



How Big Data, Comparative Effectiveness Research, and Rapid-Learning Health-Care Systems Can Transform Patient Care in Radiation Oncology

Jason C. Sanders and Timothy N. Showalter*

Department of Radiation Oncology, University of Virginia School of Medicine, Charlottesville, VA, United States

Big data and comparative effectiveness research methodologies can be applied within the framework of a rapid-learning health-care system (RLHCS) to accelerate discovery and to help turn the dream of fully personalized medicine into a reality. We synthesize recent advances in genomics with trends in big data to provide a forward-looking perspective on the potential of new advances to usher in an era of personalized radiation therapy, with emphases on the power of RLHCS to accelerate discovery and the future of individualized radiation treatment planning.

OPEN ACCESS

Edited by:

Jun Deng,
Yale University, United States

Reviewed by:

James Byunghoon Yu,
Yale University, United States
Andre Konski,
University of Pennsylvania,
United States

*Correspondence:

Timothy N. Showalter
tns3b@virginia.edu

Specialty section:

This article was submitted
to Radiation Oncology,
a section of the journal
Frontiers in Oncology

Received: 13 February 2018

Accepted: 24 April 2018

Published: 09 May 2018

Citation:

Sanders JC and Showalter TN
(2018) How Big Data, Comparative
Effectiveness Research, and
Rapid-Learning Health-Care
Systems Can Transform Patient
Care in Radiation Oncology.
Front. Oncol. 8:155.
doi: 10.3389/fonc.2018.00155

Keywords: big data, radiation oncology, comparative effectiveness research, rapid-learning health care system, personalized radiation therapy

COMPARATIVE EFFECTIVENESS RESEARCH (CER) AND BIG DATA

The Committee on CER Prioritization was created by the Institute of Medicine in 2009. They defined CER as “a strategy that focuses on the practical comparison of two or more health intervention to discern what works best for which patients and populations” (1). In essence, the goal of CER is to help answer the question “which treatment will work best, in which patient, under what circumstances?” (2). Big Data refers to data sets that are so large that they cannot be analyzed directly by individuals or traditional processing software. Big Data Analytics (BDA) is a growing field with a multitude of methods that is being utilized in various sectors from business to medicine (3). The advent of the Electronic Medical Record (EMR) has resulted in the digitalization of massive data sets of medical information including: clinic encounters, laboratory values, imaging data sets and reports, pathology reports, patient outcomes, family history, genomic, and biological data, etc.

To help with the analysis of Big Data, the NIH has created the Big Data to Knowledge (BD2K) program which has invested over \$200 million in grant awards to foster the development of methods and tools to analyze Big Data in biomedical research (4). Additionally, the BD2K program will move to make sure that biomedical Big Data is “Findable, Accessible, Interoperable, and Reusable” (4). Over the past decade, CER methodologies have become increasingly prevalent in radiation oncology research and there is much enthusiasm surrounding BDA.

RAPID-LEARNING HEALTH CARE SYSTEM (RLHCS) AND PERSONALIZED MEDICINE

The number of articles on Big Data in health care has increased exponentially from under 500 articles in 2005 to over 2500 articles in 2015 (5). As the amount of biomedical Big Data and our ability to analyze these data continues to advance, so will the implications and utilizations of the

information we are able to extract. One of the most important steps toward advancing our ability to analyze these Big Data for biomedical discovery is the creation of RLHCS, which will allow for the sharing of patient data between EMRs, ideally in real-time (6). An ideal RLHCS would take patient data that was routinely generated as part of standard patient care and compile that data into a large data system (6–8). This aggregate data would then be available for both BDA to accelerate identification of new hypotheses and CER to rapidly generate evidence through hypothesis-testing studies. Clinical data from patient records can be used readily to identify novel relationships among clinical factors and patient outcomes, or to evaluate treatment effectiveness in specific subgroups, that cannot be studied adequately in randomized, controlled trials. The extreme power of RLHCS, though, is even more exciting when one considers the possibility of adding biospecimens to accelerate discovery in genomics and proteomics. As RLHCS are created and their data sets are expanded, we will continue to identify specific genomic and proteomic data to help define cohorts and stratify patients into risk groups, treatment response groups, and potentially to help design highly tailored therapy regimens (9). In this sense, RLHCS would usher in a more fertile era for improving biomedical research than ever before. BDA and CER provide the research methodologies needed to rapidly generate evidence using RLHCS. It should be noted, however, that there are substantial practical obstacles that must be addressed to achieve the vision of RLHCS. These include patient concerns regarding privacy and security of sensitive information, interconnectivity among different health records, and regulatory barriers to the exchange of health information.

INTEGRATING A RLHCS WITH ONCOLOGY

The integration of CER, Big Data, and BDA is especially important in the field of Oncology where multiple groups are investing significant time and resources in efforts to expand the availability of data and advance the methods used to extract meaningful information from that data (4, 10–14). The American Society of Clinical Oncology started their own RLHCS, CancerLinQ, to overcome the lack of interoperability between EMRs and accomplish their goal of being able to “analyze and share data on every patient with cancer” (15). While the vision of RLHCS has not yet been fully achieved, the potential impact on society has stimulated enthusiasm toward this effort.

IMPLICATIONS FOR RADIATION ONCOLOGY

Patient Reported Outcomes (PROs)

Patient reported outcomes and quality-of-life (QoL) have become a major area of focus in health care overall, particularly in oncology. The availability of PROs within EMRs provides the foundation for a RLHCS that can be leveraged to expand insights into how cancer treatments impact patient QoL. By incorporating the PROs for massive numbers of patients, RLHCS will be

able to identify small variations and subgroups of patients that might be missed in the smaller number of patients included in traditional randomized controlled trials. These PROs and QoL domains can then be incorporated into clinical decision-making to help guide both providers and patients (16). In doing this, PROs can act as a link between the objective clinical data and the subjective patient outcomes and experiences to help improve the overall care of the patient (17). One may also conceive of potential genomics-based determinants of QoL that could be identified using BDA if RLHCS include biospecimens linked to clinical data and PROs. Finally, surveillance of a RLHCS may also be performed to identify temporal trends in PROs to estimate outcomes after implementation of new technologies.

Dose Selection and Radiosensitivity

The use of tumor-specific genes and radiosensitivity to guided treatment decisions has already been established in human papilloma virus-associated squamous-cell carcinoma of the oropharynx (18). Numerous studies have looked at identifying genes that may have implications on tumor radiosensitivity or patient toxicity (19–22). The identification of these genes and their potential implications has led to the creation of the fields of radiogenetics and radiogenomics. Efforts are currently underway to generate meaningful gene assays that will help predict tumor response to radiation. Eschrich et al. created a 10-gene model to calculate a radiosensitivity index and applied this to patients with head-and-neck, rectal, and esophageal cancer to help stratify patients into either responders or non-responders with 80% sensitivity and 82% specificity (22). Similarly, Zhao et al. retrospectively created a 24-gene assay and applied this to risk matched patients who either received postoperative radiation or no radiation following prostatectomy. Patients with a high score on the gene index who received postoperative radiation were less likely to have distant metastasis at 10 years (23). As efforts to identify genes and gene assays that may be predictors of radiosensitivity continue to be validated, we will potentially be able to integrate these findings in dose selection and toxicity prediction for individual patients based on their native and tumor genetics. Scott and colleagues have recently described a genomics-based strategy for personalizing radiation therapy dose, which would support dose de-escalation for radiosensitive tumors (24). While the clinical implication of radiosensitivity assays are still developing, big data will be key to developing future assays rapidly, as well as incorporating the genomics tools into clinical decision-making. Big data provides opportunity to refine molecular signatures based upon real-world data and to merge genomic assay results with other clinical data elements to optimize predictive analytics. A RLHCS would provide the ideal substrate for leveraging big data and CER to accelerate genomics-based discovery to make precision radiation oncology a reality.

Personalized Treatment Recommendations

Radiation oncology is unique in that treatment plans for patients are often already technically and physically personalized due

to patient-specific variations in anatomy, tumor characteristics, and stage. Since a patient's treatment plan is usually based upon a CT scan in treatment position, radiation can be considered an inherently personalized form of medicine. However, treatment planning approaches and radiation doses are generally selected based upon class solution, with technical details such as beam arrangements and dose–volume constraints adherent to generalized rules. Multiple studies have already begun to look at how BDA methods such as machine learning and neural networks can be used to aid in dose optimization and toxicity prediction modeling in radiation oncology (17, 25–27), which could provide more optimal treatment plan alternatives for individual patients. As the data and technology behind RLHCS continues to progress, we will likely be able to utilize a full spectrum of patient-specific clinical factors, PROs, genomics, patient preference, and priorities, and a menu of treatment plan alternatives in order to optimize an individual patient's radiation therapy. In order to deliver high-quality, high impact insights into radiation oncology, it is important that large datasets include detailed technical.

REFERENCES

- Institute of Medicine of the National Academies. *Initial National Priorities for Comparative Effectiveness Research*. Washington, DC: Institute of Medicine of the National Academies (2009).
- Greenfield S, Rich E. Welcome to the Journal of Comparative Effectiveness Research. *J Comp Eff Res* (2012) 1:1–3. doi:10.2217/ce.11.13
- Sivarajah U, Kamal M, Irani Z, Weerakkody V. Critical analysis of big data challenges and analytical methods. *J Bus Res* (2017) 70:263–86. doi:10.1016/j.jbusres.2016.08.001
- Margolis R, Derr L, Dunn M, Huerta M, Larkin J, Sheehan J, et al. The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: capitalizing on biomedical big data. *J Am Med Inform Assoc* (2014) 21:957–8. doi:10.1136/amiajnl-2014-002974
- de la Torre Diez I, Cosgaya HM, Garcia-Zapirain B, Lopez-Coronado M. Big data in health: a literature review from the year 2005. *J Med Syst* (2016) 40:209. doi:10.1007/s10916-016-0565-7
- Ginsburg GS, Kuderer NM. Comparative effectiveness research, genomics-enabled personalized medicine, and rapid learning health care: a common bond. *J Clin Oncol* (2012) 30:4233–42. doi:10.1200/JCO.2012.42.6114
- Ginsburg GS, Staples J, Abernethy AP. Academic medical centers: ripe for rapid-learning personalized health care. *Sci Transl Med* (2011) 3:101cm27. doi:10.1126/scitranslmed.3002386
- Abernethy AP, Etheredge LM, Ganz PA, Wallace P, German RR, Neti C, et al. Rapid-learning system for cancer care. *J Clin Oncol* (2010) 28:4268–74. doi:10.1200/JCO.2010.28.5478
- Ramsey SD, Veenstra D, Tunis SR, Garrison L, Crowley JJ, Baker LH. How comparative effectiveness research can help advance 'personalized medicine' in cancer treatment. *Health Aff (Millwood)* (2011) 30:2259–68. doi:10.1377/hlthaff.2010.0637
- Helft M. Can big data cure cancer? *Fortune* (2014) 170:70–4.
- Williams AM, Liu Y, Regner KR, Jotterand F, Liu P, Liang M. Artificial intelligence, physiological genomics, and precision medicine. *Physiol Genomics* (2018) 50(4):237–43. doi:10.1152/physiolgenomics.00119.2017
- Savage N. Big data versus the big C. *Sci Am* (2014) 311:S20–1. doi:10.1038/scientificamerican0714-S20
- Shah A, Stewart AK, Kolacevski A, Michels D, Miller R. Building a rapid learning health care system for oncology: why CancerLinQ collects identifiable health information to achieve its vision. *J Clin Oncol* (2016) 34:756–63. doi:10.1200/JCO.2015.65.0598

CONCLUSION

Much of the excitement regarding big data has centered on potential for genomic discovery, high-level radiation treatment planning, and leveraging EMRs to identify associations among factors that may provide new insights into potential causal relationships that can be further studied to accelerate progress in cancer care. Although these are certainly promising areas for discovery, we most eagerly anticipate the power of big data to connect a broad range of characteristics to accelerate evidence generation and inform personalized decision-making. We envision the use of big data and CER methods to inform the individual decisions of patients and providers by synthesizing clinical and genomic data and querying a RLHCS for the latest data on effectiveness of treatment options in relevant subgroups of patients.

AUTHOR CONTRIBUTIONS

Both authors contributed to the development and editing of the manuscript and approved the final submitted version.

- Trifiletti DM, Showalter TN. Big data and comparative effectiveness research in radiation oncology: synergy and accelerated discovery. *Front Oncol* (2015) 5:274. doi:10.3389/fonc.2015.00274
- American Society of Clinical Oncology. *Shaping the Future of Oncology: Envisioning Cancer Care in 2030, Outcomes of the ASCO Board of Directors Strategic Planning and Visioning Process, 2011–2012*. Alexandria, VA: American Society of Clinical Oncology (2012).
- Sarin R. Big data V4 for integrating patient reported outcomes and quality-of-life indices in clinical practice. *J Cancer Res Ther* (2014) 10:453–5. doi:10.4103/0973-1482.142741
- Kim KH, Lee S, Shim JB, Chang KH, Cao Y, Choi SW, et al. Predictive modeling analysis for development of a radiotherapy decision support system in prostate cancer: a preliminary study. *J Radiother Pract* (2017) 16:161–70. doi:10.1017/S1460396916000583
- Chen AM, Felix C, Wang PC, Hsu S, Basehart V, Garst J, et al. Reduced-dose radiotherapy for human papillomavirus-associated squamous-cell carcinoma of the oropharynx: a single-arm, phase 2 study. *Lancet Oncol* (2017) 18:803–11. doi:10.1016/S1470-2045(17)30246-2
- West CM, Barnett GC. Genetics and genomics of radiotherapy toxicity: towards prediction. *Genome Med* (2011) 3:52. doi:10.1186/gm268
- Torres-Roca JF, Eschrich S, Zhao H, Bloom G, Sung J, McCarthy S, et al. Prediction of radiation sensitivity using a gene expression classifier. *Cancer Res* (2005) 65:7169–76. doi:10.1158/0008-5472.CAN-05-0656
- Chistiakov DA, Voronova NV, Chistiakov PA. Genetic variations in DNA repair genes, radiosensitivity to cancer and susceptibility to acute tissue reactions in radiotherapy-treated cancer patients. *Acta Oncol* (2008) 47:809–24. doi:10.1080/02841860801885969
- Eschrich SA, Pramana J, Zhang H, Zhao H, Boulware D, Lee JH, et al. A gene expression model of intrinsic tumor radiosensitivity: prediction of response and prognosis after chemoradiation. *Int J Radiat Oncol Biol Phys* (2009) 75:489–96. doi:10.1016/j.ijrobp.2009.06.014
- Zhao SG, Chang SL, Spratt DE, Erho N, Yu M, Ashab HA-D, et al. Development and validation of a 24-gene predictor of response to postoperative radiotherapy in prostate cancer: a matched, retrospective analysis. *Lancet Oncol* (2016) 17:1612–20. doi:10.1016/S1470-2045(16)30491-0
- Scott JG, Berglund A, Schell MJ, Mihaylov I, Fulp WJ, Yue B, et al. A genome-based model for adjusting radiotherapy dose (GARD): a retrospective, cohort-based study. *Lancet Oncol* (2017) 18:202–11. doi:10.1016/S1470-2045(16)30648-9

25. Kim KH, Lee S, Shim JB, Chang KH, Yang DS, Yoon WS, et al. Mining and toxicity prediction modeling system for a clinical decision support in radiation oncology: a preliminary study. *J Korean Phys Soc* (2017) 71:231–7. doi:10.3938/jkps.71.231
26. Arimura H, Nakamoto T. Applications of machine learning for radiation therapy. *Igaku Butsuri* (2016) 36:35–8. doi:10.11323/jjmp.36.1_35
27. Nicolae A, Morton G, Chung H, Loblaw A, Jain S, Mitchell D, et al. Evaluation of a machine-learning algorithm for treatment planning in prostate low-dose-rate brachytherapy. *Int J Radiat Oncol Biol Phys* (2017) 97:822–9. doi:10.1016/j.ijrobp.2016.11.036

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Sanders and Showalter. 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 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.