Robustness evaluation of pencil beam scanning proton therapy treatment planning: a systematic review

E. Sterpin[1,2,3,*], L. Widesott[4], K. Poels[3,5], M. Hoogeman[6a, 6b], E.W. Korevaar[7], M. Lowe[8], S. Molinelli[9], F. Fracchiolla[4]

[1]KU Leuven – Department of Oncology, Laboratory of Experimental Radiotherapy, Leuven, Belgium

[2]Université catholique de Louvain – Center of Molecular Imaging, Radiotherapy and Oncology (MIRO), Brussels, Belgium

[3]Particle Therapy Interuniversity Center Leuven – PARTICLE, Leuven, Belgium

[4]Proton Therapy Center - UO Fisica Sanitaria, Azienda Provinciale per i Servizi Sanitari (APSS), Trento, Italy

[5]UZ Leuven, Department of Radiation Oncology, Leuven, Belgium

[6a] Erasmus Medical Center, Cancer Institute, Department of Radiotherapy, Rotterdam, The Netherlands

[6b] HollandPTC, Delft, The Netherlands

[7]Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, The Netherlands

[8]Christie Medical Physics and Engineering, The Christie NHS Foundation Trust, Manchester, UK

[9]Fondazione CNAO - Medical Physics Unit, Pavia, Italy

*Corresponding author: Edmond Sterpin; edmond.sterpin@kuleuven.be

Abstract

Compared to conventional radiotherapy using X-rays, proton therapy, in principle, allows better conformity of the dose distribution to target volumes, at the cost of greater sensitivity to physical, anatomical, and positioning uncertainties. Robust planning, both in terms of plan optimization and evaluation, has gained high visibility in publications on the subject and is part of clinical practice in many centers. However, there is currently no consensus on the methods and parameters to be used for robust optimization or robustness evaluation. We propose to overcome this deficiency by following the modified Delphi consensus method. This method first requires a systematic review of the literature. We performed this review using the PubMed and Web Of Science databases, via two different experts. Potential conflicts were resolved by a third expert. We then explored the different methods before focusing on clinical studies that evaluate robustness on a significant number of patients. Many robustness assessment methods are proposed in the literature. Some are more successful than others and their implementation varies between centers. Moreover, they are not all statistically or mathematically equivalent. The most sophisticated and rigorous methods have seen more limited application due to the difficulty of their implementation and their lack of widespread availability.

Introduction

Proton therapy (PT) has the potential to improve dose conformity compared to X-ray radiotherapy (XRT) due to the finite proton range and the Bragg peak [1]. However, PT is more sensitive to density changes and uncertainties encountered in clinical workflows, including CT image conversion, dose calculation, and motion. The use of scanned beams and multi-field optimization enhances dose conformation and flexibility in PT compared to treatment delivery strategies based on broad beams and patient-specific collimators/compensators. However, it also increases the risk of underdosing the target or overdosing organs at risk (OARs) due to the above-mentioned uncertainties [1]. Furthermore,

the interplay of scanned pencil beam delivery and breathing motion can result in degradation of the delivered dose distribution [2].

In XRT, safety margins around the clinical target volume (CTV) or OAR are typically used to manage uncertainties, creating the planning target volume (PTV) and planning organ at risk volume (PRV). The PTV is essential for dose prescription, treatment plan optimization, and plan evaluation based on ICRU reports 62 and 83 [3,4]. This approach is valid when the assumptions applied to the use of PTV margins are respected. This is approximately the case in XRT (with many exceptions such as breast, small lung tumors, head and neck, etc.). In PT, one of the main assumptions behind the PTV approach, that the dose distribution is unchanged by patient shifts (the static dose cloud approximation), often represents a non-realistic approximation in clinical routine. Consequently, the PT community has developed specific optimization and reporting methods [5], including beam-specific safety margins [6,7] and robust optimization techniques [8]. These techniques involve simulating uncertainties such as patient positioning and range errors and integrating them into treatment plan optimization to ensure objectives are met under both ideal and uncertain conditions. Such approaches have been covered extensively in a review by Unkelbach et al [8]. Robust optimization and evaluation have seen much wider implementation in PT than in XRT, as shown by Kaplan et al [9], though application to XRT is now starting to be introduced [10,11].

The paper of Kaplan et al. [9], published after an ESTRO physics workshop dedicated to the topic of "Plan quality assessment: dose distribution and robustness metrics", shows the variety of methodologies and metrics for robustness evaluations used in clinical workflows and the need for a consensus on how to implement them. Before the paper of Kaplan et al. [9], Hernandez et al have already stressed the importance of producing consensus for robustness analysis tools and dose evaluation metrics [12]. Within the European Particle Therapy Network (EPTN working group 5, focused on treatment planning), there is a willingness to converge to shared recommendations on robustness optimization and evaluation. This will follow a modified Delphi consensus methodology which first requires a systematic review on the topic to provide a comprehensive overview of the existing literature. The systematic review has been divided into two parts, one centered on robust optimization and the other on evaluation. The latter is the focus of the present work.

The aim of this systematic review is to provide a thorough overview of the various solutions proposed in the literature for robustness evaluation of PT treatment plans delivered with pencil beam scanning. This includes the simulation of the effects on the proton dose distribution of various errors, as well as their quantitative or visual representation. This systematic review also includes the applications of robustness evaluation according to various clinical sites.

Terminology and concepts

Definitions

We refer the reader to the definitions given in the publication of Lowe et al [13], where a glossary of terms related to uncertainties in PT is given. Here a few additional definitions and adaptations are provided:

- *Confidence level*: for a given probability density function, confidence intervals define a range within which a population parameter resides with a certain probability, the *confidence level.*
- *Scenario*: Lowe et al. define a scenario as "Potential or anticipated errors for which the dose distribution is recalculated. These may be used during robust optimization or plan evaluation. They are sometimes referred to as error scenarios." Here, we make explicit that we may also include the combined effect of errors (for instance random and systematic setup errors) over the full course of the treatment.
- *Scenario space*: A *N*-dimensional space that includes all possible scenarios. The number of dimensions, *N,* depends on the types and number of errors considered.
- *Dosimetric space*: All the dose distributions considered in robustness dosimetric analysis.

- *Range uncertainty and range error*: proton range uncertainties may come from multiple sources (image conversion to stopping powers, dose calculation, motion etc.). Strictly speaking, range errors should refer to realized errors, in the sense of a simulated range error or a measured range error. We have tried to keep this distinction in the current review. However, the literature often uses "range uncertainty" and "range errors" interchangeably, in the sense of a fixed scaling of the relative stopping powers.

The PTV/PRV concepts and robust optimization

The PTV concept, defined in ICRU 62, ensures sufficient coverage of the CTV considering uncertainties. However, ICRU does not specify a minimum threshold for the required probability of coverage, nor does it provide a methodology for assessing this probability. As stated by Kaplan et al [9], PTV margins are often based on historical choices in treatment centers, but some clinics compute margins based on quantified uncertainties and predefined probabilities of dose coverage.

While this paper does not aim to review the limitations of the PTV and PRV concepts [8,14], it acknowledges their role in enabling the prescription, optimization, and reporting of radiotherapy plans in a probabilistic manner. Under valid assumptions, the CTV should receive a predefined minimum percentage of the intended dose with a certain confidence level, while OARs should not exceed a maximum percentage/limit.

The present paper focuses on the reporting part, which could be seen as the last part of the planning process: robustness evaluation in PT. At the beginning of that planning process, we might see robust optimization instead. Robust optimization amounts to simulate explicitly uncertainties during the optimization process. This can be achieved by introducing, for instance, patient shifts and density scaling to simulate setup and range errors, respectively. A popular robust optimization approach is the worst-case optimization (minimax), which aims at minimizing the maximum value of the objective function over a set of scenarios defined by the user [15]. An extensive topical review on robust planning has been published by Unkelbach et al [8].

Systematic review description

We have followed the methodology proposed by Stoll et al [16]. Two independent experts selected the reviewed papers by using two databases. The first expert performed the search on PubMed and the second on Web of Science. Then the databases were merged and reviewed by both experts. In the case of disagreement between experts, a third expert took the final decision. In this way, the bias of using a single database has been minimized. The keywords used for the search on databases were as shown in Table 1 (the full search string can be found in the supplementary material "Logic strings used for search"). The results of the systematic review are shown in Fig. 1.

Key words	● Including: ○ Plan robustness ○ Robustness evaluation ○ Robust evaluation ○ Sensitivity analysis ○ Proton Therapy ● Excluding: ○ Proton pump
Inclusion criteria	● published between Jan 1 2016 and Oct 30 2023

	• clinical applications where a robustness evaluation method and clinically used parameters are shown and/or a new robustness evaluation methodology is introduced in literature • characterization/validation/retrospective application of existing robustness evaluation methodologies • reporting methodologies (publications where a method to report the robustness evaluation results are suggested) • describing the delivery of a pencil beam scanning (PBS) technique • Robustness evaluation of each type of uncertainty (setup, range, motion, anatomical variation, machine parameters).
Exclusion criteria	• case reports (unless describing a novel robustness evaluation methodology) • book chapters • letters • editorials • conference papers • reviews • full text unavailable • robust optimization only • describing delivery techniques including Arc therapy, FLASH (Ultra-high dose rates), mini-beams, passive scattering.
Table 1: keywords and inclusion/exclusion criteria used for the systematic review	

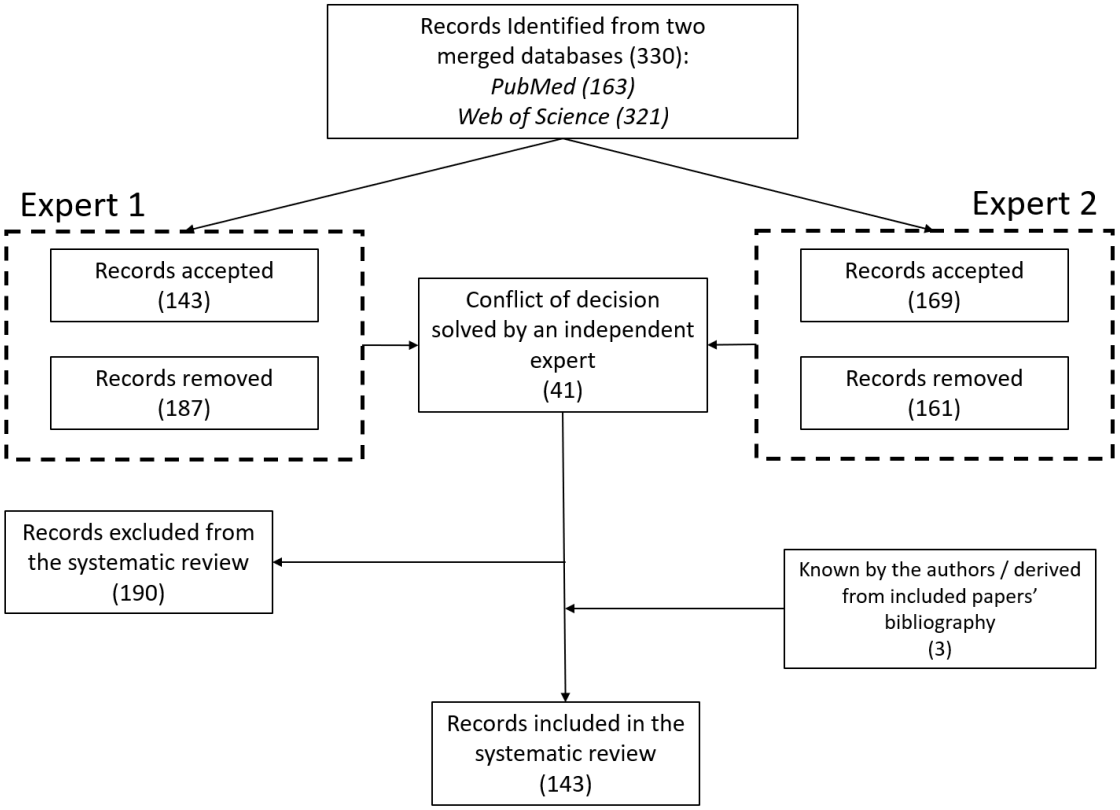

Fig. 1 Results of the systematic review. Experts 1 and 2 have performed the research using either PubMed and Web of Science, respectively.

Robustness evaluation methodologies

Evaluation against systematic setup and range errors

As reported in the survey of Kaplan et al [9], there is no agreement on which shifts should be considered in the assessment of robustness. In general, errors are defined in the cardinal directions (left-right, anterior-posterior, cranio-caudal). Those errors may be combined and may be described as lying on the surface of a cube or sphere. The simulation of setup errors in the cardinal directions alone has been reported in several studies [17,18,27–36,19,37,20–26]. The combination of errors has also been proposed [35]. Range errors may be simulated independent of setup errors [19,32,34,36–41] and/or in combination with setup errors [17,18,30–33,35,36,42–45,20,46–49,21,22,24,26–29]. Other authors have also suggested a set of scenarios, some with only setup errors simulated, and the others with only range errors [25]. Shang et al simulated two sets of scenarios for the setup errors: cardinal points and another set of diagonals [50,51]. Maes et al also added diagonal points to their robustness analysis [35]. Those were combined with range errors. Noufal et al and Boer et al evaluated on the diagonals alone, in combination with range errors [52,53], while robust optimization was performed using the cardinal points. Tommasino et al selected the eight vertices of a cube for the setup errors, with a distance compatible with the PTV margin used for optimization, and combined them with range errors [54]. Lowe et al, Albertini et al and the references cited therein presented a method where 14 possible setup errors are selected [39,55]. A scenario is eligible if the euclidean norm of each error equals the distance corresponding to an 85% confidence level of all possible distances. By taking the maximum dose deviation for each voxel over all scenarios, a worst-case dose distribution is obtained, from which error-bar volume histograms, analogous to dose volume histograms can be derived. Range errors are simulated separately, following a similar methodology.

Korevaar et al have proposed a refinement for the selection of the simulated errors, especially with respect to the range uncertainties [5]. As described by both Korevaar et al and Sterpin et al, joint probabilities of setup and range uncertainties improve the statistical consistency of the selected scenarios [5,56]. Korevaar et al suggested the joint selection of the setup and range errors on a 4D iso-probability surface. This leads to the definition of between 14 and 26 directions, which can be combined with positive and negative range errors (leading to 28 or 52 scenarios). Several recent studies have based their robustness evaluation strategies on this work [50,57,66,67,58–65].

In summary, most methods employed in the papers cited in this section are either designed according to a given commercial software or compatible with them with minimal scripting effort, if adequate. Clinical implementation is therefore not a major difficulty. However, this comes with limitations about the statistical meaning of the reported robustness. Although there is no clear agreement about what could constitute an acceptable minimal robustness evaluation method (ie, the combinations of errors that should be simulated), several authors have pointed out the conceptual issues associated to most methods and have proposed alternatives, as listed in the coming two sections.

Probabilistic robustness evaluation

A comprehensive statistical approach involves sampling errors from their distribution without preselection, encompassing both systematic and random errors. Implementing this approach requires considerable computing effort, often accomplished through a brute-force Monte Carlo method with a large number of scenarios [68–70]. The number of scenarios can be estimated using statistical convergence criteria to ensure accurate and sufficient dose calculations [56,71,72]. To expedite the process of sampling scenarios, variance reduction techniques such as importance reweighting can be employed [73]. Another approach, proposed by Perkó et al, uses polynomial chaos expansion to construct a model based on dose distributions calculated for a predefined set of errors [74]. This model enables a fast analytical calculation of the dosimetric impact of a wide range of realistic setup and

range error scenarios. When it comes to analyzing random errors, they can be calculated by either sampling or incorporating them analytically, assuming an infinite number of treatment fractions. This comprehensive approach has found application in various fields, demonstrating its effectiveness in error analysis and dose calculation scenarios [57,75,76].

Yang et al presented a comprehensive statistical evaluation of plan robustness against 625 scenarios with randomly sampled systematic setup and range errors [77], enabling the evaluation of percentiles of key dose metrics.

Wahl et al presented an analytical approximation of brute-force sampling for error scenarios [78]. Their approach used a dose calculation algorithm based on pencil beam convolution, allowing for the analytical integration of Gaussian probability densities associated with simulated uncertainties. By estimating the moments (average and co-variance) of the probability distribution, the impact of uncertainties on the dose distribution could be estimated at a lower computational cost.

Finally, Vazquez et al have developed a deep-learning based approach to predict plan robustness from nominal dose distributions [79], at very high computation speeds (within 3s). The objective was to predict the 5th and 95th percentile doses. Excellent agreement was obtained.

In contrast with the previous section, methods presented here rely on considerable implementation efforts, with software developed in-house. The methods presented enable quantified estimates of confidence levels. Important questions remain, about the importance of computing confidence levels and, if judged relevant, about the details of the methods used to compute those levels.

Integration of random setup errors and fractionation into robustness analysis

The integration of random errors is more complicated as their effect will vary with the number of fractions in a way that is not entirely predictable since it depends on the (random) sequence of errors.

A pragmatic approach is to simulate systematic setup errors as in section IV.A, but to change the magnitude of the simulated error to compensate for the effect of further random errors. Several authors have discussed the use of the simplified van Herk recipe $M_{PTV} = 2.5\Sigma + 0.7\sigma'$, where Σ represents the standard deviation of systematic errors and σ' represents the standard deviation of random errors [80]. The latter does not explicitly take into account the beam penumbra, which is assumed to be 3.2 mm and implicitly taken into account in the factor 0.7 [5,56,81]. This is somewhat implicit in several methods in the literature, which simulate setup errors that are similar to their PTV margins counterparts.

Lowe et al have integrated fractionation in the robustness analysis [39]. This has been achieved by simulating the convolution of error dose distributions obtained from 14 preselected scenarios by themselves, as many times as the number of fractions. This leads to an overall tightening of the envelope of uncertainty. Further work from the same group has also investigated the effect of range errors and fractionation for various field configurations [82].

Ribeiro et al have implemented another approach where multiple fraction treatment courses are evaluated taking systematic setup and range errors based on established values, and then sampling a random setup error for each fraction and accumulating over the full course of the treatment [83]. Teoh et al have randomly selected 3 possible realizations of systematic setup errors (within a limit of ± 3 mm) and then simulated 35 fractions with associated (sampled) random errors [84]. Sampling was performed from Gaussian distributions. Range uncertainties were not included in the analysis.

The integration of random setup errors and fractionation effects is natural in the probabilistic approaches presented in the previous section. For simpler but widely available robustness evaluation tools, it is also possible the effect of fractionation using scripting capability of commercial software or post-processing analysis. The van Herk margin recipe was also cited as a possibility. The theoretical guarantees of those more pragmatic approaches with respect to the computation of confidence levels, and the statistically sound integration of other errors like the proton range, were in general not addressed.

Integration of inter-fraction motion

The issue of inter-fraction motion from the point-of-view of treatment adaptation has been discussed in length in the recent review from Paganetti et al [85].

The impact of inter-fraction motion on robustness has been evaluated both in isolation and in combination with other errors, for instance, through a robustness evaluation of repeated images against setup and/or range errors.

Studies evaluating the effect of inter-fraction motion alone (without range and residual setup errors) have been performed for prostate [86], liver (with breath-hold) [87], sinonasal cancer [88], head and neck [89,90] and pelvic lymph node [91] irradiations. Those studies rely on the acquisition of repeated images. In contrast, Zhang et al have developed a probability model for simulating possible anatomical changes during the course of the treatment, without the need of repeated images [92]. Such a method can be used for both robust optimisation and evaluation.

Hague et al. conducted a robustness evaluation for head and neck treatments using repeated images with setup and range errors [17]. They used smaller setup errors for repeat images compared to the initial robustness evaluation on planning images, based on setup errors already being partially accounted for in the repeated images. Similar approaches were followed by other authors [59,61,62,64]. Gort et al have studied the combined effect of inter-fraction motion, residual setup errors, and range errors for cervical cancers. Voxel-wise minimum doses (i.e. taking, for each voxel, the lowest value over all error scenarios) were accumulated to evaluate the robustness of target structures, while only nominal doses were used to evaluate OARs [60]. Scandurra et al. evaluated nasopharyngeal cancer treatments by combining residual setup errors, inter-fraction motion, and range errors, accumulating voxel-wise maximum and minimum doses for each robustness evaluation on repeated images [58].

In the context of prostate irradiations, Walsh et al have performed a comprehensive analysis reflecting random and systematic rigid displacements of the prostate [93]. Setup and range errors were not incorporated in the analysis. Robustness was quantified using the difference between the nominal and the median tumor control probability evaluated from each scenario.

There is a disparity in the methods proposed/reported for investigating inter-fraction motion. The effect of inter-fraction motion is often studied alone (without additional errors), using dose accumulation. Residual setup errors and (full) range errors may also be simulated on an individual fraction basis (thus without dose accumulation). The use of residual setup errors – thus smaller than the ones used for planning images – on repeated CT or CBCT images is widely adopted. Some authors have attempted to integrate residual setup errors and range errors during the dose accumulation process. Accumulating voxel-wise worst-case minimum/maximum doses for targets/OARs was therefore suggested.

Integration of intra-fraction motion

Various studies on the treatment of moving targets have highlighted the need to perform a robustness analysis that takes into consideration the uncertainty due to intra-fraction motion [59,83]. The main source of intra-fraction motion studied is that of breathing motion [64,66,94–101], though some studies have attempted to evaluate the interplay effect in the presence of heart motion [102].

A common method to assess robustness in relation to breathing motion involves calculating the plan on each phase of a 4DCT scan and deforming the dose to a reference CT [61,63,64,100,103]. Similar evaluations have been conducted for patients in breath hold [47,87,99,104]. However, calculating the entire plan on each phase does not consider the interplay between spot delivery and patient motion. More realistic simulations have been performed by incorporating time parameters and patient breathing patterns to obtain a 4D dynamic dose distribution (4DDD), overcoming the assumption that the entire plan is delivered on a single phase [22,65,111,96,98,105–110]. Ribeiro et

al. introduced log files for treatment delivery [59,83], while Pastor-Serrano et al. emphasized the importance of large statistics of breathing patterns for realistic interplay effect assessment [112]. They also showed that lung plans based on an internal target volume (ITV) consistently failed to meet robustness requirements. An actual breathing signal measured with a pressure belt combined with delivery log files were also used by Smolders et al and Visser et al for 4D dose reconstruction [113,114], which enables evaluating the impact of motion and interplay on the dose distributions. Smolders et al have also investigated the impact of non-rigid registration and α/β uncertainties on NTCP. The impact of the former has been found one order of magnitude larger than the latter.

In this systematic review, few studies have been found to consider the uncertainties due to intrafraction motion together with setup and range uncertainties [59,66,71,94,106]. Buti et al. [115] have applied a Monte Carlo robustness evaluation developed by Souris et al for lung tumors [71]. Intra-fraction motion was simulated by computing the full treatment on every breathing phase and accumulating on the mid position CT after non-rigid registration. The scenarios were selected in the dosimetric space, favoring a statistically sounder representation of the sensitivity of the plans to uncertainties.

The simulation of intra-fraction motion alone has been widely investigated. Comprehensive simulation of intra-fraction motion, including the interplay effect, follows well-established methods. The questions remaining are more about how the uncertainties related to motion or from other sources can be adequately integrated into the process. For instance, the integration of the effect of fractionation, actual beam delivery dynamics (e.g. log files), or patient-specific breathing patterns is not systematic. Further extension to other errors like setup, couch and gantry mechanical uncertainties and range may require more sophisticated probabilistic robustness evaluation tools.

Integration of treatment delivery uncertainties

As well as uncertainties in the expected dose from a given fluence, there are some uncertainties in beam delivery which can affect the delivered dose distribution. These include machine parameters such as spot position and size.

The impact of spot positioning uncertainties on the robustness of treatment plans has been studied by Noufal et al [116]. Kopp et al have simulated a combination of uncertainties in setup, range, patient rotation and beam full width half maximum (FWHM) in air [117]. Kraan et al have studied the impact of variable spot sizes and spot spacing [118]. Errors in spot spacing were shown to have limited impact, while errors in spot size had more of an effect for larger spot sizes when the error was expressed relative to the nominal spot size.

The impact of actual treatment delivery errors can be integrated to some extent into robustness analysis by performing dose calculation using machine log-files as input for beam parameters. This has been demonstrated, for instance, by Ribeiro et al, who have integrated machine specific information in a comprehensive robustness evaluation strategy for thoracic tumors [59]. Smolders et al and Visser et al have also integrated treatment delivery log files in their 4D dose reconstruction method [113,114]. Newpower et al have developed a method based on machine learning that aims at predicting spot delivery errors as they would be recorded in the log files [49]. By combining those errors with setup and range errors, a more comprehensive representation of robustness could be achieved.

Two main themes could be distinguished during this part of the systematic review: studies focusing on the actual impact of treatment delivery errors on treatment plan quality; or studies aiming at integrating measured data about beam delivery (log-filed) into the robustness evaluation system. In practice, one could see the former as an actual robustness evaluation (how robust is my plan against errors in spot spacing?), while the latter focuses more on the monitoring of the robustness of the plan (is my plan still within specifications according to measured beam parameters?).

Integration of variable beam arrangements

Careful beam angle selection in PBS PT can be considered the first attempt to obtain a robust plan with respect to range, setup uncertainty and anatomical changes. However, evaluation methods may also be required to determine stable beam setups. This has been demonstrated, for instance, in the evaluation of the variability of the water equivalent path length (WEPL) as described by Andersen et al. for a cohort of prostate cases [91]. A similar approach was taken by Toramatsu et al [24]. Comprehensive robustness evaluation with respect to setup, range uncertainty and anatomical changes is another solution to discover the most stable beam angles as described by Feng et al [105]. The number of beams may also affect plan robustness and other important factors such as the linear energy transfer (LET) distribution, which can also be estimated as part of robustness evaluation as described by Shang et al [50]: going from 3 to 5, 7 and 9 fields in lung tumors they found a better robustness of the target coverage and low doses in lung. Also LET hot spots were significantly reduced in lung with the increase of proton beams, but a worsening of high doses in the spinal cord and heart was found.

Integration of LET and biological models

In this section, we focus on the integration of variable biological models in robustness evaluation procedures which simulate other errors, for instance setup and range errors. Bolsi et al evaluated the LET distributions in the presence of range errors [41]. Rana et al have studied the variability of LET and relative biological effectiveness (RBE) distributions with respect to setup, range, and breathing phases (without simulating the variability of the biological models themselves) [32]. Hofmaier et al have evaluated the effect of setup, range, and RBE uncertainties on the robustness of the plan using a variance-based uncertainty analysis [119]. Ödén et al introduced variable biological models into their robustness evaluation strategy [120,121]. Monte Carlo simulations were employed to simulate the effect of various radiobiological models and their uncertainties from the boundaries of dose values in each voxel according to the scenarios simulated for setup and range errors. The effect of fractionation was also considered based on the work of Lowe et al. [39]. Gutierrez et al performed separate simulations for physical errors (setup and range) and variable RBE to compare their relative effect with respect to nominal dose distributions [42].

This systematic review indicates that the integration of explicit and variable RBE models into robustness evaluation, including its interaction with other uncertainties like setup and range, yield RBE-weighted dose distributions that entail different clinical interpretations and that may directly impact clinical practice (by changing beam arrangements, robust optimization approaches, treatment fractionation schemes, etc...). The uncertainties related to the existing RBE models require caution, though.

Robustness analysis: visualization and reporting

According to our systematic review, common methods for visualizing and reporting robustness in radiation therapy plans include the display of dose-volume histogram (DVH) bands and the range of values for important dose metrics based on various setup and range error scenarios. Refinements include displaying dose-volume population histograms with color levels representing confidence levels [122]. Important DVH metrics can also be described statistically using nominal values and several features of the probability distribution of DVH points [69,70]. Obtaining a probability distribution function of the DVH points in a clinically acceptable time has been demonstrated using semi-analytical algorithms running on GPUs [69,119,122]. Another popular method is to display voxel-wise worst-case minimum (for target volumes) and maximum doses (for OARs), and the relevant associated DVH metrics [39]. The latter method enables a visual inspection of the location of potential failures in target coverage or OAR sparing. There are commercial solutions that provide such features.

Depending on the methods used, several metrics can be computed to provide an insight into the robustness level of the evaluated plan, for instance the percentage of scenarios respecting given DVH metric constraints/objectives. Confidence levels can also be computed in the scenario space. The method proposed by Korevaar et al selects scenarios on an iso-probability surface of a 4D hyper-ellipsoid defined at the 90% confidence level [5]. This hyper-ellipsoid includes 3D setup uncertainties and range uncertainties. They found a very strong correlation between CTV coverage in the robustness evaluation and PTV coverage in the nominal plan. Lowe et al have restricted setup errors within an 85% confidence level [39]. Error-bar DVHs have also been proposed from the latter work to display quantitative information about the voxel-wise worst dose errors from those pre-selected scenarios.

Review of robustness parameters used according to clinical sites

Table 2 provides a synthetic summary of the robustness methods used in publications focusing on a particular clinical site. It can be observed that most of the selected publications implemented robustness evaluation against setup and range errors (around 85%). About 21% of the papers have implemented some form of probabilistic evaluation. Most sources of uncertainties are well represented, except perhaps radiobiological effectiveness and LET which amounts for 7 of the selected publications. For each clinical site, extensive details are provided in the tables included in the supplementary material. The next sections provide a summary of the results of the systematic review for the clinical sites listed in Table 2.

Site	**Publ.**	**Worst-case**	**Prob.**	**SE**	**RE**	**Inter-F**	**Intra-F**	**IP**	**RBE-LET**
	#	(% of publications per clinical site)							
Head and neck	28	71	29	86	86	64	0	0	4
Brain and skull	10	50	50	100	80	30	10	0	20
Breast	5	80	20	100	100	40	60	0	20
Lung	30	83	23	87	87	47	80	50	7
Esophagus	10	90	10	90	90	60	70	30	0
Mediastinum	3	67	0	67	67	67	33	33	0
Abdomen and pelvis	21	76	0	81	81	48	33	29	5
All sites	**107**	**76**	**21**	**87**	**85**	**51**	**62***	**36***	**7**

Table 2 Summary of the robustness parameters used in clinical studies according to clinical sites. Extensive summary can be found in the appendix.

Publ: Publications; Prob: Probabilistic; SE: Setup Errors ; RE: Range Error; Inter-F : Inter-Fraction ; Intra-F : Intra-Fraction ; IP : Interplay effect
*Head and neck, Brain and skull were excluded from the total to compute the percentage

Head, neck, brain and skull

We found 28 articles that met our inclusion criteria focused on head and neck treatments [9,17,90,123–131,18,132–139,36,58,67,68,76,82,89]. The references, and the associated robustness evaluation methodologies and parameters can be found in Table S1.1 for head and neck. Most of the papers simulated worst-case scenarios with systematic setup errors of ±3mm and range errors of 3-3.5%. Three papers used smaller set-up errors (1mm, 2mm and 2.25 mm respectively) for their worst-case simulations [18,82,123]. Wagenaar et al have implemented a comprehensive robustness evaluation approach on repeated CTs, including hardware related uncertainties (CBCT isocenter, robotic table shifts, etc.) [67]. Some papers considered a Gaussian distribution approach for setup errors [68,76,131,134] others considered anatomical changes [36,58,67,129,137,139] and linear energy transfer (LET) as part of the robustness analysis [127].

Almost all published papers focused their attention on target coverage robustness evaluation and only a few discussed in detail the issue with respect to OAR [89,90,125,127–129,132].

Several papers related TCP and/or NTCP considerations to their robustness analysis. In Hamming-Vrieze et al [130] dose differences between nominal and scenarios were translated to TCP (2 models) and NTCP (15 models). A weighted average of the TCP/NTCP based on Gaussian distribution over the variance scenarios was calculated to assess the clinical effect of systematic errors on the population. They concluded that the estimated population impact of setup and range uncertainties on TCP/NTCP following VMAT or IMPT of oropharyngeal cancer patients was small for both treatment modalities. The use of TCP/NTCP models allows for clinical interpretation of the population effect and could be considered for incorporation in robust evaluation methods. Lalonde et al. [89] evaluated the impact of setup uncertainty reduction (SUR) and adaptation to geometrical changes (AGC) on NTCP when using online adaptive. SUR and AGC had a larger impact than AGC on reducing NTCP in online, however, SUR alone did not induce clinically meaningful toxicity reductions in any patient. Combining the benefits of both can yield significant reductions in long-term side effects. In Rojo-Santiago J et al [76], the moderate impact of geometrical and range errors on NTCPs (NTCP spreads < 3%) showed that the MBS approach is robust, therefore that the nominal plan can be generally used as a good estimator to refer patients for PT.

For brain and the surrounding structures, 10 papers fulfilled our inclusion criteria [27,42,52,57,88,140–144]. Table S1.2 provides similar information as for head and neck. Three articles simulated worst-case scenarios with systematic setup errors of ±3 mm and 3-3.5% range errors [42,140,142]. Smaller setup errors were reported in Argota-Perez et al (2 mm) and Zechner et al (reported smaller set up errors (i.e., x±0.9 mm, y±1.4 mm and z±1mm) for their worst-case simulation [27,88]. Argota-Perez has also focused on the impact of daily anatomy changes on robustness [88]. Probabilistic evaluation was considered in 5 publications [57,76,140,141,143]. Hofmaier et al also evaluated the impact of inter-observer variability (IOV) in target delineation on robustness and they found that the dominating contributions to overall uncertainty were either range uncertainty or IOV [140]. Nenoff et al reported the combined effect of anatomical and setup changes with and without daily adapted PT [143]. The performance of the Dutch robustness evaluation protocol was assessed in Rojo-Santiago et al (2021) [57] and Rojo-Santiago et al (2023) [76], by comparing the results obtained with the protocol against the probabilistic method developed by Perko et al [74]. The former paper found that the CTV D_{98} in probabilistic evaluations was on average 3% above the protocol result. The latter paper investigated in more depth the link between uncertainties, dosimetric values issued from the Dutch evaluation protocol (voxelwise minimum and maximum doses), and percentiles of dose metrics computed from the probabilistic evaluation. They concluded that the CTV $D_{99.8}$ in probabilistic

plan evaluations of VMAT and IMPT treatments lead to consistent results compared with the target D_{98} in clinical PTV-based and voxelwise minimum dose metrics, respectively.

Robustness evaluation in the head, neck, brain and skull region has been widely investigated, due to the importance of that site in PT. The methods used for robustness evaluation vary considerably from a few scenarios in a worst-case approach to sophisticated probabilistic methods. For methods based on worst-case scenarios, setup errors of 3 mm were largely dominant. For range errors, it was either 3 or 3.5%. The method used strongly dependsdepend on the tools available to the investigators. A key question is whether those sophisticated methods should eventually be used to some degree in daily clinical practice, or could be used more sparingly to calibrate more pragmatic methods like the Dutch robustness evaluation protocol.

Many research questions could be addressed. For instance, is a nominal dose distribution reliable enough to refer patients to PT through NTCP estimations? Are treatment plans robust enough in the presence of daily anatomical changes? The inclusion of LET or variable RBE models was not frequent (4% of head and neck papers; 20% of brain and skull). Yet, the brain and skull related papers both highlighted the potential clinical relevance of the variability of RBE, which may be even greater than the impact of other errors [42].

Breast

Five papers about robustness evaluation in the treatment of breast cancers were included in our review [63,95,145–147]. Essential information about robustness methods and parameters can be found in Table S1.3. All but one study used setup shifts of 5 mm in their worst-case analyses, to which different types of uncertainties were added. Ger et al did not evaluate range uncertainty, instead they evaluated if the target coverage for quality assurance (QA) CT scans (with at least two acquired during the treatment) was within the robustness criteria (e.g., the D95% of the breast as calculated on the QA CT was evaluated against the uncertainty envelope of the D95% for 3/5 mm setup errors) [145]. Jensen et al added an anatomical robustness evaluation by creating four artificial CTs: two with 3 mm and 5 mm artificial tissue extending beyond the external contour of the patient and two with 3 and 5 mm contraction of the external contour to simulate anatomical changes of the breast [146]. In Klaassen et al, robustness evaluation (using isotropic setup errors of 5 mm and a range overshoot and undershoot of 3%) was repeated for all phases of the 4D-CT, but interplay effects were not considered [63]. Oden et al calculated setup and range errors for each CT phase separately and in combination with simulated breathing motion [95]. Both a generic RBE of 1.1 and the variable RBE model proposed by Wedenberg et al were considered [148]. In addition to their standard robustness evaluation (±3 mm isocenter shifts combined with ±3.5% shift in relative proton stopping power), Kirk et al evaluated the plan sensitivity to the construction of implanted metal ports. An override applied to the contoured port was varied using proton relative stopping powers of 1, 2 and 3.38 (the largest value in their calibration curve) [147].

Important research questions were addressed in the papers related to breast cancers. However, breast was not a particular focus for developing and testing novel robustness evaluation methodologies. One paper highlighted the potential relative importance of variable RBE with respect to other errors [95]. In the papers selected in this systematic review, intra-fraction motion was investigated, excluding the interplay effect. Setup errors were either 3 or 5 mm, while range uncertainties were either 3 or 3.5%.

Other thoracic lesions (lung, esophagus, mediastinum)

A total of 43 papers were selected in this systematic review that were focused on treatment in the thorax. The details of the robustness evaluation methods and parameters can be found in Tables S1.4 for lung, S1.5 for esophagus, and S1.6 for other mediastinal tumors. Studies performed on phantoms were not included in the analysis because they did not respect the inclusion criteria. The vast majority

of selected papers regarded the treatment of lung [22,32,81,83,84,94,104,106–109,111,38,113–115,149–155,43,47,50,53,59,62,66], 10 papers were focused on the treatment of the esophagus [33,40,48,61,64,96,103,105,110,156] and 3 of them on the mediastinal region (lymphomas and thymomas) [31,99,100].

Since this group of papers is focused on treatments potentially affected by the interplay effect, we included in this analysis the methods and the parameters used to evaluate the plan robustness with respect to target motion uncertainties.

The range error considered in all these papers was either 3% or 3.5%. The setup errors varied with the tumor site and/or type. For lung cases, the vast majority used a set-up error of 5 mm (16). For repeated CTs, several of them used reduced residual setup errors for evaluation [59,62,114,155]. For the esophagus, setup error values were in the range of 3 to 8 mm, and for the mediastinal cases, these were 6 or 7 mm. For 5 papers, the plan robustness to setup errors was evaluated on weekly 4DCTs with a value of 2 mm [48,61,64,103,156].

Thirty-three papers evaluated plan robustness with respect to breathing motion. Eighteen of these used the approach illustrated in section IV.D, where a 4D dynamic dose distribution is obtained by considering the delivery time parameters and patient breathing patterns to evaluate the interplay effect. Ten papers recalculated the entire plan on each or on some phases (i.e. max inhale or max exhale) of the 4DCT.

Eight papers (1 for esophagus and 7 for lung) used some form of probabilistic approach for the plan robustness evaluation [59,81,83,84,94,96,115,156]. All the remaining papers used a worst-case scenario approach where all the uncertainties were systematic during the treatment course and the results did not account for fractionation.

The NTCPs were evaluated in several papers in the thoracic region, either as part of the robustness evaluation and reporting [150], or as an alternative to dose metrics either for the accumulated dose on repeated images [61,113,156] or the nominal doses [48,105]. In Rana et al [150], NCTP were computed alongside dose metrics during the robustness evaluation, to evaluate the impact of the worst combination of setup and range errors on NTCPs.

The number of papers addressing the thoracic region is considerable, which can be easily understood by the concerns related to motion and the highly heterogeneous media. This systematic review indicates that the research in thoracic regions has strongly focused on answering key research questions, like the importance of the interplay effect and remedies against it, or the potential superiority of PT against XRT, even in the presence of uncertainties. The impact of inter-fraction motion was also a great concern. However, the impact of RBE and LET has been sparsely studied, in only 7% of the lung-related papers and none of mediastinum and esophagus. Like in the head, neck, brain and skull region, we observe also a disparity in the methods used, from pragmatic ones, often commercially available, to more sophisticated probabilistic approaches that integrate setup errors, range errors, and breathing motion related uncertainties. But our systematic review did not reveal a study enabling a pragmatic connection between those methods, in the spirit of Rojo-Santiago et al [57,144]. In worst-case evaluation approaches, setup errors used were typically 5 mm for lung. In esophagus and mediastinum, setup errors were larger and more variable. When evaluating robustness on repeated CT, there seems a quite large agreement that the simulated setup error can be considerably reduced.

Lesions in the abdomen and pelvis (prostate, liver, pancreas, cervical cancer)

Twenty-one papers focused on abdominal lesions were selected in this review. The majority studied the treatment of prostate tumors (9 papers) [19,21,26,28,86,121,157–159], followed by liver [29,87,97,98] (4) pancreas (2) [23,101] , cervical cancer (2) [51,60] , abdominal tumors for pediatric cases (1) [30] the mobile spine and sacrum (1) [46] , and pelvic lymph node (1) [91]. Zhu et al investigated lung, esophagus, and liver but is discussed in this section [160]. A summary of the robustness methods used, and their associated parameters can be found in Table S1.7.

The robustness evaluation, where described, was performed by simulating both setup and range uncertainties. The setup error values mostly varied between 2 and 5 mm. In two papers, this value was higher (6 or 7 mm) [29,98]. Range errors of 3% or 3.5% were used in all but 2 papers where 5% was also simulated [97,160] and 1 paper where 1% was simulated [21]. Six papers evaluated the impact of the interplay effect by simulating the 4D dynamic dose distributions [30,60,97,98,101,159]. For 10 papers, the inter fraction anatomy variability was evaluated [28,30,46,60,86,87,101,157–159]. All papers performing robustness analysis adopted the worst-case scenario method. No probabilistic approaches were described in the treatment of abdominal lesions. In only one case, the impact of variable RBE models was investigated in terms of dose distribution, DVH, and NTCP comparison [121]. It highlighted a potential significant clinical impact, including in terms of patient referral to PT.

The systematic review indicates that robustness evaluation has been used in the abdomen and pelvis region to answer specific research questions. But this clinical site has not been favored for exploring more advanced probabilistic methods. For the prostate, key research questions were the robustness against inter-fraction motion and dosimetric comparison between PT and XRT. Mobile targets like the pancreas have considered robustness evaluation methods similar to the ones used in the thoracic regions, including the ones needed to quantify a potential interplay effect. Setup errors and range errors have shown greater variability compared to other locations.

Discussion

Robustness evaluation has become an important area of research in PT. Various approaches, ranging from pragmatic to comprehensive and computationally intensive methods, have been proposed for evaluating the robustness of treatment plans. Many studies in PT incorporate robustness evaluation, and clinical centers are increasingly adopting these strategies with the help of robustness evaluation tools integrated into commercial TPSs.

This systematic review found that setup and range uncertainties were widely considered in both clinical applications and methodological papers. Geometrical uncertainties, mainly related to patient positioning in the treatment room, were frequently addressed, encompassing factors such as gantry and couch rotation precision, imaging system resolution, and reliability of immobilization devices. These uncertainties were typically represented by a shift in the isocenter, incorporating the combined effect of various sources of uncertainty. Regarding range uncertainties, the reviewed clinical papers treated them as systematic uncertainties and evaluated them at their maximum magnitude (±3%/±3.5%). No intermediate values between the extreme ranges were sampled in these studies.

In contrast with a previous study by Kaplan et al [9], this review suggests that robustness parameters for setup and range are relatively well harmonized among centers for lesions localized in the same region (H&N, thoracic, brain, abdomen). The worst-case scenario is commonly used to evaluate dose to organs at risk (OAR) and target coverage in clinical studies. However, the authors of this systematic review suggest that the consistency in values and methodologies used in clinics may reflect software limitations rather than an optimal approach.

For instance, this systematic review has observed that most clinical studies did not consider the mitigation effect of fractionation and of random errors in the evaluation of plan robustness. Only a few studies considered its impact on robustness analysis [73,74,81–83] and few of them are clinically applicable due to the computational efforts required. Nonetheless, it is worth mentioning that the actual methods provided by commercial TPSs do not explicitly account for these sources of mitigation, making the robustness evaluation a limited representation of the treatment course reality. Furthermore, other factors (such as variable RBE) were considered in a minority of studies, and contouring error was not taken into account in the robustness analysis of the published works. Finally, most robustness evaluation methods used in clinical practice implement a worst-case scenario approach where extreme values of the different error types (setup and range uncertainties) are combined. Such approaches may result in selecting very unlikely scenarios for robustness analysis, as observed by Sterpin et al, Korevaar et al, and Badiu et al [5,56,81].

This systematic review found a wide range of methods used to evaluate the impact of intra- and inter-fraction motion on plan robustness. In terms of intra-fraction motion, the commonly used approach was estimation of the 4D dose distribution, allowing for interplay effects to be assessed. Regardless of whether logfiles or fixed time delivery machine parameters were considered, this method was consistently used [21,64,111,95,97,104,106–110]. On the other hand, incorporating uncertainties related to inter-fraction anatomy variations, such as those seen in the treatment of head and neck, prostate, or liver cancers, posed a greater challenge as these variations are harder to predict. In such cases, an analysis on a group of patients, involving recalculating the plan on repeated CT scans or synthetic CTs reconstructed from cone-beam CTs, was employed to assess the most effective planning approach [87,88].

In the following paragraphs, we would like to discuss some difficulties related to the definition of a shared robustness evaluation methodology standard, and establish some links to another standard - the PTV.

Despite its flaws, the PTV concept has allowed to some degree several achievements that are still missing in current applications of robustness evaluation. For instance, contouring uncertainties can be in principle included in the PTV [161], but with practical difficulties and variability during implementation [14]. In the survey of Kaplan et al., it is reported that a significant percentage of centers include the contouring error in their definition of PTV [9]. In this systematic review, no studies have explicitly included such uncertainty in either robust optimization or robustness analysis.

Moreover, typical PTV safety margin recipes enable explicitly the computation of a margin according to a certain confidence level defined over patient population probabilities [80]. Such confidence levels can also be computed during robustness evaluation. The definition of confidence levels can be performed in the scenario space (cumulative probability of scenarios considered) [5,56], or in the dosimetric space (cumulative probability of dose metrics considered) [56,74,134], according to local uncertainties.

Sterpin et al have warned about the potential shortcomings of the computation of confidence levels in the scenario space, especially if additional error types are simulated, consequently increasing the dimensionality of the scenario space [56]. One solution to these problems is to simulate the errors from their probability distribution, without restrictions, for instance using Monte Carlo simulations. Then, the confidence level can be computed in the dosimetric space, by selecting the best dose distributions for a given dose-related metric (for instance D95 for the CTV).

In clinical practice, the concept of a "worst-case" scenario requires clarification. While there is some consensus in the literature, it generally refers to the worst situation considering a specific set of scenarios or the worst combination of errors. However, it should not be mistaken for the absolute worst scenario possible, as that would be excessively conservative and difficult to define. For example, treating the wrong patient is an unlikely but not impossible scenario. The PTV, by definition, addresses this issue. When exploring errors within the PTV margin and reporting a "worst-case" scenario, it may only guarantee coverage in a certain percentage of the population, typically 90%.

Associating a confidence level to a worst-case scenario is trivial using large scenario sampling approaches [56,74,77,78]. In such cases, the conceptual link with the PTV concept can be conserved. Instead of defining a geometric margin that encompasses a given coverage probability and thus a “worst-case” line, we can define a certain confidence level for displaying clinically relevant DVH metrics. In such an approach, the worst-case can be defined as the scenario realizing the “worst” metric within the predefined confidence level.

Without this, it is not obvious how to transfer the dosimetric parameters typically accepted in the PTV/PRV context to that of the CTV/OAR robustness analysis, especially when robustness analysis is based on worst-case scenario that includes only systematic setup and range errors. With regard to target coverage, Korevaar et al [5] proposed the CTV D98 voxel-wise minimum dose to keep consistency with historic PTV-based evaluations, which was shown to be adequate in the follow-up work from Rojo-Santiago et al [57,76].

Robustness evaluation is clearly an important topic, as demonstrated to various degrees in research papers and clinical practice. The process of finding a consensus and the associated

guidelines/recommendations for ensuring a clinically relevant and consistent robustness evaluation approach per clinical site has been awaited for a while, but it will not be easy. It will have to find the right balance between, on one hand, the technical sophistication of the robustness evaluation methodology, in order to provide a trustworthy and quantified estimate of the robustness of the plan, and, on the other hand, its clinical viability from a practical point-of-view. This will be the coming task of the EPTN working group 5, according to a modified Delphi consensus methodology.

Acknowledgements and conflicts of interest

Edmond Sterpin is involved in collaborations with Ion Beam Applications and RaySearch laboratories.
Erik Korevaar is involved in collaborations with Ion Beam Applications and RaySearch laboratories.
Mischa Hoogeman is involved in collaborations with RaySearch laboratories, Varian, Siemens, Accuray, and Elekta.

References

[1] Mohan R. A review of proton therapy – Current status and future directions. Precis Radiat Oncol 2022;6:164–76. doi:10.1002/pro6.1149.

[2] Engwall E, Fredriksson A, Glimelius L. 4D robust optimization including uncertainties in time structures can reduce the interplay effect in proton pencil beam scanning radiation therapy. Med Phys 2018;45:4020–9. doi:10.1002/mp.13094.

[3] T. Landberg, J. Chavaudra, J. Dobbs, G. Hanks, K.-A. Johansson, T. Möller, J.Purdy. Prescribing, Recording and Reporting Photon Beam Therapy (Supplment to ICRU Report 50) - ICRU Report 62. vol. 74. Oxford, United Kingdom: ICRU Publications; 1999.

[4] Grégoire V, Mackie TR, De Neve W, Gospodarowicz M, Purdy JA, van Herk M, Niemierko A. Prescribing, recording, and reporting intensity-modulated photon-beam therapy (IMRT). vol. 10. 2010. doi:10.1093/jicru/ndn032.

[5] Korevaar EW, Habraken SJM, Scandurra D, Kierkels RGJ, Unipan M, Eenink MGC, et al. Practical robustness evaluation in radiotherapy – A photon and proton-proof alternative to PTV-based plan evaluation. Radiother Oncol 2019;141:267–74. doi:10.1016/j.radonc.2019.08.005.

[6] Han D, Hooshangnejad H, Chen CC, Ding K. A Beam-Specific Optimization Target Volume for Stereotactic Proton Pencil Beam Scanning Therapy for Locally Advanced Pancreatic Cancer. Adv Radiat Oncol 2021;6:100757. doi:10.1016/j.adro.2021.100757.

[7] Lin L, Souris K, Kang M, Glick A, Lin H, Huang S, et al. Evaluation of motion mitigation using abdominal compression in the clinical implementation of pencil beam scanning proton therapy of liver tumors. Med Phys 2017;44:703–12. doi:10.1002/mp.12040.

[8] Unkelbach J, Alber M, Bangert M, Bokrantz R, Chan TCY, Deasy JO, et al. Robust radiotherapy planning. Phys Med Biol 2018;63. doi:10.1088/1361-6560/aae659.

[9] Kaplan LP, Placidi L, Bäck A, Canters R, Hussein M, Vaniqui A, et al. Plan quality assessment in clinical practice: results of the 2020 ESTRO survey on plan complexity and robustness. Radiother Oncol 2022;173:254–61. doi:10.1016/j.radonc.2022.06.005.

[10] Zhang X, Rong Y, Morrill S, Fang J, Narayanasamy G, Galhardo E, et al. Robust optimization in lung treatment plans accounting for geometric uncertainty. J Appl Clin Med Phys 2018;19:19–26. doi:10.1002/acm2.12291.

[11] Ferrara E, Beldì D, Yin J, Vigna L, Loi G, Krengli M. Adaptive Strategy for External Beam Radiation Therapy in Prostate Cancer: Management of the Geometrical Uncertainties With Robust Optimization. Pract Radiat Oncol 2020;10:e521–8. doi:https://doi.org/10.1016/j.prro.2020.05.006.

[12] Hernandez V, Hansen CR, Widesott L, Bäck A, Canters R, Fusella M, et al. What is plan quality in radiotherapy? The importance of evaluating dose metrics, complexity, and robustness of treatment plans. Radiother Oncol 2020;153:26–33. doi:10.1016/j.radonc.2020.09.038.

[13] Lowe M, Gosling A, Nicholas O, Underwood T, Miles E, Chang YC, et al. Comparing Proton to Photon Radiotherapy Plans: UK Consensus Guidance for Reporting Under Uncertainty for Clinical Trials. Clin Oncol 2020;32:459–66. doi:10.1016/j.clon.2020.03.014.
[14] Biston M-C, Chiavassa S, Grégoire V, Thariat J, Lacornerie T. Time of PTV is ending, robust optimization comes next. Cancer Radiother J La Soc Fr Radiother Oncol 2020;24:676–86. doi:10.1016/j.canrad.2020.06.016.
[15] Fredriksson A. A characterization of robust radiation therapy treatment planning methods—from expected value to worst case optimization. Med Phys 2012;39:5169. doi:10.1118/1.4737113.
[16] Stoll CRT, Izadi S, Fowler S, Green P, Suls J, Colditz GA. The value of a second reviewer for study selection in systematic reviews. Res Synth Methods 2019;10:539–45. doi:10.1002/jrsm.1369.
[17] Hague C, Aznar M, Dong L, Fotouhi-Ghiam A, Lee LW, Li T, et al. Inter-fraction robustness of intensity-modulated proton therapy in the post-operative treatment of oropharyngeal and oral cavity squamous cell carcinomas. Br J Radiol 2020;93:20190638. doi:10.1259/bjr.20190638.
[18] Stock M, Gora J, Perpar A, Georg P, Lehde A, Kragl G, et al. Harmonization of proton treatment planning for head and neck cancer using pencil beam scanning: first report of the IPACS collaboration group. Acta Oncol (Madr) 2019;58:1720–30. doi:10.1080/0284186X.2019.1648858.
[19] Ong ALK, Ang KW, Master Z, Wong SMM, Tuan JKL. Intensity-modulated radiotherapy for whole pelvis irradiation in prostate cancer: A dosimetric and plan robustness study between photons and protons. Tech Innov Patient Support Radiat Oncol 2018;6:11–9. doi:10.1016/j.tipsro.2018.02.001.
[20] Hedrick SG, Walker B, Morris B, Petro S, Blakey M. Scripted spot removal in PBS proton therapy planning. J Appl Clin Med Phys 2022;23. doi:10.1002/acm2.13491.
[21] Goddard LC, Brodin NP, Bodner WR, Garg MK, Tomé WA. Comparing photon and proton-based hypofractioned SBRT for prostate cancer accounting for robustness and realistic treatment deliverability. Br J Radiol 2018;91:20180010. doi:10.1259/bjr.20180010.
[22] Liu C, Yu NY, Shan J, Bhangoo RS, Daniels TB, Chiang JS, et al. Technical Note: Treatment planning system (TPS) approximations matter - comparing intensity-modulated proton therapy (IMPT) plan quality and robustness between a commercial and an in-house developed TPS for nonsmall cell lung cancer (NSCLC). Med Phys 2019;46:4755–62. doi:10.1002/mp.13809.
[23] Sio TT, Merrell KW, Beltran CJ, Ashman JB, Hoeft KA, Miller RC, et al. Spot-scanned pancreatic stereotactic body proton therapy: A dosimetric feasibility and robustness study. Phys Medica PM an Int J Devoted to Appl Phys to Med Biol Off J Ital Assoc Biomed Phys 2016;32:331–42. doi:10.1016/j.ejmp.2015.12.009.
[24] Toramatsu C, Inaniwa T. Beam angle selection incorporation of anatomical heterogeneities for pencil beam scanning charged-particle therapy. Phys Med Biol 2016;61:8664–75. doi:10.1088/1361-6560/61/24/8664.
[25] van de Water S, van Dam I, Schaart DR, Al-Mamgani A, Heijmen BJM, Hoogeman MS. The price of robustness; impact of worst-case optimization on organ-at-risk dose and complication probability in intensity-modulated proton therapy for oropharyngeal cancer patients. Radiother Oncol 2016;120:56–62. doi:10.1016/j.radonc.2016.04.038.
[26] Butkus MP, Brovold N, Diwanji T, Xu Y, De Ornelas M, Dal Pra A, et al. Assessment of IMPT versus VMAT plans using different uncertainty scenarios for prostate cancer. Radiat Oncol 2022;17:1–12. doi:10.1186/s13014-022-02126-y.
[27] Zechner A, Ziegler I, Hug E, Lütgendorf-Caucig C, Stock M. Evaluation of the inter- and intrafraction displacement for head patients treated at the particle therapy centre MedAustron based on the comparison of different commercial immobilisation devices. Z Med Phys 2022;32:39–51. doi:10.1016/j.zemedi.2021.01.007.
[28] Xu Y, Diwanji T, Brovold N, Butkus M, Padgett KR, Schmidt RM, et al. Assessment of daily dose accumulation for robustly optimized intensity modulated proton therapy treatment of prostate cancer. Phys Medica 2021;81:77–85. doi:10.1016/j.ejmp.2020.11.035.

[29] Cozzi L, Comito T, Fogliata A, Franzese C, Tomatis S, Scorsetti M. Critical appraisal of the potential role of intensity modulated proton therapy in the hypofractionated treatment of advanced hepatocellular carcinoma. PLoS One 2018;13:1–13. doi:10.1371/journal.pone.0201992.

[30] Guerreiro F, Zachiu C, Seravalli E, Ribeiro CO, Janssens GO, Ries M, et al. Evaluating the benefit of PBS vs. VMAT dose distributions in terms of dosimetric sparing and robustness against inter-fraction anatomical changes for pediatric abdominal tumors. Radiother Oncol 2019;138:158–65. doi:10.1016/j.radonc.2019.06.025.

[31] Lideståhl A, Mondlane G, Gubanski M, Lind PA, Siegbahn A. An in silico planning study comparing doses and estimated risk of toxicity in 3D-CRT, IMRT and proton beam therapy of patients with thymic tumours. Phys Medica 2019;60:120–6. doi:10.1016/j.ejmp.2019.03.028.

[32] Rana S, Traneus E, Jackson M, Tran L, Rosenfeld AB. Quantitative analysis of dose-averaged linear energy transfer (LETd) robustness in pencil beam scanning proton lung plans. Med Phys 2022;49:3444–56. doi:10.1002/mp.15569.

[33] Liu C, Bhangoo RS, Sio TT, Yu NY, Shan J, Chiang JS, et al. Dosimetric comparison of distal esophageal carcinoma plans for patients treated with small-spot intensity-modulated proton versus volumetric-modulated arc therapies. J Appl Clin Med Phys 2019;20:15–27. doi:10.1002/acm2.12623.

[34] An Y, Shan J, Patel SH, Wong W, Schild SE, Ding X, et al. Robust intensity-modulated proton therapy to reduce high linear energy transfer in organs at risk: Med Phys 2017;44:6138–47. doi:10.1002/mp.12610.

[35] Maes D, Saini J, Zeng J, Rengan R, Wong T, Bowen SR. Advanced proton beam dosimetry part II: Monte Carlo vs. pencil beam-based planning for lung cancer. Transl Lung Cancer Res 2018;7:114–21. doi:10.21037/tlcr.2018.04.04.

[36] Yang Z, Zhang X, Wang X, Zhu XR, Gunn B, Frank SJ, et al. Multiple-CT optimization: An adaptive optimization method to account for anatomical changes in intensity-modulated proton therapy for head and neck cancers. Radiother Oncol 2020;142:124–32. doi:10.1016/j.radonc.2019.09.010.

[37] Miyazaki K, Fujii Y, Yamada T, Kanehira T, Miyamoto N, Matsuura T, et al. Deformed dose restoration to account for tumor deformation and position changes for adaptive proton therapy. Med Phys 2023;50:675–87. doi:10.1002/mp.16149.

[38] Liu C, Schild SE, Chang JY, Liao Z, Korte S, Shen J, et al. Impact of Spot Size and Spacing on the Quality of Robustly Optimized Intensity Modulated Proton Therapy Plans for Lung Cancer. Int J Radiat Oncol Biol Phys 2018;101:479–89. doi:10.1016/j.ijrobp.2018.02.009.

[39] Lowe M, Albertini F, Aitkenhead A, Lomax AJ, Mackay RI. Incorporating the effect of fractionation in the evaluation of proton plan robustness to setup errors. Phys Med Biol 2015;61:413–29. doi:10.1088/0031-9155/61/1/413.

[40] Warren S, Partridge M, Bolsi A, Lomax AJ, Hurt C, Crosby T, et al. An Analysis of Plan Robustness for Esophageal Tumors: Comparing Volumetric Modulated Arc Therapy Plans and Spot Scanning Proton Planning. Int J Radiat Oncol Biol Phys 2016;95:199–207. doi:10.1016/j.ijrobp.2016.01.044.

[41] Bolsi A, Placidi L, Pica A, Ahlhelm FJ, Walser M, Lomax AJ, et al. Pencil beam scanning proton therapy for the treatment of craniopharyngioma complicated with radiation-induced cerebral vasculopathies: A dosimetric and linear energy transfer (LET) evaluation. Radiother Oncol 2020;149:197–204. doi:10.1016/j.radonc.2020.04.052.

[42] Gutierrez A, Rompokos V, Li K, Gillies C, D'Souza D, Solda F, et al. The impact of proton LET/RBE modeling and robustness analysis on base-of-skull and pediatric craniopharyngioma proton plans relative to VMAT. Acta Oncol (Madr) 2019;58:1765–74. doi:10.1080/0284186X.2019.1653496.

[43] Wong SL, Alshaikhi J, Grimes H, Amos RA, Poynter A, Rompokos V, et al. Retrospective Planning Study of Patients with Superior Sulcus Tumours Comparing Pencil Beam Scanning Protons to Volumetric-Modulated Arc Therapy. Clin Oncol 2021;33:e118–31.

doi:10.1016/j.clon.2020.07.016.
[44] Wang X, Li H, Zhu XR, Hou Q, Liao L, Jiang B, et al. Multiple-CT optimization of intensity-modulated proton therapy – Is it possible to eliminate adaptive planning? Radiother Oncol 2018;128:167–73. doi:10.1016/j.radonc.2017.09.032.
[45] Poel R, Belosi F, Albertini F, Walser M, Gisep A, Lomax AJ, et al. Assessing the advantages of CFR-PEEK over titanium spinal stabilization implants in proton therapy - A phantom study. Phys Med Biol 2020;65. doi:10.1088/1361-6560/ab8ba0.
[46] Kroesen M, Miladinovic V, Hutschemaekers SAJ, Jacobs J, van der Vos C, Wolf AL, et al. Single-institution clinical experience using robust intensity modulated proton therapy in chordoma and chondrosarcoma of the mobile spine and sacrum: Feasibility and need for plan adaptation: Robust planning in chordoma of spine and sacrum. Radiother Oncol 2022;166:58–64. doi:10.1016/j.radonc.2021.11.020.
[47] den Otter LA, Kaza E, Kierkels RGJ, Meijers A, Ubbels FJF, Leach MO, et al. Reproducibility of the lung anatomy under active breathing coordinator control: Dosimetric consequences for scanned proton treatments. Med Phys 2018;45:5525–34. doi:10.1002/mp.13195.
[48] Hoffmann L, Mortensen H, Shamshad M, Berbee M, Bizzocchi N, Bütof R, et al. Treatment planning comparison in the PROTECT-trial randomising proton versus photon beam therapy in oesophageal cancer: Results from eight European centres. Radiother Oncol 2022;172:32–41. doi:10.1016/j.radonc.2022.04.029.
[49] Newpower MA, Chiang BH, Ahmad S, Chen Y. Spot delivery error predictions for intensity modulated proton therapy using robustness analysis with machine learning. J Appl Clin Med Phys 2023;24:1–9. doi:10.1002/acm2.13911.
[50] Shang H, Pu Y, Chen Z, Wang X, Yuan C, Jin X, et al. Impact of Multiple Beams on Plan Quality, Linear Energy Transfer Distribution, and Plan Robustness of Intensity Modulated Proton Therapy for Lung Cancer. ACS Sensors 2021;6:408–17. doi:10.1021/acssensors.0c01879.
[51] Shang H, Pu Y, Wang W, Dai Z, Jin F. Evaluation of plan quality and robustness of IMPT and helical IMRT for cervical cancer. Radiat Oncol 2020;15:34. doi:10.1186/s13014-020-1483-x.
[52] Noufal MP, Widesott L, Sharma SD, Righetto R, Cianchetti M, Schwarz M. The Role of Plan Robustness Evaluation in Comparing Protons and Photons Plans - An Application on IMPT and IMRT Plans in Skull Base Chordomas. J Med Phys 2020;45:206–14. doi:10.4103/jmp.JMP_45_20.
[53] Boer CG, Fjellanger K, Sandvik IM, Ugland M, Engeseth GM, Hysing LB. Substantial Sparing of Organs at Risk with Modern Proton Therapy in Lung Cancer, but Altered Breathing Patterns Can Jeopardize Target Coverage. Cancers (Basel) 2022;14. doi:10.3390/cancers14061365.
[54] Tommasino F, Widesott L, Fracchiolla F, Lorentini S, Righetto R, Algranati C, et al. Clinical implementation in proton therapy of multi-field optimization by a hybrid method combining conventional PTV with robust optimization. Phys Med Biol 2020;65. doi:10.1088/1361-6560/ab63b9.
[55] Albertini F, Hug EB, Lomax AJ. Is it necessary to plan with safety margins for actively scanned proton therapy? Phys Med Biol 2011;56:4399–413. doi:10.1088/0031-9155/56/14/011.
[56] Sterpin E, Rivas ST, Van den Heuvel F, George B, Lee JA, Souris K. Development of robustness evaluation strategies for enabling statistically consistent reporting. Phys Med Biol 2021;66:45002. doi:10.1088/1361-6560/abd22f.
[57] Rojo-Santiago J, Habraken SJM, Lathouwers D, Méndez Romero A, Perkó Z, Hoogeman MS. Accurate assessment of a Dutch practical robustness evaluation protocol in clinical PT with pencil beam scanning for neurological tumors. Radiother Oncol J Eur Soc Ther Radiol Oncol 2021;163:121–7. doi:10.1016/j.radonc.2021.07.028.
[58] Scandurra D, Meijer TWH, Free J, van den Hoek JGM, Kelder L, Oldehinkel E, et al. Evaluation of robustly optimised intensity modulated proton therapy for nasopharyngeal carcinoma. Radiother Oncol J Eur Soc Ther Radiol Oncol 2022;168:221–8. doi:10.1016/j.radonc.2022.01.043.
[59] Ribeiro CO, Visser S, Korevaar EW, Sijtsema NM, Anakotta RM, Dieters M, et al. Towards the

clinical implementation of intensity-modulated proton therapy for thoracic indications with moderate motion: Robust optimised plan evaluation by means of patient and machine specific information. Radiother Oncol 2021;157:210–8. doi:10.1016/j.radonc.2021.01.014.

[60] Gort EM, Beukema JC, Matysiak W, Sijtsema NM, Aluwini S, Langendijk JA, et al. Inter-fraction motion robustness and organ sparing potential of proton therapy for cervical cancer. Radiother Oncol J Eur Soc Ther Radiol Oncol 2021;154:194–200. doi:10.1016/j.radonc.2020.09.022.

[61] Anakotta RM, van der Laan HP, Visser S, Ribeiro CO, Dieters M, Langendijk JA, et al. Weekly robustness evaluation of intensity-modulated proton therapy for oesophageal cancer. Radiother Oncol J Eur Soc Ther Radiol Oncol 2020;151:66–72. doi:10.1016/j.radonc.2020.07.015.

[62] van der Laan HP, Anakotta RM, Korevaar EW, Dieters M, Ubbels JF, Wijsman R, et al. Organ sparing potential and inter-fraction robustness of adaptive intensity modulated proton therapy for lung cancer. Acta Oncol 2019;58:1775–82. doi:10.1080/0284186X.2019.1669818.

[63] Klaassen L, Petoukhova AL, Habraken SJM, Jacobs J, Sattler MGA, Verhoeven K, et al. Effect of breathing motion on robustness of proton therapy plans for left-sided breast cancer patients with indication for locoregional irradiation. Acta Oncol 2021;60:222–8. doi:10.1080/0284186X.2020.1825800.

[64] Visser S, Neh H, Oraboni Ribeiro C, Korevaar EW, Meijers A, Poppe B, et al. Assessment of a diaphragm override strategy for robustly optimized proton therapy planning for esophageal cancer patients. Med Phys 2021;48:5674–83. doi:10.1002/mp.15114.

[65] Meijers A, Knopf AC, Crijns APG, Ubbels JF, Niezink AGH, Langendijk JA, et al. Evaluation of interplay and organ motion effects by means of 4D dose reconstruction and accumulation. Radiother Oncol 2020;150:268–74. doi:10.1016/j.radonc.2020.07.055.

[66] Taasti VT, Hattu D, Vaassen F, Canters R, Velders M, Mannens J, et al. Treatment planning and 4D robust evaluation strategy for proton therapy of lung tumors with large motion amplitude. Med Phys 2021;48:4425–37. doi:10.1002/mp.15067.

[67] Wagenaar D, Kierkels RGJ, van der Schaaf A, Meijers A, Scandurra D, Sijtsema NM, et al. Head and neck IMPT probabilistic dose accumulation: Feasibility of a 2 mm setup uncertainty setting. Radiother Oncol 2021;154:45–52. doi:10.1016/j.radonc.2020.09.001.

[68] Oud M, Breedveld S, Giżyńska M, Kroesen M, Hutschemaekers S, Habraken S, et al. An online adaptive plan library approach for intensity modulated proton therapy for head and neck cancer. Radiother Oncol 2022;176:68–75. doi:10.1016/j.radonc.2022.09.011.

[69] Wahl N, Hennig P, Wieser HP, Bangert M. Analytical probabilistic modeling of dose-volume histograms. Med Phys 2020;47:5260–73. doi:10.1002/mp.14414.

[70] Wieser HP, Karger CP, Wahl N, Bangert M. Impact of Gaussian uncertainty assumptions on probabilistic optimization in particle therapy. Phys Med Biol 2020;65. doi:10.1088/1361-6560/ab8d77.

[71] Souris K, Barragan Montero A, Janssens G, Di Perri D, Sterpin E, Lee JA. Technical Note: Monte Carlo methods to comprehensively evaluate the robustness of 4D treatments in proton therapy. Med Phys 2019;46:4676–84. doi:10.1002/mp.13749.

[72] Wahl N, Hennig P, Wieser HP, Bangert M. Analytical incorporation of fractionation effects in probabilistic treatment planning for intensity-modulated proton therapy. Med Phys 2018;45:1317–28. doi:10.1002/mp.12775.

[73] Stammer P, Burigo L, J kel O, Frank M, Wahl N. Efficient uncertainty quantification for Monte Carlo dose calculations using importance (re-)weighting. Phys Med Biol 2021;66. doi:10.1088/1361-6560/ac287f.

[74] Perkó Z, Van Der Voort SR, Van De Water S, Hartman CMH, Hoogeman M, Lathouwers D. Fast and accurate sensitivity analysis of IMPT treatment plans using Polynomial Chaos Expansion. Phys Med Biol 2016;61:4646–64. doi:10.1088/0031-9155/61/12/4646.

[75] Van Der Voort S, Van De Water S, Perkó Z, Heijmen B, Lathouwers D, Hoogeman M. Robustness Recipes for Minimax Robust Optimization in Intensity Modulated Proton Therapy for Oropharyngeal Cancer Patients. Int J Radiat Oncol Biol Phys 2016;95:163–70.

doi:10.1016/j.ijrobp.2016.02.035.
[76] Rojo-Santiago J, Korevaar E, Perkó Z, Both S, Habraken SJM, Hoogeman MS. PTV-based VMAT vs. robust IMPT for head-and-neck cancer: A probabilistic uncertainty analysis of clinical plan evaluation with the Dutch model-based selection. Radiother Oncol 2023;186. doi:10.1016/j.radonc.2023.109729.
[77] Yang Z, Li H, Li Y, Li Y, Chang Y, Li Q, et al. Statistical evaluation of worst-case robust optimization intensity-modulated proton therapy plans using an exhaustive sampling approach. Radiat Oncol 2019;14:129. doi:10.1186/s13014-019-1335-8.
[78] Wahl N, Hennig P, Wieser HP, Bangert M. Efficiency of analytical and sampling-based uncertainty propagation in intensity-modulated proton therapy. Phys Med Biol 2017;62:5790–807. doi:10.1088/1361-6560/aa6ec5.
[79] Vazquez I, Gronberg MP, Zhang X, Court LE, Zhu XR, Frank SJ, et al. A deep learning-based approach for statistical robustness evaluation in proton therapy treatment planning: a feasibility study. Phys Med Biol 2023;68. doi:10.1088/1361-6560/accc08.
[80] van Herk M, Remeijer P, Rasch C, Lebesque J V. The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy. Int J Radiat Oncol Biol Phys 2000;47:1121–35.
[81] Badiu V, Souris K, Buti G, Villarroel EB, Lambrecht M, Sterpin E. Improved healthy tissue sparing in proton therapy of lung tumors using statistically sound robust optimization and evaluation. Phys Medica 2022;96:62–9. doi:10.1016/j.ejmp.2022.02.018.
[82] Malyapa R, Lowe M, Bolsi A, Lomax AJ, Weber DC, Albertini F. Evaluation of Robustness to Setup and Range Uncertainties for Head and Neck Patients Treated with Pencil Beam Scanning Proton Therapy. Int J Radiat Oncol Biol Phys 2016;95:154–62. doi:10.1016/j.ijrobp.2016.02.016.
[83] Ribeiro CO, Meijers A, Korevaar EW, Muijs CT, Both S, Langendijk JA, et al. Comprehensive 4D robustness evaluation for pencil beam scanned proton plans. Radiother Oncol J Eur Soc Ther Radiol Oncol 2019;136:185–9. doi:10.1016/j.radonc.2019.03.037.
[84] Teoh S, George B, Fiorini F, Vallis KA, Van den Heuvel F. Assessment of robustness against setup uncertainties using probabilistic scenarios in lung cancer: A comparison of proton with photon therapy. Br J Radiol 2020;93. doi:10.1259/bjr.20190584.
[85] Paganetti H, Botas P, Sharp GC, Winey B. Adaptive proton therapy. Phys Med Biol 2021;66. doi:10.1088/1361-6560/ac344f.
[86] Pedersen J, Casares-Magaz O, Petersen JBB, Rørvik J, Bentzen L, Andersen AG, et al. A biological modelling based comparison of radiotherapy plan robustness using photons vs protons for focal prostate boosting. Phys Imaging Radiat Oncol 2018;6:101–5. doi:10.1016/j.phro.2018.06.002.
[87] Yang Z, Chang Y, Brock KK, Cazoulat G, Koay EJ, Koong AC, et al. Effect of setup and inter-fraction anatomical changes on the accumulated dose in CT-guided breath-hold intensity modulated proton therapy of liver malignancies. Radiother Oncol 2019;134:101–9. doi:10.1016/j.radonc.2019.01.028.
[88] Argota-Perez R, Sharma MB, Elstrøm U V., Møller DS, Grau C, Jensen K, et al. Dose and robustness comparison of nominal, daily and accumulated doses for photon and proton treatment of sinonasal cancer. Radiother Oncol 2022;173:102–8. doi:10.1016/j.radonc.2022.05.038.
[89] Lalonde A, Bobić M, Sharp GC, Chamseddine I, Winey B, Paganetti H. Evaluating the effect of setup uncertainty reduction and adaptation to geometric changes on normal tissue complication probability using online adaptive head and neck intensity modulated proton therapy. Phys Med Biol 2023;68. doi:10.1088/1361-6560/acd433.
[90] Hirotaki K, Tomizawa K, Moriya S, Ito M, Sakae T. Impact of Anatomical Position Errors on Dose Distribution in Head and Neck Radiotherapy and Robust Image Registration Against Anatomical Changes. Anticancer Res 2023;43:1827–34. doi:10.21873/anticanres.16336.
[91] Gravgaard Andersen A, Casares-Magaz O, Petersen J, Toftegaard J, Bentzen L, Thörnqvist S, et al. Beam angle evaluation to improve inter-fraction motion robustness in pelvic lymph node irradiation with proton therapy. Acta Oncol 2017;56:846–52.

doi:10.1080/0284186X.2017.1317108.
[92] Zhang Y, Alshaikhi J, Tan W, Royle G, Bär E. A probability model for anatomical robust optimisation in head and neck cancer proton therapy. Phys Med Biol 2023;68:0–13. doi:10.1088/1361-6560/aca877.
[93] Walsh S, Roelofs E, Kuess P, Van Wijk Y, Vanneste B, Dekker A, et al. Towards a clinical decision support system for external beam radiation oncology prostate cancer patients: Proton vs. photon radiotherapy? a radiobiological study of robustness and stability. Cancers (Basel) 2018;10. doi:10.3390/cancers10020055.
[94] Shan J, Yang Y, Schild SE, Daniels TB, Wong WW, Fatyga M, et al. Intensity-modulated proton therapy (IMPT) interplay effect evaluation of asymmetric breathing with simultaneous uncertainty considerations in patients with non-small cell lung cancer. Med Phys 2020;47:5428–40. doi:10.1002/mp.14491.
[95] Ödén J, Toma-Dasu I, Eriksson K, Flejmer AM, Dasu A. The influence of breathing motion and a variable relative biological effectiveness in proton therapy of left-sided breast cancer. Acta Oncol (Madr) 2017;56:1428–36. doi:10.1080/0284186X.2017.1348625.
[96] Feng H, Shan J, Ashman JB, Rule WG, Bhangoo RS, Yu NY, et al. Technical Note: 4D robust optimization in small spot intensity-modulated proton therapy (IMPT) for distal esophageal carcinoma. Med Phys 2021;48:4636–47. doi:10.1002/mp.15003.
[97] Siregar H, Bäumer C, Blanck O, Chan M, Engwall E, Plaude S, et al. Mitigation of motion effects in pencil-beam scanning – Impact of repainting on 4D robustly optimized proton treatment plans for hepatocellular carcinoma. Z Med Phys 2022;32:63–73. doi:10.1016/j.zemedi.2020.08.001.
[98] Fracchiolla F, Dionisi F, Righetto R, Widesott L, Giacomelli I, Cartechini G, et al. Clinical implementation of pencil beam scanning proton therapy for liver cancer with forced deep expiration breath hold. Radiother Oncol 2021;154:137–44. doi:10.1016/j.radonc.2020.09.035.
[99] Righetto R, Fracchiolla F, Widesott L, Lorentini S, Dionisi F, Rombi B, et al. Technical challenges in the treatment of mediastinal lymphomas by proton pencil beam scanning and deep inspiration breath-hold. Radiother Oncol 2022;169:43–50. doi:10.1016/j.radonc.2022.02.015.
[100] Zeng C, Plastaras JP, James P, Tochner ZA, Hill-Kayser CE, Hahn SM, et al. Proton pencil beam scanning for mediastinal lymphoma: treatment planning and robustness assessment. Acta Oncol (Madr) 2016;55:1132–8. doi:10.1080/0284186X.2016.1191665.
[101] Batista V, Richter D, Chaudhri N, Naumann P, Herfarth K, Jäkel O. Significance of intra-fractional motion for pancreatic patients treated with charged particles. Radiat Oncol 2018;13:1–11. doi:10.1186/s13014-018-1060-8.
[102] Widesott L, Dionisi F, Fracchiolla F, Tommasino F, Centonze M, Amichetti M, et al. Proton or photon radiosurgery for cardiac ablation of ventricular tachycardia? Breath and ECG gated robust optimization. Phys Medica 2020;78:15–31. doi:10.1016/j.ejmp.2020.08.021.
[103] Shamshad M, Møller DS, Mortensen HR, Ehmsen ML, Jensen MF, Hoffmann L. Bone versus soft-tissue setup in proton therapy for patients with oesophageal cancer. Acta Oncol (Madr) 2022;61:994–1003. doi:10.1080/0284186X.2022.2091949.
[104] Gorgisyan J, Munck af Rosenschold P, Perrin R, Persson GF, Josipovic M, Belosi MF, et al. Feasibility of Pencil Beam Scanned Intensity Modulated Proton Therapy in Breath-hold for Locally Advanced Non-Small Cell Lung Cancer. Int J Radiat Oncol Biol Phys 2017;99:1121–8. doi:10.1016/j.ijrobp.2017.08.023.
[105] Feng H, Sio TT, Rule WG, Bhangoo RS, Lara P, Patrick CL, et al. Beam angle comparison for distal esophageal carcinoma patients treated with intensity-modulated proton therapy. J Appl Clin Med Phys 2020;21:141–52. doi:10.1002/acm2.13049.
[106] Inoue T, Widder J, van Dijk L V., Takegawa H, Koizumi M, Takashina M, et al. Limited Impact of Setup and Range Uncertainties, Breathing Motion, and Interplay Effects in Robustly Optimized Intensity Modulated Proton Therapy for Stage III Non-small Cell Lung Cancer. Int J Radiat Oncol Biol Phys 2016;96:661–9. doi:10.1016/j.ijrobp.2016.06.2454.
[107] Mastella E, Molinelli S, Pella A, Vai A, Maestri D, Vitolo V, et al. 4D strategies for lung tumors

treated with hypofractionated scanning proton beam therapy: Dosimetric impact and robustness to interplay effects. Radiother Oncol J Eur Soc Ther Radiol Oncol 2020;146:213–20. doi:10.1016/j.radonc.2020.02.025.
[108] Liu W, Schild SE, Chang JY, Liao Z, Chang YH, Wen Z, et al. Exploratory Study of 4D versus 3D Robust Optimization in Intensity Modulated Proton Therapy for Lung Cancer. Int J Radiat Oncol Biol Phys 2016;95:523–33. doi:10.1016/j.ijrobp.2015.11.002.
[109] Rana S, Rosenfeld AB. Investigating volumetric repainting to mitigate interplay effect on 4D robustly optimized lung cancer plans in pencil beam scanning proton therapy. J Appl Clin Med Phys 2021;22:107–18. doi:10.1002/acm2.13183.
[110] Yu J, Zhang X, Liao L, Li H, Zhu R, Park PC, et al. Motion-robust intensity-modulated proton therapy for distal esophageal cancer. Med Phys 2016;43:1111–8. doi:10.1118/1.4940789.
[111] Liu C, Sio TT, Deng W, Shan J, Daniels TB, Rule WG, et al. Small-spot intensity-modulated proton therapy and volumetric-modulated arc therapies for patients with locally advanced non-small-cell lung cancer: A dosimetric comparative study. J Appl Clin Med Phys 2018;19:140–8. doi:10.1002/acm2.12459.
[112] Pastor-Serrano O, Habraken S, Lathouwers D, Hoogeman M, Schaart D, Perkó Z. How should we model and evaluate breathing interplay effects in IMPT? Phys Med Biol 2021;66. doi:10.1088/1361-6560/ac383f.
[113] Smolders A, Hengeveld AC, Both S, Wijsman R, Langendijk JA, Weber DC, et al. Inter- and intrafractional 4D dose accumulation for evaluating ΔNTCP robustness in lung cancer. Radiother Oncol 2023;182:109488. doi:10.1016/j.radonc.2023.109488.
[114] Visser S, Korevaar EW, Muijs CT, Wijsman R, Langendijk JA, Pisciotta P, et al. Clinical 3D/4D cumulative proton dose assessment methods for thoracic tumours with large motion. Radiother Oncol 2023;182:109575. doi:10.1016/j.radonc.2023.109575.
[115] Buti G, Souris K, Montero AMB, Lee JA, Sterpin E. Towards fast and robust 4D optimization for moving tumors with scanned proton therapy. Med Phys 2019:4–6. doi:10.1002/mp.13850.
[116] Noufal MP, Sharma SD, Patro K, Arjunan M, Krishnan G, Tyagarajan R, et al. Impact of spot positional errors in robustly optimized intensity-modulated proton therapy plan of craniospinal irradiation. Radiol Phys Technol 2021;14:271–8. doi:10.1007/s12194-021-00625-3.
[117] Kopp B, Fuglsang Jensen M, Mein S, Hoffmann L, Nyström H, Falk M, et al. FRoG: An independent dose and LETd prediction tool for proton therapy at ProBeam® facilities. Med Phys 2020;47:5274–86. doi:10.1002/mp.14417.
[118] Kraan AC, Depauw N, Clasie B, Madden T, Kooy HM. Impact of spot size variations on dose in scanned proton beam therapy. Phys Medica 2019;57:58–64. doi:10.1016/j.ejmp.2018.12.011.
[119] Hofmaier J, Dedes G, Carlson DJ, Parodi K, Belka C, Kamp F. Variance-based sensitivity analysis for uncertainties in proton therapy: A framework to assess the effect of simultaneous uncertainties in range, positioning, and RBE model predictions on RBE-weighted dose distributions. Med Phys 2021;48:805–18. doi:10.1002/mp.14596.
[120] Ödén J, Eriksson K, Toma-Dasu I. Incorporation of relative biological effectiveness uncertainties into proton plan robustness evaluation. Acta Oncol (Madr) 2017;56:769–78. doi:10.1080/0284186X.2017.1290825.
[121] Ödén J, Eriksson K, Toma-Dasu I. Inclusion of a variable RBE into proton and photon plan comparison for various fractionation schedules in prostate radiation therapy. Med Phys 2017;44:810–22. doi:10.1002/mp.12117.
[122] Nguyen TTC, Nguyen BT, Mai N V. Robustness evaluation of Intensity Modulated Proton Therapy plans using Dose Volume Population Histogram. Phys Medica PM an Int J Devoted to Appl Phys to Med Biol Off J Ital Assoc Biomed Phys 2019;65:219–26. doi:10.1016/j.ejmp.2019.09.070.
[123] Borderías-Villarroel E, Taasti V, Van Elmpt W, Teruel-Rivas S, Geets X, Sterpin E. Evaluation of the clinical value of automatic online dose restoration for adaptive proton therapy of head and neck cancer. Radiother Oncol 2022;170:190–7. doi:10.1016/j.radonc.2022.03.011.
[124] Hedrick SG, Petro S, Ward A, Morris B. Validation of automated complex head and neck

treatment planning with pencil beam scanning proton therapy. J Appl Clin Med Phys 2022;23:1–8. doi:10.1002/acm2.13510.
[125] Leach K, Tang S, Sturgeon J, Lee AK, Grover R, Sanghvi P, et al. Beam-specific spot guidance and optimization for PBS proton treatment of bilateral head and neck cancers. Int J Part Ther 2021;8:50–61. doi:10.14338/IJPT-20-00060.1.
[126] Deiter N, Chu F, Lenards N, Hunzeker A, Lang K, Mundy D. Evaluation of replanning in intensity-modulated proton therapy for oropharyngeal cancer: Factors influencing plan robustness. Med Dosim Off J Am Assoc Med Dosim 2020;45:384–92. doi:10.1016/j.meddos.2020.06.002.
[127] Liu C, Patel SH, Shan J, Schild SE, Vargas CE, Wong WW, et al. Robust Optimization for Intensity Modulated Proton Therapy to Redistribute High Linear Energy Transfer from Nearby Critical Organs to Tumors in Head and Neck Cancer. Int J Radiat Oncol Biol Phys 2020;107:181–93. doi:10.1016/j.ijrobp.2020.01.013.
[128] Håkansson K, Smulders B, Specht L, Zhu M, Friborg J, Rasmussen JH, et al. Radiation dose-painting with protons vs . photons for head-and-neck cancer. Acta Oncol (Madr) 2020;59:525–33. doi:10.1080/0284186X.2020.1714720.
[129] Cubillos-Mesías M, Troost EGC, Lohaus F, Agolli L, Rehm M, Richter C, et al. Quantification of plan robustness against different uncertainty sources for classical and anatomical robust optimized treatment plans in head and neck cancer proton therapy. Br J Radiol 2020;93:1–9. doi:10.1259/bjr.20190573.
[130] Hamming-Vrieze O, Depauw N, Craft DL, Chan AW, Rasch CRN, Verheij M, et al. Impact of setup and range uncertainties on TCP and NTCP following VMAT or IMPT of oropharyngeal cancer patients. Phys Med Biol 2019;64. doi:10.1088/1361-6560/ab1459.
[131] Stützer K, Lin A, Kirk M, Lin L. Superiority in Robustness of Multifield Optimization Over Single-Field Optimization for Pencil-Beam Proton Therapy for Oropharynx Carcinoma: An Enhanced Robustness Analysis. Int J Radiat Oncol Biol Phys 2017;99:738–49. doi:10.1016/j.ijrobp.2017.06.017.
[132] Cubillos-Mesías M, Baumann M, Troost EGC, Lohaus F, Löck S, Richter C, et al. Impact of robust treatment planning on single- and multi-field optimized plans for proton beam therapy of unilateral head and neck target volumes. Radiat Oncol 2017;12. doi:10.1186/s13014-017-0931-8.
[133] Van Dijk L V., Steenbakkers RJHM, Ten Haken B, Van Der Laan HP, Van't Veld AA, Langendijk JA, et al. Robust Intensity Modulated Proton Therapy (IMPT) increases estimated clinical benefit in head and neck cancer patients. PLoS One 2016;11:1–14. doi:10.1371/journal.pone.0152477.
[134] Zhang Y, Alshaikhi J, Amos RA, Tan W, Anaya VM, Pang Y, et al. Pre-treatment analysis of non-rigid variations can assist robust intensity-modulated proton therapy plan selection for head and neck patients. Med Phys 2022. doi:10.1002/mp.15971.
[135] Yao W, Zhang B, Han D, Polf J, Vedam S, Lasio G, et al. Use of CBCT plus plan robustness for reducing QACT frequency in intensity-modulated proton therapy: Head-and-neck cases. Med Phys 2022. doi:10.1002/mp.15915.
[136] Borderias-Villarroel E, Huet Dastarac M, Barragán-Montero AM, Helander R, Holmstrom M, Geets X, et al. Machine learning-based automatic proton therapy planning: Impact of post-processing and dose-mimicking in plan robustness. Med Phys 2023;50:4480–90. doi:10.1002/mp.16408.
[137] Chang L, Shaaban SG, Gogineni E, Page B, Quon H, Li H, et al. Daily Head and Neck Treatment Assessment for Optimal Proton Therapy Planning Robustness. Cancers (Basel) 2023;15:1–11. doi:10.3390/cancers15143719.
[138] Bohannon D, Janopaul-Naylor J, Rudra S, Yang X, Chang CW, Wang Y, et al. Prediction of plan adaptation in head and neck cancer proton therapy using clinical, radiographic, and dosimetric features. Acta Oncol (Madr) 2023;62:627–34. doi:10.1080/0284186X.2023.2224050.
[139] Ng Wei Siang K, Both S, Oldehinkel E, Langendijk JA, Wagenaar D. Assessment of residual geometrical errors of clinical target volumes and their impact on dose accumulation for head and neck radiotherapy. Radiother Oncol 2023;188:109856. doi:10.1016/j.radonc.2023.109856.

[140] Hofmaier J, Walter F, Hadi I, Rottler M, von Bestenbostel R, Dedes G, et al. Combining inter-observer variability, range and setup uncertainty in a variance-based sensitivity analysis for proton therapy. Phys Imaging Radiat Oncol 2021;20:117–20. doi:10.1016/j.phro.2021.11.005.
[141] Zou W, Kurtz G, Nakib M, Burgdorf B, Alp M, Li T, et al. A Probability-Based Investigation on the Setup Robustness of Pencil-beam Proton Radiation Therapy for Skull-Base Meningioma. Int J Part Ther 2021;7:34–45. doi:10.14338/IJPT-20-00009.1.
[142] Ödén J, Toma-Dasu I, Witt Nyström P, Traneus E, Dasu A. Spatial correlation of linear energy transfer and relative biological effectiveness with suspected treatment-related toxicities following proton therapy for intracranial tumors. Med Phys 2020;47:342–51. doi:10.1002/mp.13911.
[143] Nenoff L, Matter M, Hedlund Lindmar J, Weber DC, Lomax AJ, Albertini F. Daily adaptive proton therapy–the key to innovative planning approaches for paranasal cancer treatments. Acta Oncol (Madr) 2019;58:1423–8. doi:10.1080/0284186X.2019.1641217.
[144] Rojo-Santiago J, Habraken SJM, Romero AM, Lathouwers D, Wang Y, Perkó Z, et al. Robustness analysis of CTV and OAR dose in clinical PBS-PT of neuro-oncological tumors: Prescription-dose calibration and inter-patient variation with the Dutch proton robustness evaluation protocol. Phys Med Biol 2023;68. doi:10.1088/1361-6560/acead1.
[145] Ger RB, Sheikh K, Gogineni E, Floreza B, Croog V, Li H, et al. Planning and Treatment Recommendations for Breast Proton Therapy From a Single Center's Experience. Adv Radiat Oncol 2023;8:101069. doi:10.1016/j.adro.2022.101069.
[146] Fuglsang Jensen M, Stick LB, Høyer M, Kronborg CJS, Lorenzen EL, Mortensen HR, et al. Proton therapy for early breast cancer patients in the DBCG proton trial: planning, adaptation, and clinical experience from the first 43 patients. Acta Oncol (Madr) 2022;61:223–30. doi:10.1080/0284186X.2021.1986229.
[147] Kirk M, Freedman G, Ostrander T, Dong L. Field-Specific Intensity-modulated Proton Therapy Optimization Technique for Breast Cancer Patients with Tissue Expanders Containing Metal Ports. Cureus 2017;9. doi:10.7759/cureus.1698.
[148] Wedenberg M, Lind BK, Hårdemark B. A model for the relative biological effectiveness of protons: The tissue specific parameter α/β of photons is a predictor for the sensitivity to LET changes. Acta Oncol (Madr) 2013;52:580–8. doi:10.3109/0284186X.2012.705892.
[149] Chieh-Wen L, Tianjun M, Tara G, Saeed A, Naichang Y, Kevin L S, et al. Dosimetric impact of tumor position displacements between photon and proton stereotactic body radiation therapy for lung cancer. J Radiosurgery SBRT 2022;8:137–46.
[150] Rana S, Rosenfeld AB. Small spot size versus large spot size: Effect on plan quality for lung cancer in pencil beam scanning proton therapy. J Appl Clin Med Phys 2022;23. doi:10.1002/acm2.13512.
[151] Rana S, Rosenfeld AB. Impact of errors in spot size and spot position in robustly optimized pencil beam scanning proton-based stereotactic body radiation therapy (SBRT) lung plans. J Appl Clin Med Phys 2021;22:147–54. doi:10.1002/acm2.13293.
[152] Borderías Villarroel E, Geets X, Sterpin E. Online adaptive dose restoration in intensity modulated proton therapy of lung cancer to account for inter-fractional density changes. Phys Imaging Radiat Oncol 2020;15:30–7. doi:10.1016/j.phro.2020.06.004.
[153] Szeto YZ, Witte MG, van Kranen SR, Sonke JJ, Belderbos J, van Herk M. Effects of anatomical changes on pencil beam scanning proton plans in locally advanced NSCLC patients. Radiother Oncol 2016;120:286–92. doi:10.1016/j.radonc.2016.04.002.
[154] Spautz S, Haase L, Tschiche M, Makocki S, Richter C, Troost EGC, et al. Comparison of 3D and 4D robustly optimized proton treatment plans for non-small cell lung cancer patients with tumour motion amplitudes larger than 5 mm. Phys Imaging Radiat Oncol 2023;27:100465. doi:10.1016/j.phro.2023.100465.
[155] Taasti VT, Hattu D, Peeters S, van der Salm A, van Loon J, de Ruysscher D, et al. Clinical evaluation of synthetic computed tomography methods in adaptive proton therapy of lung cancer patients. Phys Imaging Radiat Oncol 2023;27:100459. doi:10.1016/j.phro.2023.100459.

[156] Visser S, O. Ribeiro C, Dieters M, Mul VE, Niezink AGH, van der Schaaf A, et al. Robustness assessment of clinical adaptive proton and photon radiotherapy for oesophageal cancer in the model-based approach. Radiother Oncol 2022;177:197–204. doi:10.1016/j.radonc.2022.11.001.
[157] Tamura H, Kobashi K, Nishioka K, Yoshimura T, Hashimoto T, Shimizu S, et al. Dosimetric advantages of daily adaptive strategy in IMPT for high-risk prostate cancer. J Appl Clin Med Phys 2022;23. doi:10.1002/acm2.13531.
[158] Tilbæk S, Muren LP, Vestergaard A, Stolarczyk L, Rønde HS, Johansen TS, et al. Proton therapy planning and image-guidance strategies within a randomized controlled trial for high-risk prostate cancer. Clin Transl Radiat Oncol 2023;41. doi:10.1016/j.ctro.2023.100632.
[159] Knäusl B, Lebbink F, Fossati P, Engwall E, Georg D, Stock M. Patient Breathing Motion and Delivery Specifics Influencing the Robustness of a Proton Pancreas Irradiation. Cancers (Basel) 2023;15:1–13. doi:10.3390/cancers15092550.
[160] Zhu M, Flampouri S, Stanforth A, Slopsema R, Diamond Z, LePain W, et al. Effect of the initial energy layer and spot placement parameters on IMPT delivery efficiency and plan quality. J Appl Clin Med Phys 2023;24:1–10. doi:10.1002/acm2.13997.
[161] Rasch C, Steenbakkers R, van Herk M. Target Definition in Prostate, Head, and Neck. Semin Radiat Oncol 2005;15:136–45. doi:https://doi.org/10.1016/j.semradonc.2005.01.005.