International Journal of Radiation Oncology biology • physics

www.redjournal.org

The Profession

Needs and Challenges for Radiation Oncology in the Era of Precision Medicine

Harry Quon, MD, MS,* Todd McNutt, PhD,* Junghoon Lee, PhD,* Michael Bowers, BS,* Wei Jiang, MS,† Pranav Lakshminarayanan, PhD,* Zhi Cheng, MD, MPH,* Peijin Han, MD, MHS,* Xuan Hui, MD, ScM,* Veeraj Shah,* Joseph Moore, PhD,* Minoru Nakatsugawa, PhD,* Scott Robertson, PhD,* Emilie Cecil, MS,* Brandi Page, MD,* Ana Kiess, MD, PhD,* John Wong, PhD,* and Theodore DeWeese, MD*

*Departments of *Radiation Oncology and Molecular Radiation Sciences and †Civil Engineering, Johns Hopkins University, Baltimore, Maryland*

Received Sep 17, 2018. Accepted for publication Nov 10, 2018.

Summary

Precision radiation oncology necessitates the systematic curation of clinically relevant heterogeneity at the patient, disease state, and treatment level, in turn, relating these measures to the longitudinal health outcomes of the irradiated cancer patient to gain personalized prediction and therapeutic insights. Leveraging the database and imaging platforms intrinsic to characterizing the heterogeneity of radiation therapy to develop an effective informatics infrastructure

Modern medicine, including the care of the cancer patient, has significantly advanced, with the evidence-based medicine paradigm serving to guide clinical care decisions. Yet we now also recognize the tremendous heterogeneity not only of disease states but of the patient and his or her environment as it influences treatment outcomes and toxicities. These reasons and many others have led to a reevaluation of the generalizability of randomized trials and growing interest in accounting for this heterogeneity under the rubric of precision medicine as it relates to personalizing clinical care predictions, decisions, and therapy for the disease state. For the cancer patient treated with radiation therapy, characterizing the spatial treatment heterogeneity has been a fundamental tenet of routine clinical care facilitated by established database and imaging platforms. Leveraging these platforms to further characterize and collate all clinically relevant sources of heterogeneity that affect the longitudinal health outcomes of the irradiated cancer patient provides an opportunity to generate a critical informatics infrastructure on which precision radiation therapy may be realized. In doing so, data science—driven insight discoveries, personalized clinical decisions, and the potential to accelerate translational efforts may be realized ideally within a network of institutions with locally developed yet coordinated informatics infrastructures. The path toward realizing these goals has many needs and challenges, which we summarize, with many still to be realized and understood. Early efforts by our group have identified the

Reprint requests to: Harry Quon, MD, MS, Johns Hopkins University, Department of Radiation Oncology and Molecular Radiation Sciences, 401 N Broadway, Ste 1452, Baltimore, MD 21231. Tel: 410-502-3877; E-mail: hquon2@jhmi.edu

Proceedings from the ASTRO-NCI-AAPM Precision Medicine Workshop 2016

Conflict of interest: none.

Supplementary material for this article can be found at https://doi.org/10.1016/j.ijrobp.2018.11.017.

0360-3016/$ - see front matter
https://doi.org/10.1016/j.ijrobp.2018.11.017

among a network of institutions is a critical goal to realize precision radiation oncology.

feasibility of this approach using routine clinical data sets and offer promise that this transformation can be successfully realized in radiation oncology. © 2018 Published by Elsevier Inc.

Why Do We Need Precision Medicine in Radiation Oncology?

The practice of modern medicine is founded on the bedrock of evidence-based medicine (EBM) that was systematically advanced by investigators at McMaster University in the 1990s as the "conscientious and judicious use of current best evidence from clinical care research in the management of individual patients."[1] This pedagogy established the foundation on which published research could be systematically evaluated to teach and to guide clinical decisions because it provided confidence, afforded by statistical modeling, that the results identified in a study population were generalizable to the average patient.[2] And yet, even with the promulgation of EBM within the fabric of modern medicine, including its recognized importance in radiation oncology,[3] increasing concerns exist over the effectiveness of randomized controlled trials (RCTs).[4-8] This has led some to even proclaim a crisis is at hand.[7,9,10]

Many reasons and potential solutions have been well articulated, including the concern of inherent biases in RCTs and potential strategies.[7,9] At the core of these concerns are questions about whether the results identified for a given study population, sometimes explicitly defined by strict inclusion criteria to cultivate statistical significance,[7] are not only clinically relevant but applicable for an individual patient.[5,9,11,12] Is the heterogeneity in disease states, how a given treatment is actually executed, and are patient factors sufficiently represented and accounted for by a given positive RCT that a given treatment can be expected to work as recommended for an individual patient? And what of the side effects that are inherent in an individual patient's therapeutic decision? Clinical trials often do not extensively characterize the side effects, especially those manifesting with longer follow-up duration. This is especially important for radiation oncology, where late-manifesting normal tissue injuries require not only maturation but sufficient attention and the right measures to identify them. In fact, even the effectiveness of the cooperative research infrastructure may be insufficient to identify important late toxicities. For example, long-term evaluation of RTOG (Radiation Therapy Oncology Group) trial 9111 that examined the role of concurrent chemoradiation for laryngeal carcinomas identified an increased risk of non−cancer-related deaths in patients receiving concurrent chemotherapy without any apparent increase in late toxicities. This led these investigators to raise the possibility that fatal treatment-related events were not identified in the current system for monitoring and grading late effects.[13]

That only a fraction of patients participate in clinical trials, with 30% of radiation oncology RCTs never being completed,[14] leaves more than 90% of patients underrepresented[15] and is emblematic of concerns over the eternal validity of clinical trial results[16] and an opportunity lost to accelerate clinical care. Optimizing the clinical research paradigm argues, at least in part, for more generalizable data, especially the study of patients not otherwise eligible for clinical trials.[11] For radiation oncology, longitudinal mature data sets are especially important. Hence, although observational trials of routinely collected data (RCDs) (ie, without an a priori hypothesis) are attractive at least as an adjunct to RCTs,[8] concerns of inherent biases especially associated with treatment selection temper the enthusiasm for RCDs alone.[17]

However, proponents of this approach have hypothesized that when the data dimension is grown (spurred by rapid technological growth of wearables and the richness of omic biospecimens) to what many refer to as "clinical big data," characterizing the heterogeneity in disease states, treatment, and a patient's state will lead to transformative insights with less risk of biases.[18-20] The nascent belief then is that if this voluminous and complex longitudinal data set could be systematically curated and analyzed by advanced data mining applications,[21,22] a broad range of health care benefits may be realized,[23] including more cost-effective and personalized predictions.[24] It is the potential for novel insight discoveries that has led big data to become a disruptive force commanding, if not mandating, increasing attention from the health care domain. Our group believes that radiation oncology is particularly well suited for the clinical big data revolution in light of the quantitative nature of radiation therapy (RT) and the existing imaging and database backbone intrinsic to modern RT, which can be leveraged to establish the needed architectural framework.[23]

Although many challenges exist on the road to effectively translate data science for health care needs, arguably one of the greatest challenge health care faces is our current ability/inability to systematically curate a high-quality longitudinal data set within an information/informatics infrastructure (and its architectural framework) that places an emphasis on veracity and completeness. This is critical given that health care decisions will be a key application. Given the variability of patients, their disease states, and how health care providers provide treatment, this curation also needs to extend beyond any one institution to determine whether relationships and predictive algorithms generated are unique to a particular institution's practice or are generalizable. Thus a networked informatics infrastructure will need to be a fundamental

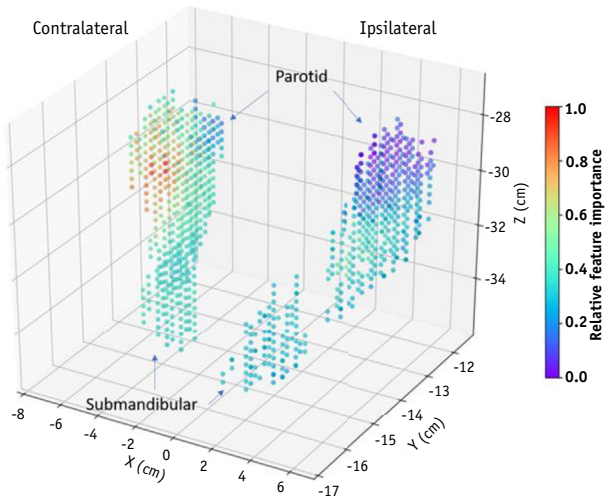

Fig. 1. Normalized dose voxel analysis of parotid and submandibular glands and their risk of grade ≥2 xerostomia based on ridge logistic regression modeling. Analysis demonstrated subvolumes of dose voxels in both the ipsilateral and contralateral parotid glands at a greater risk of grade ≥2 xerostomia.

hallmark of any transformative initiative to fully realize precision medicine in radiation oncology.

In 2007, the Oncospace project was initiated at the Johns Hopkins University. We describe how early efforts by our group suggest that this paradigm shift is promising and potentially feasible. We have demonstrated it to be effective in its local development, generating several examples of insight discovery while identifying needs and challenges to be met for this to be generalized.

What Does Precision Medicine Mean to Radiation Oncology?

In 2011 the National Academy of Science completed a committee report titled "Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease."[25] President Obama accelerated attention to this initiative several years later through announcing the precision medicine initiative. A critical consideration in the precision medicine initiative is the recognition that a multitude of measurements across multiple domains are needed. Both in-clinic measures (ie, clinical and genomic measurements) and out-of-clinic measures such as daily activities, environmental exposures, and many others collectively offer potential insights into a patient's health state and its future trajectory.

For the cancer patient, this trajectory embodies his or her cancer history. For patients receiving RT, precision therapy necessitates more than the identification of which individual or signature of molecular targets is driving radiation resistance in a given patient. Rather, identifying the optimal personalized therapeutic ratio is an important goal because this reflects the individualized risk of normal tissue injury to personalizing RT. It is this side of the equation where progress has been made. Though nascent and primarily anatomically based, they indicate the promise to realize precision RT. For example, by incorporating knowledge of the distance between planning target volumes (PTVs) and normal organs at risk (OARs), optimized and personalized RT planning based on the use of dose volume histogram (DVH) constraints is possible.[26-28]

Yet the use of planning DVH constraints itself needs to be reconsidered in our pursuit of precision RT because it is population defined and may not be representative of an individual's risk of OAR injury. OAR constraints will need to evolve to reflect measures of baseline personalized risk. The DVH is also flat in that it limits our ability to understand the risk of normal tissue injury spatially throughout an OAR. Within the Oncospace infrastructure, our group has identified the ability to collate OAR dose data, preserving its spatial information even at the dose voxel level (Fig. 1).[29] In doing so, spatial relationships within OARs and PTVs can be explored and compared with each other, to clinical outcomes and the imaging features contained in each image voxel. As a result, our group has found that the risk of recovery from RT-induced xerostomia appears to be spatially dependent and particularly affected by the dosimetry to the superior parotid subvolumes, potentially identifying more effective dosimetric constraints. These findings are consistent with preclinical studies.[30] These analyses also highlight the importance of a continuously growing inventory of National Cancer Institute Common Terminology Criteria for Adverse Events xerostomia grades, providing insights into the natural history (at least locally) of RT-induced xerostomia and indicating the value of its longitudinal curation (Fig. 2). In doing so, clear signs of improvements beginning at about 6 months after RT and a slower phase of recovery at and beyond 18 months can be identified, suggesting that normal tissue repair also occurs in human irradiated parotids.

What are the Current Challenges to Realizing Precision Radiation Oncology?

The need for an informatics infrastructure

In 2006, Pironti[31] defined an informatics infrastructure as the constellation of all the "people, processes, procedures, tools, facilities, and technology which supports the creation, use, transport, storage and destruction of information." It is thus far more than database technology. An effective informatics infrastructure is the sum of all parts and requires foremost a culture that views and values information as a commodity with its currency defined by what we can do or wish to do with the data. Our group has cultivated this culture within the fabric of routine clinical workflows, facilitated by access to web-browser customizable data forms laying the foundation for successful prospective

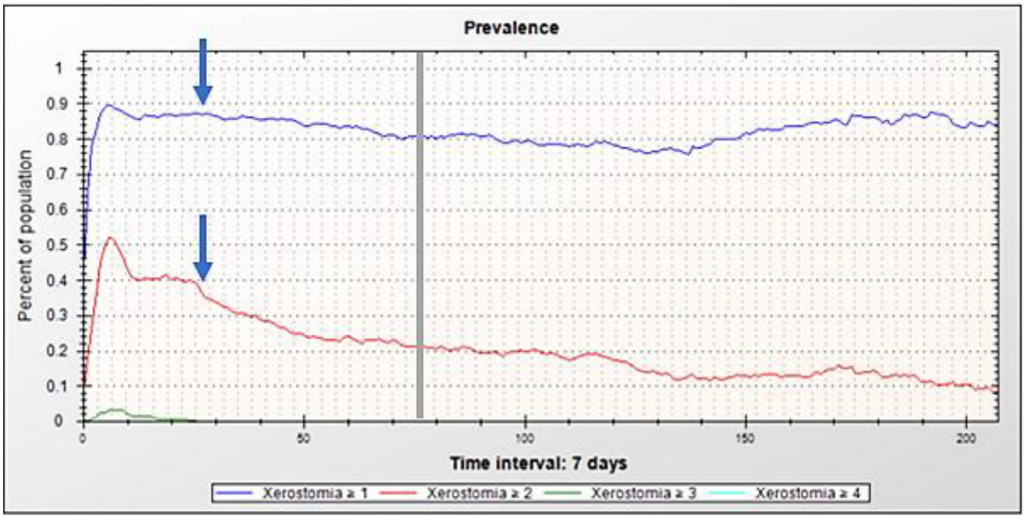

Fig. 2. Natural history of National Cancer Institute Common Terminology Criteria for Adverse Events (version 3) graded xerostomia (n = 703). Xerostomia begins to improve at 6 months after completion of radiation therapy, with a second slower phase at 18 months.

capture of structured data at the point of care. At present the data inventory houses dosimetric, laboratory, clinical toxicity, oncologic, and patient-reported outcomes (PROs) with approximately 2.6 million assessments for 4705 patients and continues to grow with each patient treated (Table 1).

Our early efforts began with database structure development populated with extraction, transformation, and loading of dosimetric and clinical data from treatment planning and record and verify applications, along with the development of web-browser functionality.[32] This has been followed by ongoing efforts addressing an important limitation of many cancer data sets: the creation of a data set with complete and accurate toxicity and oncologic outcomes that could be relied on for both clinical and translational research efforts. Health care provider data sources present unique data quality challenges given that they are not automated unlike other data sources in other sectors of society. Although the level of data generated in electronic health records is viewed as significant and exciting, much of this is unstructured, with the human determination of its value prone to inconsistencies (including semantic variations), inaccuracies, and missing information (ie, completeness).[20] Abstraction of unstructured data sets with the use of advanced data mining techniques such as natural language processing remain promising[33,34] but have the potential of creating inaccurate metadata, especially with poor data quality.

In contrast, although the use of structured data forms can aid in reducing inconsistencies and to some degree

Table 1 Clinical data inventory within Oncospace: Types of clinical data sets across 4 major tumor sites for a total of 4705 patients

Tumor site	Head and neck	Pancreas	Prostate	Thorax	Total
Vital signs	69,793	8067	3572	55,792	137,224
Performance status	12,596	835	4795	6285	24,511
Social assessment	14,616	360	849	322	16,147
Medical assessment	17,489	62	7233	1152	25,936
Examination	9032	3688	36	97	12,853
Medication	42,681	1398	3059	6459	53,597
Tumor marker	77,706	553	1811	792	80,862
RT sessions	22,287	1263	138	5863	29,551
Toxicity grading	619,157	21,511	97,206	103,144	841,018
Laboratory assessment	146,617	28,487	38,574	123,632	337,310
Quality of life patient-reported outcome assessments	297,763	15,238	122,019	13,113	448,133
Other assessments	244,785	1017	212,374	190,981	649,157
Total assessments	1,574,522	82,479	491,666	507,632	2,656,299
Total patients	1279	429	1973	1024	4705

Abbreviation: RT = radiation therapy.

Table 2 Data quality assessment of normal regions of interest: Contour integrity was assessed using either data-driven model or geometric contiguousness model

Region of interest	Integrity model or metric	No. of contours tested	No. of contours detected	No. of contours verified as errant/abnormal	Accuracy of metric or model
Bladder	Verification of bladder emptiness during CT	594	36	25	70%
Bladder	Contiguousness: extent	594	8	8	100%
Bladder	Contiguousness: region growing over volume	594	10	10	100%
Left femoral head	Femoral head distinction model	561	274 detected with ball/shaft; 76 detected with ball only	274 ball/shaft verified; 65 ball only verified	100%; 86%
Left femoral head	Contiguousness: extent	561	13	13	100%
Left femoral head	Contiguousness: region growing over volume	561	15	15	100%
Right femoral head	Femoral head distinction model	559	235 detected with ball/shaft; 115 detected with ball only	235 ball/shaft verified; 94 ball only verified	100%; 82%
Right femoral head	Contiguousness: extent	559	7	7	100%
Right femoral head	Contiguousness: region growing over volume	559	9	9	100%
Spinal cord	Slice-based extent model	1148	22	30	73%
Spinal cord	Contiguousness: extent	1148	80	80	100%
Spinal cord	Contiguousness: region growing over volume	1148	82	82	100%
Rectum	Slice-based extent model	769	12	10	83%
Rectum	Contiguousness: extent	769	14	14	100%
Rectum	Contiguousness: region growing over volume	769	19	19	100%
Prostate	Contiguousness: extent	322	1	1	100%
Prostate	Contiguousness: region growing over volume	322	1	1	100%
Brainstem	Contiguousness: extent	1140	14	14	100%
Brainstem	Contiguousness: region growing over volume	1140	14	14	100%

Abbreviation: CT = computed tomography.

incompleteness, they are not without risks of inaccuracies and are potentially more time intensive. However, given that the assessments in radiation oncology are typically repetitive in content assessment and that during a course of radiation therapy the time intervals are short, leading to rapid recall assessments, minor modifications of copied recent assessments offered an attractive efficient solution. Web-browser functionality to past assessments and other iterative tool developments in our clinics directed at facilitating efficient data entry within routine workflows have allowed us to construct an infrastructure that generates a data stream of structured outcomes including those that are patient reported. This approach is also not without concerns about propagating inaccuracies and incomplete assessments and necessitates a plan of continuous review and supervision in its development. How this approach can be integrated with electronic health records is also an unmet need. The hope then is for this large data set possesses sufficient characteristics of volume, variety, velocity, variability, and veracity (the 5Vs of big data) that will allow us to discover and characterize true relationships.

The need for data quality

This data production process raises the issue of data quality.[35] To the extent that data quality begins and ends with its production, assessing for data quality at key points along the health care production line is an important strategy. For the radiation oncology patient, points of interest may include the cancer diagnosis, staging, evaluating for inconsistencies in PTV and OAR nomenclature, contour

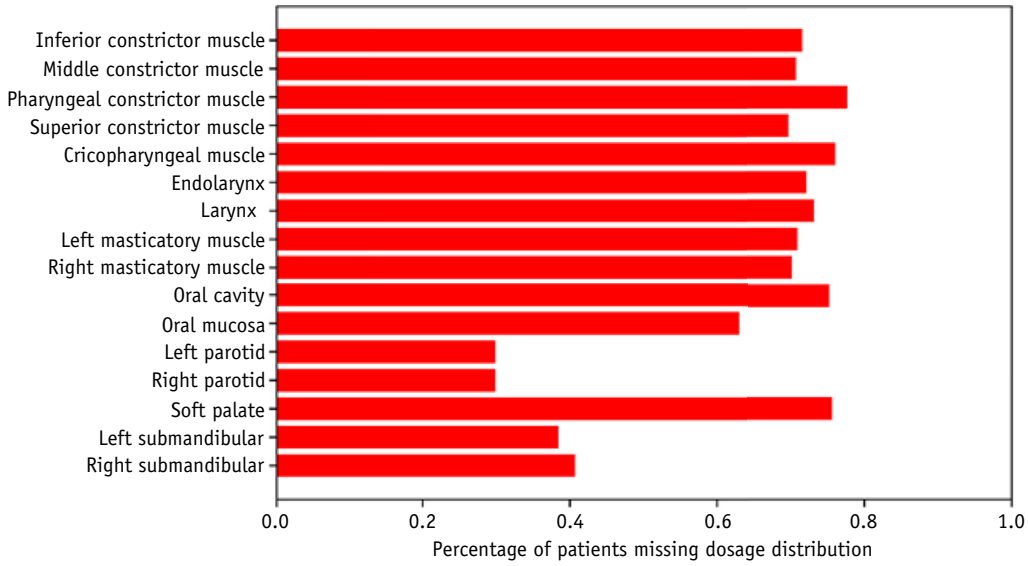

Fig. 3. Impact of missing dosimetric data on weight loss prediction model. Of an initial cohort of 1044 patients with head-and-neck cancer, a random forest algorithm could only be applied to a total of 403 patients because of the absence of normal organs.

inaccuracies, missing contour structures, toxicities assessment, tumor response rates, and patterns of relapse. Because these points of interest are health care provider dependent, data quality is a paramount consideration, as noted earlier. Integrating such a data quality management plan involves a plan to systematically data profile and manage various dimensions of poor data quality. Table 2 summarizes how effective data profiling can identify OARs that may not be accurately contoured but can affect the dosimetric interpretation or analysis.

Even more detrimental to data analysis is missing data. An incomplete set of OAR contours can significantly erode the number of patients available for analysis. Figure 3 shows this impact in the development of a weight loss prediction model in irradiated head and neck cancer patients using a random forest algorithm where 403 patients were analyzed out of an initial cohort of 1044 patients because of missing clinical and dosimetric data (Jiang W. et al, personal communication). Strategies may include the use of atlas-based segmentation to reduce the missing data. Figure E1 (available online at https://doi.org/10.1016/j.ijrobp.2018.11.017) shows how a multiatlas-based segmentation approach using the strengths of magnetic resonance imaging may provide a solution to this challenge (Lee J. et al, personal communication).

Inaccurate data are also a challenge, especially with OARs. The impact of inaccurate OARs derived from routine patient care in big data prediction models is largely unexplored. Our group has found that this inaccuracy may be quantifiable with the creation of rinds around OARs creating additional volumes that can then be analyzed and compared with the OAR alone. Results in a xerostomia prediction model suggest that these inaccuracies may be minimal. Figure E2 (available online at https://doi.org/10.1016/j.ijrobp.2018.11.017) shows how data profiling of patient weights throughout a treatment course using advanced computer science algorithmic approaches[36] can also identify potential inaccuracies and outliers and automate cleansing for data analysis.

Patient-reported outcomes have become an important clinical outcome. The electronic construction of PROs (ePROs) within an informatics infrastructure that has web browser access greatly facilitates the capture of this important data source, including its use as a clinical outcome measure. In general, when the digital construction of the PROs closely mirrors the original validated paper versions, validity is preserved, although investigators are cautioned to always consider if an ePRO correctly reflects the behavior of the paper version.[37] Challenges in the incorporation of ePROs as a component of an informatics infrastructure include the following: (1) ensuring that robust and secure Wi-Fi access in the clinics exists such that high rates of complete PROs are achieved; (2) ensuring data security, especially if the web browser access directly writes to the database containing other patient records; and (3) the abilities of the patients that also considers the time demands required to complete the ePROs, their motivation and healthstate (further elaborated below). ePRO assessments out of the clinic are attractive because they may offer an ecologic and potentially momentary assessment reflecting the potential influence of a patient's natural environment, yet they raise additional data security concerns in their integration with any informatics infrastructure.[38]

Within Oncospace, we have found that the use of electronic tablets can be successfully integrated within the routine clinical workflow, generating a voluminous inventory for analysis (Table 1). This has facilitated several hypotheses-generating discoveries, including a promising approach to classify head and neck cancer dysphagia in our

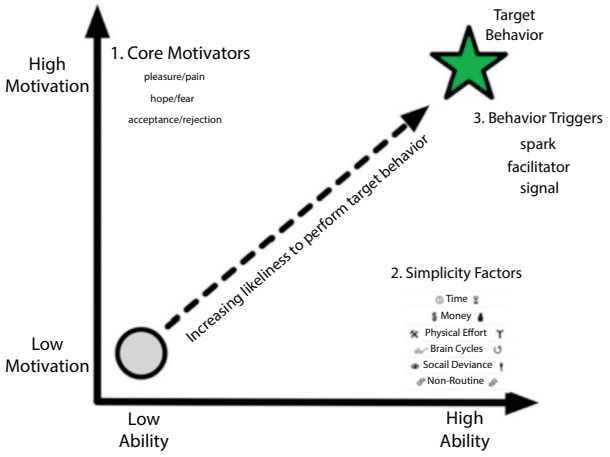

Fig. 4. Fogg behavior change model. In the Fogg model, behavior change is a product of 3 dynamic factors: motivation to change, possessing the ability to perform the behavior, or how difficult the behavior is in response to a trigger for the desired behavior.

pursuit to personalize head and neck radiation therapy (Fig. E3; available online at https://doi.org/10.1016/j.ijrobp.2018.11.017)[39] and the identification of the importance of patient-reported taste changes dictating significant weight loss after a course of head and neck radiation therapy.[40]

The rapid technological development of automated sensors and biosensors in daily wearables has been at the core of significant enthusiasm for big data analytics in the health care domain. Despite this enthusiasm, data quality considerations remain essential, as is the additional challenge of ensuring secure data transfer. In a pilot study using a smartphone app linked to Oncospace, Bluetooth-enabled biosensors linked to the app successfully allowed for out-of-clinic patient assessments to be streamed into Oncospace. However, Figure E4 (available online at https://doi.org/10.1016/j.ijrobp.2018.11.017) shows clear inaccuracies in the weight as a reminder that the enthusiasm for this data source should not negate the need for a data quality management plan.

The need to characterize tumor and normal tissue heterogeneity

Characterizing the heterogeneity of cancer and of normal OARs at risk of radiation-induced injury within the infrastructure described represents yet another major challenge to realizing precision RT. When characterizing the biological heterogeneity with traditional omic approaches, it is not whether relevant biological measures can be identified to correlate or predict for oncologic outcomes that is often in question, but rather, it is whether the results identified typically in modest cohorts of patients are reproducible and generalizable.[41] It is for these reasons that strategies to systematically curate biological measures of tumor and normal tissue heterogeneity within an informatics infrastructure, especially when networked, offer a potential solution to advance these translational efforts for precision RT. In doing so, RCDs including the spatially oriented dosimetric data within tumor volumes are available for insight discovery.

Radiation oncology is unique in that its treatment is imaging based and the routine workflow defines gross tumor volumes (GTVs) and normal OARs as regions of interest. Consequently, potentially significant opportunities exist in exploiting advanced functional and quantitative analyses, especially of routinely acquired computed tomography (CT) images that provide opportunities to understand not only the temporal but also the spatial changes in a tumor.[42] In particular, the incorporation of advanced imaging features derived from texture analysis of these regions of interest is an attractive strategy. Early work in this emerging field of study suggests that the advanced imaging features derived from CT images have promise in their prognostic significance, including for head and neck cancers, where reproducible validation has been reported across several institutions.[43-45] These observations are particularly impressive despite challenges well summarized regarding radiomics,[46] including concerns that imaging features may be affected by the variability associated with how a gross tumor volume is contoured by different health care providers, how the images are obtained and their variability across institutions, and the imaging effects of dental amalgams. These findings suggest that robust features may have been identified.

Recently our group has piloted the texture analyses of parotid and submandibular glands using planning CT images, indicating that the imaging features of these salivary glands are not only amenable to classification but significantly contribute to our modeling for grade $\geq$2 xerostomia 3 to 6 months after irradiation.[47] These findings suggest that these imaging features may be quantifying baseline salivary gland function. The effects of contour differences and dental artifacts were also analyzed. Excluding the image features that were influenced by the artifacts identified only modest degradation in toxicity prediction was identified, indicating that radiomic analysis may be a novel approach in defining a personalized therapeutic ratio.

Radiomics, as with other omic measures of tumor heterogeneity, present challenges because sample sizes are substantially smaller than the number of available covariates, especially when many of these covariates are related to each other (ie, the large *P*, small n problem).[48] Similarly, the ability to use spatially oriented dose voxels (ie, dosomics) presents similar statistical challenges in modeling high dimensionality. These include determining the optimal approaches to dimension reduction and variable/feature selection, which is very much dependent on the type of data being modeled.[49]

The need to prioritize, collaborate, and data share

It is valuable to digress from this clinically oriented discussion to understand the elements that effect changes in

human behavior, especially as one further reconsiders the challenges associated with systematically capturing health provider–based data sources. Alhough many theoretical models have been articulated, B. J. Fogg offers a design behavior change model that conceptualizes behavior change as a product of 3 dynamic factors: *motivation* to change and possessing the *ability* to perform the behavior, or how difficult the behavior is in response to a *trigger* for the desired behavior (Fig. 4).[50] For successful behavior change, high motivation can bring about a triggered behavior (change) even in the setting of a difficult task. With decreasing motivation, behavior change is likely to only occur with an easy task or one requiring limited effort or ability. Thus, it is critical to balance the goals of each local team invested in the development of an informatics infrastructure, their motivation and the data set that they seek to collect as part of the routine workflow, and what tools can be developed to efficiently facilitate this.

Because health care practices differ across the globe, our group envisions a critical need to establish a network of institutions undertaking similar data science efforts. Creative funding solutions will clearly be needed. In this vein, a federated/virtual database model mitigates the risks associated with the transfer of patient identifiers while facilitating data sharing and analyses. Nevertheless, these efforts will also raise ethical issues of consent and whether an opt-in versus an opt-out model should be selected. Opt-in models respect the ethical principles of consent yet have been found to introduce bias through patient selection[51] and paradoxically may be less ethical because consented patients incur the risk of participation in compromised quality research. Harmonizing these individual institutional efforts will also require additional needs to establish a collaborative uniform data ontogeny,[52] which will not only aid structured data capture but facilitate data integration across institutions. Data governance models with a leadership structure to include key stake holders, including academic and community centers, will also be needed. As with cooperative oncology groups, this network of institutions can then form a complementary synergistic research platform.

Conclusions

Medicine has historically been a clinical science supported by data. Despite the advances seen with EBM, the need to account for patient and disease state heterogeneity and thus precision medicine is completing the cycle toward comprehensively characterizing the experience of the individual patient. Thus, medicine and radiation oncology is at the precipice of its transformation to become a data science supported by clinicians.

References

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: What it is and what it isn't. *BMJ* 1996 Jan 13;312:71-72.
2. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA* 1992;268:2420-2425.
3. Bentzen SM. Towards evidence based radiation oncology: Improving the design, analysis, and reporting of clinical outcome studies in radiotherapy. *Radiother Oncol* 1998;46:5-18.
4. Luce BR, Kramer JM, Goodman SN, et al. Rethinking randomized clinical trials for comparative effectiveness research: The need for transformational change. *Ann Intern Med* 2009;151:206-209.
5. Ioannidis JP. Why most clinical research is not useful. *PLoS Med* 2016;13:e1002049.
6. Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005;294:218-228.
7. Every-Palmer S, Howick J. How evidence-based medicine is failing due to biased trials and selective publication. *J Eval Clin Pract* 2014;20:908-914.
8. Rothwell PM. External validity of randomised controlled trials: "To whom do the results of this trial apply?" *Lancet* 2005;365:82-93.
9. Greenhalgh T, Howick J, Maskrey N. Evidence Based Medicine Renaissance Group. Evidence based medicine: A movement in crisis? *BMJ* 2014;348:g3725.
10. Miller CG. Evidence-based medicine or ignorance-based evidence? *Int J Person Centered Med* 2011;1:633-634.
11. Ioannidis JP. How to make more published research true. *PLoS Med* 2014;11:e1001747.
12. Penston J. The irrelevance of statistics-based research to individual patients. *Int J Person Centered Med* 2011;1:240-249.
13. Forastiere AA, Zhang Q, Weber RS, et al. Long-term results of RTOG 91-11: A comparison of three nonsurgical treatment strategies to preserve the larynx in patients with locally advanced larynx cancer. *J Clin Oncol* 2013;31:845-852.
14. Nguyen TK, Nguyen EK, Warner A, Louie AV, Palma DA. Failed randomized clinical trials in radiation oncology: What can we learn? *Int J Radiat Oncol Biol Phys* 2018;101:1018-1024.
15. Unger JM, Cook E, Tai E, Bleyer A. The role of clinical trial participation in cancer research: Barriers, evidence, and strategies. *Am Soc Clin Oncol Educ Book* 2016;35:185-198.
16. Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005;2:e124.
17. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JP. Routinely collected data and comparative effectiveness evidence: Promises and limitations. *CMAJ* 2016;188:E158-E164.
18. Carlson B. P4 medicine could transform healthcare, but payers and physicians are not yet convinced. *Biotechnol Healthc* 2010;7:7-8.
19. Glance D. Evidence-based medicine is broken: Why we need data and technology to fix it. *BMJ* 2014;348:g3725.
20. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA* 2013;309:1351-1352.
21. Koh HC, Tan G. Data mining applications in healthcare. *J Healthc Inf Manag* 2011;19:65.
22. McNutt TR, Benedict SH, Low DA, et al. Using big data analytics to advance precision radiation oncology. *Int J Radiat Oncol Biol Phys* 2018;101:285-291.
23. Raghupathi W, Raghupathi V. Big data analytics in healthcare: Promise and potential. *Health Inf Sci Syst* 2014;2:3.
24. Chawla NV, Davis DA. Bringing big data to personalized healthcare: A patient-centered framework. *J Gen Intern Med* 2013;28:660-665.
25. National Research Council. Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. Washington, DC: National Academies Press; 2011.
26. Wu B, Ricchetti F, Sanguineti G, et al. Patient geometry-driven information retrieval for IMRT treatment plan quality control. *Med Phys* 2009;36:5497-5505.
27. Petit SF, van Elmpt W. Accurate prediction of target dose-escalation and organ-at-risk dose levels for non-small cell lung cancer patients. *Radiother Oncol* 2015;117:453-458.

28. Petit SF, Wu B, Kazhdan M, et al. Increased organ sparing using shape-based treatment plan optimization for intensity modulated radiation therapy of pancreatic adenocarcinoma. *Radiother Oncol* 2012;102:38-44.
29. Lakshminarayanan P. Johns Hopkins University; 2017.
30. Konings AW, Cotteleer F, Faber H, van Luijk P, Meertens H, Coppes RP. Volume effects and region-dependent radiosensitivity of the parotid gland. *Int J Radiat Oncol Biol Phys* 2005;62:1090-1095.
31. Pironti J. Key elements of an information security program. *Inf Syst Control J* 2005;1.
32. Yang W, Moore J, Quon H, et al. Browser based platform in maintaining clinical activities—use of the iPads in head and neck clinics. Journal of physics: Conference series. IOP Publishing; 2014; 012095.
33. Murff HJ, FitzHenry F, Matheny ME, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA* 2011;306:848-855.
34. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med* 2018;1: 18.
35. Loshin D. Understanding Big Data Quality for Maximum Information Usability. Cary, NC: SAS Institute Inc; 2014.
36. Dasu T, Johnson T. Exploratory Data Mining and Data Cleaning. Hoboken, NJ: John Wiley & Sons; 2003.
37. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: A meta-analytic review. *Value Health* 2008;11:322-333.
38. Stone AA, Shiffman S. Capturing momentary, self-report data: A proposal for reporting guidelines. *Ann Behav Med* 2002;24:236-243.
39. Quon H, Hui X, Cheng Z, et al. Quantitative evaluation of head and neck cancer treatment—related dysphagia in the development of a personalized treatment deintensification paradigm. *Int J Radiat Oncol Biol Phys* 2017;99:1271-1278.
40. Cheng Z, Nakatsugawa M, Hu C, et al. Evaluation of classification and regression tree (CART) model in weight loss prediction following head and neck cancer radiation therapy. *Adv Radiat Oncol* 2018;3:346-355.
41. Ioannidis JP, Khoury MJ. Improving validation practices in "omics" research. *Science* 2011;334:1230-1232.
42. Quon H, Brizel DM. Predictive and prognostic role of functional imaging of head and neck squamous cell carcinomas. Seminars in radiation oncology. Elsevier; 2012;220-232.
43. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Comm* 2014;5.
44. Leijenaar RT, Carvalho S, Hoebers FJ, et al. External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma. *Acta Oncol* 2015;54:1423-1429.
45. Parmar C, Leijenaar RT, Grossmann P, et al. Radiomic feature clusters and prognostic signatures specific for lung and head & neck cancer. *Sci Rep* 2015;5.
46. Kumar V, Gu Y, Basu S, et al. Radiomics: The process and the challenges. *Magn Reson Imaging* 2012;30:1234-1248.
47. Nakatsugawa M, Cheng Z, Goatman K, et al. Radiomic analysis of salivary glands and its role for predicting xerostomia in irradiated head and neck cancer patients. *Int J Radiat Oncol Biol Phys* 2016;96: S217.
48. Fan J, Li R. Statistical challenges with high dimensionality: Feature selection in knowledge discovery. arXiv; 2006. math/0602133.
49. Blei DM, Smyth P. Science and data science. *Proc Natl Acad Sci U S A* 2017.
50. Fogg BJ. A behavior model for persuasive design, *Persuasive '09. Proceedings of the 4th International Conference on Persuasive Technology.* New York, NY: ACM; 2009;40.
51. Junghans C, Feder G, Hemingway H, Timmis A, Jones M. Recruiting patients to medical research: Double blind randomised trial of "opt-in" versus "opt-out" strategies. *BMJ* 2005;331:940.
52. Mate S, Köpcke F, Toddenroth D, et al. Ontology-based data integration between clinical and research systems. *PloS One* 2015;10: e0116656.