

Foreward to special issue on health analytics

Farrokh Alemi

Received: 14 September 2014 / Accepted: 21 September 2014 / Published online: 9 October 2014
© Springer Science+Business Media New York (outside the USA) 2014

Chief of Performance Improvement, District of Columbia Veterans Affairs Medical Center, Affiliate Professor, George Mason University, Department of Health Care Administration and Policy

This special issue on health analytics brings together a number of articles that focus on “Big Data” and its related analytical tools. “Big Data” is loosely defined to include studies working with millions of cases, dealing with hundreds/thousands of variables or using new statistical/analytical methods that improve local (sub-group) validity. These types of studies need to be vigorously reviewed and debated and this issue is our start at doing so.

Health care organizations are collecting and focusing attention on massive amount of data. Here are some examples [1]:

1. The Personalized Medicine Institute at Moffitt Cancer Center tracks more than 90,000 patients at 18 different sites around the country.
2. In any given year, the Veteran Affairs Informatics and Computing Infrastructure collects data on more than 5.5 million patients across 153 medical centers.
3. Kaiser Permanente has a database of 9 million patients.
4. Aurora Health Care system has 1.2 million patients in its data systems.
5. The University of California’s Medical Centers and Hospitals has a database with more than 11 million patients.

6. United States Food and Drug Administration agency has combined medical records of more than 100 million individuals to track effectiveness of medications post release.

In addition, massive data gathers on the web. Patients’ preferences, market share and competitive advantages can be determined from analysis of comments and text left on the web [2, 3].

The wide availability of massive amount of data has made it easier to do analysis. It has also changed the questions scientists ask, answers they provide, and some claim it has led to a radical change in management itself [4].

The use of “Big data” is having an impact. It is changing the long lasting ideas about the value of the managerial experience versus data [5]. “Management by numbers” used to be a pejorative term, just a couple decades ago. Now it is different. Companies that get insights through analysis of big data are doing better and as a consequence attitudes towards data analysis are changing. Data driven companies are 5 % more productive and 6 % more profitable than less data driven companies [6]. There are many examples. At Mercy Hospital in Iowa City, Iowa managers who benchmark their clinicians and pay them for performance report 6.6 % improvements in quality of care [7]. Access issues aside, the VA healthcare system changed from poor to one of the best in the nation through a focus on measurement and data [8]. The use of electronic health records and its associated data have led to reductions in medication errors [9]. Managers have used electronic health records to maximize reimbursement, in ways that have surprised insurers [10]. Other managers report analyzing data within electronic health records to reduce “never events” within their facilities and to measure quality of care [11].

The availability of data has enabled managers to go beyond traditional roles and address clinical questions. For the first time, analysts are reporting on comparative effectiveness of different

Chief of Performance Improvement, District of Columbia Veterans Affairs Medical Center, Affiliate Professor, George Mason University, Department of Health Care Administration and Policy

F. Alemi (✉)
District of Columbia Veteran Affairs Medical Center,
Washington, DC, USA
e-mail: Farrokh.Alemi@va.gov

healthcare interventions. Today, analysts have data on what is best for patients and can work with their clinicians to change practices. For example, analysts have been able to examine pairs of drugs that cause a side effect not associated with either one. They found that Paxil, a widely used antidepressant, and Pravastatin, a cholesterol-lowering drug, raise patient's sugar, a problem not associated with either one [12].

Analysts are finding ways to use data within electronic health records to improve their organizations. These efforts are expected to create an unprecedented shift to more use of data. Analysis of big data faces a number of analytical issues, the simplest of which is that in large databases, every change is statistically significant: two averages that differ by a negligible amount are likely to be statistically significant when the number of observations is in millions. New data requires new methods and scientists working with these new methods need an environment to actively discuss them.

This special issue has started with a small subset of articles. In an informative article by Ajorlou, Shams, and Yang, they analyze Veterans Affairs data warehouse to improve efficiency of operations of patient centered teams [13]. The same group of authors also show how Random Forest models can be used to predict 30-day readmissions in the Veterans Affairs [14]. Random Forest models are also used by Kim, Gupta, Israni, Kasisketo evaluate organ allocation policies [15]. Liu, Traskin, Lorch, Small, and George use Random Forests and Bayesian Additive Regression Trees to examine who will benefit from neonatal intensive care units [16]. Qiu, Chinnam, Murat, Batarse, Neemuchwala, and Jordans show how a data driven reservation system for emergency rooms may reduce prolonged patient waiting times [17]. All of these articles show how machine learning and modern analytical tools can improve our insights.

Two articles in this issue are less focused on application and more on details of analyzing large data warehouses. Wells, Chagin, Li, Hu, Yu, and Kattan work addresses a fundamental problem in conducting comparative effectiveness studies in electronic health records [18]. They show how landmark time, i.e. time from a baseline event, is more helpful than the usual chronological time. Finally, Giang shows how record linkage and cleaning can be done using Bayesian machine learning tools [19].

The use of massive data is in its infancy. These papers do not constitute a complete or a representative review of the field. Readers should consider them as examples of what scientists are working on and the issues they face. As the field grows, Healthcare Management Science will be there to report on it. More novel methods or applications will emerge. New questions will be asked that can only be answered with big data. Old questions will be answered with more precision. The articles in this issue are teasers about what is likely to come in the next few years. More exciting analysis is ahead.

References

- Jaret P (2013) Mining electronic records for revealing health data. *New York Times*, January 14, 2013, <http://www.nytimes.com/2013/01/15/health/mining-electronic-records-for-revealing-health-data.html?pagewanted=all&r=0>
- Greaves F, Pape UJ, King D, Darzi A, Majeed A, Wachter RM, Millett C (2012) Associations between Internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. *BMJ Qual Saf* 21(7):600–5
- Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L (2013) Use of sentiment analysis for capturing patient experience from free-text comments posted online. *J Med Internet Res* 15(11):e239
- McAfee A, Brynjolfsson E (2012) Big data: the management revolution. *Harv Bus Rev* 90(10):60–6, 68, 128
- Eshkenazi A, APICS (2012) The Association for Operations Management. <http://www.apics.org/omnow-landing-page/2012/10/26/joining-the-big-data-revolution> accessed on October 29th 2012
- Brynjolfsson E, Hitt L, Kim H (2011) Strength in Numbers: How Does Data-Driven Decision Making Affect Firm Performance? http://www.a51.nl/storage/pdf/SSRN_id1819486.pdf
- Izakovic M (2007) New trends in the management of inpatients in U.S. hospitals—quality measurements and evidence-based medicine in practice. *Bratisl Lek Listy* 108(3):117–21
- Longman P (2010) *Best Care Anywhere: Why VA Health Care Is Better Than Yours*. Berrett-Koehler Publishers; Second Edition, Revised edition.
- Stürzlinger H, Hiebinger C, Pertl D, Traurig P (2009) Computerized Physician Order Entry - effectiveness and efficiency of electronic medication ordering with decision support systems. *GMS Health Technol Assess*. 19; 5: Doc07.
- Abelson R, Creswell J, Palmer G (2012) Medicare Bills Rise as Records Turn Electronic, *New York Times*. http://www.nytimes.com/2012/09/22/business/medicare-billing-rises-at-hospitals-with-electronic-records.html?_r=1&pagewanted=all&_moc.semityn.www Accessed October 26th 2012.
- Glaser J, Hess R (2011) Leveraging healthcare IT to improve operational performance. *Healthc Financ Manage* 65(2):82–5
- Tatonetti NP, Ye PP, Daneshjou R, Altman RB (2012) Data-driven prediction of drug effects and interactions. *Sci Transl Med* 4(125):125ra31
- Ajorlou S, Shams I, Yang K (2014) An analytics approach to designing patient centered medical homes. *Health Care Manag Sci*. doi:10.1007/s10729-014-9287-x
- Shams I, Ajorlou S, Yang K (2014) A predictive analytics approach to reducing 30-day avoidable readmissions among patients with heart failure, acute myocardial infarction, pneumonia, or COPD. *Health Care Manag Sci*. doi:10.1007/s10729-014-9278-y
- Kim S, Gupta D, Israni AK, Kasiske BL (2014) Accept/decline decision module for the liver simulated allocation model. *Health Care Manag Sci*. doi:10.1007/s10729-014-9295-x
- Liu Y, Traskin M, Lorch S, Small D, George EI (2014) Ensemble of trees approaches to risk adjustment for evaluating a hospital's performance. *Health Care Manag Sci*. doi:10.1007/s10729-014-9272-4
- Qiu S, Chinnam, RB, Murat A, Batarse B, Neemuchwala H, Jordan W (2014) A cost sensitive inpatient bed reservation approach to reduce emergency department boarding times. *Health Care Manag Sci*. doi:10.1007/s10729-014-9283-1
- Wells BJ, Chagin KM, Li L, Hu B, Yu C, Kattan MW (2014) Using the landmark method for creating prediction models in large datasets derived from electronic health records. *Health Care Manag Sci*. doi:10.1007/s10729-014-9281-3
- Giang P (2014) A machine learning approach to create blocking criteria for record linkage. *Health Care Manag Sci*. doi:10.1007/s10729-014-9276-0