



Better Patient Outcomes Through Mining of Biomedical Big Data

Clemens Suter-Crazzolara*

SAP SE, Walldorf, Germany

OPEN ACCESS

Edited by:

Eugeniu Costetchi,
Office des Publications de l'Union
Européenne, Luxembourg

Reviewed by:

Laszlo Balkanyi,
European Centre for Disease
Prevention and Control, Sweden
Caterina Rizzo,
Bambino Gesù Ospedale Pediatrico
(IRCCS), Italy

*Correspondence:

Clemens Suter-Crazzolara
sapclemens@gmail.com

Specialty section:

This article was submitted to
Digital Health,
a section of the journal
Frontiers in ICT

Received: 22 June 2018

Accepted: 13 November 2018

Published: 03 December 2018

Citation:

Suter-Crazzolara C (2018) Better
Patient Outcomes Through Mining of
Biomedical Big Data. *Front. ICT* 5:30.
doi: 10.3389/fict.2018.00030

Digitalization is changing healthcare today. Big data analytics of medical information allows diagnostics, therapy and development of personalized medicines, to provide unprecedented treatment. This leads to better patient outcomes, while containing costs. In this review, opportunities, challenges and solutions for this health-data revolution are discussed. Integration and near-instant-response analytics across large datasets can support care-givers and researchers to test and discard hypotheses more quickly. Physicians want to compare a patient to other similar patients, to learn and communicate about treatment best-practices with peers, across large cohorts and sets of parameters. Real-time interactions between physician and patient are becoming more important, allowing “live” support of patients instead of single interactions once every few weeks. Researchers from many disciplines (biomedical, payers, governments) want to interpret large anonymized datasets, to uncover trends in drug-candidate behavior, treatment regimens, clinical trials or reimbursements, and to act on those insights. These opportunities are however met by daunting challenges. Biomedical information is available in data silos of structured and unstructured formats (doctor letters, patient records, omics data, device data). Efficient usage of biomedical information is also hampered by data privacy concerns. This has led to a highly-regulated industry, as a result of which digitalization in healthcare has progressed slower than in other industries. This review concludes with examples of how integration and interpretation of big data can be used to break down data silos and pave the way to better patient outcomes, value-based care, and the creation of an intelligent enterprise for healthcare.

Keywords: patient outcomes, value-based health care, real-world evidence (RWE), intelligent hospital, data silos, data integration, big data, analytics

INTRODUCTION: THE OPPORTUNITIES AND CHALLENGES

Big data is all around us, and never has data impacted our lives in a comparable manner (Reinsel et al., 2017). According to Eric Topol (Topol, 2016), a “Gutenberg moment” in healthcare is approaching, as technology continues to progress at a rapid pace: healthcare is experiencing a revolution like the one caused by the invention of the printing press. Examples put forward by Topol include the observation that the cost of sequencing a human genome has dropped by a factor of a million in about 15 years; genomic sequencing may soon be part of standard practice. Smartphones can be regarded as mini-medical devices, capable of high speed monitoring and analytics. Such technological advances will generate large volumes of highly valuable information, leading to the democratization of healthcare: they might eventually make doctors superfluous. In this chapter the opportunities and challenges associated with this revolution are described in detail.

There is no doubt that these trends have the potential to advance our lives: actual patient outcomes can be improved. Patient outcomes can be defined as the effectiveness of the treatment of the patient for a disorder, the result of medical care—regarding mortality, morbidity and expenditure (Davies, 1994). This new look at measuring healthcare success fits models across many industries: every human activity can be broken down into quantifiable chunks, from theater visits (Logan, 2014) to customer experience (Murphy, 2014). Determining patient outcomes is not a trivial task, as cures are not black or white effects. A painkiller may completely remove the discomfort of a headache, but it is much more challenging to treat, let alone cure, a complex disease like diabetes. For that reason, healthcare providers have worked to collect so-called patient-reported outcomes measures (PROMs), which are measures of function and health status as reported by patients (Schupbach et al., 2016).

Patient outcomes are tightly linked to costs, and these costs, across the globe, continue to increase at a rapid pace 2016. Thus, there is broad consensus that patient outcomes need to improve dramatically, while containing the costs; a principle initially put forward by Porter and Lee (2013).

Improving Patient outcomes is mostly a big data challenge 2018a. The holy grail is a 360° view of the patient. The complete understanding of all medical, social and environmental information associated with an individual would lead to a perfect “machinery” for treatment and prevention. Technologically this is achievable, although it will meet challenges regarding e.g., data storage, energy cost, analytics and data privacy. However, even a partial implementation of such a system would already help to improve healthcare (Mason, 2018). In addition, much can be learned from studying entire populations. This is the area of population health, which concerns itself with the health outcomes of a group of individuals, including the distribution of such outcomes within the group (Kindig and Stoddart, 2003; Inkelas and McPherson, 2015). This approach aims to improve the health of an entire human population. Population health has specific needs toward health IT, including additional health data sets and the possibility for cross-disciplinary partnerships (Vest et al., 2016).

However, healthcare providers are for many reasons (Bresnick, 2017) vexed to reap these opportunities. First, the healthcare industry lags other industries in digital maturity. Many healthcare organizations still capture patient data in a paper-based fashion, whereas only full digitalization allows data mining. Even electronic medical records (EMR) systems are still largely digital remakes of traditional systems. Also, an emphasized focus on the security of patient data exists, often at the expense of innovation (Landi, 2018). Ironically, although physicians can get streams about stocks, Taylor Swift or Bitcoin, they can't subscribe to a patient (Choi et al., 2018).

Although there is broad consensus that big data can help improve healthcare, many challenges need to be addressed. First, unlike any other Big Data realm (CERN's Large Hadron Collider, or NASA's Hubble telescope), healthcare is the real big data sector. Why is that? It has been estimated that up to 30% of the entire world's stored data is health-related (on

the yottabyte scale) (Faggella, 2018). A single patient typically generates up to 80 megabytes yearly in imaging and EMR data (Huesch and Mosher, 2017). Next to the sheer volume of data to be analyzed, the disparate nature of the data must be addressed, which include patient demographics, laboratory results, medications, radiology, treatments, documents, but also financial and insurance information. The biggest data sources are images (used for diagnosis) and omics data, such as complete genome sequence data (Chen et al., 2018) and proteomics (Kycko and Reichert, 2014). Even data from the microbiome comes into play, as the latter impacts several human disorders (such as cancer Hartmann and Kronenberg, 2018). Per patient, thousands of data fields can be collected. IoT technology, medical devices, laboratory results, smartphones and health trackers can continuously provide real-time data. Many different metrics are needed to describe this information, e.g., age, date of birth, weight or blood concentrations—as integers, but also as kg, g/ml, count/ml, percentage of volume, etc. As the above illustrates, biological systems are vastly more complex than physical systems, the former regarded as maximalist, in comparison to the minimalist nature of the latter (Fox Keller, 2009). This necessitates alignment and cooperation between many different disciplines and dramatically impacts the mining of health data.

The biggest challenge lies in patient information that is only available as free text. Data collected from devices is available as structured information; it can be mined by software in a straightforward manner. Data in EMR systems is at least partly structured or coded. But the information in doctor letters is unstructured. Text mining or natural language processing are needed to turn this unstructured information into semantically standardized, structured data (Kreuzthaler et al., 2017).

The challenges above deal with data volume and formats. In the end the users of the data want to overcome the biggest challenges in care: to gain access to real-world data (RWD); the ability to benchmark the quality of care; unlocking, assembly, and analytics of de-identified patient medical records; to provide guidance by identifying the best, evidence-based course of care, to allow physicians to look for and identify an adverse set of events in patients and uncovering patterns to generate knowledge (Lele, 2017). To tackle this, a Logical Data Warehouse must be put in place, which must address the five Vs of big data analytics: Velocity, Volume, Value, Variety, and Veracity (Cano, 2014). Gartner (Cook, 2018) describes this challenge with the “Jobs to be Done,” with the first job taken by the analysis of terabytes of structured data, Job #2 needs to address the inclusion of little/non-structured data, and the third place is taken by the integration of the new and the old data analysis engines. Next to the big data challenges described above, the healthcare industry is confronted by more specific needs, that are explored below.

As explained, analytical software systems that support the mining of data must be able to ingest or connect many data sources. For this, data adapters must be created. Any knowledge system needs to rely heavily on ontologies (the grouping of diseases according to similarities and differences, Bertaud-Gounot et al., 2012), and coding (e.g., ICD, 2018d) and other standards, such as FHIR (Fast Healthcare Interoperability Resources, 2011), currently still a draft standard defined by

HL7 (2007-2018), which describes data formats and elements as well as application programming interfaces (APIs) for exchanging electronic medical records. In addition, extensibility of integration with other data sources and applications must be enabled. As data sources continue to evolve, more will need to be incorporated into the processes. Thus, data models must be flexible and future-proof. Once data is ingested, the health knowledge systems can provide the access to big data. Real-world evidence (RWE), can then be derived from real-world data (RWD) and provide additional value to traditional data sources (Swift et al., 2018).

Although data privacy, data security, user management and consent management may affect any industry, they are mission critical in healthcare, and on multiple levels. There is value in patient data, and indeed most data breaches happen for illicit economic purposes (2017a; 2018b). The regulation of the handling of patient data is therefore becoming more stringent, as illustrated by HIPAA (2018c) or EU-GDPR (Howell, 2018). Data privacy and informed consent management are crucial and must be complemented by enablement and education of individuals in this area (Porsdam Mann et al., 2016). This is also illustrated by the debate (Rossi, 2015; Menichelli, 2018) whether patient data should be stored behind the firewall of the organization or whether it can be held in (public) cloud- or hybrid-systems. Cloud computing is on the rise across all industries, as it allows faster innovation and reduction of cost, yet on-premise systems are often still perceived as offering better data protection. Intriguingly, data breaches seem to rest more with human- than with technology-based challenges (2017b).

Usability is another prominent challenge in the healthcare industry. Systems are ultimately used by players with different backgrounds, such as researchers from many disciplines, patients and care givers. The complexity of the massive amounts of data must remain “hidden” from the humans that use the system. In addition, users should have the option to easily collaborate on information, also in special interest groups. Thus, a special focus must be on visualization of data, in such a manner that the user can intuitively understand the information (Marcial, 2014; Dias et al., 2017).

Speed and velocity also play a role on multiple levels. Collection of (patient) data in real-time allows the data to be up-to-date at all moments, especially important for situations where quick reaction times are life critical (e.g., early warning systems in emergency rooms or outpatients monitored through mobile devices). On another level, instant responses to highly complex queries must be supported. Retrieving answers to queries across hundreds of data fields per patient lead to extended lag-times, which will negatively impact the user-experience of physicians, researchers or any other user, and will greatly affect acceptance (Raghupathi and Raghupathi, 2014).

The end-user as well as the designer of the system should be able to understand, at least at a high level, where the data comes from, and how it is analyzed. Sufficient transparency on how the data is collected and analyzed must be provided. In addition, data quality is a challenge, especially with very large, heterogenous datasets coming from many data sources. This data will contain errors, especially since a large portion is still collected

by humans. Therefore, systems must allow the detection of data inconsistencies, so that they become correctable at the source (Hasan and Padman, 2006).

STATE OF THE ART AND BEST PRACTICES

This section explores ways how the opportunities and challenges described above can be addressed.

In **Figure 1**, a possible architecture of a knowledge system for healthcare is shown as an illustration, highlighting services needed to get full value out of structured and unstructured information. Basic services (shown at the bottom of the figure) provide standard technologies that are re-usable by all analytical applications, and include e.g., functionality to support real-time, in-memory computing, geospatial functionality (e.g., to determine the location of a patient or a device), or tools for data mining. In the middle, healthcare specific services are shown. Here, three layers can be recognized: (a) a presentation layer, to ensure that the users can view relevant content (tailored to their profile), (b) functions that allow handling and extraction of health-specific information and (c) health content. A plug-in framework allows inclusion of additional data sources. Connections to EMR-systems, IoT and mobile scenarios (depicted on the left) are ensured by APIs. Finally, actual analytics is executed by applications (top), that use subsets of the possible services. As a result, a fully standardized and interoperable framework is created that can support analytics and predictive methodologies. Such a system can ideally be deployed on-premise, in public or private clouds, or combinations thereof. The following case studies use different combinations of these services.

First, the i2b2 tranSMART Foundation develops an open-source and -data community around i2b2 and tranSMART translational research platforms. The Integrating Biology and Bedside (i2b2) project (Murphy et al., 2013) is a platform for extracting, integrating, and analyzing data from electronic health records, registries, insurance claims, and clinical trials. tranSMART (Athey et al., 2013) builds on i2b2 and is a global open source community developing an informatics-based analysis and data-sharing cloud platform, for clinical and translational research. TranSMART can handle structured data from clinical trials and aligned high-content biomarker data. It provides researchers with analysis tools for advanced statistics (Canuel et al., 2015).

Within the Innovative Medicines Initiative (IMI) (Santhosh, 2018), the Harmony project¹ is a healthcare alliance about big data for Better Outcomes for medicines against hematology neoplasms. The project gathers, integrates and analyses anonymous patient data from many high-quality sources. This helps teams to define clinical endpoints and outcomes for these diseases, that are recognized by all key stakeholders.

¹HARMONY. *Healthcare Alliance for Resourceful Medicines Offensive Against Neoplasms in Hematology*. Innovative Medicines Initiative. Available online at: <https://www.imi.europa.eu/projects-results/project-factsheets/harmony> (Accessed Jun 20, 2018).

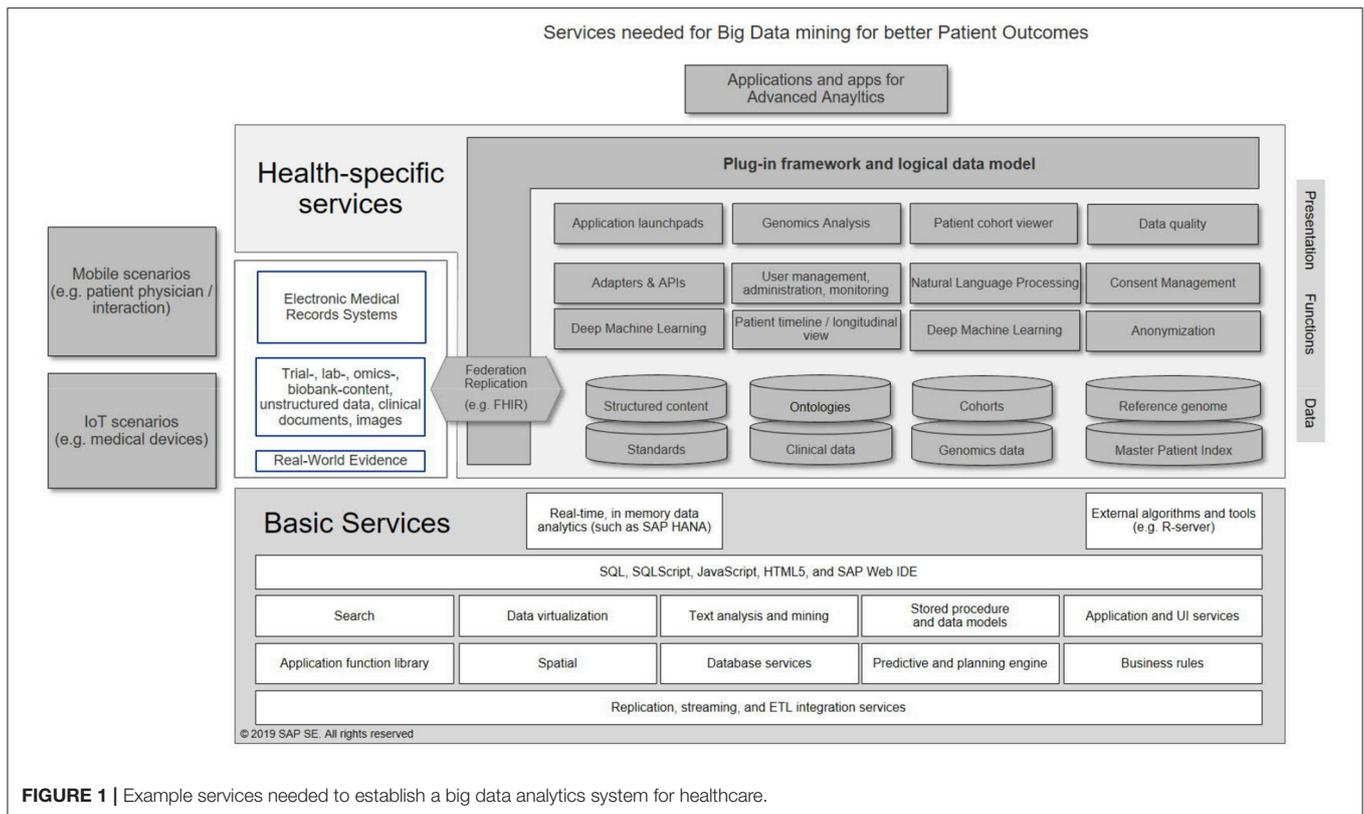


FIGURE 1 | Example services needed to establish a big data analytics system for healthcare.

The American Society of Clinical Oncology's CancerLinQ (Miller and Wong, 2018) has a focus on cancer therapy. The aspiration of CancerLinQ is to build a real world, big data learning system beyond its network of 100+ community oncology practices, and to offer a holistic view of the cancer patient's journey, to support quality improvement and discovery. CancerLinQ taps into information that exists beyond the limited cohort of data within traditional clinical trials and medical oncology. CancerLinQ has engaged the community to incorporate the perspectives of the oncology care team, to create one of the largest sources of real-world evidence in oncology. This also highlights the need for interdisciplinary working groups, consisting of parties with areas of expertise, such as IT professionals that create the knowledge systems, subsequently used by researchers from specific fields for data mining, who in turn support medical professionals.

Some approaches are focused on highly specific domains. ProteomicsDB² (Schmidt et al., 2018) is a protein-centric, in-memory system for the exploration of quantitative mass spectrometry-based proteomics data. ProteomicsDB currently holds 8.8Tb of data and comes with analysis pipelines for exploration of protein expression across hundreds of tissues, body fluids and cell lines. Other quantitative omics data, such as transcriptomics data, protein-protein interaction information, and drug-sensitivity/selectivity data can be included into

analyses. Queries across this data resource are carried out in real-time, allowing more information to be gathered per unit time than with classical databases. In general, in-memory analytic tools can provide dramatic speed increases (Firnkorner et al., 2014).

In healthcare, delayed responses can be lethal. A real-time, direct interaction with patients is crucial, be it through medical devices in intensive care units, or smartphones carried by outpatients. One application is the case of viral outbreaks. As the World Health Organization observed, the critical determinant of epidemic size is the speed of implementation of control measures. Real-time outbreak response systems can help improve timeliness of measures, as shown by the SORMAS³ project in the case of the spread of the Ebola virus. Based on the input from field workers (key actors in viral containment), the combination of cloud-based and in-memory database technology enables interactive data capture and analyses. The front-ends for the data entry are smartphone based, ideal in remote areas. Such systems allow real-time, bidirectional information exchange between field workers and the emergency center, automated status reports and GPS tracking (Fährnich et al., 2015).

Healthcare is, like all other industries, impacted by new big data technologies. Artificial Intelligence (AI) and Machine Learning (ML) provide more profound insights into disease (Gupta and Qasim, 2017; Haegerich, 2018) as illustrated by the following examples. Neurological disorders are a challenging

²<https://www.ProteomicsDB.org>

³https://sormasorg.helmholtz-hzi.de/About_SORMAS.html

group of diseases, as both diagnosis and prognosis pose difficult problems, with many factors influencing the course of the disease (physical, social, hereditary, etc.), further hampered by scarcer longitudinal patient data and the variability in the definition of outcomes (Janssen et al., 2018). For schizophrenia, a very grave disorder, diagnosis surprisingly still relies on interviews of the patient and/or relatives. The usage of neuroimaging data introduces the hurdle of complex dimensionality. Machine-learning techniques are especially suited to tackle this group of highly challenging diseases, and can provide more empirical insights in cause and progression (Dluhoš et al., 2017).

For these reasons, machine learning is beginning to impact the prevention and treatment of cardiovascular disease (Johnson et al., 2018), cancer (Rabbani et al., 2018), or diabetes (Contreras and Vehi, 2018). Image interpretation seems to be a low hanging fruit; however, creating an ML algorithm may be surprisingly easy, but understanding the data structures and statistics is often difficult. In addition, it is still challenging to proof that patient outcomes can be improved and/or costs contained with these methods (Dreyer and Geis, 2017). Still, machine learning may be overhyped - but the technology is ready for prime time, if its limitations are recognized (Hutson, 2018). What is missing, is physician's trust over whether AI is reliable and worthy of adoption (Byers, 2018).

Finally, quantum computing offers the possibility to process extremely large amounts of data in real-time and is predicted to impact areas as diverse as medical imaging, disease screening, drug development and health data protection (Raudaschl, 2017).

CONCLUSIONS

In the sections above, the challenges and opportunities of big data analysis for healthcare were discussed. In addition, some real-life examples of how this can be implemented were put forward. The intelligent enterprise for healthcare can only be created if digitalization is fully embraced, and advanced analytics is applied to the challenge of improving business performance (Quin, 1999; 2018e). This will ultimately enable value-based healthcare (2017c). In the end, it is expected that the analysis of Big Data will continue to drive better patient outcomes (Berg, 2015; Slaweck, 2018) although some caution has to be taken in consideration (Househ et al., 2017). The entire healthcare arena will continue to change, as new disruptive technologies emerge, costs continue to balloon, and patients demand control of their health experience. Data privacy must continue to be guaranteed and improved. Within such a big data and Big Analytics setting, the human aspect must also continue to play a central role. Caregivers need to be enabled to not just use advanced data systems, but also need to consider the patient holistically (age, activity, social setting and emotional station) (Monegain, 2018). After all, an individual's health does not depend only on the data related to that person.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

- (2007-2018). *Health Level Seven International*. Available online at: <http://www.hl7.org/> (Accessed Jun 20, 2018).
- (2011). *Introducing HL7 FHIR*. Available online at: <https://www.hl7.org/fhir/summary.html> (Accessed Jun 20, 2018).
- (2016). *Global Health Expenditure Database*. World Health Organization. Available online at: http://apps.who.int/nha/database/Regional_Averages/Index/en (Accessed Jun 20, 2018).
- (2017a). *Top Three Causes of Data Breach are Expensive*. Calyptix Security. Available online at: <https://www.calyptix.com/top-threats/top-3-causes-data-breach-expensive/> (Accessed Jun 20, 2018).
- (2017b). *Top Three Causes of Data Breach are Expensive*. Calyptix Security. Available online at: <https://www.calyptix.com/top-threats/top-3-causes-data-breach-expensive/> (Accessed Jun 20, 2018).
- (2017c). *What Is Value-Based Healthcare? NEJM Catalyst*. Available online at: <https://catalyst.nejm.org/what-is-value-based-healthcare/> (Accessed Jun 20, 2018).
- (2018a). *Healthcare Big Data and the promise of Value-Based Care*. NEMJ Catalyst. Available online at: <https://catalyst.nejm.org/big-data-healthcare> (Accessed Jun 20, 2018).
- (2018b). *2016 Reported Data Breaches Expose Over 4 Billion Records*. RBS. Available online at: <https://www.riskbasedsecurity.com/2017/01/2016-reported-data-breaches-expose-over-4-billion-records/> (Accessed Jun 20, 2018).
- (2018c). *Health Insurance Portability and Accountability Act*. California Department of Healthcare Services. Available online at: <http://www.dhcs.ca.gov/formsandpubs/laws/hipaa/Pages/1.00WhatIsHIPAA.aspx> (Accessed Jun 20, 2018).
- (2018d). *World Health Organization. ICD Purpose and Uses*. Classification. Available online at: <http://www.who.int/classifications/icd/en/> (Accessed Jun 20, 2018).
- (2018e). *The Intelligent Enterprise*. SAP. Available online at: <https://www.sap.com/products/intelligent-enterprise.html> (Accessed Jun 20, 2018).
- Athey, B. D., Braxenthaler, M., Haas, M., and Guo, Y. (2013). tranSMART: an open source and community-driven informatics and data sharing platform for clinical and translational research. *AMIA Jt. Summits Transl. Sci. Proc.* 2013, 6–8.
- Berg, J. (2015). *3 Ways Big Data is Improving Healthcare Analytics*. HealthcareITNews. Available online at: <https://www.healthcareitnews.com/blog/3-ways-big-data-improving-healthcare-analytics> (Accessed Jun 20, 2018).
- Bertaud-Gounot, V., Duvauferrier, R., and Burgun, A. (2012). Ontology and medical diagnosis. *Inform Health Soc Care* 37, 51–61. doi: 10.3109/17538157.2011.590258
- Bresnick, J. (2017). *Top 10 Challenges of Big Data Analytics in Healthcare*. Health IT Analytics. Available online at: <https://healthitanalytics.com/news/top-10-challenges-of-big-data-analytics-in-healthcare> (Accessed Jun 20, 2018).
- Byers, J. (2018). *HIMSS18: Separating AI Hype From Reality*. Available online at: <https://www.healthcarediver.com/news/himss18-artificial-intelligence-ai-machine-learning/518504/> (Accessed Jun 20, 2018).
- Cano, J. (2014). *The V's of Big Data: Velocity, Volume, Value, Variety, and Veracity*. XSI. Available online at: <https://www.xsnet.com/blog/bid/205405/the-v-s-of-big-data-velocity-volume-value-variety-and-veracity> (Accessed Jun 20, 2018).
- Canuel, V., Rance, B., Avillach, P., Degoulet, P., and Burgun, A. (2015). Translational research platforms integrating clinical and omics data: a review of publicly available solutions. *Brief. Bioinform.* 16, 260–274. doi: 10.1093/bib/bbu006
- Chen, H. Z., Bonneville, R., and Roychowdhury, S. (2018). Implementing precision cancer medicine in the genomic era. *Semin Cancer Biol.* doi: 10.1016/j.semcancer.2018.05.009. [Epub ahead of print].

- Choi, K., Gitelman, Y., and Asch, D. A. (2018). Subscribing to your patients — reimagining the future of electronic health records. *N. Engl. J. Med.* 378, 1960–1962. doi: 10.1056/NEJMp1800874
- Contreras, I., and Vehi, J. (2018). Artificial intelligence for diabetes management and decision support: literature review. *J. Med. Internet Res.* 20:e10775. doi: 10.2196/10775
- Cook, H. (2018). *The Logical Data Warehouse and its Jobs to be Done*. Gartner Blog Network. Available online at: <https://blogs.gartner.com/henry-cook/2018/01/28/the-logical-data-warehouse-and-its-jobs-to-be-done> (Accessed Jun 20, 2018).
- Davies, A. R. (1994). Patient defined outcomes. *Qual. Health Care* 3 (Suppl.) 6–9. doi: 10.1136/qshc.3.Suppl.6
- Dias, C. R., Pereira, M. R., and Freire, A. P. (2017). Qualitative review of usability problems in health information systems for radiology. *J. Biomed. Inform.* 76, 19–33. doi: 10.1016/j.jbi.2017.10.004
- Dluhoš, P., Schwarz, D., Cahn, W., Van Haren, N., Kahn, R., Španiel, F., et al. (2017). Multi-center machine learning in imaging psychiatry: a meta-model approach. *Neuroimage* 155, 10–24. doi: 10.1016/j.neuroimage.2017.03.027
- Dreyer, K. J., and Geis, J. R. (2017). When machines think: radiology's next frontier. *Radiology* 285, 713–718. doi: 10.1148/radiol.2017171183
- Faggella, D. (2018). *Where Healthcare's Big Data Actually Comes From*. Available online at: <https://www.techemergence.com/where-healthcares-big-data-actually-comes-from> (Accessed Jun 20, 2018).
- Fähnrich, C., Denecke, K., Adeoye, O. O., Benzler, J., Claus, H., Kirchner, G., et al. (2015). Surveillance and Outbreak Response Management System (SORMAS) to support the control of the Ebola virus disease outbreak in West Africa. *Euro. Surveill.* 20:21071. doi: 10.2807/1560-7917.ES2015.20.12.21071
- Firnkor, D., Knaup-Gregori, P., Lorenzo Bermejo, J., and Ganzinger, M. (2014). Alignment of high-throughput sequencing data inside in-memory databases. *Stud. Health Technol. Inform.* 205, 476–480.
- Fox Keller, E. (2009). “It is possible to reduce biological explanations to explanations in Chemistry and/or Physics,” in *Contemporary Debates in Philosophy of Biology*, eds F. J. Ayala and R. Arp (Hoboken, NJ: Wiley-Blackwell Publishing Ltd.), 19–31. doi: 10.1002/9781444314922.ch1
- Gupta, M., and Qasim, M. (2017). *Advances in AI and ML are Reshaping Healthcare*. SAP News Center. Available online at: <https://techcrunch.com/2017/03/16/advances-in-ai-and-ml-are-reshaping-healthcare/> (Accessed Jun 20, 2018).
- Haegerich, S. (2018). *How AI Will Push the Frontiers of Modern Medicine*. Available online at: <https://news.sap.com/how-ai-will-push-the-frontiers-of-modern-medicine/> (Accessed Jun 20, 2018).
- Hartmann, N., Kronenberg, M. (2018). Cancer immunity thwarted by the microbiome. *Science* 360, 858–859. doi: 10.1126/science.aat8289
- Hasan, S., and Padman, R. (2006). Analyzing the effect of data quality on the accuracy of clinical decision support systems: a computer simulation approach. *AMIA Annu. Symp. Proc.* 324–328.
- Househ, M. S., Aldosari, B., Alanazi, A., Kushniruk, A. W., and Borycki, E. M. (2017). Big data, big problems: a healthcare perspective. *Stud. Health Technol. Inform.* 238, 36–39.
- Howell, D. (2018). *Five Ways the GDPR Will Change Healthcare*. Future Health Index. Available online at: <https://www.futurehealthindex.com/2018/03/29/gdpr-will-change-healthcare/> (Accessed Jun 20, 2018).
- Huesch, D., and Mosher, T. J. (2017). *Using It or Losing It? The Case for Data Scientists Inside Health Care*. NeJM Catalyst. Available online at: <https://catalyst.nejm.org/case-data-scientists-inside-health-care/> (Accessed Jun 20, 2018).
- Hutson, M. (2018). Has artificial intelligence become alchemy? *Science* 360:478. doi: 10.1126/science.360.6388.478
- Inkelas, M., and McPherson, M. E. (2015). Quality improvement in population health systems. *Healthcare* 3, 231–234. doi: 10.1016/j.hjdsi.2015.06.001
- Janssen, R. J., Mourão-Miranda, J., and Schnack, H. G. (2018). Making individual prognoses in psychiatry using neuroimaging and machine learning. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 3, 798–808. doi: 10.1016/j.bpsc.2018.04.004
- Johnson, K. W., Torres Soto, J., Glicksberg, B. S., Shameer, K., Miotto, R., Ali, M., et al. (2018). Artificial intelligence in cardiology. *J. Am. Coll. Cardiol.* 71, 2668–2679. doi: 10.1016/j.jacc.2018.03.521
- Kindig, D., and Stoddart, G. (2003). What is population health? *Am. J. Public Health.* 93, 380–383. doi: 10.2105/AJPH.93.3.380
- Kreuzthaler, M., Martínez-Costa, C., Kaiser, P., and Schulz, S. (2017). Semantic technologies for re-use of clinical routine data. *Stud. Health Technol. Inform.* 236, 24–31.
- Kycko, A., and Reichert, M. (2014). Proteomics in the search for biomarkers of animal cancer. *Curr. Protein Pept. Sci.* 15, 36–44. doi: 10.2174/1389203715666140221110945
- Landi, H. (2018). *Study: Healthcare Lags Other Industries in Digital Transformation, Customer Engagement Tech*. Healthcare Informatics. Available online at: <https://www.healthcare-informatics.com/news-item/patient-engagement/study-healthcare-lags-other-industries-digital-transformation-customer> (Accessed Jun 20, 2018).
- Lele, C. (2017). *The Growing Importance of Real World Data*. Available online at: <http://www.pharmexec.com/growing-importance-real-world-data> (Accessed Jun 20, 2018).
- Logan, B. (2014). *Brian Logan. Pay-Per-Laugh: The Comedy Club That Charges Punters Having Fun*. The Guardian. Available online at: <https://www.theguardian.com/stage/2014/oct/14/standup-comedy-pay-per-laugh-charge-barcelona> (Accessed Jun 20, 2018).
- Marcial, L. (2014). *Usability in Healthcare: A 'Wicked' Problem*. User Experience Magazine. Available online at: <http://uxpamagazine.org/usability-in-healthcare/> (Accessed Jun 20, 2018).
- Mason, R. (2018). *How IT Can Reshape Patient Care*. CIO. Available online at: <https://www.cio.com/article/3287652/healthcare/how-it-can-reshape-patient-care.html> (Accessed Jun 20, 2018).
- Menichelli, V. (2018). *Six Considerations For Big Data And Analytics*. Digitalist Magazine. Available online at: <http://www.digitalistmag.com/cio-knowledge/2018/06/12/6-considerations-you-need-to-make-for-big-data-analytics-06174507> (Accessed Jun 20, 2018).
- Miller, R. S., and Wong, J. L. (2018). Using oncology real-world evidence for quality improvement and discovery: the case for ASCO's CancerLinQ. *Future Oncol.* 14, 5–8. doi: 10.2217/fo-2017-0521
- Monegain, B. (2018). *Precision Medicine: 'We Want to Make Sure People Feel Respected'*. Clinical Ethicist Says Healthcare IT News. Available online at: <http://www.healthcareitnews.com/news/precision-medicine-we-want-make-sure-people-feel-respected-clinical-ethicist-says> (Accessed Jun 20, 2018).
- Murphy, L. (2014). *Understanding Your Customer's Desired Outcome*. Sixteen Ventures. Available online at: <https://sixteenventures.com/customer-success-desired-outcome> (Accessed Jun 20, 2018).
- Murphy, S. N., Weber, G., Mendis, M., Gainer, V., Chueh, H. C., Churchill, S., et al. (2013). Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J. Am. Med. Assoc.* 307, 124–130. doi: 10.1136/jama.2009.000893
- Porsdam Mann, S., Savulescu, J., and Sahakian, B. J. (2016). Facilitating the ethical use of health data for the benefit of society: electronic health records, consent and the duty of easy rescue. *Philos. Trans. A Math. Phys. Eng. Sci.* 374:20160130. doi: 10.1098/rsta.2016.0130
- Porter, M. E., and Lee, T. H. (2013). *The Strategy That Will Fix Health Care*. Harvard Business Review. Available online at: <https://hbr.org/2013/10/the-strategy-that-will-fix-health-care> (Accessed Jun 20, 2018).
- Quin, J. B. (1999). *Strategic Outsourcing: Leveraging Knowledge Capabilities*. MIT Sloan Management Review. Available online at: <http://sloanreview.mit.edu/article/strategic-outsourcing-leveraging-knowledge-capabilities/> (Accessed Jun 20, 2018).
- Rabbani, M., Kanevsky, J., Kafi, K., Chandelier, F., and Giles, F. J. (2018). Role of artificial intelligence in the care of patients with nonsmall cell lung cancer. *Eur. J. Clin. Invest.* 48:e12901. doi: 10.1111/eci.12901
- Raghupathi, W., and Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Inf. Sci. Syst.* 2:3. doi: 10.1186/2047-2501-2-3
- Raudaschl, A. (2017). *Quantum Computing and Health Care*. BMJ Technology Blog. Available online at: <http://blogs.bmj.com/technology/2017/11/03/quantum-computing-and-health-care/> (Accessed Jun 20, 2018).
- Reinsel, D., Gantz, J., and Rydning, J. (2017). *Data Age 2025: The Evolution of Data to Life-Critical. Don't Focus on Big Data; Focus on the Data That's Big*.

- IDC/Seagate. Available online at: <https://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>
- Rossi, B. (2015). *The great IT myth: is cloud really less secure than on-premise?* Information Age. Available online at: <http://www.information-age.com/great-it-myth-cloud-really-less-secure-premise-123459135/> (Accessed Jun 20, 2018).
- Santhosh, R. (2018). *10 Years of Europe's Partnership for Health*. Innovative Medicines Initiative . Available online at: <https://www.imi.europa.eu/> (Accessed Jun 20, 2018).
- Schmidt, T., Samaras, P., Frejno, M., Gessulat, S., Barnert, M., Kienegger, H., et al. (2018). ProteomicsDB. *Nucleic Acids Res.* 46, D1271–D1281. doi: 10.1093/nar/gkx1029.
- Schupbach, J., Chandra, A., and Huckman, R. S. (2016). *A Simple Way to Measure Health Care Outcomes*. Harvard Business Review. Available online at: <https://hbr.org/2016/12/a-simple-way-to-measure-health-care-outcomes> (Accessed Jun 20, 2018).
- Slawewski, C. M. (2018). *How Can Big Data Lead to Better Outcomes?* Proceedings: DIA Europe. Available online at: <https://globalforum.diaglobal.org/issue/june-2018/how-can-big-data-lead-to-better-outcomes/> (Accessed Jun 20, 2018).
- Swift, B., Jain, L., White, C., Chandrasekaran, V., Bhandari, A., Hughes, D. A., et al. (2018). Innovation at the Intersection of Clinical Trials and Real-World Data Science to Advance Patient Care. *Clin. Transl. Sci.* 11, 450–460. doi: 10.1111/cts.12559
- Topol, E. (2016). *The Patient Will See You Now: The Future of Medicine is in Your Hands*. New York, NY: Basic Books.
- Vest, J. R., Harle, C. A., Schleyer, T., Dixon, B. E., Grannis, S. J., Halverson, P. K., et al. (2016). Getting from here to there: health IT needs for population health. *Am. J. Manag. Care.* 22, 827–829.

Conflict of Interest Statement: CS-C was employed by SAP SE.

Copyright © 2018 Suter-Crazzolara. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.