Radiotherapy and Oncology 153 (2020) 43–54

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Review Article

# From multisource data to clinical decision aids in radiation oncology: The need for a clinical data science community

Joanna Kazmierska [a,b,1], Andrew Hope [c,1], Emiliano Spezi [d,e], Sam Beddar [f,g], William H. Nailon [h], Biche Osong [i], Anshu Ankolekar [i], Ananya Choudhury [i], Andre Dekker [i], Kathrine Røe Redalen [j,2], Alberto Traverso [i,2,*]

[a] Radiotherapy Department II, Greater Poland Cancer Centre; [b] Electroradiology Department, University of Medical Sciences, Poznan, Poland; [c] Princess Margaret Cancer Centre, Toronto, Canada; [d] School of Engineering, Cardiff University, United Kingdom; [e] Department of Medical Physics, Velindre Cancer Centre, Cardiff, United Kingdom; [f] Department of Radiation Physics, The University of Texas MD Anderson Cancer Center; [g] The UTHealth Graduate School of Biomedical Sciences, Houston, United States; [h] Department of Oncology Physics, University of Edinburgh, United Kingdom; [i] Department of Radiation Oncology (Maastro), GROW School for Oncology, Maastricht University Medical Centre+, The Netherlands; [j] Department of Physics, Norwegian University of Science and Technology, Trondheim, Norway

ARTICLE INFO

*Article history:*
Received 15 July 2020
Received in revised form 25 September 2020
Accepted 26 September 2020
Available online 13 October 2020

*Keywords:*
Artificial intelligence
Big data
Data science
Personalized treatment
Radiotherapy
Shared decision making

ABSTRACT

Big data are no longer an obstacle; now, by using artificial intelligence (AI), previously undiscovered knowledge can be found in massive data collections. The radiation oncology clinic daily produces a large amount of multisource data and metadata during its routine clinical and research activities. These data involve multiple stakeholders and users. Because of a lack of interoperability, most of these data remain unused, and powerful insights that could improve patient care are lost. Changing the paradigm by introducing powerful AI analytics and a common vision for empowering big data in radiation oncology is imperative. However, this can only be achieved by creating a clinical data science community in radiation oncology. In this work, we present why such a community is needed to translate multisource data into clinical decision aids.



*The clinic as a learning health care system*

Many large commercial enterprises are redirecting their business approaches to exploit the new knowledge they can gain from the data they collect daily. This strategy arose from the need to mine a large amount of diverse data, often unstructured and coming from multiple sources: so-called "big data." Whereas initially big data seemed to present an obstacle, now it is becoming more evident that leveraging massive data collections using novel techniques can reveal previously undiscovered knowledge [1,2]. These techniques include analytic methods spanning from traditional statistics and hypothesis testing to more advanced algorithms inspired by machine learning (ML), a branch of artificial intelligence (AI), in which powerful computational systems augment our brain's learning capacity by employing complex mathematical algorithms to reveal patterns in data, mainly for the purpose of generating new knowledge [3,4]. AI is not a new concept in oncology. Recent reviews described two major applications of AI in the medical field: automation and decisions' augmentation [5,6]. The former includes applications such as auto contouring of both organs at risk (OAR)s and target volumes; the latter covers the whole spectrum of decision support systems. However, by comparing applications in the enterprise domain, AI is often referred to as "data analytics". In this view, two powerful applications of AI in medicine are often forgotten, namely the ability of AI to retrieve

*Abbreviations:* AAPM, American Association of Physicists in Medicine; AI, artificial intelligence; ASTRO, American Society for Radiation Oncology; CARO, Canadian Association of Radiation Oncology; CRC, patient association for colorectal cancer; CT, computed tomography; EFOMP, European Federation of Organisations in Medical Physics; ESTRO, European Society for Radiotherapy and Oncology; HER, electronic health records; FARO, The Federation of Asian Organizations for Radiation Oncology; FDA, Foods and Drug Administration; I2B2, Informatics for Integrating Biology and the Bedside; ML, machine learning; PHT, Personal Health Train; PROM, patient-reported outcome measures; RANZCR, Royal Australian and New Zealand College of Radiologists; RCTs, randomized clinical trials; RWE, real world evidence.

* Corresponding author at: Department of Radiation Oncology (Maastro), GROW School for Oncology, Maastricht University Medical Centre+, Maastricht, The Netherlands.
*E-mail address:* alberto.traverso@maastro.nl (A. Traverso).

[1] Shared first authors.
[2] Shared senior authors.

https://doi.org/10.1016/j.radonc.2020.09.054
0167-8140/

data belonging to multiple sources and spread across different locations. Most of the big-companies' data analytics include powerful tool that can collect data from multiple sources, such as for example social media or health wearables. This part of AI is very useful, because of the intrinsic nature of multisource data: they are sparse and unstructured. The second powerful aspect of AI is that, after having retrieved multisource data, automated QA can be performed. This aspect of AI is often forgotten, but in radiation oncology data quality is fundamental for applying AI in the clinic. After automated QA, which starts from unstructured data (often referred to as "veracity" of big data), data are transformed into a network of knowledge, the so-called "linked-data", resulting in new knowledge. We believe this broader view of AI is key to a new era in our healthcare systems.

Translating this new learning paradigm into radiation oncology will improve the classification of disease and reveal new ways to improve cancer treatment and predict patients' clinical events. Unfortunately, the radiation oncology community lags far behind in the adoption of big data approaches for providing the patient-centered, individualized care often referred to as personalized medicine [7]. With the term radiation oncology community, we do not only refer to radiation oncologists, but the extensive community concerned with treating cancer patients with radiotherapy. The treatment is rarely only consisting of radiotherapy but is often multimodal and consisting of radiotherapy in combination with chemotherapy, immunotherapy and/or surgery involving multiple professionals such as radiologists, pathologists, surgeons, medical physicists, and medical- and radiation oncologists.

While the community agrees that the future of medicine as a whole and radiation oncology in particular is in learning health care systems, where data are transformed into new knowledge as part of clinical routine, there remain gaps in our ability to rapidly learn from data generated in the clinic during the course of patient care [5]. By definition a learning health care system is a system that has been designed to generate and apply the best evidence generated from a collaborative effort among patients and care providers." The central point of a learning healthcare system is that knowledge discovery becomes a natural outgrowth of patient care. A learning healthcare system is meant to push forward evidence-based medicine by: a) fast translation from knowledge produced in clinical research to clinical practice; (b) empowerment of a shared responsibility culture between the different stakeholders involved in the clinic; and (c) facilitating engagement of patients and doctors for evidence production and dissemination [8].

In radiation oncology, we still mainly learn by narrowing and simplifying our research questions, in the process often moving them far from the complexity of real-world clinical practice. For example, most support for clinical decisions comes from clinical trial data. On one hand, clinical trials can provide high-quality data, but on the other hand, they have several major drawbacks: a) their exclusion of patients with complex cases that do not fit their strict inclusion criteria; b) their narrow focus on just one research question or a limited number of questions that often determine the choice of collecting specific variables; c) their high cost; d) the long time required to recruit sufficient patients to reach statistically significant results; and e) their infrequent exploration of how combinations of several factors might influence patients' outcomes. Conversely, patients produce a vast amount of data, from diagnosis to treatment and follow-ups. Only a small percentage of this data is actively used to produce new insights that can push our clinical practice and lessons learnt from clinical trials towards personalized medicine. In this view, big data empowered with AI is not a strategy to substitute randomized clinical trials, but rather a strategy to augment the knowledge from clinical trials. For example, AI can be used to explore multisource big data to better stratify patients and optimize clinical trials enrollment by defining group of patients for which the introduction of a new treatment is more likely to be found beneficial. Finally, by leveraging multisource big data a large spectrum of prognostic as well as confounding factors can be examined. This data integrates and considerably expand the original collections from randomized clinical trials. A recent communitarian effort is being carrying on boosting the efficacy of RCTs. This effort foresees the possibility to increase the "pragmaticism" of RCTs. A detailed review [78] pointed out the prominent role of AI in supporting this translation. By combining the expertise brought by clinical data scientists and medical doctors it is possible to use robust, validated and well understood AI tools to improve trial success rates starting from trial design and preparation (e.g. a better recruitment strategy) to execution." A recent investigation providing updating guidelines for more "pragmatic" RCTs, the SPIRIT-AI [79], is supporting the above-mentioned transition and it is currently adopted in recent clinical trial protocols that included AI-driven intervention. These guidelines were needed considering the increasing number of RCTs making use of AI tools. It is important to highlight how specific attention and dedicated methodologies need to be adopted when performing casual inferencing from both randomized clinical trials and multisource data [9]. The same issues that exist for casual inferencing from observational studies, such as the presence of confounding factors, sampling selection and cross-population biases also exist for inferencing from aggregated data (e.g. multisource data). A recent study [10] recommended the extension of parametric causal inferencing mathematical models specifically developed for clinical trials to non-parametric models specifically developed for aggregated data. The authors claimed that this methodology cannot be separated from the AI analytical tools used to process that data. This is indeed a key point, which need to be combined with the need of improving data quality (see coming statements) for robust casual inferencing.

The fuel of a learning health care system in radiation oncology is the data that are generated every day in the clinic. This requires us to reimagine the clinic as a source of big data. Currently, most of the big data generated in the clinic are wasted as a source of research because we have been unable to equip the clinic with big analytics. Nevertheless, we cannot support a learning paradigm in radiation oncology solely by borrowing technologies from other fields, such as business enterprises. The path towards a learning health care system in radiation oncology needs to pass several milestones, which are summarized in Fig. 1. All these milestones represent shifts in our concept of traditional clinical medicine. These milestones are:

- M1: Understanding the clinic as a source of big data: Where do the data come from and why are the data "big?" Data are not only produced directly by daily routine clinical activities, but also indirectly, for example when researchers process clinical data. This produces a combination of data and metadata, which are logically connected but might be sparse, even within the same institution. Combining and reunifying these data is the largest challenge to be tackled.
- M2: Identifying data types and involved stakeholders. The generation of multisource big data involves different professions and users. It is fundamental to identify not only the users and stakeholders of these data but also the major constraints that limit the interoperability of these data. Interoperability and data housekeeping are the keys for boosting data quality. High-quality data will have a strong impact on the robustness and integrity of our data-driven clinical decisions.
- M3: Defining which data analytics can be used to extract unique insights from multisource data. After we learn how to correctly retrieve, curate, mine, and combine multisource big data, it is

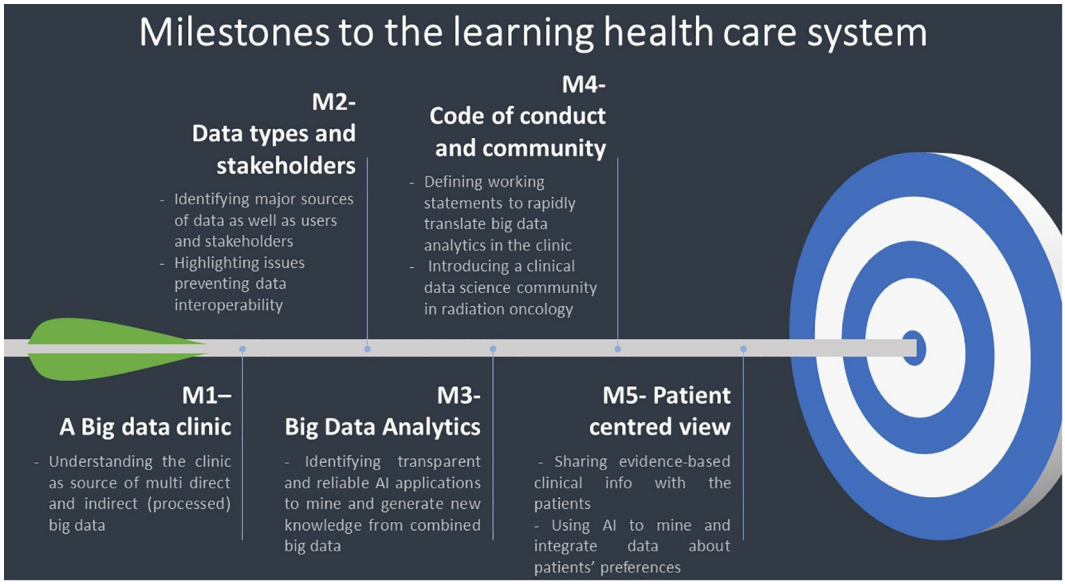


**Fig. 1.** Overview of the milestones for moving the clinic towards a learning health care system. Five milestones have been identified. Milestones 1 and 2 involve developing a deep understanding of clinical data as a source of big data and metadata and the need to involve stakeholders and users and to address issues that limit data interoperability. M3 introduces robust and collaborative AI-driven analytics for the development of clinical decision aids. M4 introduces a new clinical data science community for radiation oncology to harmonize existing initiatives and define a code of conduct. M5 introduces a patient-centered view and decision-making processes for the learning health care system.

possible to use AI as the engine to burn the data fuel. However, use of AI per se does not guarantee success. Strong transparency and robust methodology will enable meaningful applications of AI to discover new knowledge in the data. This methodology should comprise both analytics for verifying data quality, as well as methods for tackling the issues related to causal inferencing from aggregated data.

- M4: Defining working statements and a code of conduct to rapidly translate data analytics into the clinic as decision aids. To fill the gap between AI developments and their translation into the clinic as decision aids, a global effort to involve all the professional figures, stakeholders, and users identified at M2 is needed. This effort requires the creation of a clinical data science community in radiation oncology. Such a community would not replace previous efforts or already-existing focused work and task groups, but instead act as a harmonizer by defining a code of conduct and a shared vision.
- M5: A patient-centered learning health care system. The brand-new learning health care system has to be made patient-centered by a) developing AI analytics to include patient perspective data; b) improving the expandability of AI analytics, and c) using decision aids in combination with shared decision making.

## Multisource data, data types, and stakeholders (M1 and M2)

The clinic is a source of big data. Common data types include medical images, electronic health records (EHRs), and patient-reported outcomes [11]. However, the clinic also indirectly produces metadata associated with traditional data types from the algorithms that process data [7]. Examples are quantitative imaging biomarkers and radiomic data (large amounts of features extracted from medical images and analyzed using data characterization algorithms), which generate predictive or prognostic factors from source data. In Table 1, we summarize the main types of these highly variable multisource data and provide descriptions of the commonly available formats, the data owners, stakeholders, and users, and issues with or barriers to interoperability of the data.

### *AI to empower multisource data (M3)*

One of the largest issues faced when dealing with multisource big data is that the ability to process these data is beyond our human brain capacity. However, recent developments in AI algorithms have emerged as attractive and much-needed tools to empower multisource data analysis. AI and ML have created opportunities to build powerful computational facilities and a surge in data sharing, data collection, and advanced data mining algorithms. The use of ML algorithms in radiation oncology is rapidly growing; their main applications are quality assurance, organ segmentation, treatment planning, image guidance, motion tracking, and treatment response modeling. However, radiation oncology has not yet fully exploited the enormous potential of AI for analyzing multisource data that integrate variables from time-dependent sources, such as sequential quantitative imaging or genetic biomarkers. These developments could change the classical paradigms for radiotherapy by automating and optimizing clinical processes and quality control to provide decision support for personalized patient care, for instance by altering radiotherapy prescriptions and fractionation schedules. Hence, AI-based analysis of multisource data could dramatically change the way radiotherapy is approached and will likely play a central role in the future development of personalized, precision medicine.

Despite the great potential of AI, the current situation in radiation oncology is that only a small percentage of the data collected is used for decision-making in the clinic owing to several obstacles that hinder the sharing, processing, and deployment of data in the clinic. By throwing these data in the "trash," we risk losing unique insights that could radically change our clinical practice. We need to realize that human capabilities are not sufficient to process big clinical data and that clinicians need the help of AI to fully translate the large amounts of data collected in the clinic into decision-making about routine clinical practice. Artificial intelligence in clinical care is recently being recognized as a medical device by the FDA, with applications spanning from medical imaging analysis, clinical decision aids and tools to optimize patient care [80,81]. These applications not only apply to the USA, but similar evidence is seen in Europe and Asia. The FDA has developed a

**Table 1**
Summary of the main data types available for multisource data analysis.

| Data type | Description and Common format | Stakeholders | Data users | Major issues for Interoperability |
|---|---|---|---|---|
| EHRs | EHRs are computerized medical information systems that collect, store, and display patient information. | Patients<br>Hospitals | Clinicians<br>Nurses<br>Medical physicists<br>Researchers<br>Radiation therapists (RT/RTT) and dosimetrists<br>Administrative staff | Free-text entries |
| Outcome measures | Data on survival, recurrence, and toxicities are commonly found in the EHR, but when part of clinical trials, data can also be found in spreadsheets or electronic/paper-based case report forms. | Patients<br>Hospitals<br>Clinical trial units<br>Study funders<br>Regulatory agencies | Clinicians<br>Researchers<br>Research nurses<br>Clinical data coordinators | Lack of standardization Free-text entries Consistency and completeness of collected data |
| Laboratory data | Software and databases used to manage and store results from laboratory tests and pathology data | Patients<br>Hospitals | Clinicians<br>Nurses<br>Engineers<br>Researchers | Lack of standardization in acquisition and analysis<br>Non digital format<br>Image storage, management, transmission and sharing |
| Genomics | Separate databases for large-scale genomic data | Patients<br>Hospitals | Clinicians<br>Researchers | |
| Medical images | Medical images are acquired for diagnosis, staging, and treatment planning. The most common modalities include PET, CT, CBCT, MRI, and ultrasonography. Medical images are regulated by a commonly accepted standard (DICOM). | Patients<br>Hospitals | Clinicians<br>Medical physicists<br>Radiation therapists<br>Researchers | Lack of standardization in acquisition and analysis<br>Duplication of data within same Institution |
| Radiotherapy TPS & Verification Systems | Dose-volume histograms, metrics for radiation dose delivered to the tumors and organs at risk at single treatments and over the whole treatment course, are saved in the TPS.<br>Plan information<br>Dose distribution Treatment delivery data | Patients<br>Hospitals | Clinicians<br>Medical physicists<br>Radiation therapists | Institutional and clinician bias in treatment<br>Data accessibility and full use of data for data analytics |
| Patient-reported outcomes | Patient-reported outcomes, such as treatment-induced side effects, can be found in the EHR if part of standard treatment. For clinical trials, there may be various types of electronic or paper forms. | Patients<br>Hospitals | Clinicians<br>Nurses<br>Researchers | Lack of standardization<br>Free-text entries Consistency and completeness of collected data |
| National cancer registries | Population-based registries of cancer incidence, treatment, and outcomes are often recorded in national databases. | Health authorities | Clinicians<br>Researchers<br>Government | Institutional bias diagnosis, treatment, and follow-up.<br>Free-text entries. |
| Nonmedical information | Environment, income, socioeconomic status, race, ethnicity, education, housing | Local government | Government<br>Researchers | Information bias |

*Abbreviations:* EHR, electronic health record; PET, positron emission tomography; CT, computed tomography; CBCT, cone beam computed tomography; MRI, magnetic resonance imaging; TPS, treatment planning system.

complete product lifecycle for AI applications, like what was conceived for medical devices [82]. A nice example of combining big data with AI is presented in the study by Mayo et al, where a decision support system is used to improve dose delivery to spare health tissues [83].

### *A patient-centered clinical data science community in radiation oncology (M4 and M5)*

The key to success at achieving the above-mentioned milestones is the creation of a new community: one focused on clinical data science in radiation oncology. Because multiple stakeholders are involved, the problems of big data cannot be solved by only one professional discipline; instead, they will require a joint effort bringing together broad, multidisciplinary expertise and including clinicians, medical physicists, data scientists, biologists, patients, and other stakeholders. However, instead of proposing an independent community, we recommend building upon already-existing working groups and task groups that touch these professional roles. To coordinate these communities and working groups and to speed the realization of a learning health care system, we propose the development of a collection of statements and a code of conduct. Finally, we underline the importance of introducing tools that enable not only the collection and elaboration of data reporting patients' perspectives, but also a synergy between clinicians, decision aids, and patients.

### *Vision and statements*

With this position paper, we offer a basis for shifting the current paradigm in radiation oncology towards the clinic as a learning health care system. In the subsequent sections of the paper, we will elaborate on five supporting statements that are fundamental to reach the above milestones. For each statement, we have identified already-existing activities, efforts, or smaller communities that will be our main interlocutors in coordinating these efforts within the community. An overview of the statements is provided in Fig. 2.

### *Statement 1: FAIR principles for data management plans*

Over the past ten decades, numerous patient registries and databases have been established worldwide. However, few people know of their existence, let alone how to access the information in them. This lack of exposure and accessibility limits the power and, hence, the potential benefits of ML/AI tools, since a model's performance directly correlated with the amount of information it learns (trains) on, as seen in Statement 4. The need to improve the infrastructures that support the use and reuse of all these pieces of information (multisource data) in their respective silos is therefore paramount. To lay the groundwork for accomplishing this, a diverse group of stakeholders—both private and academic—jointly designed a set of principles referred to as the FAIR data principles [12]. The FAIR principles urge that all data sets should be FAIR:

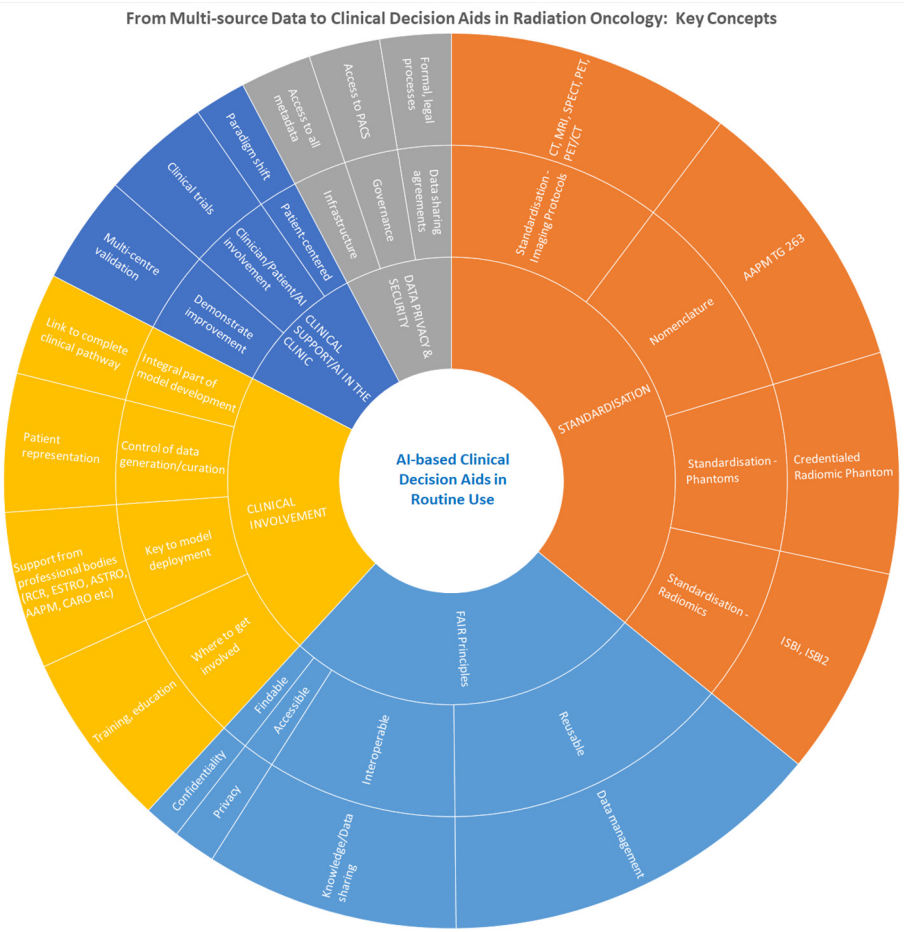


**Fig. 2.** Overview of the vision of the community, the milestones and the code of conduct statements.

findable, accessible, interoperable, and reusable, while respecting data privacy and patient confidentiality principles. The goal is to improve data (re)use by providing detailed metadata descriptions that are readable to both humans and machines, thereby making data findable, accessible, and interoperable. Therefore, for multisource data to work as intended, all data sources must adhere to the FAIR Guiding Principles and respect data privacy and confidentiality.

With the dawn of this new, data-centric era, data can be regarded as the new oil. Like crude oil, which differs in its physical and chemical characteristics from region to region, data also differs from one source or format to another. Because multisource data stem from various sources and are collected using various methods, information can be incomplete, inconsistent, biased, or imprecise. Therefore, in using multisource data, answers to the following five W-questions will facilitate the transparency of the data-generation procedure:

- Who generated the data?
- Why were the data generated?
- When were the data generated?
- Where were the data generated?
- What generated the data?

The phrase "you are what you eat" applies not only to humans but to models as well. The data science implication of this axiom means that when a model feeds (trains) on "bad" data, the resulting model is inevitably bad—in other words, less accurate—and the converse is almost always true. Therefore, no matter how complicated a model might be, it will never catch extra information in the data more effectively than a simple, explainable model would. Ensuring the integrity and quality of multisource data is even more essential if inferencing is envisaged.

In a multisource AI-driven radiotherapy system, clear guidelines need to be laid out for data stewardship. The health care organization should incorporate metadata containing data provenance at the source, and proper data lineages should be maintained at all stages of data management. Adopting health care standards that address data stewardship at the source and enable an audit trail from data acquisition to data curation will ensure better traceability for AI models.

Finally, data must be interoperable. Data from one source should be semantically as well as syntactically interpretable across different systems. Health care data structure and exchange standards like HL7 FHIR, OHDSI OMOP, and XDS provide the means to structure data in globally acknowledged and accepted formats. Use of clinical coding terminology systems and vocabulary (WHO ICD, ICF, SNOMED CT, LOINC, etc.) should be encouraged. The focus of adopting these standards should be to make the implementations of shared terminologies and vocabularies as generic as possible while permitting specificity as needed. In all cases, including those in which adoption of a health care data exchange standard is not possible, data should be sufficiently supported by metadata.

### *Data quality and effects on AI applications*

Data quality assurance is an essential exercise at all stages of data curation, although the definition of "quality" is context dependent and adopting a single measure to gauge quality is challenging. The elements of data quality are accuracy, complete-

ness, consistency, credibility, and timeliness. By accuracy, we mean that the intended value of the data is both correct and unambiguous. A very preliminary way of ensuring accuracy at the source is by using validation rules at the time of data acquisition. However, as data are shared across domains, a validation rule can itself become inconsistent, thereby increasing the chance of the data's being inaccurate. Jack Olson, in his book *Data Quality: The Accuracy Dimension*, argues that data can never be 100% accurate [13]. This is because the content of data can be validated against permissible values but not against the actual occurrence. He gives the example of how the value "brown" for eye color can be a valid entry but not an accurate one, simply because the person's eye color may actually be blue. Thus, inaccuracy can create bias in AI systems that may affect clinical outcomes. However, the very AI systems that demand quality data for better decision support may in fact contribute to improving the quality of data at the source. While AI is largely seen as a tool to extract value from data, it can also act as an instrument to add value back to the data.

## Statement 2: Standardization of methodologies

### *Standardization of radiomic algorithms*

The lack of standardization in image processing methods and quantitative radiomic feature extraction, as well as the lack of validated and verifiable reference values, is hampering the clinical implementation of quantitative radiomic imaging biomarkers [14,15]. A lack of standardized, consistent, clear, and sufficient detail in reporting radiomic features, in addition to intrinsic issues with repeatability and reproducibility [16], make radiomic findings difficult to reproduce [17] and trust. Standardization of radiomic algorithms and image processing pipelines is essential for the development of the field and should be strongly encouraged. The approach proposed by the Image Biomarker Standardization Initiative [18], which includes the standardization of a set of 174 radiomic features, the definition of a general radiomics image processing scheme, and the publication of imaging data sets and reference feature values (https://theibsi.github.io), is the most advanced effort to date. It is expected to continue and to be widely accepted and used in the future [19].

### *Standardization of image acquisition and phantoms*

To be useful, quantitative imaging biomarkers must be both repeatable and reproducible [20]. Radiomic features are affected by acquisition, reconstruction, and image preprocessing settings [21–25]. This applies to all imaging modalities. Standardization and harmonization of imaging procedures are essential requirements for the development of robust, repeatable, and valid imaging biomarkers. The standardization of imaging protocols for all imaging modalities (computed tomography [CT], magnetic resonance imaging, magnetic resonance spectroscopy, single-photon emission CT, and positron emission tomography) should be strongly encouraged within the same institution and across different institutions, as it would facilitate the interoperability of quantitative imaging biomarkers [26,27]. This would be particularly important when assessing treatment or tumor response on a large scale and as part of clinical trials [28]. The standardization of imaging protocols should be accompanied by the development of new phantoms specifically designed to address the challenge of providing reproducible reference values for textural features [22].

### *Standardization of nomenclature within the radiation oncology community*

The adoption of a standard radiation oncology nomenclature would enable and facilitate extraction and sharing of all types of data from EHRs across different institutions, states, provinces, countries, and continents. Such a standardized nomenclature would support large international clinical trials, ease collaborations across borders, and contribute to improvements in clinical practice and patient care [29]. Moving forward, it is essential that new clinical trial protocols use standardized nomenclatures for capturing their data. The question remains which standard nomenclature should be used. At present, the standardized nomenclature for radiotherapy proposed by the American Association of Physicists in Medicine (AAPM) Task Group (TG) 263 [30] seems to be the most likely standard to become accepted and widely used. The Global Quality Assurance of Radiation Therapy Clinical Trials Harmonization Group (https://rtqaharmonization.org) has recently unified the contouring of organs at risk by compiling, in line with AAPM TG 263 and the American Society for Radiation Oncology (ASTRO), guidance for delineation and a standard nomenclature for integration into clinical trial protocols [29].

### *Statement 3: Privacy-preserving collaborative big data infrastructures*

FAIR data principles ensure that data are syntactically and semantically interoperable, thereby promoting seamless data sharing among health care providers. While sharing patient-level data for better decision making is important, protecting patient privacy is essential. Ethical, legal, and societal issues regarding data sharing bar hospitals and clinics from sharing data. When there is too little data shared, ML and AI technologies starve themselves with little or no data.

However, if we cannot bring data to the algorithms, it is possible to send algorithms to the data. Infrastructures built around these data silos can connect and provide a way to send algorithms to the data sources. While the data stays well protected within the jurisdiction of the healthcare provider or the patient themselves, the algorithms via the infrastructure can fetch results. This way privacy of patient data is protected while at the same time, research is promoted. This section explores big data infrastructures that enables privacy preserving collaborative research.

We talk about two types of infrastructure: centralized collaborative big data infrastructures and federated big data infrastructures. In centralized infrastructures, different hospitals and healthcare providers enter a collaboration and upload patient data to a secured centralized repository [31]. Researchers can use data from the repository either train their algorithms and perform analysis. Additionally, the centralized repository can also provide a compute environment where algorithms can be sent and computation performed. A researcher initiating a data analysis process will have no direct access to the data and can only retrieve the result of the analysis. However, it is important to mention that the data will still be located outside individual hospitals. As such, patient's consent in sharing data to a central repository and use for secondary purposes needs to be addressed properly. Another initiative, Informatics for Integrating Biology and the Bedside (I2B2), aims to integrate data from different biomedical disciplines and to deliver these data to researchers. I2B2 provides tools and frameworks for merging and linking genomic and biological data to clinical data in a health data warehouse [32,33]. Similarly, the HMO Cancer Research Network connects more than 11 million patient records from 14 health care providers in a virtual data warehouse [34].

Federated big data infrastructure emphasizes keeping the data at the source while pushing the analytics to the source. Each hospital maintains a local data repository to which researchers can

send their algorithms and from which they can fetch results. In a collaborative environment, each hospital would act as an individual data provider, generating sets of results that can then be aggregated to obtain global results. An example is the Personal Health Train (PHT) [35]. PHT shifts the focus from sharing data to sharing algorithms to the source of the data, essentially within the jurisdictional environment of the hospital. A researcher using the infrastructure is agnostic about the data schema and distribution at the source and as such relies heavily upon the FAIR data principles. Each hospital hosts a data station containing FAIR data, and provides a computation environment for the train (metaphor for package containing algorithms and data retrieval query). PHT is platform independent and the researcher can autonomously choose the technology for implementing the algorithm (e.g., Python, R, Matlab, Java). The communication between the data stations and researchers occurs through a secured and centralized message broker. Study showed that PHT is scalable so that federated, privacy-preserving analyses involving many thousands of patients can be conducted [36–39].

Another example is DataSHIELD, a collaborative and privacy-preserving data analysis environment connecting multiple hospitals. This infrastructure enables researchers to send algorithms to the data without having to retrieve data locally. Unlike PHT, DataSHIELD sends algorithms packaged in the R statistical programming environment to an Opal database hosted at each hospital [40,41]. PCORnet is a network of several clinical research institutes that supports pragmatic trials and comparative effectiveness research across one or several of the participating institutes [42]. More recently, MedCO provided a privacy preserving federated data analysis platform [43,44]. MedCO focusses on keeping data at the source and provides multi-party homomorphic encryption to all data sources, providing an additional layer of security and privacy. OpenSAFELY initiative in the UK enables trusted analysts to run large scale computation on pseudonymised patient records inside environments managed by electronics health records software company [45].

It is important to mention that creating a collaborative environment connecting many different hospitals and clinics while preserving patient data privacy is a multifaceted challenge. Keeping data at the source may not be sufficient when the amount of data is small. The infrastructure needs to be adaptive, flexible, scalable, and secure. It should be transparent to the patient and to society in general to maintain trust. A balance between respecting privacy and creating maximal societal benefits needs to be ensured. While the data need to be FAIR, the analysis should be fair, accurate, confidential, and transparent (FACT).

Finally, it is important to acknowledge the important role that legal and professional bodies have in ensuring that there are appropriate legislative frameworks in place that public and commercial stakeholders can adopt and follow. At the heart of any centralized or federated multi-source data science initiative in radiation oncology must therefore be full engagement with regional data protection regulations such as those set out in the European Union General Data Protection Regulations (GDPR), which are now the cornerstone of data sharing initiatives in Europe.

## Statement 4: Involvement of clinicians in the data science community

*The clinician is an essential member of the data science community*

For decades, decisions in medicine have been based on clinical guidelines that are carefully developed and based on the highest-level evidence from large randomized controlled trials. Recently, individualized approaches to treatment have become an increasingly compelling research area. This trend is particularly prominent in oncology, where the discovery of new prognostic and predictive factors including viral infection, hypoxia signatures, driver mutations, and many others has enabled more precise treatment selection to match the characteristics of each patient and tumor [46]. However, greater personalization makes generating level-1 evidence difficult, if not impossible, as the number of matched patients in each subgroup decreases, ultimately coming down to a single individual.

Predictive modelling using AI and multisource data offers a way to address this conundrum. The multidisciplinary field of clinical ML attracts researchers from diverse disciplines, including clinicians, computer scientists, medical physicists, and biostatisticians. Unfortunately, these different research communities often work in isolation, with separate jargon, specialized publications, and hermetic knowledge. Often, groups of scientists access partial data but lack the full clinical context or a complete understanding of the limitations of the data (e.g., embedded treatment effects) because their expertise may not lie in the clinical domain. Thus, to overcome these obstacles to the clinical implementation of AI tools, close cooperation among specialties is mandatory.

The role of clinician is—and will remain—crucial to the clear definition of a relevant clinical problem and the identification of appropriate prediction targets, e.g. biologically relevant mechanisms (hypoxia, gene expression) or cancer- and treatment-specific outcomes like survival, relapse, and treatment toxicity. Clinicians must be involved in both baseline data review and model generation to detect garbage-in garbage-out situations arising from malformed or poorly designed models. Data science approaches often highlight previously known clinical factors that are already used clinically to select patients for treatment, which can confound the interpretation of outcomes data. AI models that detect and latch onto novel details of individual patient cases are required so that these approaches can supplement, rather than reiterate, current clinical practice.

Finally, and perhaps most importantly, clinicians will be the end users of any deployed multisource-based ML tool; they will be the ones to interpret the output of such tools to make responsible decisions about patient care and provide feedback to improve the database. Most clinicians trust their own experience and intuitions developed over years of practice and might find it difficult to rely on a model's prediction, especially if they do not understand the reasoning behind it. Thus, close collaboration between algorithm developers and clinicians is necessary to create models that clinicians can trust. For example, many studies focus on the interpretability of data science tools, the lack of which is one of the key obstacles to the wide clinical adoption of predictive models. Ultimately, the clinician is a critical bridge between the patient and the treatment team. This bridge is even more important as it allows patient preferences to be integrated into the planning process and may in turn change the way the ML/AI model is deployed. One example would be the development of multiple pareto-optimal treatment plans that integrate patient preferences into the final decision-making to select the outcomes most valuable to the patient. Furthermore, this integration stands to bring about a shift in the training of medical professionals in radiotherapy; for example, knowledge about how AI tools work and how to use them in personalized medicine will replace skills in, for instance, delineation of organs at risk [47,48].

*The clinician should be involved in data generation and data curation*

The constantly increasing power of computers has made collecting and analyzing large amounts of data relatively easy and allows the building of searchable and expandable databases for research, modeling, and generation of new hypotheses. Often these data sets are used only once and then discarded or kept internally

by the originating institution, which limits the power and capabilities of ML/AI for using these data sets. Therefore, the clinician must play a crucial role in designing and maintaining dedicated databases for ML model training. Choosing the relevant features for a clinical problem, considering the defined outcomes, and identifying possible biases are all still in the domain of clinical expertise [49]. For example, clinical decisions about radical versus palliative approaches in localized advanced head and neck tumors are to some extent affected by the clinician's personal experience of successful treatments.

The data must be understood by the clinician before any modeling can take place, and clinicians are more willing to use models if the input features are aligned with evidence-based practice [50]. An example of such a model based on data routinely collected and updated every hour from electronic records of intensive care unit patients has been described by Thorsen-Meyer et al. [51]. However, models should not only use patient characteristics known to be important, but also uncover previously unknown associations. Here again, the clinician can help distinguish a truly novel predictive variable from biases, data set artefacts, or confounding factors. Additionally, clinicians can easily provide feedback in case of a false prediction and follow up with a misidentified patient, especially in cases with an unusual trajectory or medical history [50].

To enable searchable and expandable databases for modeling, research, and generation of new hypotheses, all data sets must be shared (adhering to the FAIR principles described above) and accessible for other institutions to use for training, validation, or additional analysis. In this respect, several approaches may help clinicians become more engaged in data collection and curation. The most important is integration of this process into standard clinical workflows and standard operating procedures. This will allow clinical data, treatment planning information, diagnostic imaging, and outcomes to be seamlessly collected as part of clinical practice. Such integrated data collection will incentivize physicians to contribute high-quality data on all patients. Even simple synoptic endpoint collection can provide a powerful backbone for large data set generation. When leveraged properly, rapid learning and automatic data collection will be crucial for clinicians in this era of fast progress in new therapeutics as well as technology and information overload. Crucially, simple methods to share data safely are necessary. If such methods are implemented properly, clinicians can derive visible benefits from sharing their efforts, which will convince them to contribute willingly. Examples of this could include simple quality assurance and second opinions, rapid outcome estimates, speed up evaluation of new technologies or automated workflow acceleration. Moreover, there is evidence that publication where associated data are shared in accordance with FAIR principles are cited more frequently [52], creating an additional incentive for the clinicians.

Publication with open FAIR data available for readers should also be promoted as such by journal editors, meaning safe repositories have to be provided for authors. Building a culture of data sharing, not only within research institutions, but also hospitals and biotechnology companies is the most important challenge for the future. Policy makers will have an important role to play in creating a global structured policy of data sharing. A good example is the Final Report and Action Plan from the European Commission's Expert Group on FAIR data "Turning FAIR into reality", which paves a way to build infrastructure, recognition of obstacles and benefits and creates incentives for European research institutions to participate in data sharing. General concept of transparency and data sharing concerns also pharmaceutical companies, where process of sharing data obtained in clinical trials is still in it's infancy, mostly due to lack of policies of data sharing. A step forward to change this situation has been done by Miller et al who developed dedicated score -The Good Pharma Scorecard - to monitor of transparency in process of sponsored research and data sharing process. To have such policy in place is very important for industry itself due to increasing pressure of external stakeholders including patients and clinicians to speed up gathering knowledge and evidence by transparent collection of data [53]. Another initiative with potential to facilitate routine clinical data collection and sharing is the Real World Data (RWD) framework developed and proposed by the US Food and Drug Administration. The aim is to collect post-approval data from electronic health records (EHR) and other clinical data repositories to generate Real World Evidence (RWE) of risk and benefits of currently approved products. Such approach promotes shared learning and encourages stakeholders to use RWE in their research, as well as to use common data models, unified terminology and data encoding for different sources.

Clinicians are crucial not only for defining clinical problems and relevant outcomes, but also for the dynamic expansion and adaptation of databases, accounting for biases and unusual patient trajectories. We cannot forget that all this work should be focused on individual patient benefits but may also provide population-level benefits if resources are constrained.

## The clinician should be involved in all steps of model development and deployment

An increasing number of clinicians is interested in cooperating with AI/ML scientists [2]. However, others remain reluctant and do not yet trust AI models, preventing deployment of these tools in the clinic [54]. One commonly stated reason for this distrust is unsatisfactory predictive performance, especially of prognostic models. However, what level of predictive accuracy is clinically acceptable is unclear. Moreover, accurate prediction of complex endpoints like overall survival is very difficult, even for an experienced clinician. A related question is how much better than a human a model must perform to be considered useful, especially if the human baseline is low. Many published models perform well for well-defined, simple outcomes such as prediction of local control or extracapsular extension in involved head and neck lymph nodes [55,56].

The most-used metric to measure predictive performance on a binary classification task is the area under the receiver operating characteristic curve. To generate predictions for new data, a single operating point needs to be selected. The standard approach is to give equal weights to specificity and sensitivity, but in many real clinical situations the cost of error may vary and may differently affect patients' outcomes. For example, a model that incorrectly suggests a patient will have a very high risk of toxicity may deprive the patient of the possibility to receive curative therapy. Only by knowing the holistic clinical picture can one decide how to define the expected parameters of models, including the desired specificity and sensitivity of predictions. The role of the clinician in defining these optimal operating points is crucial. Clinicians should also be involved in the model review and development to prevent 'blind alleys' and other problems.

The accuracy of a model is said to be inversely proportional to its explainability [57]. The trade-off between explainability and accuracy is still an unsolved problem, as the best-performing models based on deep learning are "black boxes"—synonymous with a lack of transparency and understanding—and are the least explainable. The results of successful attempts to improve the explainability of neural networks in health care were published by Yang [58].

However, even an explainable model may not be clinically applicable/actionable if the output is not additive or meshed into existing clinical approaches. It is vital that all models grow from and additively expand to fit existing clinical knowledge. Rediscov-

ering, for instance, that a tumor's size predicts patient outcomes can be avoided by incorporating clinical knowledge early in the problem domain and establishing target areas to enhance [59]. Hence, ideally, models should be organically integrated with practice to provide continuous feedback, allowing clinicians to monitor and understand their effects and limitations.

Any model, no matter how accurate and interpretable in the development stage, needs to be thoroughly validated in a controlled trial before clinicians can trust it.

Validating AI models will require new trial design, especially in terms of endpoints and evaluation criteria. Nagendran et al found only two completed and published randomized controlled trials (RCTs), of AI algorithms in gastroenterology and ophthalmology [60,61,62], while the FDA has approved more than 16 deep learning algorithms in ophthalmology, radiology and cardiology. The most often used endpoint in such studies is the performance of AI/ML tools on some metric (e.g. receiver operating characteristic) versus human experts. However, ever if the AI outperforms human experts, it is not clear whether replacing the clinician's experience by an automated algorithm translates into benefit for patients in real-world use. Additionally, many clinical tasks have no well-defined ground truth, making an objective, direct comparison difficult. Endpoints such as performance of clinician supported by an algorithm, improved workflow efficiency or time and financial savings could better reflect the real-world impact of clinical AI. Another challenge in model validation is how to evaluate and update models under data distribution shift, for example when the treatment guidelines or patient population change. Policy changes, such as the regulatory framework for AI/ML Software as a Medical Device recently proposed by the FDA, can help build clinician trust in AI models by enforce transparency and continuous performance monitoring as part of the approval process.

Good performance in a retrospective test set is not sufficient; for example, a model can perform well in data from the institution where it was developed in but fail when tested on data from a different hospital, despite seemingly identical input data and targets. This can happen because of covariate shift—a change in the distribution of input data between different institutions, such as different CT scanner models and protocols—and because of unobserved confounders that have real impact on outcomes, like the quality of the health care system, the provision of supportive care, and even the approach of individual clinician treating the patient. Continuously reporting model accuracy at each deployed institution with individual physicians' feedback will be critical to maintaining and monitoring models and will help to ensure that physicians maintain confidence in the approach.

To fully take advantage of this opportunity, clinician involvement is necessary at every step—from formulating the problem, through the selection of appropriate input data and prediction targets, to model validation in a prospective clinical trial.

Data collection, curation and model development will require both financial and human resources. Recent rapid progress in AI has led to increased public interest and expectations of many stakeholders, including patients, regulators and governments. Resource allocation to AI research and implementation — both on central level, like EC and local institutional boards — is therefore expected to increase in the near future.

Moreover, systemic solutions and infrastructure like automatic data collection, rapid learning systems, standardized format of collection, easy retrieval and seamless integration with the clinical workflow will reduce the load on clinicians and make the shift smooth and effective.

To be trustworthy and actionable, the predictions need to be interpretable, although the balance between interpretability and accuracy is still a subject of research. Interpretation and prospective validation, with identification of biases and confounding factors, preferably in a clinical trial, will increase clinician trust and allow for deployment of ML tools in the clinic.

## Statement 5: From AI to a patient-oriented view

Realizing AI's potential in clinical practice calls for a patient-centered perspective in the development, design, and implementation of AI tools. First, AI tools must be oriented towards addressing clinical questions that matter to patients. Second, the output of AI tools must be integrated into decision aids that present relevant clinical information in a format that is clear, understandable, and actionable [63]. Finally, AI-based model outputs must be explainable and be combinable with patient preferences in a shared decision-making process.

### *AI to enable retrieving patient data*

Orienting AI towards the patient perspective involves determining what is relevant to patients in clinical terms as well as how they experience their condition. Certain aspects of these data are routinely collected as patient-reported outcome measures (PROMs) and stored in patients' EHRs. It is unclear to what extent PROMs are analyzed and used [64], particularly due to doctors' lack of time, resources, and expertise [65]. This presents an opportunity for AI/ML techniques to harness and analyze EHR data and PROMs to identify relationships between treatments and patient-relevant outcomes [66].

Aside from patient data stored in hospital records, increasing amounts of data are also generated externally, as more patients, through use of the internet, are taking an active role in managing their health decisions. Consequently, the role of patient organizations is also shifting from providing information to building platforms and online communities in which patients can share experiences and knowledge that go beyond the data captured in PROMs. For instance, the patient organization PatientsLikeMe, an online community that connects over 650,000 patients across nearly 3000 health conditions, is based on the principle of seeing the patient as a person rather than as a disease, and accordingly collects data on patients' definitions of health and outcomes. PatientsLikeMe is actively involved in AI initiatives to generate insights from this vast and rich data source [67]. Initiatives such as these can help target AI tools to clinically relevant questions.

### *Clinical decision aids for doctors and patients*

AI tools must be built into decision aids that support shared deliberation and decision-making processes between patients and doctors. Decision aids provide a means to inform patients about their conditions, reflect on their own values, and weigh their treatment options in the context of their preferences. Poor design is one of the main factors that hinders decision aid uptake [68]. Implementation of AI-enhanced decision aids is more likely to succeed when development follows a user-centered design process that takes into account end-users' contexts, needs, goals, and decision-making [69]. We have previously emphasized the importance of including doctors in the development process, as well as the patient focus in determining relevant clinical questions. Once developed, it is critical that decision aids be embedded into the clinical workflow, for instance through integration into the hospital's EHR system, to minimize the amount of time and manual work required in entering a patient's data [70]. In addition, integrating the decision aid into the clinical consultation itself can pave the way for data-driven shared decision-making in which AI-based recommendations are discussed in the context of the doctor's clinical knowledge/experience and the patient's preferences.

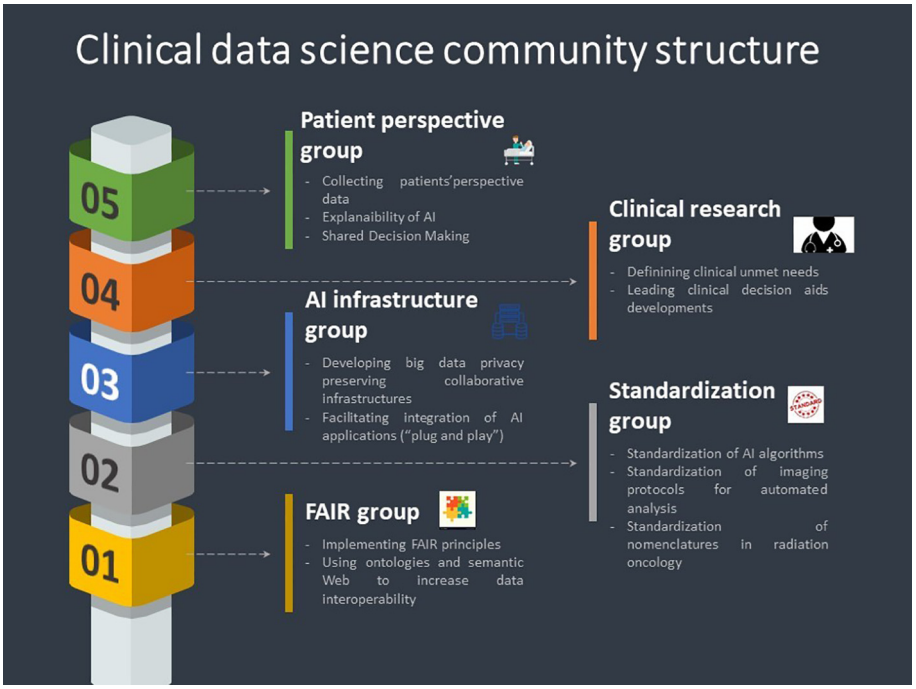


**Fig. 3.** Overview of the structure of the new data science community in radiation oncology. In line with the presented milestones, five major working groups are identified: FAIR principles group (M1); standardization group (M 2); AI applications and big data collaborative infrastructures group (M3); clinical research and definition of unmet clinical needs group (M4); and patient-centered decision aids and shared decision-making group (M5). The role of each group is to coordinate with similar existing task forces and working groups from European and American societies active in radiation oncology.

*Bringing AI into the consultation*

Traditional shared decision-making consists of a two-way information exchange between doctors and patients; doctors share their clinical expertise on the treatment options and their benefits and risks, and patients share their values and preferences [71]. When both sides understand each others' perspectives, they can deliberate on the available options from a common ground and make a choice that is rooted in the best clinical evidence as well as the patient's individual circumstances. The introduction of AI-based decision aids represents a third angle from which treatment information can be personalized according to the patient's individual characteristics [72]. The "black box" nature of certain AI tools may make it challenging for the doctor to articulate the reasoning behind a given diagnosis or treatment recommendation [73], so the explainability of the AI tool is a crucial factor in ensuring that decision-making does not shift back towards paternalism [74]. This includes interpreting AI model outputs, such as risk estimates or prognoses, and communicating them in a way that is understandable to patients [75].

Moreover, little is known about the patient perspective on receiving AI-supported care. Preliminary findings from skin cancer screening suggest that patients are open to the potential of AI in improving care quality as long as it functions as decision *support* rather than decision replacement [76] and the doctor-patient dyad is maintained [58]. More research is needed to understand the shift in roles and responsibilities that accompanies AI implementation and how to use AI models to empower patients. In particular, the perspectives of social scientists and anthropologists are needed to bring AI into alignment with human decision-making [77].

## Discussion

In this paper, we identified the barriers that are currently limiting the adoption of big data analytics in the clinic toward the development of a learning health care system. The main barrier is the ability to handle the large amount of data and metadata produced in the clinic as result of daily clinical and research activities. As we discussed, this big data involves different stakeholders and users and presents significant interoperability issues. We therefore identified the need to analyze the major sources of multisource data and metadata and the limits on their interoperability (milestones M1 and M2).

We next discussed how, when it becomes capable of fully connecting these sparse multisource data, AI will provide powerful analytics to develop data-driven clinical decision aids (milestone M3). However, because of the variety of data types and stakeholders involved, multiple professionals need to be involved and coordinated. For this reason, we presented the need to define a clinical data science community in radiation oncology, to act as harmonizer of the different professional figures with a common vision sustained by a code of conduct and working statements and with a strong orientation toward patient-centered care (milestone M5). This community will not be an independent actor, but will build upon already existing communities, efforts, and working groups. Clinicians will have a prominent leading role both in determining the requirements of the technical developments and in continuously interfacing with the more technical professionals. This is meant to guarantee that technical developments are in line with unmet clinical needs. We envision this community to be fully embedded within the major global radiation oncology societies, such as ESTRO, ASTRO, CARO, AAPM, EFOMP, RANZCR and FARO and to include patient societies such as CRC and PatientsLikeMe.

Our future activities will be to engage with the above-mentioned societies to define working groups, as briefly depicted in Fig. 3.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We thank the participants of the 3rd ESTRO Physics Workshop in Budapest, 2019, for their involvement in the discussions that lead to this article.

We thank ESTRO for supporting us in the organization of the workshop.

Editorial support was provided by Amy Ninetto of Scientific Editing Services, Research Medical Library, The University of Texas MD Anderson Cancer Center.

## References

[1] Hofer IS, Halperin E, Cannesson M. Opening the black box: understanding the science behind big data and predictive analytics. Anesth Analg 2018;127:1139–43.
[2] Cirillo D, Valencia A. Big data analytics for personalized medicine. Curr Opin Biotechnol 2019;58:161–7.
[3] Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Aff 2014;33:1123–31.
[4] Krumholz HM. Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. Health Aff 2014;33:1163–70.
[5] Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. J Family Med Prim Care 2019;8:2328. https://doi.org/10.4103/jfmpc.jfmpc_440_19.
[6] Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer 2018;18:500–10.
[7] National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease 2011 National Academies Press (US) Washington (DC).
[8] Budrionis A, Bellika JG. The Learning Healthcare System: Where are we now? A systematic review. J Biomed Inform 2016;64:87–92.
[9] Stuart EA. Matching methods for causal inference: A review and a look forward. Statist Sci 2010;25:1–21.
[10] Bareinboim E, Pearl J. Causal inference and the data-fusion problem. PNAS 2016;113:7345–52.
[11] Traverso A, Dankers FJWM, Wee L, van Kuijk SMJ. Data at Scale. In: Kubben P, Dumontier M, Dekker A, editors. Fundamentals of Clinical Data Science. Cham: Springer International Publishing; 2019. p. 11–7. https://doi.org/10.1007/978-3-319-99713-1_2.
[12] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3. https://doi.org/10.1038/sdata.2016.18.
[13] Olson JE. Data quality: the accuracy dimension. San Francisco: Morgan Kaufmann; 2003.
[14] Yip SSF, Aerts HJWL. Applications and limitations of radiomics. Phys Med Biol 2016;61:R150–66.
[15] Buvat I, Orlhac F, Soussan M. Tumor texture analysis in PET: Where do we stand? J Nucl Med 2015;56:1642–4.
[16] Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: a systematic review. Int J Radiat Oncol Biol Phys 2018;102:1143–58.
[17] Zwanenburg A. Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. Eur J Nucl Med Mol Imaging 2019;46:2638–55.
[18] Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology 2020;295:328–38.
[19] Depeursinge A, Andrearczyk V, Whybra P, van Griethuysen J, Müller H, Schaer R, et al. Standardised convolutional filtering for radiomics. ArXiv:200605470 [Cs, Eess] 2020.
[20] Morin O, Vallières M, Jochems A, Woodruff HC, Valdes G, Braunstein SE, et al. A deep look into the future of quantitative imaging in oncology: a statement of working principles and proposal for change. Int J Radiat Oncol Biol Phys 2018;102:1074–82. https://doi.org/10.1016/j.ijrobp.2018.08.032.
[21] Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. Magn Reson Imaging 2004;22:81–91. https://doi.org/10.1016/j.mri.2003.09.001.
[22] Mackin D, Ger R, Dodge C, Fave X, Chi P-C, Zhang L, Yang J, Bache S, Dodge C, Jones AK, Court L. Effect of tube current on computed tomography radiomic features. Sci Rep 2018;8. https://doi.org/10.1038/s41598-018-20713-6.
[23] Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, Rodriguez-Rivera E, Dodge C, Jones AK, Court L. Measuring Computed Tomography Scanner Variability of Radiomics Features:. Invest Radiol 2015;50:757–65.
[24] Berenguer R, Pastor-Juan MDR, Canales-Vázquez J, Castro-García M, Villas MV, Mansilla Legorburo F, Sabater S. Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters. Radiology 2018;288:407–15.
[25] Whybra P, Parkinson C, Foley K, Staffurth J, Spezi E. Assessing radiomic feature robustness to interpolation in 18F-FDG PET imaging. Sci Rep 2019;9. https://doi.org/10.1038/s41598-019-46030-0.
[26] Ger RB, Zhou S, Chi P-C, Lee HJ, Layman RR, Jones AK, et al. Comprehensive investigation on controlling for CT imaging variabilities in radiomics studies. Sci Rep 2018;8. https://doi.org/10.1038/s41598-018-31509-z.
[27] Branco LRF, Ger RB, Mackin DS, Zhou S, Court LE, Layman RR. Technical Note: Proof of concept for radiomics-based quality assurance for computed tomography. J Appl Clin Med Phys 2019;20:199–205.
[28] Moran J, Molineu A, Kruse J, Oldham M, Jeraj R, Galvin J, et al. Guidance for the physics aspects of clinical trials. AAPM 2018.
[29] Melidis C, Bosch WR, Izewska J, Fidarova E, Zubizarreta E, Ulin K, et al. Global harmonization of quality assurance naming conventions in radiation therapy clinical trials. Int J Radiat Oncol Biol Phys 2014;90:1242–9.
[30] Mayo CS, Moran JM, Bosch W, Xiao Y, McNutt T, Popple R, et al. American Association of Physicists in Medicine Task Group 263: Standardizing Nomenclatures in Radiation Oncology. Int J Radiat Oncol Biol Phys 2018;100:1057–66.
[31] Rangarajan S, Liu H, Wang H, Wang C-L. Scalable architecture for personalized healthcare service recommendation using big data lake. In: Beheshti A, Hashmi M, Dong H, Zhang WE, editors. Service research and innovation. Cham: Springer International Publishing; 2018. p. 65–79. https://doi.org/10.1007/978-3-319-76587-7_5.
[32] Welcome | i2b2 Research Data Warehouse n.d. https://i2b2.cchmc.org/ (accessed March 13, 2019).
[33] Boussadi A, Zapletal E. A fast healthcare interoperability resources (FHIR) layer implemented over i2b2. BMC Med Inf Decis Making 2017;17. https://doi.org/10.1186/s12911-017-0513-6.
[34] Cancer Research Network (CRN) n.d. https://healthcaredelivery.cancer.gov/crn/ (accessed July 7, 2020).
[35] Beyan O, Choudhury A, van Soest J, Kohlbacher O, Zimmermann L, Stenzhorn H, Karim MR, Dumontier M, Decker S, da Silva Santos LOB, Dekker A. Distributed analytics on sensitive medical data: the personal health train. Data Intellegence 2020;2:96–107.
[36] Deist TM, Jochems A, van Soest J, Nalbantov G, Oberije C, Walsh S, et al. Infrastructure and distributed learning methodology for privacy-preserving multi-centric rapid learning health care: euroCAT. Clin Transl Radiat Oncol 2017;4:24–31.
[37] Deist TM, Dankers FJWM, Ojha P, Scott Marshall M, Janssen T, et al. Distributed learning on 20 000+ lung cancer patients – The Personal Health Train. Radiother Oncol 2020;144:189–200.
[38] Shi Z, Zhovannik I, Traverso A, Dankers FJWM, Deist TM, Kalendralis P, et al. Distributed radiomics as a signature validation study using the Personal Health Train infrastructure. Sci Data 2019;6. https://doi.org/10.1038/s41597-019-0241-0.
[39] Choudhury A, van Soest J, Nayak S, Dekker A. Personal health train on FHIR: A privacy preserving federated approach for analyzing FAIR data in healthcare. In: Bhattacharjee A, Borgohain SKR, Soni B, Verma G, Gao X-Z, editors. Machine learning, image processing, network security and data sciences. Singapore: Springer; 2020. p. 85–95. https://doi.org/10.1007/978-981-15-6315-7_7.
[40] Gaye A, Marcon Y, Isaeva J, LaFlamme P, Turner A, Jones EM, et al. DataSHIELD: taking the analysis to the data, not the data to the analysis. Int J Epidemiol 2014;43:1929–44. https://doi.org/10.1093/ije/dyu188.
[41] DataSHIELD – New Directions and Dimensions n.d. https://datascience.codata.org/articles/10.5334/dsj-2017-021/ (accessed July 9, 2019).
[42] PCORnet, the National Patient-Centered Clinical Research Network. PCORnet n.d. https://pcornet.org/ (accessed March 18, 2019).
[43] MedCo | Collective protection of medical data n.d. https://medco.epfl.ch/ (accessed September 15, 2020).
[44] Sav S, Pyrgelis A, Troncoso-Pastoriza JR, Froelicher D, Bossuat J-P, Sousa JS, et al. POSEIDON:Privacy-Preserving Federated Neural Network Learning. ArXiv:200900349 [Cs] 2020.
[45] OpenSAFELY n.d. https://opensafely.org/ (accessed September 15, 2020).
[46] Baumann M, Krause M, Overgaard J, Debus J, Bentzen SM, Daartz J, et al. Radiation oncology in the era of precision medicine. Nat Rev Cancer 2016;16:234–49.
[47] Lustberg T, van Soest J, Gooding M, Peressutti D, Aljabar P, van der Stoep J, et al. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. Radiother Oncol 2018;126:312–7.
[48] Thompson RF, Valdes G, Fuller CD, Carpenter CM, Morin O, Aneja S, et al. Artificial intelligence in radiation oncology: A specialty-wide disruptive transformation? Radiother Oncol 2018;129:421–6.
[49] El Naqa I, Ruan D, Valdes G, Dekker A, McNutt T, Ge Y, Wu QJ, Oh JH, Thor M, Smith W, Rao A, Fuller C, Xiao Y, Manion F, Schipper M, Mayo C, Moran JM, Ten Haken R. Machine learning and modeling: Data, validation, communication challenges. Med Phys 2018;45:e834–40.
[50] Tonekaboni S, Joshi S, McCradden MD, Goldenberg A. What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use. ArXiv:190505134 [Cs, Stat] 2019
[51] Thorsen-Meyer H-C, Nielsen AB, Nielsen AP, Kaas-Hansen BS, Toft P, Schierbeck J, et al. Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. Lancet Digital Health 2020;2:e179–91.

[52] Piwowar HA, Day RS, Fridsma DB. Sharing detailed research data is associated with increased citation rate. PLoS ONE 2007;2. https://doi.org/10.1371/journal.pone.0000308.
[53] Miller J, Ross JS, Wilenzick M, Mello MM. Sharing of clinical trial data and results reporting practices among large pharmaceutical companies: cross sectional descriptive study and pilot of a tool to improve company practices. BMJ 2019;366:. https://doi.org/10.1136/bmj.l4217l4217.
[54] Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. WIREs Data Mining Knowl Discov 2019;9. https://doi.org/10.1002/widm.2019.9.issue-410.1002/widm.1312.
[55] Boldrini L, Cusumano D, Chiloiro G, Casà C, Masciocchi C, Lenkowicz J, et al. Delta radiomics for rectal cancer response prediction with hybrid 0.35 T magnetic resonance-guided radiotherapy (MRgRT): a hypothesis-generating study for an innovative personalized medicine approach. Radiol Med 2019;124:145–53. https://doi.org/10.1007/s11547-018-0951-y.
[56] Kann BH, Hicks DF, Payabvash S, Mahajan A, Du J, Gupta V, Park HS, Yu JB, Yarbrough WG, Burtness BA, Husain ZA, Aneja S. Multi-Institutional Validation of Deep Learning for Pretreatment Identification of Extranodal Extension in Head and Neck Squamous Cell Carcinoma. JCO 2020;38:1304–11.
[57] Luo Yi, Tseng H-H, Cui S, Wei L, Ten Haken RK, El Naqa I. Balancing accuracy and interpretability of machine learning approaches for radiation treatment outcomes modeling. BJR|Open 2019;1:20190021. https://doi.org/10.1259/bjro.20190021.
[58] Yang Y, Tresp V, Wunderle M, Fasching PA. Explaining Therapy Predictions with Layer-Wise Relevance Propagation in Neural Networks. 2018 IEEE International Conference on Healthcare Informatics (ICHI), New York, NY: IEEE; 2018, p. 152–62. https://doi.org/10.1109/ICHI.2018.00025.
[59] Welch ML, McIntosh C, Haibe-Kains B, Milosevic MF, Wee L, Dekker A, Huang SH, Purdie TG, O'Sullivan B, Aerts HJWL, Jaffray DA. Vulnerabilities of radiomic signature development: The need for safeguards. Radiother Oncol 2019;130:2–9.
[60] Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ 2020;368. https://doi.org/10.1136/bmj.m689.
[61] Lin H, Li R, Liu Z, Chen J, Yang Y, Chen H, et al. Diagnostic efficacy and therapeutic decision-making capacity of an artificial intelligence platform for childhood cataracts in eye clinics: a multicentre randomized controlled trial. EClinicalMedicine 2019;9:52–9.
[62] Wang Pu, Berzin TM, Glissen Brown JR, Bharadwaj S, Becq A, Xiao X, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut 2019;68:1813–9.
[63] Middleton B, Sittig DF, Wright A. Clinical decision support: a 25 year retrospective and a 25 year vision. Yearb Med Inform 2016;25:S103–16.
[64] Barr PJ, Berry SA, Gozansky WS, McQuillan DB, Ross C, Carmichael D, et al. No date for the PROM: the association between patient-reported health events and clinical coding in primary care. J Patient Rep Outcomes 2020;4. https://doi.org/10.1186/s41687-020-0183-5.
[65] Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on patientslikeme. J Med Internet Res 2010;12. https://doi.org/10.2196/jmir.1549.
[66] Should Artificial Intelligence Augment Medical Decision Making? The Case for an Autonomy Algorithm. AMA Journal of Ethics 2018;20:E902-910. https://doi.org/10.1001/amajethics.2018.902.
[67] Okun S, Wicks P. DigitalMe: a journey towards personalized health and thriving. Biomed Eng Online 2018;17. https://doi.org/10.1186/s12938-018-0553-x.
[68] Agoritsas T, Heen AF, Brandt L, Alonso-Coello P, Kristiansen A, Akl EA, et al. Decision aids that really promote shared decision making: the pace quickens. BMJ 2015:g7624. https://doi.org/10.1136/bmj.g7624.
[69] Witteman HO, Dansokho SC, Colquhoun H, Coulter A, Dugas M, Fagerlin A, et al. User-centered design and the development of patient decision aids: protocol for a systematic review. Syst Rev 2015;4. https://doi.org/10.1186/2046-4053-4-11.
[70] Lenert L, Dunlea R, Del Fiol G, Hall LK. A model to support shared decision making in electronic health records systems. Med Decis Making 2014;34:987–95.
[71] Barry MJ, Edgman-Levitan S. Shared decision making — The pinnacle of patient-centered care. N Engl J Med 2012;366:780–1.
[72] Schork NJ. Artificial intelligence and personalized medicine. In: Von Hoff DD, Han H, editors. Precision Medicine in Cancer Therapy, vol. 178. Cham: Springer International Publishing; 2019. p. 265–83. https://doi.org/10.1007/978-3-030-16391-4_11.
[73] Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco California USA: ACM; 2016, p. 1135–44. https://doi.org/10.1145/2939672.2939778.
[74] McDougall RJ. Computer knows best? The need for value-flexibility in medical AI. J Med Ethics 2019;45:156–60.
[75] Miller DD. The medical AI insurgency: what physicians must know about data to practice with intelligent machines. npj Digit Med 2019;2. https://doi.org/10.1038/s41746-019-0138-5.
[76] Pearce C, McLeod A, Rinehart N, Whyte R, Deveny E, Shearer M. Artificial intelligence and the clinical world: a view from the front line. Med J Aust 2019;210. https://doi.org/10.5694/mja2.v210.S610.5694/mja2.50025.
[77] Irving G, Askell A. AI safety needs social scientists. Distill 2019;4. https://doi.org/10.23915/distill10.23915/distill.00014.
[78] Harrer S, Shah P, Antony B, Hu J. Artificial intelligence for clinical trial design. Trends Pharmacol Sci 2019;40:577–91.
[79] Rivera SC, Liu X, Chan A-W, Denniston AK, Calvert MJ. SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. BMJ 2020;370:m3210. https://doi.org/10.1136/bmj.m3210.
[80] Houy N, Le Grand F. Personalized oncology with artificial intelligence: The case of temozolomide. Artif Intell Med 2019;99:101693. https://doi.org/10.1016/j.artmed.2019.07.001.
[81] He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019;25:30–6.
[82] Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. npj Digit Med 2020;3. https://doi.org/10.1038/s41746-020-00324-0.
[83] Mayo CS, Mierzwa M, Moran JM, Matuszak MM, Wilkie J, Sun G, et al. Combination of a big data analytics resource system with an artificial intelligence algorithm to identify clinically actionable radiation dose thresholds for dysphagia in head and neck patients. Adv Radiat Oncol 2020. https://doi.org/10.1016/j.adro.2019.12.007.