Original Report

Development of a model web-based system to support a statewide quality consortium in radiation oncology

Jean M. Moran PhD a,⁎, Mary Feng MD a, Lisa A. Benedetti MS b, Robin Marsh CMD a, Kent A. Griffith MPH, MS c, Martha M. Matuszak PhD a, Michael Hess MSI d, Matthew McMullen MS e, Jennifer H. Fisher MS f, Teamour Nurush PhDb g, Margaret Grubb MS a, Stephen Gardner MS h, Daniel Nielsen MS a, Reshma Jagsi MD, DPhil a, James A. Hayman MD, MBA a, Lori J. Pierce MD a, On Behalf of the Michigan Radiation Oncology Quality Consortium

aDepartment of Radiation Oncology, University of Michigan, Ann Arbor, Michigan
bDepartment of Radiation Oncology, William Beaumont Hospital, Royal Oak, Michigan
cDepartment of Biostatistics, University of Michigan, Ann Arbor, Michigan
dSchool of Information, University of Michigan, Ann Arbor, Michigan
eRadiation Oncology, St. Joseph Mercy Hospital, Ypsilanti, Michigan
fJohnson Family Center for Cancer Care, Mercy Health Partners, Muskegon, Michigan
g21st Century Oncology of Michigan, Farmington Hills, Michigan
hRadiation Oncology Department, Henry Ford Health System, Detroit, Michigan

Received 16 May 2016; revised 23 September 2016; accepted 10 October 2016

Abstract

Purpose: A database in which patient data are compiled allows analytic opportunities for continuous improvements in treatment quality and comparative effectiveness research. We describe the development of a novel, web-based system that supports the collection of complex radiation treatment planning information from centers that use diverse techniques, software, and hardware for radiation oncology care in a statewide quality collaborative, the Michigan Radiation Oncology Quality Consortium (MROQC).

Methods and materials: The MROQC database seeks to enable assessment of physician- and patient-reported outcomes and quality improvement as a function of treatment planning and delivery techniques for breast and lung cancer patients. We created tools to collect anonymized data based on all plans.

Results: The MROQC system representing 24 institutions has been successfully deployed in the state of Michigan. Since 2012, dose-volume histogram and Digital Imaging and Communications in

Supplementary material for this article (http://dx.doi.org/10.1016/j.prro.2016.10.002) can be found at www.practicalradonc.org.

Sources of support: This work was supported by funding from Blue Cross Blue Shield of Michigan and Blue Care Network.

Conflicts of interest: None.

⁎ Corresponding author. Department of Radiation Oncology, University of Michigan Health System, 1500 E. Medical Center Drive, Ann Arbor, MI 48109-5010.
E-mail address: jmmoran@med.umich.edu (J.M. Moran).

http://dx.doi.org/10.1016/j.prro.2016.10.002
1879-8500/© 2016 American Society for Radiation Oncology. Published by Elsevier Inc. All rights reserved.
Introduction

Concerns about the ability to apply the results of clinical trials to the broader community have generated growing interest in comparative effectiveness research in all areas of medicine, including oncology. Simultaneously, recognition of the power of collecting “big data” to allow for a learning health care system for oncology care has led to an acute need to compile complex radiation treatment planning information for subsequent analyses.

Unfortunately, existing cancer registries and claims-based datasets lack the detail to explore many of the important questions about comparative effectiveness in radiation oncology. For example, the Surveillance, Epidemiology, and End Results database has shown limited accuracy even in the determination of whether or not radiation therapy (RT) was delivered, let alone the administered techniques. Reliable analyses of questions, such as the role of intensity modulated RT (IMRT) in treatment, are impossible without detailed data regarding patient anatomy, dose distributions, complexity of treatment plans, and outcomes. There have been prior efforts to create registries that capture the physics component of patient data such as the National Radiation Oncology Registry. Others are evaluating the data exchange needs for supporting large enterprise research in radiation therapy. Although the National Institute of Health supports an infrastructure for multi-institutional clinical trials, we are unaware that any systematic database exists for comparing treatment plans and efficacy for common cancers such as breast and lung cancer in a clinical setting across multiple facilities in the United States.

Here, we have developed and improved an observational database that captures complex medical physics details, treatment planning information, delivered dose, and outcomes for radiation therapy of lung and breast cancer patients for the Michigan Radiation Oncology Quality Consortium (MROQC). MROQC is a collaborative consortium, funded by Blue Cross Blue Shield of Michigan, and coordinated by members of the Department of Radiation Oncology at the University of Michigan. The consortium has grown from 4 to 24 participating radiation oncology sites across Michigan. The consortium’s customized registry includes the variations in patient care seen across practices, variations often excluded in data taken from clinical trials.

Data capture from the consortium and analysis are for quality purposes to identify predictive factors for toxicity, defined for breast cancer patients as moist desquamation and breast pain and for lung cancer patients as esophagitis and pneumonitis. The physics data focused on information that could contribute to the quality of RT, including both toxicity and tumor control. Because of known challenges regarding the lack of interconnectivity between treatment planning and delivery systems, we developed methods to obtain high-integrity data independent of the treatment planning system or treatment management system manufacturer and software versions.

This report describes the development of the collaborative consortium, the web-based platform for data collection, and the audit process to provide guidance to others actively engaged in the development of similar large-scale registries for radiation therapy. A primary goal of the consortium is to use the database to identify patients who may benefit from IMRT as opposed to 3-dimensional planning when considering patient and physician reported outcomes. The consortium has several projects involving patient outcomes in which the acquired physics data will be correlated with physician assessments and patient-reported outcomes.

Methods

Consortium and initial pilot

Each center identifies a physician, physicist, and administrator as per the participation requirement. Funding supports the majority of the data entry for all eligible patients, typically done by a clinical research associate and a dosimetrist. The acquired data relate to the treatment and toxicity for patients with early-stage breast cancers irradiated to the whole breast (with or without lymph nodes) and patients treated definitively or postoperatively for lung cancer. Eligible patients with breast cancer are those with invasive breast cancer or ductal carcinoma in situ treated with breast-conserving surgery and whole breast irradiation where the use of neoadjuvant or concurrent chemotherapy and/or treatment to the regional nodes is allowed. Patients are excluded if treated with postmastectomy RT or accelerated partial breast irradiation, or if they have previous thoracic irradiation or bilateral treatment. Eligible patients with lung cancer are newly diagnosed lung cancer patients (small cell and non-small cell) to be treated curatively regardless of the use or timing of chemotherapy. Postoperative patients are...
eligible along with those treated with definitive radiation. Patients are excluded if they have metastatic disease or prior thoracic irradiation, or are treated with stereotactic body radiation therapy. All patients are entered into the registry and if eligible, the required data are added.

The coordinating center made recommendations for specific elements of data collection which were piloted and revised by an initial 4 consortium institutions. The emphasis was on physics data elements considered to be relevant to clinical and patient reported outcomes. In addition to data in Digital Imaging and Communications in Medicine (DICOM) format, other tools were created for accurate and efficient data entry. There are web-based forms along with dose-volume histogram (DVH) and DICOM-RT data collection tools.

**System design**

Each institution is required to obtain institutional review board approval for this project. Data collection is required to be secure, deidentified, and Health Insurance Portability and Accountability Act compliant. Each patient has a unique local identifier known only to his or her institution and a consortium-wide unique identifier used for all analyses.

The institutional, breast, and lung technical forms are shown in Appendices E1 through E3 (available as supplementary material online only at www.practicalradonc.org) and were implemented as web-based forms using Lime Survey (LimeSurvey Project/Carsten Schmitz, 2012; LimeSurvey: An Open Source survey tool, http://www.limesurvey.org). Branching logic is used as necessary for efficient data entry of nested questions. Discrete selections are preferred to text entry to minimize entry inconsistencies. The patient-specific forms focus on simulation, planning, treatment guidance, and dose information including the prescription.

When the consortium began, we initially limited practice changes. For example, Figure 1 shows differences in contouring a specific breast volume compared to using a 95% isodose surface as a surrogate. This was considered reasonable given the consortium’s focus on acute toxicities, such as erythema and desquamation, which could be due to higher doses.

Because separation and breast size have been shown to be relevant when assessing toxicity for breast cancer patients, the separation is required to be measured on a central axial computed tomography scan midway through the breast to be treated.

Also required are the DICOM-RT plan file and DVHs for key structures. We have a dedicated anonymization server that runs the anonymization process. Files are uploaded there, sanitized, and then placed in the database. The anonymized fields are displayed in a searchable text format to allow the data abstractor to confirm and attest that no protected health information (PHI) was present.

Although reviewing the accuracy of contoured structures on image datasets is outside the project’s scope, published anatomical atlases are posted to a knowledgebase for providers. The database was built to capture variability in different practice patterns across the state to identify areas for improvement in patient care, similar to other collaborative quality initiatives.

Because no automated tools for multi-institutional data collection exist, the data abstractor interacts with different systems, as shown in Figure 2. Data collection tools developed among consortium members are shared on the consortium knowledgebase. One institution modified an in-house script for their treatment planning system to aid others in extracting DVHs easily. Another site developed documents to track clinical flow that were then posted to the consortium knowledgebase.

Data are stored centrally in MySQL and NoSQL databases maintained by the coordinating center. Each night, data are extracted for analysis in SAS and stored elsewhere. The servers hosting the database and the SAS servers are in a tier II data center as defined by the Uptime Institute (https://uptimeinstitute.com/tiers).

**Data privacy**

The design of the data submission and data storage components focused on ensuring compliance with Health Insurance Portability and Accountability Act requirements. The coordinating center does not have access to patient identifiers. Data submission is done with 2-factor authentication (such as a token key and password). The anonymized
DICOM-RT plan files are uploaded for all patients. The process requires that only DICOM-RT plan files are present for PHI considerations. If other files are present, the upload fails and an automated error message is generated for both the abstractor and database support team.

Plan data and DVHs

Plan data

Users create a single zip file of all DICOM-RT plan files for each patient. The number of segments per unique gantry angle, total monitor units or treatment time, presence of wedges, beam energy, and other plan metrics are extracted and stored in the database upon upload using custom Matlab code. The IMRT delivery type categories are step-and-shoot or segmental multileaf collimator (MLC), dynamic MLC, TomoTherapy, and volumetric modulated arc therapy. The segmental MLC technique beams are divided by ≤5 segments (fields with control points and monitor units for delivery) and >5 segments.

DVHs

For the DVHs collected, the user selects the type of DVH (cumulative or differential), the dose and volume units, and the volume of each structure (used to convert data entered in % to mL). Only numeric dose and volume data are accepted that are plotted along with key metrics such as volume and mean doses (Fig 3) for review by the data abstractor after submission.

Consortium results and lessons learned

The consortium results are reviewed by the coordinating center and at the regular in-person meetings (3 times per year). The consortium’s physics group meets by teleconference between the full consortium meetings as needed.

Institutional questionnaire

Among the 20 institutions participating by December 2015, 4 treatment planning systems were represented. All
institutions reported 4-dimensional computed tomography capability, with several institutions having a device for breath hold control for use at simulation and/or treatment delivery. The primary algorithm type for photon beams was a model-based system (94%) versus a pencil beam or other semi-empirical algorithm (6%). Eight institutions were credentialed with the Imaging and Radiation Oncology Core (IROC) - Houston lung phantom and 10 with the head and neck phantom for IMRT credentialing. The primary algorithm type for electron beams was pencil beam (61%) compared with Monte Carlo (39%). Eight institutions had multiple algorithms available for clinical use.

Patient-specific web-based forms

Data quality is monitored by the physics and statistics team at the coordinating center. Initial data errors included transposition of the total dose and the dose per fraction, incorrect DVH type, and miscategorization of initial and boost plans. Logical checks, such as allowed ranges for numeric input and color-coded completeness indicators, were engineered to improve the quality of the entered data. Limits were used to flag incorrect entries such as input in cGy instead of Gy. Upper bounds were chosen to include the range of dose prescriptions, but were not so high as to allow a mismatch in the reported units. Over time, the database evolved to make data collection streamlined and robust.

We have limited changes that would significantly alter the data analysis. Forms for breast and lung were changed to collect the planning type information so that field-in-field techniques were adequately represented. The questions regarding targets for patients with lung cancer were changed (Fig 4) because the original wording based on the International Commission on Radiation Units & Measurements 62 definitions was inadequate.

Infrastructure needs

Several studies have identified the need for interconnectivity between the treatment planning system, treatment management system, and hospital systems’ electronic medical record systems for toxicities to create registry databases for radiation oncology.6,7 We created a registry despite the lack of interconnectivity through adequately funded staffing, consortium-wide engagement to ensure the collection of quality data, and dedicated personnel at the coordinating center. A helpdesk e-mail was available on the consortium website, and 665 support tickets have been generated since the launch of website data collection through March 2015. Of those, 64% were automatically generated by errors made in preparing the files for upload. The causes of the automated tickets were no files found (33%), incompatible file name (22%), unzipped files (20%), non-DICOM file found (16%), and nonplan file found (9%).

Audits

Annual physics audits are performed at each institution by coordinating center personnel. The audit consists of reviewing 12 cases from the preceding year, 6 of each
cancer type, which are selected at random by the statistical team to match the institution’s proportion of cases treated by each delivery type (IMRT or 3-dimensional). With the auditor present but shielded from seeing PHI, information is reviewed using the source systems for the patient’s information. The checks include confirmation of laterality, motion management methods at simulation and treatment if used, plan type, contours created, dose calculations, separation on a mid-breast axial slice, image guidance, prescription, DVH data, and plan data.

For the 2015 data, the number of discrepancies ranged from 2 to 64 in the audited data of each institution (of 650-906 audited data points). Each institution received a report on its accuracy rate along with the list of discrepancies with corrections requested within 1 month. Then, coordinating center personnel reviewed audit cases to verify that corrections were made.

In addition to reviewing data points from the randomly selected cases, auditors found systematic errors at 9 institutions. One institution had excluded supraclavicular nodal plans from the DICOM data submission. Another institution submitted DVH data with a very small dose bin size such that DVHs were truncated. If a systematic error was identified, coordinating center personnel queried all of the data from that institution and provided a list of affected patients to the data abstractor for corrections.

An accuracy rate of >90% was determined for the audited institutions. Common sources of error during the initial audits included plan reporting, incomplete or missing contours, and concordance between physics data elements.

Using data collection to assess practice patterns

Through the data collection, we have learned about practices and techniques used for eligible breast and lung cancer patients. For example, Figure 5 shows the use of motion management techniques for the treatment of breast (Fig 5A) and lung (Fig 5B) patients. The analysis shows that devices for breath hold are used fairly frequently for patients with left-sided breast cancer (30% of those patients) but not for those with lung cancer (1.7%).

Variations in whether or not structures are contoured for treatment planning are shown in Table 1. An analysis of these results is beyond the scope of this manuscript; however, these types of data are invaluable in analyzing the impact of treatment techniques and dose on outcomes for these patients.

For treatment planning, we observed that contours are frequently created prospectively for patients with breast cancer for the ipsilateral lung and heart for 97% and 83% of patients, respectively. With respect to dose calculation accuracy, model-based dose calculations are used for photon and electron algorithms for 94% and 39% of the plans, respectively.

As of December 2015, data have been submitted for approximately 90% (5775 patients) of the 6436 eligible cases. These data can be used to measure many endpoints, including adherence to recommended guidelines. For example, the prescription dose and fractionation information acquired through the physics web-based forms was used to determine rates of hypofractionation for patients with breast cancer who met the eligibility criteria specified by American Society for Radiation Oncology’s Choosing Wisely campaign.15

Discussion

This report has summarized the experiences of a statewide collaborative quality consortium in developing the infrastructure necessary to collect treatment planning information for assessing and potentially improving the quality of radiation oncology care. Other quality collaborative groups have been developed in fields outside radiation oncology and have demonstrated significant improvements in the quality of patient care.13,16 The consortium provides an excellent model for development of robust systems that
capture complex radiation treatment planning information from different sites that are diverse in many ways. This model is applicable to development of an analogous database in any state or country.

Within radiation oncology, an infrastructure which supports data collection for the National Institutes of Health-sponsored trials, including patient images and DICOM-RT plan data to support robust quantitative analyses, has been developed. The Integrating the Healthcare Enterprise-Radiation Oncology initiative has been working steadily on connectivity among different systems in radiation oncology for safe quality care for many years. The National Radiation Oncology Registry had created a registry for prostate cancer patients, with some of the initial efforts focused on creating a data dictionary and identifying the data to be collected. Others have developed in-house software systems to automate data collection at their own institutions. For example, the OncoSpace framework addressed data integrity challenges regarding individual treatment plans and delivered dose by building a data-mining framework that can be connected to toxicity data. They demonstrated its application to 684 head and neck cancer patients treated in their clinic. In spite of the lack of interconnectivity across systems with the database, the consortium was able to gather evidence within a customized registry that represents the variations in patient care that are often deliberately excluded from clinical trials or that are challenging to investigate outside of an individual institution.

In this work, we provided our experiences building a system that includes details about the simulation, planning, and treatment delivery alongside physician and patient information as a model for other statewide databases. Skripacak et al eloquently laid out the requirements to facilitate automated data collection for federated databases for research purposes. In the consortium, we have tackled many of those issues to build a functional system for quality assessment customized to breast and lung cancer patients in the state of Michigan. We strongly agree that a standard framework for data collection is essential. In our registry, we

![Figure 5](image)

Figure 5: The distribution of (A) the use of a breath hold technique for left-sided breast cancer patients (n = 2392 patients) and (B) the use of a motion management method for lung cancer patients (n = 1035 patients).
are collecting information beyond what is included in DICOM-RT such as the use of motion management to assess the current use of state of the art techniques.

Our database has provided information about different practice patterns that is unavailable without a registry or based on DICOM-RT data alone. Our consortium found that 30% of patients with treatment to the left breast are treated using deep inspiration breath hold with a device. For patients with lung cancer, we discovered that an internal target volume was created for 45% of patients compared with no motion management for 52% of the patients. Guidance from the American Association of Physicists in Medicine Task Group 76 on the management of respiratory motion has been available since 2006. This is the first evidence to our knowledge of the variation in the use of devices for motion management between patients with breast and lung cancer. It is also the first demonstration in a statewide setting of the low adoption of formal motion management methods. The use of the deep inspiration breath hold technique in multi-institutional settings was previously unknown, and these data are being used to guide further quality improvement projects by the consortium. Similarly, information regarding the use of modern algorithms for electron calculations for patients has otherwise been unknown outside of clinical trials.

In our consortium, we also developed an audit process to support overall data integrity. These assessments led to the development of an automated data checker that was launched in September 2014. The abstractor is able to run it at the conclusion of data entry for each patient to confirm that no errors were present for the supported fields. This is valuable and timely feedback for manually entered data.

Analyses of these data can help improve RT outcomes for patients in 3 ways: (1) by assessing toxicity, dose prescription compliance with national guidelines, and adherence to planning limits for organs at risk; (2) by correlating practice variations with outcomes; and (3) by revealing additional patient characteristics and treatment variations that improve outcomes. Many ongoing quality improvement projects are using the database. Primary endpoints have ranged from: (1) efficacy of IMRT reducing moist desquamation and pain for patients with breast cancer, (2) methods that reduce esophagitis and pneumonitis for patients with lung cancer, and (3) the use of motion assessment for all patients. Other MROQC projects include a quantitative evaluation of the use of IMRT techniques compared to conformal techniques among practices in the state of Michigan and analysis of dose-based correlations to physician and patient-reported outcomes. The consortium continues to actively collect data for eligible patients.

Commercial systems can facilitate these and similar efforts with enhanced tools by using standard names and supporting composite DVHs (ie, all treatment plans are created, stored, and exported for structures with same names). The American Association of Physicists in Medicine TG-263 is currently discussing and compiling the standard names for these entities. These challenges must also be addressed for adaptive therapy in the broader radiation therapy community. The consortium model can apply to other interconnectivity efforts for clinical trials and federated databases.

**Conclusions**

MROQC has successfully created a consortium for quality improvement in radiation oncology. A web-based system allows for consistent data entry independent of the treatment planning or delivery systems. This system was engineered with logical checks and in-person audits, which resulted in high-quality data until better interconnectivity and more automation is available for data collection. As a consortium, we collectively developed and shared solutions for the data collection. We used the web-based system to collect data for almost 5800 patients by consortium members as of December 2015.

The engagement of a multidisciplinary team from multiple institutions has been essential to the consortium’s success. Example patterns of care data that can be evaluated include variations in the use of model-based dose calculation algorithms, plan complexity in IMRT, image-guided RT strategies, and the performance and use of motion assessment information during treatment planning and delivery. Through collaboration and assessment, these data can be used to enable the crucial analysis of the relationship between treatment technique specifics, patient outcomes, and toxicities.

**Acknowledgments**

The authors acknowledge the contributions of Jennifer Steers (dose-volume histogram analysis tool), James Irre...
References

10. MITA. Digital Imaging and Communications in Medicine (DICOM), Parts 1-16. 2015.