the biggest challenge is social rather than technical. Her data suggest that health care providers view data innovation and the accompanying predictive analysis as resource-intensive to an extent that exceeds the benefits provided.

Predictive analytics hold great potential, but from the perspective of providers in everyday practice, data management requires more time, personnel, and financial resources, not less. Neff's paper uses the example of a primary care physician who is not as interested in routine pedometer readings as in simply encouraging sedentary patients to become more active. Conversely, technology-savvy patients who come to clinic armed with quantified-self data want providers to review and value the data. Some devices and self-quantified applications are regulated by the FDA, especially those deemed to present risk, but most applications are not consistently regulated and the reliability of the data collected is unknown. Until the challenges of integration and operability in practice are solved, a great deal of work remains to integrate data into providers' practices such that they will be useful in various settings and contexts. Research-related implications are discussed below in greater detail.

BIG DATA IN RESEARCH

Perhaps the research mission of colleges and schools of pharmacy has been the most impacted by big data and computational science. This is true for virtually all of the disciplines, including the laboratory sciences, clinical and translational sciences, and the social and administrative sciences. A challenging aspect of preparing this portion of the Argus Commission report is that there is not a taxonomy of terms that adequately encompasses these evolving areas of research. It would seem, however, that virtually every area of what has classically been called "the pharmaceutical sciences" has been or will be influenced and changed as access to data, computing power, and analytical methods evolve in this era of research and discovery. Characterization of this evolution will include brief summaries of key topics.

Causation versus Correlation

In Big Data[3], Mayer-Schonberger and Cukier devote substantial attention to how correlation may trump causation in the conduct of research using big data sources. They note that in the small data age much research was driven by hypotheses, and scientists worked to carefully collect and analyze small data sets. Hypotheses are derived from a variety of theories from the natural and social sciences that are repeatedly tested to "help explain and/or predict the world around us."

Is theory "dead" in the big data era? The authors debunk this notion and rather note that big data is founded on theory – statistical, mathematical, and computer science theories, specifically. They continue, "big data may offer a fresh look and new insights precisely because it is unencumbered by the conventional thinking and inherent biases implicit in the theories of a specific field." So while big data science will not make theories and theoretical frameworks obsolete, these authors do note it will "fundamentally transform the way we make sense of the world." The Argus Commission notes that this is exactly the purpose of science, and, therefore, our research and research training in the biomedical and pharmaceutical sciences must be carefully re-examined.

Data Sources

The amount of data available for research is expanding rapidly through both "organic" means (i.e., data captured secondary to day to day activities) and purposeful efforts to make data for research more accessible. Ninety percent of all data have been created during the last two years. Three billion people will soon be online, creating close to eight zettabytes of data by 2015.[16] One zettabyte, or ZB, is one sextillion or 10^{26} bytes. There are both public and private initiatives underway to expand access to data for research purposes. In a recent viewpoint published in JAMA, authors Mega et al note that "clinical research has reached a pivotal moment, not only with the exponential expansion of tools for data capture as well as data sources, but also with the opportunity to reevaluate how to integrate the information to optimize medical decision making."[17]

Big Data 2 Knowledge (BD2K) is a trans-NIH initiative established to enable biomedical research as a digital research enterprise, to facilitate discovery and support new knowledge, and to maximize community engagement.[18] The BD2K initiative addresses four major aims that, in combination, are meant to enhance the utility of biomedical big data:

- To facilitate broad use of biomedical digital assets by making them discoverable, accessible, and citable.
- To conduct research and develop the methods, software, and tools needed to analyze biomedical big data.
- To enhance training in the development and use of methods and tools necessary for biomedical big data science.
- To support a data ecosystem that accelerates discovery as part of a digital enterprise.

Overall, the focus of the BD2K initiative is the development of innovative and transforming approaches as well as tools for making big data and data science a more prominent component of biomedical research. Several current funding opportunities under the BD2K initiative have been posted to the NIH web site.

ClinicalTrials.gov was established as part of FDA modernization legislation passed in 1997 and is a web-based resource that provides patients, their family members, health care professionals, researchers, and the public with easy access to information on publicly and privately supported clinical studies on a wide range of diseases and conditions.[19] The web site is maintained by the National Library of Medicine (NLM) at the National Institutes of Health (NIH).

The ClinicalTrials.gov registration requirements were expanded after Congress passed the FDA Amendments Act of 2007 (FDAAA). Section 801 of FDAAA (FDAAA 801) requires more types of trials to be registered and additional trial registration information to be submitted. The law also requires the submission of results for certain trials. This led to the development of the ClinicalTrials.gov results database, which contains information on study participants and a summary of study outcomes, including adverse events.[20] The results database was made available to the public in September 2008. As of January 2015, 182,394 studies from all 50 states and over 150 countries were in ClinicalTrials.gov.

A private sector initiative was launched in May 2013 to provide investigators the opportunity to request access to de-identified patient-level data from private sector clinical trials. Originally the initiative was solely sponsored by GlaxoSmithKline and limited to data from their clinical trials. However, the database had expanded by early 2014 to include data from 10 companies. In early 2014 over 1,200 studies were listed on the site. Researchers can submit requests to access data through a publicly available web site, and those requests are reviewed by an independent panel of scientists. According to Strom et al, early utilization of such data sources "can lead to conclusions about the types of patients who should receive a given treatment that differ from the conclusions drawn by the original investigators."[21]

The International Society for Pharmacoeconomics and Outcomes Research (ISPOR), which was founded by leading health outcomes researchers 20 years ago this year, maintains an international digest of databases for relevant research listed by country.[22] Over 350 sources of data are currently listed, including over 150 public and private databases from the United States. ISPOR has included plenary sessions and related programs on big data and data analyses in programming for the past several years.

Another big data source has been under development by the Food and Drug Administration for several years. FDA launched its Sentinel Initiative in 2008 with the aim of developing and implementing a proactive system that will complement existing systems that the Agency has in place to track reports of adverse events linked to the use of its regulated products.[23] Subsequently they also initiated a "Mini-Sentinel" pilot program, the Agency's first step towards building a nationwide rapid-response electronic safety surveillance system for drugs and other medical products.

In actuality, the Mini-Sentinel database is not so "mini"; the pilot project, which took two years to develop, includes 17 data partners across the U.S. and encompasses the data of nearly 100 million patients. Mini-Sentinel evaluations will help scientists better understand potential safety issues associated with FDA-approved medical products. Importantly, scientists can get responses to their questions in a matter of weeks, as compared to months or even longer, using traditional surveillance methods.

One additional unique source of data that AACP has come to appreciate is maintained by the National Minority Quality Forum (NMQF). NMQF President Gary Puckrein, PhD, has frequently presented at AACP meetings. The Forum's nationwide database is a tremendous resource for researchers, corporations, and health-care organizations that need to understand and pinpoint disparities at the zip-code level in order to target chronic-illness interventions for specific populations.[24]

A unique aspect of the National Minority Quality Forum's overall efforts to improve self-care, prevention, treatment, management, and monitoring has been the use of the Forum database to map health data in the United States. These web-based resources enable users to view and compare information graphically—nationwide, by state, by congressional district, and by state legislative district. The indexes and atlases provide focus and perspective for educational, advocacy, and public-affairs initiatives. The indexes and atlases are available to advocacy groups, policy makers, and sponsoring organizations through a password-protected link from the Forum web site. They are updated periodically as new data become available.

Examples of Current Research Programs

As part of their work on this year's report the Argus Commission interviewed a small number of AACP members identified as leaders of programs making substantial commitments to big data research and education. One such program is the University of Maryland, Baltimore (UMB) where the School of Pharmacy is the lead in an all-academic health center-wide partnership with Optum Labs.[25] Optum Labs was co-founded in January 2013 by Optum (the health services business of UnitedHealth Group) and Mayo Clinic as the industry's first open, collaborative research and innovative center.[26] Based in Cambridge, MA, Optum Labs enables a community of healthcare stakeholders to conduct cutting-edge research by sharing information assets, technologies, knowledge tools, and scientific expertise, with the primary goal of improving patient care and patient value. Optum Labs' data and resources are made available to partners to support collaborative, publication-oriented research, and clinical translation of innovation. Partners include, but are not limited to, several organizations (AARP, Medical Group Management Association), corporations (Merck, Novartis), health plans (Lehigh Valley Health Network), and universities (Harvard, Brown, Johns Hopkins), each of which brings a range of assets to the collaboration.

Eleanor Perfetto, Ph.D., professor in the Department of Pharmaceutical Health Services Research, leads the collaborative for the UMB. She notes that their work will largely, though not exclusively, focus on Alzheimer's disease and healthy aging research. The collaboration will "enhance and augment UMB's existing research and informatics resources with the data, tools, expertise, and infrastructure available at Optum Labs."[25]

Past AACP President and former Argus chair Dean Marilyn Speedie was also asked to provide insights into how big data science has become part of the research programs at the University of Minnesota. She noted areas that have thrived for quite a few years such as high throughput drug candidate screening. More recent research includes the work of a linguist who completed a fellowship in health informatics who utilizes language patterns available in large databases for the early identification of individuals with Alzheimer's disease and to detect cognitive side effects of CNS drugs such as antiepileptic agents. Genomics, epigenetics, and proteomics scientists also are dealing with huge data bases, usually in collaboration with the computer science and math faculty.[27]

Patsy Babbit, PhD, Professor in the Department of Bioengineering and Therapeutics at the School of Pharmacy at the University of California – San Francisco (UCSF), noted that big data sciences have been evolving over a period of several decades with the maturation of bioinformatic sciences, analytical methods, and ever more powerful computational power.[28] This has influenced research in many individual fields and has driven much multidisciplinary collaboration. UCSF joined with UC – Berkeley and UC – Santa Cruz to create the Quantitative Biosciences 3 Consortium to promote cross-campus, interdisciplinary research using quantitative approaches to tackle the most difficult challenges.[29] Partnering across institutions helped address the high cost of much of the necessary equipment required for this work and sought the synergy and productivity that could be achieved through such a collaboration. The announcement in January 2015 of the appointment of Dr. Atul Butte to lead the new UCSF Institute for Computational Health Sciences signals the continued commitment of this leading research institution to maximize the ability of researchers and clinicians to integrate big data analytics and knowledge into practice, research, and education.[30]

BIG DATA IN PHARMACY EDUCATION

The Bill and Melinda Gates Foundation published its Postsecondary Success advocacy plan. The first priority identified was educational data and information. Their recommendation is "(c)reate a national data infrastructure that enables consistent collection and reporting of key performance metrics for all students in all institutions that are essential to promoting the change needed to reform the higher education system."[31]

The magnitude of the data sets used in education tend to be much smaller than those described previously for health care and research, yet educational datasets may well represent big data. The 2013-14 AACP Academic Affairs Committee[32] produced a good overview of the use of big data and analytics in pharmacy education, specifically delineating the differences between "academic analytics" and "learning analytics." Academic analytics are those that help administration, typically at the full university level or higher (e.g., consortial level) to identify effectiveness of units and/or processes. These analyses may help make decisions on resource allocation, changing programs to better meet their desired goals, etc. Fain[33] described an effort currently underway at the consortial level (the University Innovation Alliance) where innovations may be assessed and shared. This group is also looking at pooling the same data from the member schools to identify trends or gaps in data. Colleges/schools often use academic analytics to look at their student populations to identify admissions criteria from predictors of programmatic success as well as to identify predictors of success within their student bodies. This latter area often transitions to "learning analytics" when the results are used to aid the success of an individual student.

Learning analytics is the "micro" equivalent of the "macro" academic analytics focusing on assessing factors influencing an individual student's performance or learning. Kolowich[34] reported on several learning management systems (LMS) that collect student data (e.g., participation in various on-line activities, access of on-line resources, activity during lecture). Analyzing these LMS data against student performance allows the identification of certain predictors of individual success in a course. Programs cited were Student Explorer®, ECoach®, and Lecture Tools®. Online testing software can capture student performance on assessments and match these to desired course and programmatic outcomes. Most of these software packages also have tools to increase faculty efficiency and ease with identifying and communicating with students exhibiting behaviors that affect their performance.[34] Often, these analytics can: occur in real time (e.g., classroom programs showing student visits to course site tools); produce both individual student and aggregated data; indicate student progress or lack thereof over time; identify the preferred learning habits/styles of individual students or a class; and assist in identifying at-risk students through more than grades and before it is too late to recover.

One huge impact that the availability of learning analytics big data will have on education is the push that programs will feel to move to totally digital processes that thereby produce digital data that can be analyzed. Currently, most schools and universities do gather and store vast amounts of student data. However, individual student performance data, and especially learning behavior data within courses, is not available electronically. With analytics giving faculty the ability to better assist students in succeeding in individual courses or entire programs, the academy will see pressures to have the raw data available for these analytics to mine. The use of digital programs within the classroom will likely meet with resistance from many faculty.

When discussing the generation and analysis of big data relevant to potential pharmacy students, enrolled pharmacy students, and pharmacy graduates, it is essential to recognize that student information, academic records, student assistance information, and disciplinary processes merit access only to the extent appropriate to legal and ethical restraints. Similarly, private information related to all individuals must be preserved to the extent dictated by law and ethics.

How do pharmacy colleges and schools gain the attention of high school and of college students deciding on career direction in today's sometimes-commoditized and highly competitive health professions education "marketplace?" In what ways can we use information, including big data where appropriate, to identify and recruit students to pharmacy education and careers? Huemme discusses marketing to consumers. As mobile sensing devices, remote sensing, cameras, microphones, and wireless networks grow, so does the data being generated. The WhatsApp messaging program now owned by Facebook provides insight into consumer behavior in social domains. Understanding such behavior requires analysis of consumer response to commercial messages, including descriptive, diagnostic, predictive, and prescriptive assessments.[35] In what ways can the pharmacy academy use social media to promote pharmacy careers, including career direction, as well as pre-pharmacy and pharmacy curricula and specific pharmacy college selection?

Digitalization of our pharmacy recruitment, admissions, and educational program performance and outcome information should provide the pharmacy academy with big data resources that can help guide our recruitment and selection of the most appropriate candidates for admission based on pooled student performance data. Pooled digitalized big data should also be useful to pharmacy colleges and schools in assessing and modifying pre-pharmacy and pharmacy curricula based on national trend data from a number of diverse sources which potentially could include data such as de-identified reaccreditation information, students' attitudinal and outcomes data (e.g., test, course, curricular, college processes), faculty member and preceptor attitudinal and outcomes data, and attitudinal data from those employing pharmacy graduates. Big data generated from employers, graduates, pharmacists, preceptors, other health professionals, and the public could foster a more clearly defined standard for professionalism and can potentially help direct pharmacy students and college advisors in making career focus and pathway decisions in a timelier manner than decisions which are often not made until the final year in the Pharm.D. program.

Advancements in big data analysis allow us to improve health-care decision making processes. These may include improvements in simulations, disaster management and emergency care, acute and ambulatory care, medication therapy management, patient health outcomes, and preventative care. The pharmacy academy should seek to optimize its involvement in efforts to utilize big data to improve its participation in health care education, service, and scholarly efforts.

The AACP Transformative Community Engagement Award annually recognizes a college for exemplary activities in provision of care within the communities within the scope of its influence. The assessment processes for such recognition include evaluation of the college's mission, leadership, budgetary commitments and marketing, strategic planning, faculty and student development, curricular engagement, strategic partnering, and service delivery and impact relevant to the communities of concern. Such community partnerships help to eradicate community perceptions of "ivory tower" colleges and universities. Big data can be generated by many methods currently being utilized in assessing communities' attitudes and behaviors and in developing public trust in our societal institutions, such as the measured response in Boston following the bombing during the Boston Marathon in 2013. This was based on a Citizens Connect program which permitted the search for suspects to progress in an orderly manner. Use of such data sets could provide opportunities for health professions programs to more meaningfully transform our communities as partners.[36]

There are a number of services available to the public that may generate big data sets of interest to the pharmacy academy and the communities we serve. Examples may include health care, including pharmacy services, locations and information, such as hours of service, availability of needed services such as delivery, retail clinic, or telemedicine availability, and transportation available to access care various locations. Additionally, online information about health, wellness, and medication information may all influence health care behaviors involving pharmacist care. Big data may also foster community pharmacy research, such as the influence of health and wellness, medication therapy management, and nutrition counseling services on obesity and diabetes mellitus within communities vs. communities lacking such services. A few individuals with expertise in data organization and technology may provide communities with help in organization and interpretation of big data.[37]

Preparing Faculty and Future Researchers

The expansion of data in terms of volume and variety over the past decade is truly transforming research and clinical practice. In her interview UCSF Professor Babbit noted that there must be changes in education at both the professional doctoral and graduate levels to equip future clinicians and scientists to thrive in the big data environment. The National Science Foundation and NIH have encouraged the development of new training programs and have been emphasizing the importance of creating new and diverse career pathways for individuals with bioinformatics and statistical training.

UCSF has brought together five different graduate programs in the Quantitative Biosciences Program. The programs include Bioengineering, Biological and Medical Informatics, Biophysics, Biology and Chemical Biology, and Pharmaceutical Sciences and Pharmacogenomics. The consortium promotes curriculum development and faculty and student engagement in the quantitative biosciences. It provides a common physical space for graduate students across the five programs to meet, discuss, and debate ideas. It helps support the financial needs of graduate students in the quantitative biosciences, thus helping free these students to focus their energies on addressing important biological questions.

Dr. Babbit noted that the International Society for Computational Biology (ICSB) and related organizations host many of the conferences attracting researchers in relevant fields. The ISCB web site[38] also hosts a job board which lists a fascinating array of positions around the world at universities, research institutes, and corporations. This includes numerous postdoctoral training positions. Certainly, the market for well-prepared scientists equipped to participate in and lead this era of research will grow and become highly competitive for the foreseeable future.

Dr. Philip Bourne assumed the role of Associate Director for Data Sciences at NIH in 2014. He leads the development of the overall NIH vision in Data Science and coordinates across the 27 Institutes and Centers in support of biomedical research as a digital enterprise. In a recent blog post, Dr. Bourne noted that "Training is needed at all levels, from the branding of biology (from K12 education on up) as an analytical science to retraining of established biomedical researchers to understand and contribute to the emerging digital enterprise. A first step is to catalog what is already being taught to determine what is lacking in specific areas. Such training should engage experts from a variety of disciplines such as computer science, behavioral and social sciences, statistics, mathematics, and others. New physical and virtual training facilities will likely be needed to complete coverage of relevant topics. Educational content itself should be part of the digital enterprise and hence be catalogued and findable using appropriate metadata standards."

The 2014-16 Research and Graduate Affairs Committee chaired by University of Maryland Dean Natalie Eddington was charged to critically examine graduate education in the pharmaceutical sciences. Dean Eddington confirmed that the computational and analytical skills required of scientists to be successful will be included in their analysis. Their report, which will be issued in spring 2016, will provide much needed guidance for the broad array of graduate programs across the colleges and schools of pharmacy.

POLICY IMPLICATIONS IN THE ERA OF BIG DATA

Clearly, colleges and schools of pharmacy and their faculty have found and will continue to find that their work both creates and utilizes big data of numerous types from many sources. The ability to access, analyze, and apply such data to our work is essential. The Argus Commission felt it was important to consider policy implications related to their charge and accessed a report to President Obama from his Council of Advisors on Science and Technology. "Big Data and Privacy: A Technological Perspective"[39] offers several recommendations germane to academic pharmacy. Certainly the collection and analysis of data have the potential to impact the privacy of individuals and groups in ways unimaginable at times when current laws and regulations were promulgated. It is essential that those engaged in the generation and use of big data remain knowledgeable of applicable and emerging laws and regulations and contribute their insights into policies and procedures that will insure the protection of our students, our faculty, our patients, and those who are involved with our research and assessment activities. As stated in the recommendations to the President, policies at every level should focus more on the actual uses of big data and intended outcomes of that use rather than on data collection, storage and analysis, or specific technological solutions.

CONCLUSIONS

The role of big data will be increasingly important in the pharmaceutical and clinical sciences at all levels of our service, education, and research agendas. Investment in the data sciences should be a part of all colleges' and schools' strategic efforts. Academic pharmacy has a tremendous opportunity to be a major contributor to the big data movement by investing in the faculty who wish to conduct big data research from the molecular drug discovery level to impacting populations and health care overall. Given the growing number of public access data sources, researchers from established as well as emerging research institutions have few limits on engaging in the big data research arena. Collaborations among and between institutions is strongly encouraged.

Big data analysis will also have profound impact on how colleges and schools assess student outcomes and evaluate curricula, as well as in developing successful researchers within our graduate research programs and competent pharmacists from our professional degree programs. Opportunities are endless where colleges and schools of pharmacy can play a role in this endeavor. Big data can serve as excellent platform for interprofessional education and research programs. Our engagement in the big data processes must be undertaken with the understanding that the term is imprecise at best, sometimes reflecting primarily the size of the dataset, while in other cases representing the size of the efforts involved in data gathering and use.

REFERENCES

[1] The Myth Encyclopedia. http://www.mythencyclopedia.com/Am-Ar/Argus.html. Accessed April 23, 2015.

[2]Butte A. American Association of Colleges of Pharmacy 2014 science plenary: research in education and care in the era of big data. https://www.youtube.com/watch?v=XjQ2lbherj0&feature=youtube. Accessed April 24, 2015.

[3] Mayer-Schonberger V, Cukier K. *Big Data: A Revolution that Will Transform How We Live, Work and Think*. New York: Mariner Books; 2013.

[4] Roski J, Bo-Linn GW, Andrews TA. Creating value in health care through big data: opportunities and policy implications. *Health Affairs*. 2014;33(7).

[5] Daschle TA. Academic medicine in a transformational time. *Acad Med*. 2015;90(1).

[6] Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. *JAMA*. 2014;311(24):2479-80.

[7] Olsen LA, Aisner D, McGinnis JM, eds. The learning healthcare system: workshop summary (IOM roundtable on evidence-based medicine). Washington, DC: The National Academies Press, 2007. http://www.nap.edu/catalog.php?record_id=11903. Accessed April 24, 2015.

[8] IOM (Institute of Medicine). *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America*. Washington, DC: The National Academies Press, 2013.

[9] Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA*. 2013; 309(13):1351-52.

[10] Lin YK, Brown RA, Yang HJ, Li SH, Lu HM, Chen H. Data mining large-scale electronic health records for clinical support. *IEEE Intell Syst*. 2011;26:87-90.

[11] Miller, K. 2012. Leveraging social media for biomedical research: how social media sites are rapidly doing unique research on large cohorts, biomedical computation review. http://biomedicalcomputationreview.org/content/leveraging-social-media-biomedical-research. Accessed April 24, 2015.

[12] Shaikh AR, Butte AJ, Schully SD, Dalton WS, Khoury MJ, Hesse BS. Collaborative medicine in the age of big data: the case of cancer. *J Med Internet Res.* 2014;16(4):e101.

[13] Big Data, Better Health. Smarter Care Starts with Your Smartphone. https://ginger.io. Accessed April 24, 2015.

[14] Amarasingham R, Patzer RE, Huesch M, Nguyen NQ, Xie B. Implementing electronic health care predictive analytics: considerations and challenges. *Health Affair*. 2014;33(7):1148-1154.

[15] Neff G. Why big data won't cure us. *Big Dat*a. 2013;1(3):117–123.

[16] Hansen MM, Miron-Shatz T, Lau AYS, Paton C. Big data in science and healthcare: a review of recent literature and perspectives. In: *IMIA Yearbook of Medical Informatics*. 2014;9:93-95.

[17] Mega JL, Sabatine MS, Antman EM. Population and personalized medicine in the modern era. *JAMA*. 2014;312(19):1969-70.

[18] National Institutes of Health: Data Science. Big data to knowledge (BD2K). http://datascience.nih.gov/bd2k. Accessed April 24, 2015.

[19] National Institutes of Health: ClinicalTrials.gov. ClinicalTrials.gov background. https://clinicaltrials.gov/ct2/about-site/background. Accessed April 24, 2015.

[20] National Institutes of Health: ClinicalTrials.gov. About the results database. https://clinicaltrials.gov/ct2/about-site/results. Accessed April 24, 2015.

[21] Strom BL, Buyse M, Hughes J, Knoppers BM. Data sharing, Year 1 – Access to data from industry-sponsored clinical trials. *NEJM*. 2014;371:2052-2054.

[22] International Society for Pharmacoeconomics and Outcomes Research. ISPOR international digest of databases. http://www.ispor.org/DigestOfIntDB/CountryList.aspx. Accessed April 24, 2015.

[23] Food and Drug Administration. FDA's sentinel initiative. http://www.fda.gov/Safety/FDAsSentinelInitiative/default.htm. Accessed April 24, 2015.

[24] National Minority Quality Forum. Forum databases. http://www.nmqf.org/forum-databases. Accessed April 24, 2015.

[25] Carroll M. Partnership led by the school of pharmacy will increase the scope and impact of alzheimer's disease and healthy aging research. https://rxsecure.umaryland.edu/apps/news/story/view.cfm?id=502&CFID=5776718&CFTOKEN=d1e6350a275ee37d-3CEA0E8E-BAFD-ED66-B820ABDFDC791035. Accessed April 24, 2015.

[26] Optum Labs. Industry collaborations and partners. https://www.optum.com/optumlabs/our-partners.html. Accessed April 24, 2015.

[27] M.Speedie, email, April 24, 2015.

[28] P. Babbit, telephone conversation, February 28, 2015

[29] University of California San Francisco. UCSF taps Atul Butte to lead big data center. http://www.ucsf.edu/news/2015/01/122656/ucsf-taps-atul-butte-lead-big-data-center. Accessed April 24, 2015.

[30] Bill and Melinda Gates Foundation. Postsecondary success advocacy priorities. http://postsecondary.gatesfoundation.org/wp-content/uploads/2015/03/PS-ADV-Priorities-V1.pdf. Accessed April 24, 2015.

[31] Cain J, Conway JM, DiVall MV, et al. Report of the 2013-2014 academic affairs committee. *Am J Pharm Educ*. 2014;78(10):Article S23.

[32] Fain P. University innovations alliance kicks off with big completion goals. *Inside Higher Ed.* September 17, 2014. https://www.insidehighered.com/news/2014/09/17/university-innovation-alliance-kicks-big-completion-goals. Accessed May 5, 2015

[33] Kolowich S. Can universities use data to fix what ails the lecture? *The Chronicle of Higher Education*. August 11, 2014. http://chronicle.com/article/Can-Colleges-Use-Data-to-Fix/148307/. Accessed May 5, 2015.

[34] Educause. 7 things you should know about analytics. Educause Learning Initiative; 2010. http://net.educause.edu/ir/library/pdf/ELI7059.pdf. Accessed April 24, 2015.

[35] Huemme C. Leveraging big data and creative content for customer engagement. http://www.ayantek.com/leveraging-big-data-and-creative-content-customer-engagement. Accessed April 24, 2015.

[36] Lusk K. The rewards and risks of big data: urban seminar panelists discuss the open city. http://www.bu.edu/ioc/the-rewards-and-risks-of-big-data-urban-seminar-panelists-discuss-the-open-city/. Accessed April 24, 2015.

[37] Thakuriah P. Urban informatics, big data and community engagement: a study of enabling organizations. http://digitalinfomediary.wordpress.com/. Accessed April 24, 2015.

[38] International Society for Computational Biology. ISCB – membership job board. http://www.iscb.org/iscb-careers. Accessed April 24, 2015.

[39] President's Council of Advisors on Science and Technology. Big data and privacy: a technological perspective. Executive Office of the President. https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf. Accessed April 24, 2015.

RECOMMENDATIONS AND SUGGESTIONS

Recommendation 1: AACP should assist member colleges and schools with the identification and use of relevant data analytic technologies, facilitate the development of consortia to advance collaboration, and serve as a secure repository for de-identified data that could be used to identify innovative practices in teaching, learning and assessment.

Recommendation 2: AACP should provide guidance to member colleges and schools to assure that any student's, employee's, or other individual's information included in big data reporting and analyses be limited to that information that can be legally and ethically divulged.

Recommendation 3: AACP and member colleges and schools, in partnership with individuals with big data expertise, should assess methods of using social media and other potential marketing tools to identify, engage, recruit, and retain students to programs offered by the pharmacy academy and provide the academy with resources to apply such processes.

Recommendation 4: AACP and member colleges and schools, in partnership with individuals with big data expertise, should develop data-driven systems and processes to guide member colleges and schools in areas such as:

- the recruitment and selection of the most appropriate candidates for admission based on pooled student performance data;
- assessing and modifying pre-pharmacy and pharmacy curricula; and
- directing student professionalism development and career advising for pharmacy students.

Recommendation 5: AACP and member colleges and schools should facilitate the use of data analytics to increase the health of individuals and communities through education, research, and service, and such efforts should contribute to the promotion and tenure process for faculty and professional staff.

Recommendation 6: AACP should lead the planning and development of an interprofessional summit of health professions educators to fully explore the impact of big data on education, research, and practice.

Suggestion 1: All colleges and schools should assess their current involvement in the big data sciences based upon review of the Argus report and other relevant information subsequently made available and should identify within their strategic plans where they might further invest to maximize opportunities within their individual programs, across their campus, and within the communities they serve.

Suggestion 2: Colleges and schools of pharmacy should participate fully in their campuses' efforts in big data science and, where institutions have not yet pursued efforts in big data, colleges and schools of pharmacy should provide the leadership to promulgate this opportunity.