Particle therapy in head and neck cancer

Development and evaluation of an online three-level proton vs photon decision support prototype for head and neck cancer – Comparison of dose, toxicity and cost-effectiveness

Qing Cheng a, 1, Erik Roelofs a, 1, Bram L.T. Ramaekers b, Daniëlle Eekers a, Johan van Soest a, Tim Lustberg a, Tim Hendriks a, Frank Hoebers a, Hans Paul van der Laan c, Erik W. Korevaar c, Andre Dekker a, Johannes A. Langendijk c, Philippe Lambin a, b, *

a Department of Radiation Oncology (MAASTRO Clinic), GROW – School for Oncology and Developmental Biology, Maastricht University Medical Center; b Department of Clinical Epidemiology and Medical Technology Assessment (KEMTA), Maastricht University Medical Center; and c Department of Radiation Oncology, University of Groningen, University Medical Center Groningen, The Netherlands

A R T I C L E I N F O

Article history:
Received 22 May 2015
Received in revised form 1 December 2015
Accepted 5 December 2015
Available online 26 February 2016

Keywords:
Proton therapy
Decision support system
Quantitative comparison
Mean dose extraction
Toxicity prediction
Cost-effectiveness estimation

A B S T R A C T

To quantitatively assess the effectiveness of proton therapy for individual patients, we developed a prototype for an online platform for proton decision support (PRODECIS) comparing photon and proton treatments on dose metric, toxicity and cost-effectiveness levels. An evaluation was performed with 23 head and neck cancer datasets.

Due to the continuous development of new cancer treatments and the sophistication of existing radiotherapy, it has become increasingly challenging to identify the best treatment for a specific patient [1]. A multifactorial clinical decision support system (CDSS) could help meet this challenge when combining clinical, dosimetric and cost variables (e.g. information about the patient or tumour) with expert knowledge (e.g. on a specific treatment modality) to make a quantitative treatment comparison [2–7]. Such a tool would facilitate individualised radiotherapy treatment.

Given its favourable dose distribution, proton therapy is expected to be less toxic and more effective than photon therapy [8–10]. As a result, many oncology centres around the world have introduced proton therapy over the last decade [11]. However, planning studies show that not all patients would benefit from this more expensive treatment [12,13]. Clinical data-exchange platforms have been developed previously to justify patient stratification for a fair and efficient use of the treatment [14–16]. However, its cost-effectiveness compared to conventional photon radiotherapy is yet unevaluated for many cancers [17–19].

Dutch health authorities have agreed upon the need for a model-based indication methodology to select patients eligible for proton radiotherapy [20–22]. Supplementary Fig. 1 illustrates a Dutch decision tree regarding proton therapy reimbursement. It determines whether a patient is expected to benefit sufficiently from proton therapy justifying reimbursement of the treatment costs. For an effective and efficient evaluation of these aspects, a CDSS is needed that supports the claim whether or not proton therapy is expected to have a clinical benefit in a given patient.

We postulate that such a CDSS should have at least three levels. The first dosimetric level should evaluate whether a radiotherapy plan meets predefined dosimetric threshold for a patient’s organs at risk (OARs). The second toxicity level should estimate whether the probability of radiation induced normal tissue toxicity for the patient is different between different treatment plans. The third cost-effectiveness level should evaluate if the extra costs for a certain increase in effectiveness does not exceed a threshold set by society. The effectiveness is defined in quality-adjusted life years (QALYs), which are calculated by estimating the quality and quantity of life extended by a medical intervention [23].

Corresponding author at: Department of Radiation Oncology (MAASTRO), GROW, Maastricht University Medical Centre, Dr. Tanslaan 12, NL-6229 ET Maastricht, The Netherlands.
E-mail address: philippe.lambin@maastro.nl (P. Lambin).

1 These authors have had an equal contribution to the manuscript.
To this end, we developed an online, three-level photon vs. proton CDSS prototype named PRODECIS (PROton DECIsion Support). In this study, we evaluated the system's performance for patients with head and neck cancer (HNC). Data are provided online on www.cancerdata.org [24].

Materials and methods

We designed a modular CDSS (Fig. 1) to support the decision between proton and photon therapy. The system was implemented in Java SE 7 (Oracle, Redwood Shores, CA, USA) and Matlab 2010b (Mathworks, Natick, MA, USA) and designed to import photon and proton treatment plans in DICOM-RT format. A PHP webform was created to upload the data and additionally ask for clinical parameters of the complication models. All patient information and results were anonymously stored in a MySQL Workbench 6.0 (Oracle, Redwood Shores, CA, USA) database.

Computation services were separated into three levels. On the dosimetric level, we adopted in-house dose-volume histogram (DVH) metrics calculation algorithms to extract mean doses from both photon and proton plans. On the toxicity level, we used a number of validated late toxicity prediction models using the TRIPOD Type 4 standard [25] (e.g., regression models [23,26,27]). On the cost-effectiveness level, we incorporated published Markov models1 ([23]) to assess the QALY and costs of the treatment.

Experimental setup

To test the system, we used datasets from a ROCOCO cohort of 25 HNC patients for whom both photon and proton plans were available [13]. First, on the dosimetric level we computed the dose to the supraglottis area, the superior pharyngeal constrictor muscle (PCM), and the ipsi- and contralateral parotid glands. Then, on the toxicity level we estimated the normal tissue complication probability (NTCP) for xerostomia and swallowing dysfunction at 6 and 12 months after therapy, using the models published in previous work [23,26,27]. Since the parotid gland location was indicated with left or right in the given datasets, we defined the ipsi- and contralateral parotid glands as receiving higher or lower doses, respectively. Finally, on the cost-effectiveness level we used a Markov model constructed for HNC patients [23] with pretreatment RTOG grade 2–swallowing dysfunction and xerostomia. The model is described in Supplementary Table 1.

Threshold definition

For the purpose of treatment comparison, we collected various thresholds to define clinical benefit. On the dose comparison level, from expert opinions and literature, we defined a clinical benefit when a plan met clinical, desirable OAR mean dose thresholds being parotid gland <26 Gy, superior PCM <50 Gy and supraglottis area <50 Gy [28–30].

On the toxicity level, based on the CTCAEv4.0 toxicity criteria, we considered clinical benefit as a predicted reduction in probability of grade 2+ toxicity of >10%. In addition, we used the definition of a “complication profile” where, for each patient, the toxicity probability reductions exceeding 5% were summed and clinical benefit was set at a total reduction of 15% or more [31].

On the cost-effectiveness level, we set the acceptable cost per additional QALY derived from the Markov model at €80,000. This is the official threshold proposed in the Netherlands by the Dutch Council for Public Health and Care [32].

Statistics

We used two-tailed Wilcoxon signed rank tests to determine whether the differences between plans were significant. P-values of less than 0.05 were considered significant.

Results

System development

The PRODECIS prototype was successfully built on a pipelined image processing framework [33] from within our institute. For scaling purposes, each level of computations was encapsulated into a module and was then installed identically in two parallel pipelines (A and B in Fig. 1). After the whole plan of a treatment was transferred, the respective pipeline began computing. Once both computation pipelines were done, the results were delivered to the third pipeline, comparing the multilevel results with the defined threshold per level. Finally, the comparison results were emailed back to the user. From the 25 datasets, the calculations did not succeed for two, due to DICOM compatibility issues. For every patient, it took approximately five minutes for a computer with standard specifications (Intel® Core™ i5-3210M CPU processor with 2.5 GHz, 4 GB memory) to finish all given tasks.

Experiment results

The system proved successful in the automatic evaluation of proton treatment eligibility according to the model-based approach and predefined thresholds. The number of cases where proton therapy ranked higher as well as average outcomes for both modalities are summarised in Supplementary Table 2. In Fig. 2, the individual results are shown for toxicity and cost-effectiveness, relative to the defined thresholds.

On the dosimetric level, proton therapy significantly lowered doses to the OARs, except for the superior PCM. For the latter, only the proton plans stayed below the thresholds for 2 cases, whereas these were 4, 5 or 12 when considering the supraglottis area, ipsi- or contralateral parotid glands, respectively.

On the toxicity level, proton therapy significantly reduced all toxicities. On average, the probability of swallowing dysfunction 6 months after treatment was reduced from 37% to 28% and from 23% to 18% at 12 months. The probability of xerostomia was reduced for all 23 cases after treatment: from 48% to 25% at 6 months and from 46% to 23% at 12 months. With combined toxicity thresholds, protons outperformed photons for 23 cases at 6 months and 21 cases at 12 months.

On the cost-effectiveness level, we observed an increase in QALY for all the patients in their proton therapy plans, although it was also significantly more expensive. Using the nationally accepted criterion of €80,000 per QALY gained, proton therapy was found to be cost-effective for 8 of the 23 patients.

Discussion

We successfully developed and evaluated the PRODECIS prototype to comply with the Dutch model and added the option to evaluate cost-effectiveness. The study shows that, given nationally accepted guidelines for 15% reduction of a complication profile including swallowing dysfunction and xerostomia, all patients would benefit from proton therapy after 6 months and 91% after 12 months, while 35% would be considered cost-effective at a threshold of 80,000€ per gained QALY. Although a CDSS was previously applied [34,35], we have not found any application that could make quantitative decision-making about photon vs. proton therapy at three levels.

1 Available on www.predictcancer.org.
A key characteristic of the system is its parallel pipeline structure, which allows easy extension by reusing the modular code. Another important feature is the dynamic selection of models based on the tumour type. Such flexibility enables the system to rapidly adapt to different user requests and incorporate new insights from the oncology society. Provided the availability of relevant prediction models, future studies could perform systematic experiments to search for an optimal outcome among multiple treatment options at any anatomical site. The third system feature is its use of the Markov model. It consists of health stages in terms of toxicity RTOG grade and translates toxicity probabilities into transitions between health stages (Supplementary Fig. 2). Through the transition of a patient’s health state after treatment, the model estimates the costs and effects of the treatment. An adapted version, referred to as micro-simulation, was developed to predict survival of patients with non-small-cell lung cancer [36]. A recently published study [37] shows the same approach to estimate cost-effectiveness of the use of spacers when treating prostate cancer.

Multiple advantages of using the PRODECIS CDSS are foreseen. First, it provides the opportunity for a clinician to make a model-based decision following the Dutch guidelines. Second, it allows clinics to quantitatively prioritise the limited treatment slots and allocate them to the patients expected to gain the most from proton therapy [17,38]. Third, it quantifies clinical evidence for health insurance policy development. Furthermore, it can help in evaluating the cost-effectiveness of deploying a new technology. A final point to note is that consent for data-exchange to the proposed online system can readily be asked from the patient who is being considered for proton therapy and has a direct benefit of the reuse of its data.

However, these advantages will only be achieved when the following concerns are addressed sufficiently. As the system is still in an early stage, extension to a fully operational system offering user management is required to account for audit trails, for instance. As with the MISTIR platform, security measures for encrypted data transfer are to be provided [14]. Furthermore, the system offers a single-shot evaluation and currently lacks proper case management to retrieve previous comparison outcomes for re-evaluation. Similar to the ReCompare system [15], the PRODECIS platform is targeted towards referring photon therapy centres, accepting previously calculated photon treatment plans for comparisons. However, PRODECIS also uses user-provided proton plans, whereas ReCompare uses the proton plan generated by the operating proton therapy centre (PTC) itself. Such service can optionally be provided by the staff backing PRODECIS, but as the comparison is performed automatically using standardised models and thresholds an inde-
pendent evaluation of plan quality and prediction of complication rates is offered to other PTCs as well.

A prerequisite for the automatic numerical evaluation of PRODECIS is that the uploaded treatment plans should adhere to strict protocols, specifying contours per tumour group. The use of so-called “umbrella protocols” and international naming convention guidelines will facilitate data exchange in a reusable fashion [39–42]. Although quality assurance methods are implemented, such as contour name mapping, major violation of the protocol will prohibit evaluation, requiring corrections by the user.

It is foreseen that with current efforts from both community as well as industry, instead of calculations based on user-provided treatment plans, automatic plan generation could be applied [43,44]. As an alternative to automatic planning, the estimation of DVH parameters might be reliable and fast, given a sufficiently large historical database of patients with the best planning [as in the study of proton therapy plan optimisation [45]]. Patient-selection using a comprehensive matching mechanism based on essential patient characteristics including clinical aspects, tumour location and organ distribution is considered to be incorporated into the PRODECIS system as shown in Supplementary Fig. 3. This will greatly improve the workflow, avoiding the resource-intensive bottleneck of double treatment planning.

A critical factor of the model-based selection method is the quality of the treatment plans under evaluation. Therefore, we expect realistic clinical-grade (thus not “beyond-state-of-the-art”) treatment plans that would be administered to the patient in real practice. This means planning protocols need to be up-to-date and in line with the technical possibilities of the treatment options. As for the experiment, the published proton plans for 25 HNC patients are not considered as current standard anymore. We have now produced robust treatment plans (unpublished data), which produce not dramatically different but more realistic proton plans, where in some cases the differences are clinically relevant. To further evaluate the system, we will experiment with external datasets from different centres using different treatment techniques.

Furthermore, in the current prototype, the system only considered those toxicities for which reliable NTCP-models were available and that connected to the cost-effectiveness model. Additional models can easily be added including more OARs such as oral cavity, brainstem, or area postrema to predict acute and late radiation-induced toxicities, which may likely be reduced by proton therapy as well and could mean that the cost-effectiveness of protons will be underestimated. For instance, the first comparison of IMRT versus IMPT among oropharyngeal cancer patients treated with chemo-radiation in the MD-Anderson Cancer centre [46] showed a significant decrease of required tube feeding during the course of radiation when IMPT was used. In this regard, direct measurement of QALY’s in prospective data registration programs is needed to obtain better insight into the cost-effectiveness of protons. In addition, to maximise system utility, it is highly desirable to use toxicity models that consider multiple stages. e.g. A reduction of grade 4–5 toxicities is of utmost clinical benefit, but the number of patients is too low to train such a model reliably, which requires international data pooling or rather distributed learning systems [47–49].

The HNC Markov model adopted in this system depended on acceptable costs, which vary from country to country and even from hospital to hospital. It also depended on toxicity estimation models that were regressed without patients’ biomedical data. Furthermore, previous interventions such as surgery or chemotherapy were not included in the system, which will bias the complication predictions. Therefore, service at this level is a proof of principle and not conclusive.

The next step will be to include genetic biomarkers of radiosensitivity to further improve the prediction of late toxicities [50,51]. We aim to continuously update the system with additional models that apply to other diseases and are scalable to other countries. Finally, patient-specific data such as molecular information, patient-reported outcomes and personal preference should be incorporated to truly improve the level of personalisation in decision support systems.

Conflict of interest

None.

Acknowledgments

We thank MAASTRO’s software development team for their support. Furthermore, we acknowledge the involvement of colleagues from the clinic’s departments, the data management team and treatment planning. The authors acknowledge financial support from the EU 7th framework programme (ARTFORCE – no. 257144, REQUIRE – no. 601826), an NCI Pre-Seed Grant (no. 93612005) and Kankeronderzoekfonds Limburg from the Health Foundation Limburg.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.radonc.2015.12.029.

References


