Published in final edited form as:

Int J Radiat Oncol Biol Phys. 2016 July 1; 95(3): 873–879. doi:10.1016/j.ijrobp.2016.03.006.

Overview of the American Society for Radiation Oncology– National Institutes of Health–American Association of Physicists in Medicine Workshop 2015: Exploring Opportunities for Radiation Oncology in the Era of Big Data

Stanley H. Benedict, PhD^{*}, Karen Hoffman, MD[†], Mary K. Martel, PhD[†], Amy P. Abernethy, MD, PhD[‡], Anthony L. Asher, MD, FAANS, FACS[§], Jacek Capala, PhD^{||}, Ronald C. Chen, MD, MPH[¶], Bhisham Chera, MD[¶], Jennifer Couch, PhD[#], James Deye, PhD[#], Jason A. Efstathiou, MD, DPhil^{**}, Eric Ford, PhD^{††}, Benedick A. Fraass, PhD, FAAPM, FACR, FASTRO^{‡‡}, Peter E. Gabriel, MD, MSE^{§§}, Vojtech Huser, MD, PhD^{|||}, Brian D. Kavanagh, MD, MPH, FASTRO^{¶¶}, Deepak Khuntia, MD^{##}, Lawrence B. Marks, MD, FASTRO[¶], Charles Mayo, PhD, FAAPM^{***}, Todd McNutt, PhD^{†††}, Robert S. Miller, MD, FACP, FASCO^{†††}, Kevin L. Moore, PhD^{§§§}, Fred Prior, PhD^{||||}, Erik Roelofs, MS^{¶¶¶}, Barry S. Rosenstein, PhD^{###}, Jeff Sloan, PhD^{****}, Anna Theriault, B.Sc^{††††}, and Bhadrasain Vikram, MD[#]

^{*}University of California at Davis, Sacramento, California [†]University of Texas MD Anderson Cancer Center, Houston, Texas [‡]Duke University, Durham, North Carolina [§]American Association of Neurological Surgeons, Rolling Meadows, Illinois ^{||}Clinical Radiation Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland [¶]University of North Carolina School of Medicine, Chapel Hill, North Carolina [#]National Cancer Institute, National Institutes of Health, Bethesda, Maryland ^{**}Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts ^{††}University of Washington, Seattle, Washington ^{‡‡}Cedars-Sinai Medical Center, Los Angeles, California ^{§§}Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania ^{|||}Lister Hill National Center for Biomedical Communications, National Library of Medicine, National Institutes of Health, Bethesda, Maryland ^{¶¶}University of Colorado, Denver, Colorado ^{##}Varian Medical Systems, Palo Alto, California ^{***}University of Michigan, Ann Arbor, Michigan ^{†††}The Johns Hopkins University School of Medicine, Baltimore, Maryland ^{†††}American Society of Clinical Oncology, Alexandria, Virginia ^{§§§}University of California, San Diego, La Jolla, California ^{||||}University of Arkansas for Medical Sciences, Little Rock, Arkansas ^{¶¶¶}Department of Radiation Oncology (MAASTRO), GROW - School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, the Netherlands ^{###}Icahn School of Medicine at Mount Sinai, New York, New York ^{****}Mayo Clinic, Rochester, Minnesota ^{††††}Elekta, Sunnyvale, California

Reprint requests to: Stanley H. Benedict, PhD, University of California at Davis, Sacramento, CA 95817. shbenedict@ucdavis.edu.

Conflict of interest: none.

Introduction: Exploring Big Data Impact on Radiation Oncology

Big data research refers to the collection and analysis of large sets of data elements and interrelationships that are difficult to process with traditional methods. It can be considered a subspecialty of the medical informatics domain under data science and analytics. This approach has been used in many areas of medicine to address topics such as clinical care and quality assessment (1–3). The need for informatics research in radiation oncology emerged as an important initiative during the 2013 National Institutes of Health (NIH)–National Cancer Institute (NCI), American Society for Radiation Oncology (ASTRO), and American Association of Physicists in Medicine (AAPM) workshop on the topic “Technology for Innovation in Radiation Oncology” (4). Our existing clinical practice generates discrete, quantitative, and structured patient-specific data (eg, images, doses, and volumes) that position us well to exploit and participate in big data initiatives. The well-established electronic infrastructure within radiation oncology should facilitate the retrieval and aggregation of much of the needed data. With additional efforts to integrate structured data collection of patient outcomes and assessments into the clinical workflow, the field of radiation oncology has a tremendous opportunity to generate large, comprehensive patient-specific data sets (5).

However, there are major challenges to realizing this goal. For example, existing data are presently housed across different platforms at multiple institutions and are often not stored in a standardized manner or with common terminologies to enable pooling of data. In addition, many important data elements are not routinely discretely captured in clinical practice. There are cultural, structural, and logistical challenges (eg, computer compatibility and workflow demands) that will make the dream of big data research difficult.

The big data research workshop provided a forum for leaders in cancer registries, incident report quality-assurance systems, radiogenomics, ontology of oncology, and a wide range of ongoing big data and cloud computing development projects to interact with peers in radiation oncology to develop strategies to harness data for research, quality assessment, and clinical care. The workshop provided a platform to discuss items such as data capture, data infrastructure, and protection of patient confidentiality and to improve awareness of the wide-ranging opportunities in radiation oncology, as well as to enhance the potential for research and collaboration opportunities with NIH on big data initiatives.

The goals of the workshop were as follows:

1. To discuss current and future sources of big data for use in radiation oncology research,
2. To identify ways to improve our current data collection methods by adopting new strategies used in fields outside of radiation oncology, and
3. To consider what new knowledge and solutions big data research can provide for clinical decision support for personalized medicine.

The workshop sessions, speakers, and titles are listed in Table 1. Each session is briefly summarized in the following sections.

Novel Big Data Resources in Development That Are Not Unique to Radiation Oncology

Genomics

The potential of big data resources from radiogenomics and pathologic and genetic data was explored. Radiogenomics is the study of the link between germline genotypic variations and the large clinical variability observed in response to radiation therapy (RT). The aim of radiogenomics is to identify the alleles that underlie the inherited dissimilarities in phenotype through the performance of genome-wide association studies, with the aim to enhance precision medicine through development of a predictive assay to help personalize and optimize cancer treatment with radiation (6). An advantage to the performance of genome-wide association studies is that this approach avoids having to make any a priori assumption of the genes that are important for outcomes resulting from RT. These studies are considered to generate big data because thousands of subjects are involved in ongoing studies, with each subject producing gigabytes of clinical and genotyping information. To accomplish this research, the Radiogenomics Consortium (RGC) was established in 2009; the RGC is an NCI-NIH-supported Cancer Epidemiology Consortium through the Epidemiology and Genomics Research Program (<http://epi.grants.cancer.gov/Consortia/single/rgc.html>) consisting of 194 investigators at 112 institutions in 26 countries. The goal of the RGC is to bring together collaborators to pool samples and data for increased statistical power of radiogenomics studies. Through the RGC, the size of radiogenomics studies is now in the range of 10,000 subjects.

Initiatives to capture broad clinical data

Clinical records include vast amounts of data that are potentially useful for clinical care and research. However, extracting and storing these data accurately, and in a meaningful manner for all stakeholders (eg, patients, physicians, and researchers), are most challenging and are largely not currently done. The rapid adoption of electronic health record (EHR) systems, as incentivized by the federal government, has created new opportunities to gather such data, both for real-time clinical use and for population-based research. However, presently, the EHR systems do not readily support many of these activities. There is a great need to better capture high-quality patient-level data, both structured and unstructured, to support these activities.

One approach to better organizing oncology information is CancerLinQ (Cancer Learning Intelligence Network for Quality)—a new initiative of the American Society of Clinical Oncology (www.cancerlinq.org). Its approach is to aggregate data through direct feeds from EHRs and practice management systems in oncology practices (7). Future data sources may include cancer registry data, claims data, and genomic databases and could also involve clinical trial matching, development of risk stratification models, and postmarket surveillance. When CancerLinQ is fully realized, the observational data obtained from the aggregated experience of millions of cancer patients will have the potential to transform clinical research by enabling insights and hypothesis generation on a scale not achievable through conventional clinical trials. A separate but similar approach is in development by Flatiron Health (New York, NY) and focuses on extraction of clinical data (including key

radiation and medical oncology parameters) from patient records to enable research to define optimal care (<http://www.flatiron.com/>).

As cancer care evolves toward increased precision in disease characterization and therapeutic intervention, the number of “like” patients who can be grouped together may become ever smaller (as the number of relevant variables being considered increases). This is actually a “small data” problem that necessitates two approaches to achieving larger data sets for investigation: (1) increasing the number of analyzable patients by collecting data more consistently; and (2) aggregating data across institutions to achieve sample sizes with sufficient power for meaningful analysis. The ability to combine data sets from heterogeneous source systems and maintain unambiguous, shared meaning (so-called semantic interoperability) requires the use of standardized methods for capturing and encoding clinical knowledge, referred to as the *ontology of oncology*. Encoding or tagging clinical data sets with concepts from ontology is a powerful means of achieving semantic interoperability and should be more widely pursued in the oncology community. In addition, effort is needed to enhance existing ontology to include greater coverage of radiation oncology concepts (8–10). For example, an ongoing AAPM initiative hopes to standardize anatomic structure names (Task Group 263).

Challenges considered from this session include the following:

1. How can we collect information from the many thousands of cancer patients who received RT who have been genotyped by NCI and other sources but whose treatment, dosimetric, and outcomes data are often incomplete and located in fragmented and disparate databases?
2. What is the most reasonable approach to harmonize toxicity data coming from large numbers of cohorts whose clinical information was collected using a variety of evaluative instruments?
3. How can we translate the big data gained from radiogenomics studies into a clinically useful assay to predict which patients are at greatest risk of the development of toxicities after RT with a goal to personalize and optimize treatment?

Big Data Resources Currently Available and/or in Development Within Radiation Oncology

General

Large amounts of discrete quantitative data presently exist in our radiation oncology–specific EHRs (eg, Mosaic [Elekta, Stockholm, Sweden] and Aria [Varian Medical Systems, Palo Alto, California]) and, similarly, in our treatment planning systems. For many of these items, the vendors have provided the ability to extract, transform, and process information from these databases directly. It is expected and hoped that these systems will evolve to also incorporate broader data from other aspects of a patient’s record. Presently, most radiation oncology organizations are still tackling issues with their transactional databases and are learning how to use the data in their systems effectively (eg, to improve operational

efficiency, understand costs, and enhance outcomes). The recent development of knowledge management platforms may represent an effective “big data platform” for the future (11). Ideally, these platforms will integrate information from specific radiation oncology systems and provide associated data management and advanced analytics solutions, essentially enabling big data research. Cloud-based platforms are promising for cost-effective scaling of solutions to multiple institutions. However, compared with more traditional architectures, these approaches introduce additional challenges for ensuring security and privacy and achieving buy-in from stakeholders charged with protecting patient information.

Imaging

There is much potentially valuable information embedded in medical images, including radiation treatment planning images. However, there are no widely used standards to facilitate the definition and pooling of image-based data. The Cancer Imaging Archive (TCIA) (12) provides a good framework to address this challenge. The Cancer Imaging Program is 1 of 4 programs in the Division of Cancer Treatment and Diagnosis of the NCI. TCIA serves as a resource to support NCI-funded research activities. The TCIA team has defined standard operating procedures for data acquisition, de-identification, and curation and recently extended these to support complex RT data sets. TCIA has addressed unique informatics problems inherent in collection, curation, quality assurance, and publication of longitudinal imaging and treatment records linked to volumetric RT data sets with the goal of enabling quantification of radiobiological responses to therapeutic interventions (13). Such tools might enable big data research that relates image-based metrics to outcomes.

Quality

In the context of big data initiatives, it is important to consider what data might be obtained concerning the safety-critical processes in RT. One such stream of information is through incident reports, and these are now being collected in an organized fashion through the Radiation Oncology Incident Learning System (RO-ILS), sponsored by ASTRO and AAPM (14). The program has experienced a steady growth in both number of participants and use of the system. Future directions for RO-ILS and incident learning in general will include the following: increasing use and swaying the culture of safety, exploiting the data-mining tools emerging in other areas of big data, enhancing the processing and categorization of data through improved ontology and automated approaches such as machine learning and natural language processing to handle the large amount of narrative information, and linking the incident learning data to outcomes measures.

Action items from this session include the following:

1. There is a need to develop “knowledge management” systems that are scalable and flexible, use a common ontology, and leverage data (both within and external to radiation oncology systems) to create predictive decision support solutions.
2. Further support is needed for the DICOM (Digital Imaging and Communications in Medicine) standards for complex RT data sets across all planning and treatment delivery platforms.

3. There is a need to promote a culture of safety that advances (and is mutually reinforced by) comprehensive incident reporting to improve quality assurance and safety of RT.
4. Careful curation of data should be performed so that they are reusable by other scientists.
5. The creation of high-quality, publicly available data collections is needed to foster research reproducibility and allow new avenues of research by combining high-quality data sets from multiple completed projects.

Strategies to Optimize Big Data Within Radiation Oncology and Integrate Outside Resources: How Do We Help Patients?

Building a data culture in radiation oncology

Randomized controlled trials (RCTs) have high internal validity for defining causal relationships among variables for the patients and settings fitting the constraints of the trial. RCTs are costly to run, usually provide answers years after peak interest in the question, and may be challenging to generalize to routine practice. Patient-centered outcomes research efforts have heightened awareness of the value of large data sets, reflecting the diversity of “real-world” treatments to benchmark the external validity of RCTs, characterize practice patterns, and generate new and better-formulated hypotheses. Obtaining these data sets requires changing our culture and processes so that the value of data entry and curation is not just what must be done to enable treatment of a given patient to proceed but also to ensure that automated and accurate aggregation of information is a priority. This improves our ability to use data-driven approaches to solve current patient care problems as well as to facilitate participation in clinical trials.

Pooling radiation oncology data internationally

Integrating and optimizing big data collection internationally will increase the number of analyzable subjects and provide more intersubject diversity (which will better enable assessment of the impact of various variables). Investigators from Europe have developed the Euregional Computer Assisted Theragnostics project (EuroCAT) at MAASTRO Clinic in the Netherlands (www.eurocat.info). This is a collaborative project involving a distributed learning network of 5 RT institutes in the Netherlands, Germany, and Belgium. EuroCAT is superseded by successive computer-assisted theragnostics projects reaching global coverage with additional nodes in Europe, Australia, China, India, and the United States, encompassing 19 nodes at the moment. The project has successfully used innovative information technologies to learn from each other's data without the data having to leave the institution (15, 16). This novel and secure approach was achieved by adding to the institutes' medical records a semantic interoperable layer, in which local terms are converted into concepts from a well-defined ontology (eg, NCI Thesaurus or ROO [Radiation Oncology Ontology]; <http://biportal.bioontology.org/ontologies/ROO>). This renders the clinical data from a native language into an anonymous, encoded “machine-readable” format. It is then possible to retrieve information via a semantic portal and facilitate reasoning by applying data mining and artificial intelligence techniques to extract personalized models that can

accurately predict how treatment decisions affect a patient's chance of survival and toxicity. Validation of these models can be achieved by using the global data pooling network through dividing the data into training and validation data sets. A number of validated models are publicly available (www.predictcancer.org) to support decisions made by physicians and patients.

Registry-based data

Several registries (eg, Surveillance, Epidemiology, and End Results [SEER]) have been used to facilitate big data research in the areas of quality assessment, comparative effectiveness, public health, and clinical decision support and insight discovery and are resources for oncology-based research. For example, many studies have used such registries to examine practice patterns. If a high proportion of patients receive nonstandard treatment, these studies can inform health policy and efforts needed to address these gaps and/or perhaps disparities in care. A primary goal of comparative effectiveness research is to provide data directly comparing the patient outcomes associated with, for example, different treatment options (17). Often, randomized trials are not feasible to address comparative effectiveness questions, and big data sources, such as registries, are needed. There are also potential advantages and disadvantages to using clinical trials versus registry data for comparative effectiveness research. For example, registry-based studies are usually much faster and less expensive than traditional clinical trials (18).

In an effort to take advantage of the wide array of biomedical big data, the NIH and NCI have created several new programs designed to aid researchers in accessing data and software, analyzing data, and more easily sharing their own data and software. Examples include the NIH Big Data to Knowledge program (<https://datascience.nih.gov/bd2k>) and the NCI's Informatics Technology for Cancer Research program (<http://itcr.nci.nih.gov/>). Tools, methods, and resources developed through these programs are shared with the larger research community through collections and videos and will ultimately be available through the eRA Commons (electronic Research Administration available to the public). Informatics Technology for Cancer Research tools and educational resources can be found at the National Cancer Informatics Program Hub (<https://nciphub.org/groups/itcr/collections/itcr-training-assets>); the NIH Big Data to Knowledge program has supported the development of a data discovery index (<https://biocaddie.org/>) and training programs including curricula and course development (<https://datascience.nih.gov/bd2k/funded-programs/enhancing-training>). In addition, there are programs designed specifically to support training of the next generation of biomedical data scientists and data science-savvy biomedical researchers.

Cancer genesis

Cancer initiation and development are complex, with multiscale processes involving alterations and interactions across spatial and temporal scales. Systems biology, supported through the Integrative Cancer Biology Program (<http://icbp.nci.nih.gov/>), brings together mathematics and experimental biology to iteratively model and understand these processes. The Physical Sciences in Oncology program supports teams of quantitative researchers partnered with experimental biologists to generate data at challenging scales such as data on the physical forces associated with cell growth, tumor metastasis, and a quantitative

approach to modeling and understanding cancer biology (<http://physics.cancer.gov/>). A particularly challenging but promising approach takes advantage of the orthogonal data and mathematical and statistical methods used to model the underlying mechanisms of cancer initiation and progression together with models, data, and insights from population outcomes research through the NCI's "Bridging the Gap Between Cancer Mechanism and Population Science" (<http://grants.nih.gov/grants/guide/pa-files/PA-13-081.html>).

Important opportunities from this session include the following:

1. Widening the potential for interlinkage of cancer data registries and developing strategies to include analytics for a broad range of treatment approaches (widely variable dose/volume strategies);
2. Developing technology and adopting a culture change to enable interinstitutional pooling of data to form large analyzable databases;
3. Engaging with legislative and regulatory groups to find effective and inexpensive electronic methods to gather long-term follow-up data on survival, recurrence, and patient-reported outcomes while still respecting the need to protect patient health care information; and
4. Understanding and identifying the key clinical decisions and questions where big data can be most useful.

Summary and Discussion

"How do we help patients?" was the overriding theme of the workshop. The 1.5 days of presentations, panel discussions, and breakout groups focused on improving the big data strategies either currently available, in development, or outside of radiation oncology, with the intent to provide better patient quality in RT. The promise of big data in radiation oncology is to provide improved access to our collective experience treating patients so as to improve care for new and future patients. This improvement can take the form of actions such as reducing geographic disparities in care; ensuring continual quality improvement for individual practices; and ideally, personalizing treatments based on the outcomes of prior, similar patients. Each of these objectives requires different levels and resolution of clinical data that may be contained in registries, electronic medical records, tissue banks, and treatment planning and imaging systems. Furthermore, our aim is to leverage big data to formulate a rapid-learning health management system that can assist in providing knowledge to practicing clinicians, augmenting and improving the care they can provide to their patients.

Our next workshop will further deal with many of the concepts, ideas, and proposals presented in this workshop and the previous one. Organization for the next ASTRO-NIH workshop, "Precision Medicine in Radiation Oncology: Personalizing Radiation Treatment," is under way; it will be held at the NIH campus on June 16–17, 2016.

Acknowledgments

The Organizing Committee for the American Society for Radiation Oncology–National Institutes of Health–American Association of Physicists in Medicine workshop “Exploring Opportunities for Radiation Oncology in the Era of Big Data,” August 13–14, 2015, included Stanley H. Benedict, PhD (Chair), Karen Hoffman, MD (Co-chair), Mary K. Martel, PhD, Brian D. Kavanagh, MD, John Bayouth, PhD, Dan Low, PhD, Lawrence B. Marks, MD, Benedick A. Fraass, PhD, Todd McNutt, PhD, Charles Mayo, PhD, Jacek Capala, PhD, Bhadrasain “Vik” Vikram, MD, and James Deye, PhD. The program committee is deeply indebted to the National Cancer Institute for its support of the meeting in providing resources and hosting the meeting at the National Institutes of Health. We are also grateful to the American Society for Radiation Oncology and the American Association of Physicists in Medicine for their administrative support and financial assistance, without which this meeting would not have been possible.

References

1. Hersh WR. Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance. *Am J Manag Care*. 2007; 13(Pt 1):277–278. [PubMed: 17567224]
2. Safran C, Bloomrosen M, Hammond WE, et al. Toward a national framework for the secondary use of health data: An American Medical Informatics Association white paper. *J Am Med Inform Assoc*. 2007; 14:1–9. [PubMed: 17077452]
3. Shortliffe, EH.; Cimino, JJ., editors. *Biomedical Informatics: Computer Applications in Health Care and Biomedicine (Health Informatics)*. 3rd. New York: Springer; 2012.
4. Chetty IJ, Martel MK, Jaffray DA, et al. Technology for innovation in radiation oncology. *Int J Radiat Oncol Biol Phys*. 2015; 93:485–492. [PubMed: 26460989]
5. Lambin P, Roelofs E, Reymen B, et al. ‘Rapid Learning health care in oncology’—An approach towards decision support systems enabling customised radiotherapy. *Radiother Oncol*. 2013; 109:159–164. [PubMed: 23993399]
6. Rosenstein BS, West CM, Bentzen SM, et al. Radiogenomics: Radiobiology enters the era of big data and team science. *Int J Radiat Oncol Biol Phys*. 2014; 89:709–713. [PubMed: 24969789]
7. Schilsky RL, Michels DL, Kearbey AH, et al. Building a rapid learning health care system for oncology: The regulatory framework of CancerLinQ. *J Clin Oncol*. 2014; 32:2373–2379. [PubMed: 24912897]
8. Hoehndorf R, Schofield PN, Gkoutos GV. The role of ontologies in biological and biomedical research: A functional perspective. *Brief Bioinform*. 2015; 16:1069–1080. [PubMed: 25863278]
9. Asher AL, Speroff T, Dittu RS, et al. The National Neurosurgery Quality and Outcomes Database (N2QOD): A collaborative North American outcomes registry to advance value-based spine care. *Spine (Phila Pa 1976)*. 2014; 39(Suppl 1):S106–S116. [PubMed: 25299254]
10. Asher AL, Devin CJ, Mroz T, et al. Clinical registries and evidence-based care pathways: Raising the bar for meaningful measurement and delivery of value-based care. *Spine (Phila Pa 1976)*. 2014; 39(Suppl 1):S136–S138. [PubMed: 25077915]
11. Moore KL, Kagadis GC, McNutt TR, et al. Vision 20/20: Automation and advanced computing in clinical radiation oncology. *Med Phys*. 2014; 41:010901. [PubMed: 24387492]
12. Prior FW, Clark K, Commean P, et al. TCIA: An information resource to enable open science. *Conf Proc IEEE Eng Med Biol Soc*. 2013; 2013:1282–1285. [PubMed: 24109929]
13. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *J Digit Imaging*. 2013; 26:1045–1057. [PubMed: 23884657]
14. Hoopes DJ, Dicker AP, Eads NL, et al. RO-ILS: Radiation Oncology Incident Learning System: A report from the first year of experience. *Pract Radiat Oncol*. 2015; 5:312–318. [PubMed: 26362705]
15. van Soest, JPA.; Dekker, ALAJ.; Roelofs, E., et al. Application of machine learning for multicenter learning. In: Naqa, IE.; Li, R.; Murphy, MJ., editors. *Machine Learning in Radiation Oncology*. New York: Springer International; 2015. p. 71-97. Available at: http://link.springer.com/chapter/10.1007/978-3-319-18305-3_6; 2015 [Accessed September 14, 2015]

16. Roelofs E, Dekker A, Meldolesi E, et al. International data-sharing for radiotherapy research: An open-source based infrastructure for multicentric clinical data mining. *Radiother Oncol.* 2014; 110:370–374. [PubMed: 24309199]
17. Robertson S, Moore J, Cheng S, et al. A data-mining framework for large scale analysis of dose-outcome relationships in a database of irradiated head and neck (HN) cancer patients. *Med Phys.* 2015; 42:4329–4337. [PubMed: 26133630]
18. Roelofs E, Persoon L, Nijsten S, et al. Benefits of a clinical data warehouse with data mining tools to collect data for a radiotherapy trial. *Radiother Oncol.* 2013; 108:174–179. [PubMed: 23394741]

Table 1

Workshop lectures and participants

Lecture title	Presenters
Keynote Address 1: What Lies Ahead for the NIH Laboratory for Informatics: Selected Informatics Challenges Ahead for Big Data	Vojtech Huser, MD, PhD, NIH
Presentations on Novel Big Data Resources in Development That Are <i>Not Unique</i> to Radiation Oncology	
Moderator: Lawrence B. Marks, MD	
The Potential of Big Data Resources for Radiogenomics and Pathologic/Genetic Data	Barry Rosenstein, PhD, Mount Sinai
Organizing the World's Oncology Information and Making It Useful for Patients, Physicians, Life Sciences and Researchers	Amy Abernethy, MD, Duke University
Update on ASCO and CancerLinQ	Robert S. Miller, MD, FACP, ASCO
The Ontology of Oncology: Language of Quantitative Cancer	Peter Gabriel, MD, University of Pennsylvania
Overview of Neurological Surgery Registries: N2QOD/AANS Spine Registry and the AANS-ASTRO Radiosurgery Registry	Anthony L. Asher, MD, FAANS, FACS, American Association of Neurological Surgeons, Rolling Meadows
Presentations on Big Data Resources Available and in Development <i>Within</i> Radiation Oncology	
Moderator: Kevin Moore, PhD	
National Radiation Oncology Registry (NROR) Pilot Update	Jason Efstathiou, MD, DPhil, Harvard
How Do We Get Big Data out of Aria and Mosaic? What Are the Elekta and Varian Roadmaps?	Jack Challis, Elekta Deepak Khuntia, MD, Varian
Image Based Data Collection and Management, Including RT Planning Data	Fred Prior, PhD, Washington University
Radiation Oncology Incident Reporting and Learning Systems: RO-ILS	Eric Ford, PhD, University of Washington
Presentations on Strategies to Optimize Big Data Within Radiation Oncology and Integrate Outside Resources: "How Do We Help Patients"	
Moderator: Todd McNutt, PhD	
Overview of the Euregional Computer Assisted Theragnostics Project (EuroCAT) at MAASTRO Clinic (Netherlands)	Erik Roelofs, MS, MAASTRO
Update on the Informatics Technology for Cancer Research (ITCR) at the NCI and BD2K Initiative at the NIH	Jennifer Couch, PhD, NIH
Linking Broad Based Data From Registries for Quality Assessment, Comparative Effectiveness, and Public Health	Ronald Chen, MD, MPH, University of North Carolina
Linking Detailed Data From Institutional Databases for Clinical Decision Support and Insight Discovery	Todd McNutt, PhD, Johns Hopkins University
Keynote Address 2: Automating Discovery of What Works: Statistical Methods to Identify Patterns and Interdependencies	Jeff A. Sloan, PhD, Mayo Clinic

Abbreviations: AANS, American Association of Neurological Surgeons; ASCO = American Society of Clinical Oncology; ASTRO = American Society for Radiation Oncology; BD = big data; NCI = National Cancer Institute; NIH = National Institutes of Health; N2QOD = National Neurosurgery Quality and Outcomes Database; RT = radiation therapy.