



Harnessing data science to advance radiation oncology

Vogelius, Ivan R.; Petersen, Jens; Bentzen, Søren M.

Published in:
Molecular Oncology

DOI:
[10.1002/1878-0261.12685](https://doi.org/10.1002/1878-0261.12685)

Publication date:
2020

Document version
Publisher's PDF, also known as Version of record

Document license:
[CC BY](https://creativecommons.org/licenses/by/4.0/)

Citation for published version (APA):
Vogelius, I. R., Petersen, J., & Bentzen, S. M. (2020). Harnessing data science to advance radiation oncology. *Molecular Oncology*, 14(7), 1514–1528. <https://doi.org/10.1002/1878-0261.12685>

REVIEW ARTICLE

Harnessing data science to advance radiation oncology

Ivan R. Vogelius^{1,2} , Jens Petersen³ and Søren M. Bentzen⁴

1 Department of Oncology, Rigshospitalet, Copenhagen, Denmark

2 Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

3 Department of Computer Science, University of Copenhagen, Denmark

4 Department of Epidemiology & Public Health, Greenebaum Cancer Center, University of Maryland Baltimore, MD, USA

Keywordsartificial intelligence; data science;
radiotherapy**Correspondence**I. R. Vogelius, Department of Oncology,
Rigshospitalet, Copenhagen, Denmark
E-mail: ivan.richter.vogelius@regionh.dk(Received 15 November 2019, revised 27
February 2020, accepted 1 April 2020)

doi:10.1002/1878-0261.12685

Radiation oncology, a major treatment modality in the care of patients with malignant disease, is a technology- and computer-intensive medical specialty. As such, it should lend itself ideally to data science methods, where computer science, statistics, and clinical knowledge are combined to advance state-of-the-art care. Nevertheless, data science methods in radiation oncology research are still in their infancy and successful applications leading to improved patient care remain scarce. Here, we discuss data interoperability issues within and across organizational boundaries that hamper the introduction of big data and data science techniques in radiation oncology. At the semantic level, creating common underlying models and codification of the data, including the use of data elements with standardized definitions, an ontology, remains a work in progress. Methodological issues in data science and in the use of large population-based health data registries are identified. We show that data science methods and big data cannot replace randomized clinical trials in comparative effectiveness research by reviewing a series of instances where the outcomes of big data analyses and randomized trials are at odds. We also discuss the modern wave of machine learning and artificial intelligence as represented by deep learning and convolutional neural networks. Finally, we identify promising research avenues and remain optimistic that the data sources in radiation oncology can be linked to yield important insights in the near future. We argue that data science will be a valuable complement to, but not a replacement of, the traditional hypothesis-driven translational research chain and the randomized clinical trials that form the backbone of evidence-based medicine.

1. Introduction

Data science is a multidisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data (Wikipedia, 2019). Data science is emerging as a new paradigm in the biomedical sciences, distinct from conventional theoretical and

empirical science. In this new paradigm, patterns detected in large sets of clinical data provide a means to understand the nature of disease and its response to therapy, either alone or by representing a bedside-to-bench inverse translation, in which hypotheses are derived from clinical outcome data and then later studied in detail in the laboratory or tested in controlled clinical trials.

Abbreviations

API, application programming interface; DL, deep learning; EHR, electronic health record; PRO, patient-reported outcome; R&V, electronic record-and-verify systems; RGC, Radiogenomics Consortium; SNP, single nucleotide polymorphism.

A successful data science project should combine computer science, statistical knowledge, and domain knowledge from the field of interest.

Radiotherapy is a mainstay in modern anticancer therapy and is indicated in more than half of cancer patients at some point during disease management (Lievens and Grau, 2012). Modern radiotherapy is a computer-intensive and technology-heavy discipline with regulatory requirements for documenting and verifying radiotherapy exposure and should therefore be well suited for data science advances.

In the present review, we focus on data science in radiation oncology, as defined by the combination of data analytics and big data sources with the ultimate aim of gaining insights into improved therapy for future cancer patients. The data science paradigm can be contrasted with the translational research chain paradigm (Fig. 1), and in this review, we define data science as shown in Fig. 1. It should be mentioned that the workflow in radiation oncology is complex and involves many tasks that are likely to be successfully automated or at least partly supported by tools developed using machine learning techniques (Meyer *et al.*, 2018; Thompson *et al.*, 2018). Although there are some promising parallels between these efforts and the data science approach discussed here, it is beyond the scope of this review to cover the automation of radiation treatment planning and delivery in any detail.

Data science is, in many ways, a compelling concept in the field of radiation oncology. First and foremost, a data science approach seeks to analyze data on human patients treated as part of clinical routine. Pre-clinical tumor models show many important biological differences to spontaneous tumors in humans. With current advances in the understanding of the importance of host–disease interactions (Hanahan and Weinberg, 2011), in particular with the rise of immunoncology, including immunotherapy–radiotherapy combinations, the limitations of relying mainly on *in vitro* assays and small-animal models to develop new therapies are clear. Secondly, adverse events of radiation therapy depend not only on the detailed deposition of dose in time and space but also on patient-level comorbidities, comedications, and patient age (Bentzen, 2006). These are cofactors that are difficult to represent adequately in preclinical models. Also, the analysis of routine clinical data is a promising complement. Indeed, we have seen examples where human outcome data analysis has been used to describe the detrimental impact of radiotherapy on the immune system (Shiraishi *et al.*, 2018; Terrones-Campos *et al.*, 2019; van Rossum *et al.*, 2020). Such

findings might may serve as an important context for preclinical studies that aim to explore the beneficial effect of radiation for immune therapy (Durante and Formenti, 2019; Formenti and Demaria, 2013). Thirdly, although the controlled clinical trial is very likely to remain as the gold standard in evidence-based radiation oncology (Bentzen, 1998; Bentzen and Yarnold, 2014), there are numerous types of questions that cannot easily be subjected to clinical trial methodology. Big data from population-level registries are increasingly emerging as an important complement to trial outcomes, as discussed further below. In the case of radiation oncology, however, the big data are unfortunately often not all that big.

In this review, we discuss the data sources involved in radiation oncology data science, the methodological hurdles to consider when using big data sources, and potential solutions and promising future research avenues. We finish with an important discussion of the clinical utility of the knowledge obtained from data science in radiation oncology.

2. Data sources and missing links in radiation oncology

Large databases of treatment and outcome data have long been available for healthcare providers and researchers in multiple countries and regions across the globe. Databases provide essential reimbursement mechanisms for healthcare providers across public and insurance-based systems alike. Furthermore, most developed countries have national cancer registries with mandatory reporting of incident cases.

At the hospital level, electronic health records (EHRs) have long been in routine use in developed countries, and few will disagree that EHRs will expand their future role as an integral part of the hospital data infrastructure. Finally, focusing on the radiotherapy providers, electronic record-and-verify (R&V) systems are mandatory for documenting detailed radiation exposure and for improving patient safety. Detailed exposure data for all patients are stored and must be retrievable for clinical use in the common event of a future indication for re-irradiation. These different data sources, together with their key strengths and weaknesses, are shown in Fig. 2.

We note a substantial challenge in radiotherapy data science: the immediate loss of detailed radiotherapy exposure data when moving data from the departmental R&V database to other data sources. Hospital EHRs will typically carry information on prescription dose, time, and delivered number of fractions, but not the 3D dose distribution data available from R&V

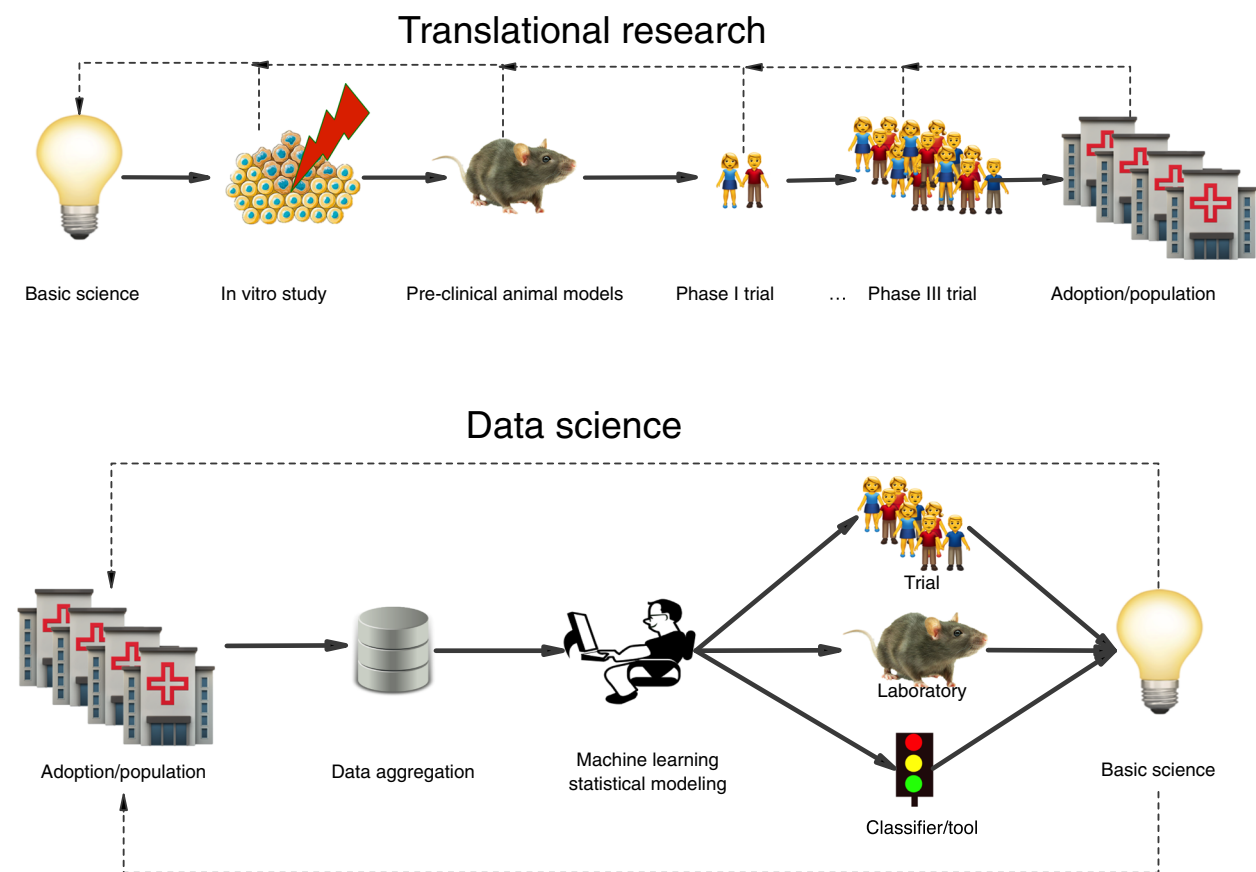


Fig. 1. The translational research chain versus the data science approach.

records. When moving beyond the hospital to claims-based registries or tumor registries, radiotherapy is often only recorded as a one-bit yes/no item, which, with few exceptions, preclude meaningful inference with regard to the optimal use of radiotherapy. Conversely, the R&V databases lack the crucial long-term outcome data from the larger systems (Fig. 3) and will often be limited in terms of the number of individuals treated. Even when patient-level links exist between these databases, developing an ontology for defining structures and for reporting doses for R&V records, and for defining specific long-term outcomes on the registry side, remains work in progress.

Powerful examples of the value of establishing a widely adopted ontology include the TNM classification for staging of cancer, developed between 1943 and 1952 by P. Denoix at the Institute Gustave-Roussy. This ontology was subsequently adopted by the Union for International Cancer Control (UICC), the International Federation of Gynecology and Obstetrics (FIGO), and the American Joint Committee for Cancer (AJCC) (UICC, 2017). Another example is the development of the Common Terminology Criteria

for Adverse Events (CTCAE), which has been widely adopted by many single or multimodality cancer studies (Trotti *et al.*, 2003). In radiation oncology, there have also been commendable efforts to standardize organ delineation for radiation therapy planning and the corresponding nomenclature used in routine clinical care (Duane *et al.*, 2017; Landberg *et al.*, 2016; Offersen *et al.*, 2015).

Nevertheless, ensuring the interoperability of databases requires us to do more. We need to define a set of minimum data elements to record radiation modality and delivery technique, which would be an important step forward (Hayman *et al.*, 2019). However, the effort needed to achieve a full, meaningful annotation of complex 3D imaging and exposure data is substantial. It should be recognized that manual annotations are often inconsistent even within a department and that they are generally restricted to the normal structures necessary for treating the patient in question; further details needed for research cannot be expected to be reliably annotated in clinical routine databases. Having said that, the retrospective estimation of dosimetry might be possible in the absence of 3D

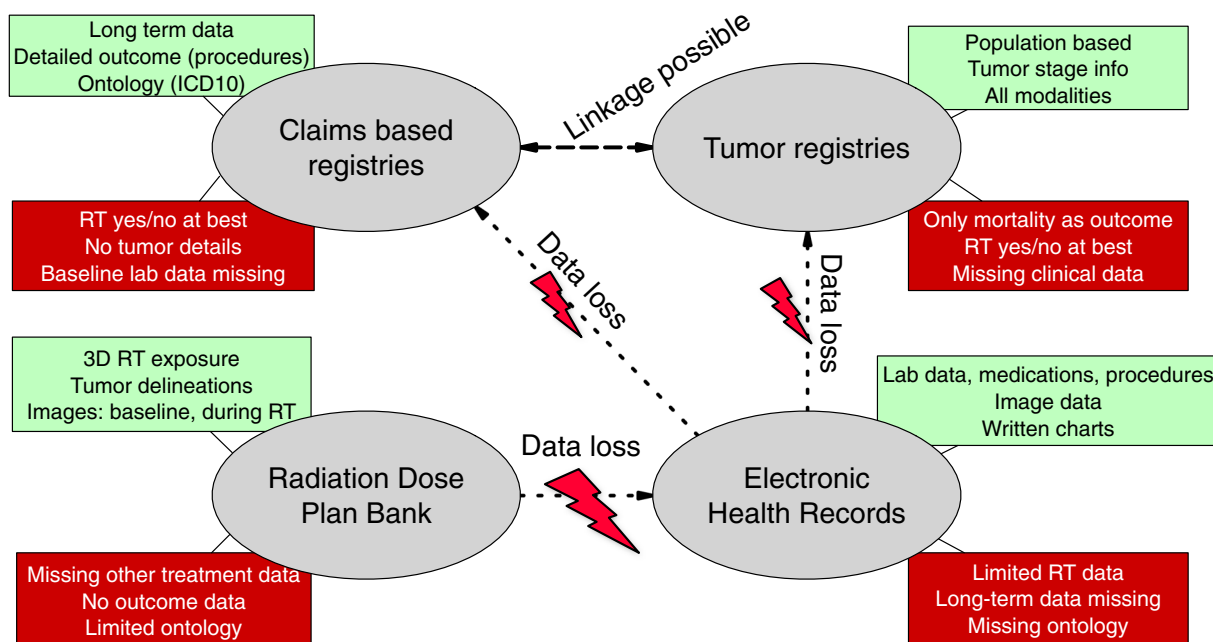


Fig. 2. Key sources of data and points of radiation data loss. From lower left and counterclockwise, radiation dose plan banks are currently available in all modern institutions as record-and-verify systems that contain image data and 3D radiation dose exposure, but no follow-up data. Hospital-based EHR systems contain more detail on other treatments, but do not contain detailed, granular radiotherapy data; such data are often reduced to prescription dose/fractionation. Radiotherapy data are further lost when these data are moved to large claims-based registries or tumor registries, where data on various aspects of long-term outcome are available. Red boxes: examples of main shortcomings of data source. Green boxes: examples of main strengths.

dosimetry data in selected cases: For example, cardiac exposure after breast and lymphoma treatment can be reasonably reliably quantified from 2D portal images or treatment descriptors in historical series (Darby *et al.*, 2013; Maraldo *et al.*, 2015; van Nimwegen *et al.*, 2017). Modern radiotherapy planning is becoming increasingly individualized, however, and therefore, such retrospective dosimetry will, in most cases, be unreliable with current radiotherapy (Maraldo *et al.*, 2012). The studies of cardiac exposure demonstrate another main strength of data science: the potential ability to study relatively rare, high-grade toxicity endpoints of high clinical relevance, such as major coronary events (Darby *et al.*, 2013). Most institutional series will not have enough power to resolve any relevant exposure–risk relationships for high-grade toxicity events. Consequently, low-grade clinical endpoints (such as low-grade radiation pneumonitis for lung exposure; Marks *et al.*, 2010a; Marks *et al.*, 2010b) often form the basis for recommended dose constraints, even when the disease itself has a dismal prognosis.

As an alternative to using the registries mentioned here, multiple institutions have successfully combined data in population-based studies (Thorsen *et al.*, 2016)

in which the numbers of cases reach several thousands, but in which dosimetry data remain limited. Finally, the concept of federated learning has been proposed and demonstrated as a means by which to combine insights from institutional series without the logistical and legal complexities of sharing patient data (Defraene *et al.*, 2019; Jochems *et al.*, 2017). However, the numbers of patients and the dosimetric granularity offered by current published examples of federated learning remain limited, as shown in Fig. 3. We now leave the data source discussion to focus on challenging aspects of the data analysis itself.

3. Applying data science to radiation oncology: The methodological challenges

Here, we discuss the methodological challenges of applying data science to radiation oncology. It should be emphasized that the covariates, as well as the endpoints that are analyzed, should be defined by clinical relevance or scientific interest and not by availability. When analyzing and publishing the data science material, it is relevant to keep some terms and concepts in mind. The safest route to appease a peer reviewer is to

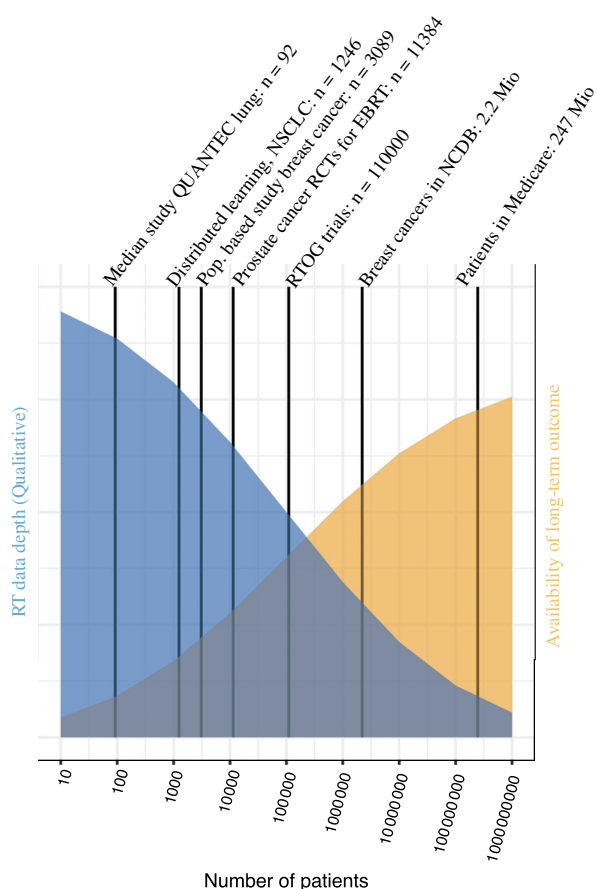


Fig. 3. Schematic illustration of loss of granularity of radiotherapy data when moving from single institutional series to the largest of available datasets. This graph provides a schematic illustration of the level of radiotherapy data granularity versus sample sizes across selected published studies and available databases in the United States as examples. QUANTEC (Marks *et al.*, 2010a) is an example of a federated learning model that aims to bridge institutional series (Defraene *et al.*, 2019). Also shown is a population-based series of breast cancer patients without detailed dosimetry, but with information available about whether internal mammary nodes were included in the radiotherapy target (Thorsen *et al.*, 2016). The graph also shows the number of patients in randomized trials of external beam radiotherapy (EBRT) for prostate cancer (Vogelius and Bentzen, 2018; Widmark *et al.*, 2019), the number of patients in Radiation Therapy Oncology Group (RTOG) trial databases (personal communication), and the number of breast cancer patients in the National Cancer Database (NCDB) and Medicare. Where long-term outcomes are available in the large series (to the right), radiotherapy information is often reduced to one bit of information (radiotherapy given or not) in these studies (McGale *et al.*, 2016). Abbreviations: QUANTEC: Quantitative Analyses of Normal tissue Effects in the Clinic. RT, radiotherapy; RCT, randomized controlled trial.

argue that the objective of a data science publication is to elucidate associations in data. A limitation of this approach, however, is that the presence of associations rarely has any clinical relevance. The next step would

be to perform prediction of outcome. Predictive models rely on one or more covariates that together give an (ideally) robust assessment of the probability of a given endpoint (Collins *et al.*, 2014, 2015). Note that the covariates need not be causally related to the endpoint in question to provide an accurate outcome prediction. The downside of predictive modeling without causal content is that the use of further refined radiotherapy to modify a dose distribution-related covariate might fail to provide the expected clinical benefit if other associations are broken, for example, when moving from photon to proton therapy. Causal relationships are generally preferable to both association studies and predictions without causal inference, but inference methods rely on avoiding bias. Here, the most robust method is the randomized controlled trial, but other methods exist where randomization is not feasible (Pearl, 2009).

Although causal inference is challenging, and prone to residual bias and confounding even when exercising the best possible care (cf. Fig. 4), such methods are likely to be more efficient when combined with improved, prospective outcome reporting, across a change of treatments (e.g., when moving from photon to proton therapy). Such prospective data registration programs are in place at a few leading institutions, with the Dutch-coordinated effort in head-and-neck cancer radiotherapy across modalities and institutions as a prime example (Verdonck-de Leeuw *et al.*, 2019). Patient-reported outcomes (PROs) are also of great interest in relation to toxicity assessments, and some data suggest that they can improve routine clinical care (Basch *et al.*, 2017). More research is needed, however, into the relationship that exists between PROs, and physician-assessed and analytical toxicity endpoints.

In order to be relevant and acceptable to radiation oncologists, predictive models must account for already well-established risk factors. In terms of the risk of toxicity, several clinical covariates have already been established (Appelt *et al.*, 2014; Thor *et al.*, 2019; Vogelius and Bentzen, 2012), and should be accounted for in the modeling. When it comes to disease-control endpoints, adjustment for disease stage is clearly necessary (and often available in tumor registries), but more detailed knowledge of disease burden and other prognostic factors might be important as well, in order to test the added utility of novel prognostic assays. For example, tumor volume and HPV status have been shown to provide more robust predictors of survival than radiomics features in head-and-neck cancer (Ger *et al.*, 2019). Similar examples exist of clinical factors that confound image-derived features in lung cancer

(Davey *et al.*, 2020). An important general discussion of such problems can be found in Welch *et al.* (2019).

Some authors have argued in favor of registry-based studies by highlighting the many obstacles that are encountered when designing and conducting large, randomized controlled trials (Quon *et al.*, 2019). The problem remains, however, that even after adjusting for available known covariates, hidden biases might still exist. This is exactly why the randomized controlled trial remains the gold standard of generating medical evidence in comparative effectiveness research. This does not preclude, however, mathematical modeling and data science methods as a component of evidence-based medicine.

4. Mathematical modeling and evidence-based medicine

Mathematical modeling is an attractive method for data-driven inference, where evidence from randomized trials is not available, trials are not feasible, or as a tool to describe heterogeneity in treatment effect. But how do we build confidence in a predictive model that supports clinical decisions? A key element to this end is external validation, that is, a quantitative assessment of a model's performance in an independent dataset. There is a rich statistical literature on the validation of models (Collins *et al.*, 2014, 2015). However, the term validation itself is not consistently defined (Altman and Royston, 2000), especially in radiation oncology, where exposure variables are often correlated. Still, most validation studies tend to conclude that the model tested is indeed valid! But there is no consensus among radiation oncology modelers as to what this exactly means. And validation is only meaningful within a carefully specified domain; models are most often not generalizable across domains. For example, a dose–volume–response model developed from photon therapy outcomes in adults may have been ‘validated’ in a similar setting but may not perform well in a cohort of patients treated with proton therapy, or in a population of, say, pediatric patients.

Using the literature, we have compared data science approaches against the gold standard of the randomized controlled trial in some radiotherapy-relevant cases (Ang *et al.*, 2014; McGale *et al.*, 2016; Pignon *et al.*, 2009; Zandberg *et al.*, 2018), where the effect size of an intervention has been estimated using both methods (Fig. 4). In many cases, the randomized controlled trial and registry-based effect estimates do not fall on the diagonal line of agreement, even when considering their confidence intervals. Or they are in direct contradiction (as shown by the gray zones in Fig. 4).

The meta-analysis by Pignon *et al.* (2009) shows that the coordinated synthesis of trial outcomes can yield statistical power that is comparable to that of registry analysis but without the associated risk of bias. It is remarkable that the confidence intervals of well-conducted meta-analyses are often comparable to those of registry studies, despite fewer patients (Pignon *et al.*, 2009). This is an indication that statistical power in registry studies is not limited by sampling variation alone, but by an overdispersion of effect sizes between individuals in the data, resulting from heterogeneity of treatment effects. It should be emphasized that for registry studies, it is the management of bias/confounding that dominates over statistical sampling uncertainty. *P*-values, which were developed to compare small sample sizes, often look impressive at face value in registry studies, but should largely be ignored because bias/confounding is the real concern (cf. Fig. 4). Effect sizes with confidence intervals are much more informative, but even these should be tempered by a careful analysis of possible bias.

The limitations of data science should thus be recognized in comparative effectiveness research. However, data science is an important and very relevant supplement to the paradigm of translational research, in particular for questions that are not amenable to being investigated by randomized controlled trials. This includes the previously mentioned example of a dose–response relationship for cardiotoxicity. In the next section, we discuss the dominant methodological development in data science, artificial intelligence, or, more precisely, deep learning (DL) methods.

5. Artificial intelligence: Deep learning

Machine learning is a branch of artificial intelligence, in which a mathematical model is built based on a sample dataset, known as the ‘training dataset’. This field of data science is being revolutionized by DL methods, a term that is typically associated with the training of multiple-layered/deep neural networks. DL techniques are increasingly being explored in radiotherapy, for purposes such as treatment planning, setup, verification, adaptation, or follow-up in clinics and in research (Meyer *et al.*, 2018; Sahiner *et al.*, 2019). In contrast to methods that combine selected and designed features and classifiers, DL utilizes end-to-end representation learning, in which features and their combinations are learned at multiple hierarchies jointly to solve a particular prediction task directly from data. This is one of the reasons for this approach's success. DL most often involves learning from suitably sampled training data

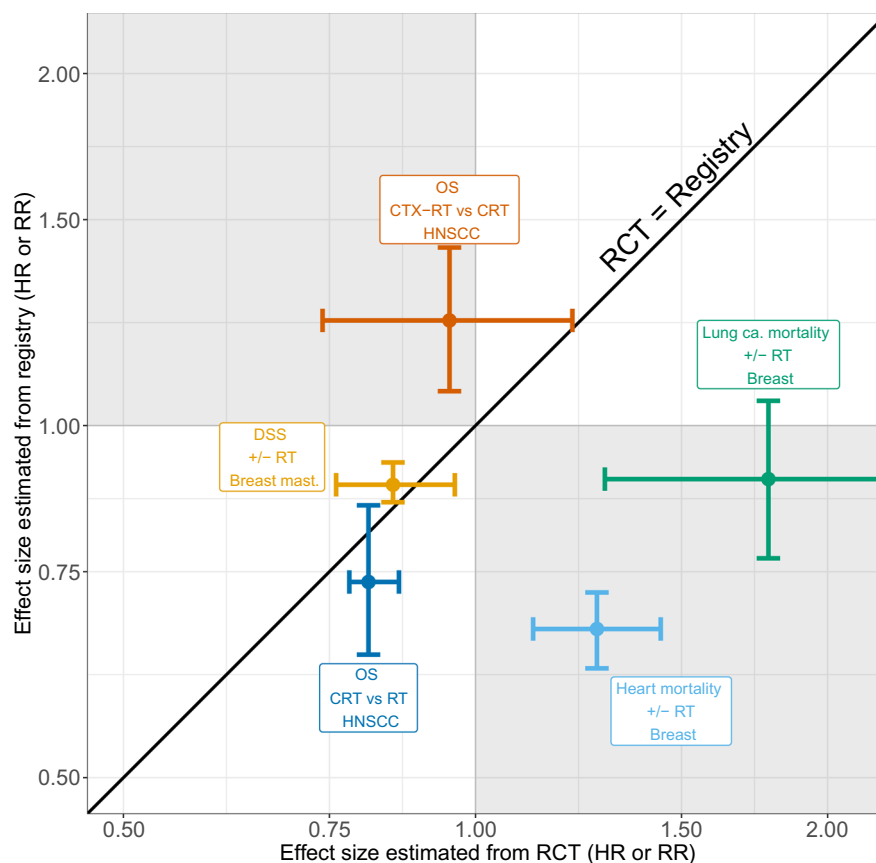


Fig. 4. A comparison of effect-size estimates from randomized controlled trials and registry-based analyses. The schematic shows published effect-size estimates from randomized controlled trials (x-axis) and registry-based analyses (y-axis). Concordant effect sizes are indicated by the black identity line. We see examples of registry-based studies over- and underestimating effects, as well as being relatively in agreement. Data are collected from Ang *et al.* (2014), McGale *et al.* (2016), Pignon *et al.* (2009), and Zandberg *et al.* (2018), and reanalyzed/replotted for this analysis by SMB and IRV. CRT, chemoradiotherapy; CTX-RT, cetuximab + RT; DSS, disease-specific survival; HNSCC, head-and-neck squamous cell cancer; OS, overall survival.

with ground-truth annotation provided by human experts or observed clinical outcome, so-called supervised learning. Less frequent applications involve semisupervised and even unsupervised learning, in which learning is done by finding inherent structures in the data with limited or no ground truth provided (see Table 1 for a glossary of terms).

Employed DL models can easily fit to both relevant and spurious signals and even to random noise in the training data (Zhang *et al.*, 2016), and the performance or ability to generalize to unseen data is therefore the relevant measure of success. The generalization error of a DL model, given a problem and training data, is something that must be empirically estimated, and occasional failures should be expected and recognized as an inherent property of the DL approach. As DL methods become an integral

part of clinical workflow and research, these risks must be properly assessed. It is not enough to assess a DL model's performance on measures, such as average similarity of segmentations and reconstructed images, with ground truths that are often clinically irrelevant. This is because DL methods are surprisingly good at recognizing and replicating even complex signals that often appear in the training data, while potentially failing to recognize obvious but rare signals that occur clinically (Meyer *et al.*, 2018). Estimates of prediction uncertainty in DL models could be used to raise warning flags in such cases. Such estimate prediction techniques include Monte Carlo dropout (Kendall and Gal, 2017), model ensembles (Lakshminarayanan *et al.*, 2017), and variational autoencoders (Hu *et al.*, 2019; Kohl *et al.*, 2018), which function by allowing many different predictions to be generated for each

Table 1. Glossary of data science terms.

Application programming interface (API)	Communication protocol that allows external communication with software or server. In this field, APIs allow researchers to write code (scripts) to query radiotherapy databases to extract features from (large numbers of) individual patients' scan or dosimetry data
Artificial neural networks (ANNs)	An ANN is a network of artificial neurons, connected such that output from a given neuron forms the input to one or more neurons in the next 'layer'. Passing input data through many successive such layers allows for complex transformations, that is, complex mathematical functions that link a set of inputs to a specific output
Artificial neuron	The artificial neuron is the basic building block of an ANN. It is a mathematical function that takes multiple real-valued inputs, each of which is multiplied by a weight. These weighted inputs are then summed and put into a so-called activation function that outputs a real value. The activation function is typically a nonlinear function, for example, a sigmoid function
Deep learning (DL)	A type of learning that uses multiple ANN layers to progressively extract higher level features from the raw input
Federated learning (a.k.a., distributed learning)	This approach entails training a model simultaneously on several datasets that reside on different servers while communicating model data (such as goodness-of-fit data and regression coefficients) rather than exchanging the data itself
Generalization error	Generalization errors are calculated by metrics that quantify the amount of error a prediction algorithm makes on a set of previously unseen data
High-dimensionality datasets	This is a general term used to describe datasets that contain large numbers of features per patient, including genomic data and image features
Machine learning (ML)	The study of how computers learn from data to solve problems. ML is also used to refer to algorithms or systems that learn from data how to solve a task, as opposed to being explicitly programmed how to do so
Multiple-layered network/deep neural networks	These are neural networks that consist of many layers of neurons between the input and output, such that the output of one layer becomes input for the next
Ontology	Representation, formal naming, and definition of the data in a field of research, examples are tumor characteristics (e.g., UICC staging), organ delineation/naming, dose descriptors, and disease/procedural codes
Record-and-verify databases	Databases that were originally invented to document treatments and reduce risk of errors, and that have evolved into complete information systems that contain image data, planned dose matrices, and detailed delivery data. They usually have some sort of application programming interface
Semisupervised learning	Machine learning from input data, where only a subset of input data is paired with output data, that is, an approach that mixes supervised and unsupervised learning
Single-layer model	This term describes conventional regression models that could be seen to provide a 'single layer': In these models, a single mathematical descriptor (e.g., logistic function or Cox model) connects input data to outcome prediction. It is used for illustration here, but it is not an often-used term
Supervised learning	The task of learning a function that maps an input to an output, based on example input–output pairs. Regression models are examples of this approach
Tall datasets	These are 'Big data' datasets where the number of cases (individuals, patients) is much larger than the number of features per case. Examples are population-based cancer registries or claims databases
Unsupervised learning	This approach finds patterns in datasets without preexisting labels, that is, based solely on the structure of the input data, which is also known as self-organization. Hierarchical clustering is an example of such a method
Wide datasets	'Big data' datasets, in which the number of features (data items) per case is much larger than the number of cases. Examples include data from genomics or proteomics or from medical imaging

data point. Under the critical assumption that the training set is representative of the underlying true distributions, such methods should be able to convey the variation of predictions to the end user, whether they are a scientist or clinician. However, more work is needed to bridge this critical gap of conveying uncertainty information when using these novel modeling approaches.

A related critique about DL methods is the 'black-box' nature of its output, which can result in a lack of transparency or even interpretability (Vollmer *et al.*,

2020). Because DL methods consist of deep hierarchies of nonlinear functions, whose many parameters are entirely learned from data, understanding their inner workings and predictions is not straightforward. This contrasts with conventional regression models that base their predictions on handcrafted features, and on the statistical significance and the associated effect size (e.g., relative risk or hazard ratio) for each feature, which can be reported to understand the drivers of the model. Understanding why and how DL models work is an active research area, as is conveying such

information to the end user (Gilpin *et al.*, 2018). For instance, possible insights into the inner workings of trained DL models can be gained by perturbing inputs and investigating the possible consequences, or by generating heat maps of the importance of a particular part of the input for a given prediction (Ancona *et al.*, 2017; Sahiner *et al.*, 2019).

The major factor that limits the performance of DL methods in medicine is the lack of good-quality reference data to learn from, thus emphasizing the importance of bridging the gaps in the data sources discussed earlier in the review. The performance of DL models grows with the amount of data in the training set, and although the performance of individual DL models can become saturated after a certain amount of data is included, further improvements can typically be made by extending model architectures (Hestness *et al.*, 2017; Sun *et al.*, 2017). Making public datasets that comprise protected health information is complicated by patient privacy concerns and laws, but has the potential to contribute enormously to advancing the field. The website grand-challenge.org hosts a large number of challenges and datasets in medical imaging, including the StructSeg2019 segmentation for radiotherapy planning challenge 2019 (Structseg, 2019). However, there is still a limited number of challenges and datasets for radiotherapy applications available.

6. Promising solutions and leaps forward

It is clear, that data loss is a primary obstacle to applying data science to radiation oncology, in particular the loss of detailed information about radiation exposure and image data when linking from dose plan to outcome registries. However, promising tools exist to overcome this data loss.

For instance, modern radiotherapy record-and-verify systems now provide users with relatively accessible interfaces to enable them to interact with data on a database level, using a so-called application programming interface (API). APIs are emerging as automation procedures, and they enable the reporting of dose–volume data for available annotated body regions in the dose planning systems (Cai *et al.*, 2019; Cardan *et al.*, 2019). APIs have also been used to link lung exposure to vital status registries (Stervik *et al.*, 2020). While these methods are still limited to structures delineated at the time of treatment planning and stored in the database, it is still a substantial leap forward from datasets that record dates, times, and prescription dose, but not dose distribution (Rubinstein and Warner, 2018).

Another approach to improve the interoperability of distinct datasets is the establishment of large consortia of collaborators. One example is the Early Breast Cancer Trialists' Collaborative Group (EBCTCG), which combines randomized trials data to achieve statistical power comparable to registry studies (Fig. 4). Another successful example is the Radiogenomics Consortium (RGC). Radiogenomics is the scientific study of the link between early or late radiation toxicity and common genetic variations, such as single nucleotide polymorphisms (SNPs). SNPs occur on average once in every 1000 nucleotides in the human genome, which means that there are roughly 4–5 million SNPs in an individual patient's genome. Due to the high dimensionality of SNP datasets, the early literature on SNP predisposing for radiation toxicity was dominated by false-positive SNPs that were not validated in independent data series (Barnett *et al.*, 2012a). This issue, combined with the inherently large number of covariates that affect the phenotypic presentation and the relatively low prevalence of many single-nucleotide variants of potential interest, means that this research field requires relatively large sample sizes to achieve sufficient statistical power to detect clinically relevant effect sizes (Barnett *et al.*, 2012a). To this end, RGC was established in 2009 (West and Rosenstein, 2010). It consists of a large volunteer research network, which currently comprises 222 members in 33 countries across 133 institutions. The RGC has successfully identified SNPs involved in radiation toxicity (Andreassen *et al.*, 2016; Kerns *et al.*, 2019), which are of potential value in guiding therapy decisions in individual patients (Bergom *et al.*, 2019).

Federated learning is another collaborative approach but one that removes the need to share protected health information. However, current publications from federated learning collaborations are limited to sample sizes that are smaller than the most prominent consortia (cf. Fig. 3).

Federated learning, consortia, or the automated linkage of record-and-verify databases to registries all suffer from a limited ontology of organ structure delineations, as well as limited ontology for procedures and endpoints. Adjusting for between-data-series variation is often an important contribution of consortia. For example, RGC has contributed to the field's methodology by providing methods to standardize toxicity scores between data series (Barnett *et al.*, 2012b). The limited availability of organ delineations on routinely treated patients is a challenge that remains, but automated delineation approaches are being actively pursued by both commercial and academic researchers as a way to counteract the continuously increasing

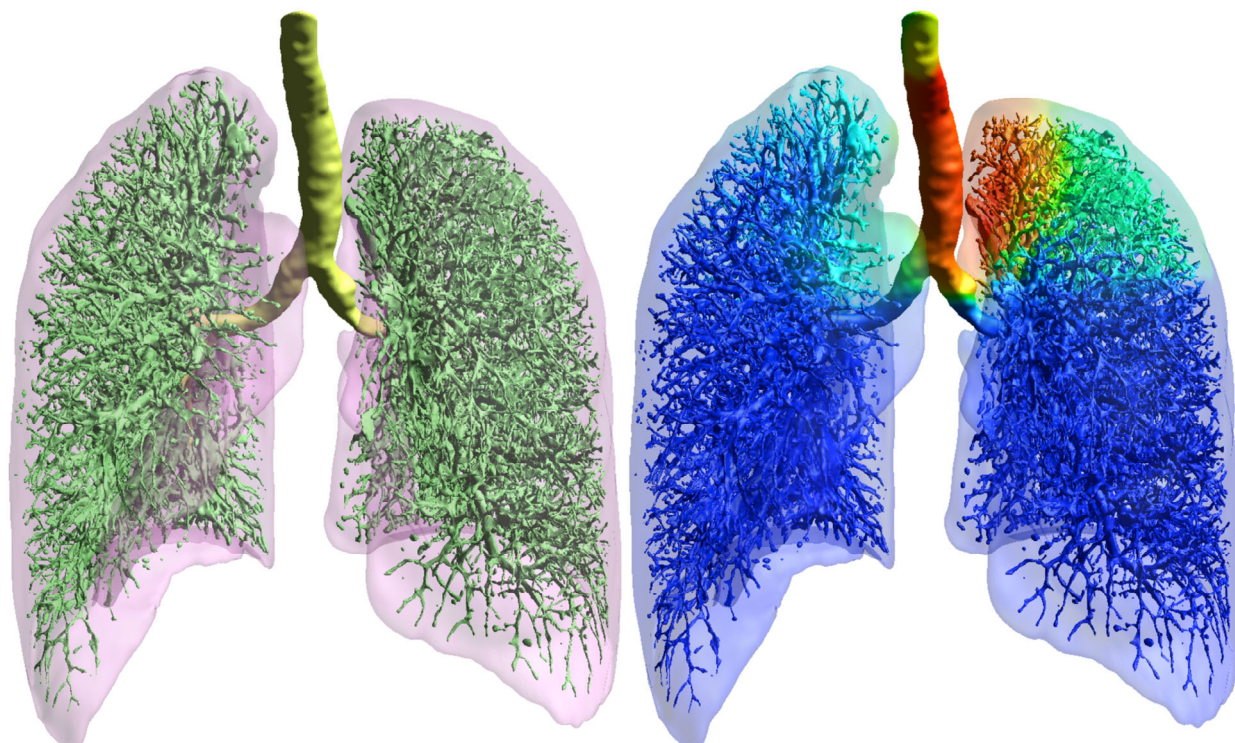


Fig. 5. A DL algorithm to analyze radiation exposure in routine clinical setting. Output from a DL algorithm in the form of a 3D U-Net architecture (Ronneberger *et al.*, 2015). The algorithm was trained on a dataset of manual annotations of lung substructures (vessels and airways). Subsequently, the DL algorithm performed annotations on a previously unseen routine, planning CT scan from a record-and-verify system yielding the airway and vessel annotations in green shades depicted on the left. Right: The annotated CT scan is overlaid with a 3D dose color wash to show the potential of automating access to detailed radiotherapy exposure data, which may subsequently be connected to outcome data from the larger registries. Note that some of the smaller vessels are exaggerated in size due to partial volume effects.

burden of substructure delineation in clinical practice (Zhu *et al.*, 2019). These methods are expected to improve in performance in coming years, and can also be used to increase the granularity of organ exposure in all of the above settings. In addition, structure delineation might come to include much more detailed substructures beyond those that are of current relevance for guideline-driven radiation dose planning. Figure 5 shows an example of automated airway and vessel annotation on a routine scan from record-and-verify database, which clearly exceeds the detail that could ever be achieved on substantial patient numbers with manual segmentation.

It should be emphasized that despite the difficulties associated with data linkage and interpretation, a fundamental truth is that the data are there and the data are accessible. We are no longer limited by manual tape switching or by the need to burn and send DVD's, as we were just 10 years ago. We are in an era where it is becoming realistic to perform large-scale analysis of detailed exposure data across many

institutions. Despite the necessary efforts to secure patient confidentiality and legal collaboration agreements, there are good reasons to be optimistic about the ability to harness data science in radiation oncology.

7. Towards clinical utility

Turning to the discussion of clinical utility of data science findings (Liu *et al.*, 2019), it should be emphasized that the key word in data science is science rather than data. The success of data science in radiation oncology will ultimately be measured by the clinical utility of the AI tools that are developed or by the novel scientific insights that are uncovered by using such methods, rather than by the simple ability to generate yet another prognostic model or by referring to the amount of data analyzed.

To that end, the methodological challenges must be overcome, and defining analyses and endpoints by availability must be superseded by choosing covariates

and endpoints based on clinical relevance and insight. There is an urgent need for cancer-specific endpoints rather than overall survival, which is used in even the best current tumor registries.

Uncertainty, bias and generalizability should be assessed critically to enable the end user clinician to judge the validity of the model. This is not a new insight (Wyatt and Altman, 1995), but it becomes even more important in the era of neural networks and big data science. The treating physician must fulfill the role of a ‘learned intermediary’ when using decision aids generated by data science and in order to fulfill that role the predictive model must have face validity and provide adequate descriptions of uncertainty. Clinical credibility is certainly still relevant. This is also an educational challenge: Radiation oncologists must be taught how to read, understand, and critically appraise papers that report data science results.

There are two examples of clinically relevant problems that can only be solved by big datasets and where data science is particularly promising: re-irradiation and detailed pattern-of-failure analysis. The retreatment of patients is a very common challenge in radiation oncology, yet our knowledge of organ tolerance is very limited. This is a complex problem where the intensity of the first treatment, the time interval from the first treatment, and the full time line of other anticancer medications are all expected to impact the risk of adverse events. Similarly, our knowledge of the detailed pattern of failure after primary radiation is currently limited to small, opportunistic series that fail to reflect the frequency of cancer recurrence after radiotherapy as a clinical problem (Chao *et al.*, 2003; Due *et al.*, 2014; Nygard *et al.*, 2018). This is a major limiting factor in harnessing the recent technological developments in radiation oncology to provide a dose-painting approach to radiotherapy, as envisioned decades ago by Ling *et al.* (2000) and Bentzen (2005). Clearly, modern radiotherapy equipment has the ability to meet these needs, but the biological knowledge has to come from data science, and the resulting data-driven hypotheses for treatment improvements will subsequently need comparative effectiveness testing in controlled clinical trials.

8. Conclusion

In conclusion, data science has an important role to play in radiation oncology and we are currently seeing just the first wave of that influence. It should be remembered, however, that data science should be multidisciplinary, and as such, it should involve statistical capabilities, and computational and domain knowledge from clinical radiation oncology.

Clearly, the translational research chain has limitations that are well described in the literature, including the reproducibility crisis and, equally important but less appreciated, the risk of falsely rejecting a good target prior to entering the human testing phase (Baumann *et al.*, 2001; Begley and Ellis, 2012). The addition of big data and data science will supplement, but not replace, the translational research chain. For data science to achieve its potential in radiation oncology, however, several breakthroughs are needed to overcome the limitations mentioned above.

In our view, as shown in Fig. 1, data science should complement translational science and traditional methods should be used to verify conclusions whenever possible. Data science methods applied to large curated datasets linking high-dimensionality biomarker data with clinical outcome data have the potential to provide tangible benefits to our future patients. Domain knowledge, however, is a key ingredient needed to harness the power of these methods.

9. Acknowledgements

The authors acknowledge support from Kirsten and Freddy Johansen’s award and National Cancer Institute (Grant No. P30 CA 134274-04). SMB was supported by funds through the Maryland Department of Health’s Cigarette Restitution Fund Program. IRV and JP received research funding from Varian Medical Systems.

Conflict of interest

The authors declare no conflict of interest.

References

- Altman DG and Royston P (2000) What do we mean by validating a prognostic model? *Stat Med* **19**, 453–473.
- Ancona M, Ceolini E, Öztireli C and Gross M (2017) A Unified View of Gradient-Based Attribution Methods for Deep Neural Networks, NIPS 2017-Workshop on Interpreting, Explaining and Visualizing Deep Learning. ETH Zürich, Zürich.
- Andreassen CN, Rosenstein BS, Kerns SL, Ostrer H, De Ruyscher D, Cesaretti JA, Barnett GC, Dunning AM, Dorling L and West CM (2016) Individual patient data meta-analysis shows a significant association between the ATM rs1801516 SNP and toxicity after radiotherapy in 5456 breast and prostate cancer patients. *Radiother Oncol* **121**, 431–439.
- Ang KK, Zhang Q, Rosenthal DI, Nguyen-Tan PF, Sherman EJ, Weber RS, Galvin JM, Bonner JA,

- Harris J, El-Naggar AK *et al.* (2014) Randomized phase III trial of concurrent accelerated radiation plus cisplatin with or without cetuximab for stage III to IV head and neck carcinoma: RTOG 0522. *J Clin Oncol* **32**, 2940–2950.
- Appelt AL, Vogelius IR, Farr KP, Khalil AA and Bentzen SM (2014) Towards individualized dose constraints: adjusting the QUANTEC radiation pneumonitis model for clinical risk factors. *Acta Oncol* **53**, 605–612.
- Barnett GC, Coles CE, Elliott RM, Baynes C, Luccarini C, Conroy D, Wilkinson JS, Tyrer J, Misra V, Platte R *et al.* (2012a) Independent validation of genes and polymorphisms reported to be associated with radiation toxicity: a prospective analysis study. *Lancet Oncol* **13**, 65–77.
- Barnett GC, West CM, Coles CE, Pharoah PD, Talbot CJ, Elliott RM, Tanteles GA, Symonds RP, Wilkinson JS and Dunning AM (2012b) Standardized total average toxicity score: a scale-and grade-independent measure of late radiotherapy toxicity to facilitate pooling of data from different studies. *Int J Radiat Oncol Biol Phys* **82**, 1065–1074.
- Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C and Schrag D (2017) Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. *JAMA* **318**, 197–198.
- Baumann M, Bentzen SM, Doerr W, Joiner MC, Saunders M, Tannock IF and Thames HD (2001) The translational research chain: is it delivering the goods? *Int J Radiat Oncol Biol Phys* **49**, 345–351.
- Begley CG and Ellis LM (2012) Drug development: raise standards for preclinical cancer research. *Nature* **483**, 531–533.
- Bentzen SM (1998) Towards evidence based radiation oncology: improving the design, analysis, and reporting of clinical outcome studies in radiotherapy. *Radiother Oncol* **46**, 5–18.
- Bentzen SM (2005) Theragnostic imaging for radiation oncology: dose-painting by numbers. *Lancet Oncol* **6**, 112–117.
- Bentzen SM (2006) Preventing or reducing late side effects of radiation therapy: radiobiology meets molecular pathology. *Nat Rev Cancer* **6**, 702–713.
- Bentzen SM and Yarnold J (2014) A toast to the silver anniversary of Clinical Oncology: a quarter of a century of advances in evidence-based radiation dose fractionation. *Clin Oncol (R Coll Radiol)* **26**, 599–601.
- Bergom C, West CM, Higginson DS, Abazeed ME, Arun B, Bentzen SM, Bernstein JL, Evans JD, Gerber NK, Kerns SL *et al.* (2019) The implications of genetic testing on radiation therapy decisions: a guide for radiation oncologists. *Int J Radiat Oncol Biol Phys* **105**, 698–712.
- Cai B, Altman MB, Reynoso F, Garcia-Ramirez J, He A, Edward SS, Zoberi I, Thomas MA, Gay H, Mutic S *et al.* (2019) Standardization and automation of quality assurance for high-dose-rate brachytherapy planning with application programming interface. *Brachytherapy* **18**, 108–114 e101.
- Cardan RA, Covington EL, Fiveash JB and Pople RA (2019) Using a whiteboard web application for tracking treatment workflow metrics for dosimetrists and physicians. *Med Dosim* **45**, 73–76.
- Chao KS, Ozyigit G, Tran BN, Cengiz M, Dempsey JF and Low DA (2003) Patterns of failure in patients receiving definitive and postoperative IMRT for head-and-neck cancer. *Int J Radiat Oncol Biol Phys* **55**, 312–321.
- Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, Voysey M, Wharton R, Yu LM, Moons KG *et al.* (2014) External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol* **14**, 40.
- Collins GS, Reitsma JB, Altman DG and Moons KG (2015) Transparent reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* **162**, 55–63.
- Darby SC, Ewertz M, McGale P, Bennet AM, Blom-Goldman U, Bronnum D, Correa C, Cutter D, Gagliardi G, Gigante B *et al.* (2013) Risk of ischemic heart disease in women after radiotherapy for breast cancer. *N Engl J Med* **368**, 987–998.
- Davey A, van Herk M, Faivre-Finn C, Mistry H and McWilliam A (2020) Is tumour sphericity an important prognostic factor in patients with lung cancer? *Radiother Oncol* **143**, 73–80.
- Defraene G, Dankers F, Price G, Schuit E, van Elmpt W, Arredouani S, Lambrecht M, Nuytens J, Faivre-Finn C and De Ruyscher D (2019) Multifactorial risk factors for mortality after chemotherapy and radiotherapy for non-small cell lung cancer. *Radiother Oncol*. pii: S0167-8140(19)33078-6. [Epub ahead of print]. <https://doi.org/10.1016/j.radonc.2019.09.005>
- Duane F, Aznar MC, Bartlett F, Cutter DJ, Darby SC, Jagsi R, Lorenzen EL, McArdle O, McGale P, Myerson S *et al.* (2017) A cardiac contouring atlas for radiotherapy. *Radiother Oncol* **122**, 416–422.
- Due AK, Vogelius IR, Aznar MC, Bentzen SM, Berthelsen AK, Korreman SS, Loft A, Kristensen CA and Specht L (2014) Recurrences after intensity modulated radiotherapy for head and neck squamous cell carcinoma more likely to originate from regions with high baseline [18F]-FDG uptake. *Radiother Oncol* **111**, 360–365.
- Durante M and Formenti S (2019) Harnessing radiation to improve immunotherapy: better with particles? *Br J Radiol* **93**, 20190224.

- Formenti SC and Demaria S (2013) Combining radiotherapy and cancer immunotherapy: a paradigm shift. *J Natl Cancer Inst* **105**, 256–265.
- Ger RB, Zhou S, Elgohari B, Elhalawani H, Mackin DM, Meier JG, Nguyen CM, Anderson BM, Gay C, Ning J *et al.* (2019) Radiomics features of the primary tumor fail to improve prediction of overall survival in large cohorts of CT- and PET-imaged head and neck cancer patients. *PLoS ONE* **14**, e0222509.
- Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M and Kagal L (2018) Explaining explanations: an overview of interpretability of machine learning. In 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), pp. 80–89. IEEE, Piscataway, NJ.
- Hanahan D and Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* **144**, 646–674.
- Hayman JA, Dekker A, Feng M, Keole SR, McNutt TR, Machtay M, Martin NE, Mayo CS, Pawlicki T and Smith BD (2019) Minimum data elements for radiation oncology: an ASTRO consensus paper. *Pract Radiat Oncol* **9**, 395–401.
- Hestness J, Narang S, Ardalani N, Damos G, Jun H, Kianinejad H, Patwary M, Ali M, Yang Y and Zhou Y (2017). Deep learning scaling is predictable, empirically. arXiv preprint arXiv:1712.00409.
- Hu S, Worrall D, Knekt S, Veeling B, Huisman H and Welling M (2019) Supervised uncertainty quantification for segmentation with multiple annotations. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 137–145. Springer, Berlin.
- Jochems A, Deist TM, El Naqa I, Kessler M, Mayo C, Reeves J, Jolly S, Matuszak M, Ten Haken R, van Soest J *et al.* (2017) Developing and validating a survival prediction model for NSCLC patients through distributed learning across 3 countries. *Int J Radiat Oncol Biol Phys* **99**, 344–352.
- Kendall A and Gal Y (2017) What uncertainties do we need in Bayesian deep learning for computer vision? In Advances in Neural Information Processing Systems, pp. 5574–5584. Neural Information Processing Systems Foundation, Inc. (NIPS): San Diego, CA.
- Kerns SL, Fachal L, Dorling L, Barnett GC, Baran A, Peterson DR, Hollenberg M, Hao K, Narzo AD and Ahsen ME (2019) Radiogenomics Consortium Genome-Wide Association Study meta-analysis of late toxicity after prostate cancer radiotherapy. *J Natl Cancer Inst* **112**(2), 179–190. <https://doi.org/10.1093/jnci/djz075>
- Kohl S, Romera-Paredes B, Meyer C, De Fauw J, Ledsam JR, Maier-Hein K, Eslami SA, Rezende DJ and Ronneberger O (2018) A probabilistic u-net for segmentation of ambiguous images. In Advances in Neural Information Processing Systems, pp. 6965–6975. Neural Information Processing Systems Foundation, Inc. (NIPS): San Diego, CA.
- Lakshminarayanan B, Pritzel A and Blundell C (2017) Simple and scalable predictive uncertainty estimation using deep ensembles. In Advances in Neural Information Processing Systems, pp. 6402–6413. Neural Information Processing Systems Foundation, Inc. (NIPS): San Diego, CA.
- Landberg T, Chavaudra J, Dobbs J, Gerard J-P, Hanks G, Horiot J-C, Johansson K-A, Möller T, Purdy J, Suntharalingam N *et al.* (2016) Report 62. *J Int Comm Radiat Units Measur* **os32**. <https://doi.org/10.1093/jicru/os32.1.Report62>
- Lievens Y and Grau C (2012) Health economics in radiation oncology: introducing the ESTRO HERO project. *Radiother Oncol* **103**, 109–112.
- Ling CC, Humm J, Larson S, Amols H, Fuks Z, Leibel S and Koutcher JA (2000) Towards multidimensional radiotherapy (MD-CRT): biological imaging and biological conformality. *Int J Radiat Oncol Biol Phys* **47**, 551–560.
- Liu Y, Chen P-HC, Krause J and Peng L (2019) How to read articles that use machine learning: users' guides to the medical literature. *JAMA* **322**, 1806–1816.
- Maraldo MV, Brodin NP, Vogelius IR, Aznar MC, Munck Af Rosenschold P, Petersen PM and Specht L (2012) Risk of developing cardiovascular disease after involved node radiotherapy versus mantle field for Hodgkin lymphoma. *Int J Radiat Oncol Biol Phys* **83**, 1232–1237.
- Maraldo MV, Giusti F, Vogelius IR, Lundemann M, van der Kaaij MAE, Ramadan S, Meulemans B, Henry-Amar M, Aleman BMP, Raemaekers J *et al.* (2015) Cardiovascular disease after treatment for Hodgkin's lymphoma: an analysis of nine collaborative EORTC-LYSA trials. *Lancet Haematol* **2**, e492–e502.
- Marks LB, Bentzen SM, Deasy JO, Kong FM, Bradley JD, Vogelius IS, El Naqa I, Hubbs JL, Lebesque JV, Timmerman RD *et al.* (2010a) Radiation dose-volume effects in the lung. *Int J Radiat Oncol Biol Phys* **76**, S70–76.
- Marks LB, Ten Haken RK and Martel MK (2010b) Guest editor's introduction to QUANTEC: a users guide. *Int J Radiat Oncol Biol Phys* **76**, S1–S2.
- McGale P, Cutter D, Darby SC, Henson KE, Jagsi R and Taylor CW (2016) Can observational data replace randomized trials? *J Clin Oncol* **34**, 3355–3357.
- Meyer P, Noblet V, Mazzara C and Lallement A (2018) Survey on deep learning for radiotherapy. *Comput Biol Med* **98**, 126–146.
- Nygard L, Vogelius IR, Fischer BM, Kjaer A, Langer SW, Aznar M, Persson GF and Bentzen SM (2018) A competing risk model of first failure site after definitive (chemo) radiation therapy for locally advanced non-small cell lung cancer. *J Thorac Oncol* **13**, 559–567.

- Offersen BV, Boersma LJ, Kirkove C, Hol S, Aznar MC, Biete Sola A, Kirova YM, Pignol JP, Remouchamps V, Verhoeven K *et al.* (2015) ESTRO consensus guideline on target volume delineation for elective radiation therapy of early stage breast cancer. *Radiother Oncol* **114**, 3–10.
- Pearl J (2009) Causal inference in statistics: an overview. *Statist Surv* **3**, 96–146.
- Pignon JP, le Maitre A, Maillard E, Bourhis J, MACH-NC Collaborative Group (2009) Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): an update on 93 randomised trials and 17,346 patients. *Radiother Oncol* **92**, 4–14.
- Quon H, McNutt T, Lee J, Bowers M, Jiang W, Lakshminarayanan P, Cheng Z, Han P, Hui X and Shah V (2019) Needs and challenges for radiation oncology in the era of precision medicine. *Int J Radiat Oncol Biol Phys* **103**, 809–817.
- Ronneberger O, Fischer P and Brox T (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation, pp. 234–241. Springer International Publishing, Cham.
- Rubinstein SM and Warner JL (2018) CancerLinQ: origins, implementation, and future directions. *JCO Clin Cancer Inform* **2**, 1–7.
- Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, Summers RM and Giger ML (2019) Deep learning in medical imaging and radiation therapy. *Med Phys* **46**, e1–e36.
- Shiraishi Y, Fang P, Xu C, Song J, Krishnan S, Koay EJ, Mehran RJ, Hofstetter WL, Blum-Murphy M, Ajani JA *et al.* (2018) Severe lymphopenia during neoadjuvant chemoradiation for esophageal cancer: a propensity matched analysis of the relative risk of proton versus photon-based radiation therapy. *Radiother Oncol* **128**, 154–160.
- Stervik L, Pettersson N, Scherman J, Behrens CF, Ceberg C, Engelholm S, Gunnarsson K, Hallqvist A, Nyman J, Persson GF *et al.* (2020) Analysis of early respiratory-related mortality after radiation therapy of non-small-cell lung cancer: feasibility of automatic data extraction for dose–response studies. *Acta Oncol* 1–8. <https://doi.org/10.1080/0284186X.2020.1739331>
- Structseg (2019) Automatic structure segmentation for radiotherapy planning challenge. <https://structseg2019.grand-challenge.org/>
- Sun C, Shrivastava A, Singh S and Gupta A (2017) Revisiting unreasonable effectiveness of data in deep learning era. In Proceedings of the IEEE International Conference on Computer Vision, pp. 843–852. The Computer Vision Foundation.
- Terrones-Campos C, Ledergerber B, Vogelius IR, Specht L, Helleberg M and Lundgren J (2019) Lymphocyte count kinetics, factors associated with the end-of-radiation-therapy lymphocyte count, and risk of infection in patients with solid malignant tumors treated with curative-intent radiation therapy. *Int J Radiat Oncol Biol Phys* **105**, 812–823.
- Thompson RF, Valdes G, Fuller CD, Carpenter CM, Morin O, Aneja S, Lindsay WD, Aerts HJ, Agrimson B and Deville C Jr (2018) Artificial intelligence in radiation oncology: a specialty-wide disruptive transformation? *Radiother Oncol* **129**, 421–426.
- Thor M, Deasy J, Iyer A, Bendau E, Fontanella A, Apte A, Yorke E, Rimner A and Jackson A (2019) Toward personalized dose-prescription in locally advanced non-small cell lung cancer: validation of published normal tissue complication probability models. *Radiother Oncol* **138**, 45–51.
- Thorsen LB, Offersen BV, Dano H, Berg M, Jensen I, Pedersen AN, Zimmermann SJ, Brodersen HJ, Overgaard M and Overgaard J (2016) DBCG-IMN: a population-based cohort study on the effect of internal mammary node irradiation in early node-positive breast cancer. *J Clin Oncol* **34**, 314–320.
- Trotti A, Colevas AD, Setser A, Rusch V, Jaques D, Budach V, Langer C, Murphy B, Cumberlin R and Coleman CN (2003) CTCAE v3.0: development of a comprehensive grading system for the adverse effects of cancer treatment. *Semin Radiat Oncol* **13**, 176–181.
- UICC (2017) TNM History, Evolution and Milestones. Geneva, Switzerland: Union for International Cancer Control.
- van Nimwegen FA, Ntentas G, Darby SC, Schaapveld M, Hauptmann M, Lugtenburg PJ, Janus CPM, Daniels L, van Leeuwen FE, Cutter DJ *et al.* (2017) Risk of heart failure in survivors of Hodgkin lymphoma: effects of cardiac exposure to radiation and anthracyclines. *Blood* **129**, 2257–2265.
- van Rossum PSN, Deng W, Routman DM, Liu AY, Xu C, Shiraishi Y, Peters M, Merrell KW, Hallemeier CL, Mohan R *et al.* (2020) Prediction of severe lymphopenia during chemoradiation therapy for esophageal cancer: development and validation of a pretreatment nomogram. *Pract Radiat Oncol* **10**, e16–e26.
- Verdonck-de Leeuw IM, Jansen F, Brakenhoff RH, Langendijk JA, Takes R, Terhaard CHJ, Baatenburg de Jong RJ, Smit JH and Leemans CR (2019) Advancing interdisciplinary research in head and neck cancer through a multicenter longitudinal prospective cohort study: the NETHERlands Quality of life and Biomedical Cohort (NET-QUBIC) data warehouse and biobank. *BMC Cancer* **19**, 765.
- Vogelius IR and Bentzen SM (2012) A literature-based meta-analysis of clinical risk factors for development of radiation induced pneumonitis. *Acta Oncol* **51**, 975–983.
- Vogelius IR and Bentzen SM (2018) Dose response and fractionation sensitivity of prostate cancer after

- external beam radiation therapy: a meta-analysis of randomized trials. *Int J Radiat Oncol Biol Phys* **100**, 858–865.
- Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, Cumbers S, Jonas A, McAllister KSL, Myles P *et al.* (2020) Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ* **368**, l6927.
- Welch ML, McIntosh C, Haibe-Kains B, Milosevic MF, Wee L, Dekker A, Huang SH, Purdie TG, O’Sullivan B, Aerts H *et al.* (2019) Vulnerabilities of radiomic signature development: the need for safeguards. *Radiother Oncol* **130**, 2–9.
- West C and Rosenstein BS (2010) Establishment of a radiogenomics consortium. *Int J Radiat Oncol Biol Phys* **76**, 1295–1296.
- Widmark A, Gunnlaugsson A, Beckman L, Thellenberg-Karlsson C, Hoyer M, Lagerlund M, Kindblom J, Ginman C, Johansson B, Bjornlinger K *et al.* (2019) Ultra-hypofractionated versus conventionally fractionated radiotherapy for prostate cancer: 5-year outcomes of the HYPO-RT-PC randomised, non-inferiority, phase 3 trial. *Lancet* **394**, 385–395.
- Wikipedia (2019) Data Science. https://en.wikipedia.org/wiki/Data_science
- Wyatt JC and Altman DG (1995) Commentary: Prognostic models: clinically useful or quickly forgotten? *BMJ* **311**, 1539–1541.
- Zandberg DP, Cullen K, Bentzen SM and Goloubeva OG (2018) Definitive radiation with concurrent cetuximab vs. radiation with or without concurrent cytotoxic chemotherapy in older patients with squamous cell carcinoma of the head and neck: analysis of the SEER-medicare linked database. *Oral Oncol* **86**, 132–140.
- Zhang C, Bengio S, Hardt M, Recht B and Vinyals O. (2016) Understanding deep learning requires rethinking generalization. arXiv preprint arXiv:1611.03530.
- Zhu W, Huang Y, Zeng L, Chen X, Liu Y, Qian Z, Du N, Fan W and Xie X (2019) AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Med Phys* **46**, 576–589.