Practical robustness evaluation in radiotherapy - A photon and proton-proof alternative to PTV-based plan evaluation

Korevaar, Erik W; Habraken, Steven J M; Scandurra, Daniel; Kierkels, Roel G J; Unipan, Mirko; Eenink, Martijn G C; Steenbakkers, Roel J H M; Peeters, Stephanie G; Zindler, Jaap D; Hoogeman, Mischa

Published in:
Radiotherapy and Oncology

DOI:
10.1016/j.radonc.2019.08.005

Publication date:
2019

Practical robustness evaluation in radiotherapy

Korevaar, Erik W.; Habraken, Steven J.M.; Scandurra, Daniel; Kierkels, Roel G. J.; Unipan, Mirko; Eenink, M; Steenbakkers, Roel J.H.M.; Peeters, Stephanie G.; Zindler, Jaap; Hoogeman, Mischa

Published in:
Radiotherapy and Oncology

Publication date:
2019

Highlights

- A practical method is described to assess robustness of CTV dose in both photon and proton treatments
- By a calibration procedure consistency with historic PTV-based evaluations is achieved
- The method has been clinically introduced in the Dutch proton centres and replaced PTV-based evaluations, solving its inaccuracies caused by the ‘static dose cloud approximation’

Practical robustness evaluation in radiotherapy – A photon and proton-proof alternative to PTV-based plan evaluation

Authors

Erik W. Korevaar[1], Steven J.M. Habraken[2,3], Daniel Scandurra[1], Roel G.J. Kierkels[1], Mirko Unipan[4], Martijn G.C. Eenink[2], Roel J.H.M. Steenbakkers[1], Stephanie G. Peeters[4], Jaap D. Zindler[2,3], Mischa Hoogeman[2,3], Johannes A. Langendijk[1]

Affiliations

1. Department of Radiation Oncology, University Medical Center Groningen, University of Groningen
2. Holland Proton Therapy Center, Delft, The Netherlands
3. Department of Radiation Oncology, Erasmus Medical Center Cancer Institute, Rotterdam, The Netherlands
4. Proton Therapy Centre South-East Netherlands (ZON-PTC), Maastricht, The Netherlands

Corresponding author

Erik W. Korevaar
Department of Radiation Oncology
University Medical Center Groningen
PO Box 30001, 9700 RB Groningen
The Netherlands
Phone +31 6 3162 3737
E-mail e.w.korevaar@umcg.nl

Abstract

Background and purpose

A planning target volume (PTV) in photon treatments aims to ensure that the clinical target volume (CTV) receives adequate dose despite treatment uncertainties. The underlying static dose cloud approximation (the assumption that the dose distribution is invariant to errors) is problematic in intensity modulated proton treatments where range errors should be taken into account as well. The purpose of this work is to introduce a robustness evaluation method that is applicable to photon and proton treatments and is consistent with (historic) PTV-based treatment plan evaluations.

Materials and methods

The limitation of the static dose cloud approximation was solved in a multi-scenario simulation by explicitly calculating doses for various treatment scenarios that describe possible errors in the treatment course. Setup errors were the same as the CTV-PTV margin and the underlying theory of 3D probability density distributions was extended to 4D to include range errors, maintaining a 90% confidence level. Scenario dose distributions were reduced to voxel-wise minimum and maximum dose distributions; the first to evaluate CTV coverage and the second for hot spots. Acceptance criteria for CTV D98 and D2 were calibrated against PTV-based criteria from historic photon treatment plans.

Results

CTV D98 in worst case scenario dose and voxel-wise minimum dose showed a very strong correlation with scenario average D98 ($R^2 > 0.99$). The voxel-wise minimum dose visualised CTV dose conformity and coverage in 3D in agreement with PTV-based evaluation in photon therapy. Criteria for CTV D98 and D2 of the voxel-wise minimum and maximum dose showed very strong correlations to PTV D98 and D2 ($R^2 > 0.99$) and on average needed corrections of -0.9% and +2.3%, respectively.

Conclusions

A practical approach to robustness evaluation was provided and clinically implemented for PTV-less photon and proton treatment planning, consistent with PTV evaluations but without its static dose cloud approximation.

Key words: Photon therapy; Proton therapy; Setup error; Range error; Margins; Robustness evaluation; Comparative planning

Introduction

The use of margins in photon radiotherapy is a long established and universally adopted method to provide adequate target coverage under the presence of uncertainties. The CTV-PTV margin provides a geometrical buffer zone around the target within which the desired dose is achieved for the majority of treatments; criteria of 95% of the prescription dose in 90% of the patient population has found general appeal [1,2]. The suitability of a geometrically-expanded buffer zone arises from the (relative) insensitivity of megavoltage photon dose distributions to density changes in the beam path. By and large, the biggest risk to a photon dose distribution is a geometrical miss – a translation of the CTV relative to the beam. Therefore, the static dose cloud approximation (dose distribution is invariant to errors) inherent in the PTV concept, while not perfect, has worked well [3].

This fundamental assumption is not valid for proton therapy (PT) [4,5] due to the relationship between the proton range and the medium-dependent stopping power. On top of the inherent uncertainties in predicting the stopping power [6,7], errors in proton range arise from beam path density changes due to patient set-up or anatomy differences [8]. This range uncertainty has been addressed by range adapted PTVs, where an extra margin is added to the distal portion of the PTV in the beam path. This works reasonably well for delivery techniques where the dose is uniform for each of the beams [9]. With the advent of pencil beam scanning, intensity modulated or multi-field optimised proton therapy (IMPT) became widely available [10]. Just as its intensity modulated photon radiotherapy counterpart, this technique allows uniform target coverage to be achieved using non-uniform beam doses, leading to potentially steep inter-beam dose gradients patched together within the target. This is where the use of a geometric buffer zone breaks down, as now the range uncertainties can lead to dose differences within the CTV itself and not just at the periphery [11].

One method which has been developed to overcome the limitations of the static dose cloud approximation is robust optimisation, whereby deviations from the nominal (i.e. error-free) dose distribution are calculated explicitly for each scenario and minimised through the use of planning objectives [12–14]. A scenario is commonly called ‘error scenario’, although in this work terminology ‘treatment scenario’ is used, see next section. Since becoming available in commercial treatment planning systems, robust optimisation has demonstrated the ability to improve target coverage robustness and/or organ at risk (OAR) sparing compared to PTV-based planning, in both proton [15–21] and photon therapy [22–25]. It must be noted however, that robust target coverage is generally one of many variables in the optimisation objective function and therefore is in competition with other objectives, such as OAR dose. Therefore, there must be some evaluation of the optimised plan that quantifies whether coverage criteria have been satisfied under the specified uncertainty conditions. In PTV-based planning there are established metrics for target coverage, such as PTV V95% ≥ 98% prescription dose [26], which ensures a certain probability of tumour control as well as consistency of practice and dose reporting world-wide. There is currently no such established metric for reporting on plan robustness, the consequence being that any two treatment plans with similar target coverage in the nominal scenario may behave very differently in the scenarios. One could assess the dose distribution of each scenario but in clinical practice this is unfeasible. Methods of summarising the dose deviations in the scenarios have been investigated, including the use of dose volume histogram uncertainty bands [27–29], error-bar distributions [30] and root-mean-square volume histograms [16]. While each of these methods are useful for relative comparisons of

robustness, clearly defined target coverage acceptance criteria, such as V95% ≥ 98% for the PTV, are lacking and while error-bar distributions provide 3D spatial information of where local dose deviations may occur, it is not clear how coverage can be assessed in a PTV-like way. These limitations are not trivial and lead to difficulties for physicians and physicists to assess the quality of a robustly optimised plan. Furthermore, without an established robustness metric, consistent with PTV-based evaluation, how can the quality of PTV-based treatment plans be compared to robustly optimised plans, especially in proton therapy where the PTV concept is no longer adequate? This question is particularly relevant in the new era of model-based patient selection whereby the effectiveness of different treatment modalities is compared. In the Dutch system, PT can be indicated if a significant clinical benefit can be expected relative to conventional photon treatments; in this case, a difference in the normal tissue complication probability (NTCP) [31,32]. However, in order for this comparison to be fair, the target coverage must be comparable.

The aim of this work is to introduce a robustness evaluation method that addresses the two main issues highlighted above: (i) overcome the limitations of current robustness evaluation methods, and (ii) ability to universally compare target coverage under uncertainties for any treatment modality (i.e. photons and protons) without the use of a PTV. Perhaps most importantly, the proposed method is consistent with current PTV practice, which provides continuity for a radiotherapy community that relies heavily on PTV-based clinical data [1,26,33].

Materials and methods

1. *Outline and terminology*

The scenario-based evaluation consists of the following steps:

i. Scenario-based evaluation method: treatment scenarios are created by sampling deviations from the generated treatment plan. The dose distribution is calculated for each treatment scenario.
ii. Evaluation dose distribution: Scenario dose distributions are combined into evaluation dose distributions (described below), so as to provide concise information for daily clinical decision making.
iii. Acceptance criteria calibration: Dose metrics are evaluated based on acceptance criteria, defined in a calibration procedure using a set of historic treatment plans.

2. Scenario-based evaluation method

Following the definition of Tilly [34], a treatment scenario describes a fractionated treatment with a systematic setup error and for each fraction a random setup error. In view of computational demands, sparse sampling is used, i.e. robustness is determined for only a limited number of well-chosen worst-case scenarios. More specifically, errors are drawn from known probability distributions and sampled at a 90% confidence level, as commonly used in CTV-PTV margin recipes [2] and close to the value of 85% proposed in [35].

Setup errors are modelled as rigid, uncorrelated, and normally distributed 3D translations of the planning CT, with standard deviations Σ_x, Σ_y and Σ_z. From integration of the 3D Gaussian distribution

$G(x; \Sigma_x^2) \bullet G(y; \Sigma_y^2) \bullet G(z; \Sigma_z^2)$ with $r_3^2 = (x/\Sigma_x)^2 + (y/\Sigma_y)^2 + (z/\Sigma_z)^2$, we find that $r_3 = 2.5$ corresponds to a 90% confidence level (figure 1) with r_3 the length of the error vector in 3D.

Random variations in patient setup σ, resulting in dose blurring and shrinkage of the volume treated to 95% of the prescribed dose (i.e. the V95%), are taken into account as an additional systematic contribution. Following [2]: $M = 2.5\Sigma + 1.64\sqrt{(\sigma^2 + \sigma_p^2)} - 1.64\,\sigma_p$ with σ_p the penumbra width, for σ_p = 3.2 mm, the systematic shift of the 95% isodose line caused by random variations can be approximated by 0.7σ. This way, again for computational efficiency, a multi-fraction treatment scenario is approximated as a treatment with only a systematic setup error.

Variations other than rigid setup errors, such as non-rigid deformations, anatomical changes, intrafraction motion [36] and rotations (to the extent that they are not accounted for by translations) are beyond the scope of this paper.

Under the static dose cloud assumption, the scenario-based approach is equivalent to a PTV-based evaluation, as illustrated in figure 2, and previously found by Harrington [37].

For PT, range errors must also be taken into account, usually modelled by scaling HUs of the planning CT. Assuming that range errors (d) are not correlated with setup errors and are normally distributed, range and setup errors can be combined into a 4D Gaussian distribution $r_4^2 = (x/\Sigma_x)^2 + (y/\Sigma_y)^2 + (z/\Sigma_z)^2 + (d/\Sigma_d)^2 = r_3^2 + (d/\Sigma_d)^2$, where Σ_d is the standard deviation of range errors (figure 1). Here we find that $r_4 = 2.8$ corresponds to a 90% confidence level, with r_4 the length of the error vector in 4D. Some relevant combinations of setup and range errors are listed in figure 1b. It is recommended to use the same setup errors in treatment scenarios for protons and photons, provided both modalities have comparable setup accuracy. From Figure 1b, it can be seen that the range error should then be reduced from the common value of 1.5 to 1.24 times the standard deviation to keep the confidence level at 90%. For example, a 3.5% range error that corresponds to $d/\Sigma d = 1.5$ reduces to 2.9%, an adjustment that is less than differences in clinical practice between institutes [6]. An example set of 14 treatment scenarios with rigid shifts of the planning CT, corresponding to directions defined on a cube, is shown in figure 3. Adding positive and negative range errors to each setup error results in 28 scenarios for protons and 14 for photons.

For each treatment scenario a dose distribution is calculated, which is referred to as the scenario dose.

3. Evaluation dose distribution

In order to provide meaningful yet simple methods to guide plan assessment in daily clinical practice, the multiple scenario dose distributions can be 'summarised' into an evaluation dose distribution. For example, the voxel-wise minimum dose [38] is the composite of minimum dose values per voxel from all scenarios (figure 4b). Alternatives include the voxel-wise mean dose (figure 4c) or to simply select the worst-case scenario dose, i.e. the dose of the scenario with worst CTV coverage (figure 4a). For each type of evaluation dose distribution, its suitability for assessment of target coverage was determined by investigating the correlation between the CTV V95% (or D98) to the average V95 (or D98) of all treatment scenarios. This was done for 842 proton and 150 photon plans (supplement). The voxel-wise maximum dose is another evaluation dose distribution which is more suited to identifying potential hot spots.

4. Acceptance criteria calibration

The transition from PTV-based to scenario-based evaluation imposes the risk of introducing a systematic change in patient treatment. This can be resolved by a calibration procedure in which correlations between evaluation parameters for both methods are analysed for a set of historic treatment plans. In conventional radiotherapy, PTV V95% ≥ 98% is commonly used [26]. McGowan et al. [39] created a database with metrics of historically treated patients as reference for new patient plans. Following a similar approach, correlations between the PTV- and scenario-based metrics were determined for conversion of established acceptance criteria to the new evaluation metrics for various photon cases (5 lung, 6 oesophagus, 8 breast and chest wall, 24 head and neck and 13 other indications).

Results

Target coverage in each of the three types of evaluation dose distributions showed high correlations using the CTV D98 metric (table 1 and figures S3, S4b), with a slope close to unity for both photon and proton plans. In comparison, the V95 loss metric showed weaker correlations and slopes that varied between the different evaluation distributions and modalities (figures S1, S4a). The variation in slopes can be explained by the way these dose distributions are constructed (figure S2). Of these, the voxel-wise minimum dose correlated best for both modalities (R^2 0.684 and 0.844 for proton and photon plans, respectively).

Qualitatively, evaluation of the CTV coverage in the voxel-wise minimum dose generally showed good agreement with PTV evaluations in the nominal dose distribution in slice-by-slice reviews (conceptually explained in figure 2 and demonstrated on patient examples in figures 5a, 5b). As a consequence, the voxel-wise minimum CTV D98 metric was clinically introduced for target coverage assessment and used in the second part of this study. The voxel-wise maximum dose evaluations indicated possible dose increases in cases with severe density inhomogeneities, like VMAT in the thoracic region, also shown in figure 5b.

Figure 5c shows examples of calibration of CTV D98 criteria of the voxel-wise minimum dose against D98 in PTV evaluations for various indications. A strong correlation was found with a slope just 0.9% below unity. Voxel-wise maximum dose calibration against D2 in PTV-based evaluation showed a strong correlation as well with a slope 2.3% above unity (figure 5d). Scenario-based evaluations increased D2-D98 inhomogeneity by about 3%, due to the effect of setup error on dose variations.

Discussion

Scenario-based robustness evaluation improved accuracy by removing the static dose cloud approximation in PTV-based evaluation and kept consistency with historic PTV-based evaluations. D98 of three types of evaluation dose distributions (voxel-wise minimum, voxel-wise mean and worst scenario) were found to be highly correlated with scenario average metrics. Of these, the voxel-wise minimum dose was selected as it provided valuable spatial information, such as conformity of the high dose to the target, and allowed clinicians to judge the clinical significance of any under-dosage just as with the traditional PTV approach. The voxel-wise maximum dose showed locations of possible hot spots in the target and can also be used for serial type OARs (e.g. spinal cord and optical nervous system) to replace planning at risk volume evaluations recommended by the ICRU.

For voxel-wise minimum and voxel-wise maximum dose distributions, only small corrections to PTV-based criteria were needed, which was defined in a calibration procedure (Figure 5c). This indicates that in photon treatments the static dose cloud approximation is largely true. In terms of model-based patient selection, the calibration allowed an un-biased comparison of organ at risk doses between proton and photon plans under comparably robust target coverage. As such, it forms the basis of the Dutch national consensus guidelines for proton plan evaluation. For practical reasons it was chosen to allow using PTV-based evaluation for photon plans and scenario-based evaluation in proton plans. Ideally robustness evaluation is performed as the standard method in both photon and proton planning since then results can be compared directly. In a transition phase from PTV-based to scenario-based evaluation in photon planning one could perform both evaluations to (i) gain familiarity and confidence in scenario-based evaluation and (ii) to obtain data for the calibration method.

A challenge is that scenario-based evaluation is computationally more expensive than PTV-based evaluation where speed is essential in a busy clinic. To keep calculation times clinically acceptable, several approximations were made:

- Monotonous increase of dosimetric errors with setup error size was assumed and no 'intermediate' scenarios were included. Although Casiraghi et al. found that, for IMPT, scenarios at cut-off values were representative [40], it is recommended to verify this per treatment site.
- Random setup errors were converted into systematic errors for protons the same way as for photons, based on penumbra width σ_p. Although proton and photon beams may have comparable penumbras, proton penumbras vary between machines and with airgap for range shifter beams.
- As random errors cause dose blurring, hot-spots will reduce in magnitude, and for evaluation of risk of hot-spots the setup error shift should be reduced. McKenzie et al. proposed a margin of 1.3Σ that was either reduced or increased with 0.5σ to include the effect of dose blurring [41].

These approximations may be accounted for by acceptance criteria adjustment in the calibration procedure.

Sensible error magnitudes should be defined per institution specific to their own circumstances, as these depend on patient setup protocols and CT calibration method. In this way the calibration to current PTV methods will result in target coverage criteria that are neither too strict nor too loose.

A limitation of the voxel-wise minimum and maximum dose evaluation is a lack of information of the number of scenarios involved. There is no distinction between a hot or cold dose region that occurs in a single or in many scenarios. A solution could be to report the percentage of scenarios in which dose criteria are fulfilled, as well as the scenario average (as done in this work) to reduce sensitivity to single, unfavourable, scenarios.

Although the effect of non-rigid deformations and rotations was not explicitly taken into account, high correlations can be expected with setup and range uncertainties. Our clinical experience is that IMPT plans that were optimised and evaluated to be robust against setup/range errors were usually robust against non-rigid anatomical changes seen on repeat CT scans. Nevertheless, the described

method can easily be extended to explicitly include repeated CTs and 4DCT scans to account for anatomical changes and breathing motion [42]. Further developments are to remove the approximation of random setup errors as a systematic error and introduce explicit dose calculations of fractionated, probabilistic treatment scenario sets. This would improve accuracy but requires speed-ups for clinical routine usage [43,44]. Finally, a probabilistic approach is a logical development towards a higher goal of robustness evaluation in terms of expected biological outcome.

In conclusion, to the best of our knowledge, the presented Dutch consensus of robustness evaluation is the first to provide comparable robustness of photon and proton treatments for planning comparison, and is in use in daily clinical practice. Centres that clinically introduce robustness evaluation should consider calibration of acceptance criteria against historic treatments to avoid introduction of systematic changes in treatments.

Conflicts of interest

None

Acknowledgements

The authors are indebted to many who contributed by sharing their view on robustness evaluation and especially thank Sanne van Dijk, Antje C. Knopf, Stefan Both, Hendrik P. Bijl, Frank Hoebers, Danielle Eekers and the European Particle Therapy Network Work package 5.

References

[1] ICRU. Prescribing, Recording and Reporting Photon Beam Therapy. ICRU Report 50. vol. os26. Journal of the ICRU; 1993.

[2] 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. doi:10.1016/S0360-3016(00)00518-6.

[3] Karlsson K, Lax I, Lindbäck E, Poludniowski G. Accuracy of the dose-shift approximation in estimating the delivered dose in SBRT of lung tumors considering setup errors and breathing motions. Acta Oncol (Madr) 2017;56:1189–96. doi:10.1080/0284186X.2017.1310395.

[4] Moyers MF, Miller DW, Bush DA, Slater JD. Methodologies and tools for proton beam design for lung tumors. Int J Radiat Oncol Biol Phys 2001;49:1429–38.

[5] Engelsman M, Kooy HM. Target volume dose considerations in proton beam treatment planning for lung tumors. Med Phys 2005;32:3549–57. doi:10.1118/1.2126187.

[6] Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. Phys Med Biol 2012;57:R99-117. doi:10.1088/0031-9155/57/11/R99.

[7] Lomax AJ. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 1: the potential effects of calculational uncertainties. Phys Med Biol 2008;53:1027–42. doi:10.1088/0031-9155/53/4/014.

[8] Lomax AJ. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: the potential effects of inter-fraction and inter-field motions. Phys Med Biol 2008;53:1043–56. doi:10.1088/0031-9155/53/4/015.

[9] Park PC, Zhu XR, Lee AK, Sahoo N, Melancon AD, Zhang L, et al. A beam-specific planning target volume (PTV) design for proton therapy to account for setup and range uncertainties. Int J Radiat Oncol Biol Phys 2012;82:e329-36. doi:10.1016/j.ijrobp.2011.05.011.

[10] Lomax AJ, Boehringer T, Coray A, Egger E, Goitein G, Grossmann M, et al. Intensity modulated proton therapy: a clinical example. Med Phys 2001;28:317–24. doi:10.1118/1.1350587.

[11] Kraan AC, Van De Water S, Teguh DN, Al-Mamgani A, Madden T, Kooy HM, et al. Dose uncertainties in IMPT for oropharyngeal cancer in the presence of anatomical, range, and setup errors. Int J Radiat Oncol Biol Phys 2013;87:888–96. doi:10.1016/j.ijrobp.2013.09.014.

[12] Unkelbach J, Chan TCY, Bortfeld T. Accounting for range uncertainties in the optimization of intensity modulated proton therapy. Phys Med Biol 2007;52:2755–73. doi:10.1088/0031-9155/52/10/009.

[13] Fredriksson A, Forsgren A, Hårdemark B. Minimax optimization for handling range and setup uncertainties in proton therapy. Med Phys 2011;38:1672–84. doi:10.1118/1.3556559.

[14] Chen W, Unkelbach J, Trofimov A, Madden T, Kooy H, Bortfeld T, et al. Including robustness in multi-criteria optimization for intensity-modulated proton therapy. Phys Med Biol 2012;57:591–608. doi:10.1088/0031-9155/57/3/591.

[15] 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:e0152477.

[16] Liu W, Frank SJ, Li X, Li Y, Park PC, Dong L, et al. Effectiveness of robust optimization in intensity-modulated proton therapy planning for head and neck cancers. Med Phys 2013;40:051711. doi:10.1118/1.4801899.

[17] Liu W, Frank SJ, Li X, Li Y, Zhu RX, Mohan R. PTV-based IMPT optimization incorporating planning risk volumes vs robust optimization. Med Phys 2013;40:21709. doi:10.1118/1.4774363.

[18] 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;12:1–10. doi:10.1186/s13014-017-0931-8.

[19] Li H, Zhang X, Park P, Liu W, Chang J, Liao Z, et al. Robust optimization in intensity-modulated proton therapy to account for anatomy changes in lung cancer patients. Radiother Oncol 2015;114:367–72. doi:10.1016/j.radonc.2015.01.017.

[20] Chang JY, Li H, Zhu XR, Liao Z, Zhao L, Liu A, et al. Clinical implementation of intensity modulated proton therapy for thoracic malignancies. Int J Radiat Oncol Biol Phys 2014;90:809–18. doi:10.1016/j.ijrobp.2014.07.045.

[21] Liu W, Mohan R, Park P, Liu Z, Li H, Li X, et al. Dosimetric benefits of robust treatment planning for intensity modulated proton therapy for base-of-skull cancers. Pract Radiat Oncol 2014;4:384–91. doi:10.1016/j.prro.2013.12.001.

[22] 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.

[23] Archibald-Heeren BR, Byrne M V., Hu Y, Cai M, Wang Y. Robust optimization of VMAT for lung cancer: Dosimetric implications of motion compensation techniques. J Appl Clin Med Phys 2017;18:104–16. doi:10.1002/acm2.12142.

[24] Jensen CA, Roa AMA, Johansen M, Lund J-A, Frengen J. Robustness of VMAT and 3DCRT plans toward setup errors in radiation therapy of locally advanced left-sided breast cancer with DIBH. Phys Med 2018;45:12–8. doi:10.1016/j.ejmp.2017.11.019.

[25] Wagenaar D, Kierkels RGJ, Free J, Langendijk JA, Both S, Korevaar EW. Composite minimax robust optimization of VMAT improves target coverage and reduces non-target dose in head and neck cancer patients. Radiother Oncol 2019;136:71–7. doi:10.1016/j.radonc.2019.03.019.

[26] ICRU. Prescribing, Recording, and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT). ICRU Report 83. vol. 10. Journal of the ICRU; 2010. doi:10.1093/jicru/ndq025.

[27] Trofimov A, Unkelbach J, DeLaney TF, Bortfeld T. Visualization of a variety of possible dosimetric outcomes in radiation therapy using dose-volume histogram bands. Pract Radiat Oncol 2012;2:164–71. doi:10.1016/j.prro.2011.08.001.

[28] Wang X, Zhang X, Li X, Amos R a., Shaitelman SF, Hoffman K, et al. Accelerated partial-breast irradiation using intensity-modulated proton radiotherapy: Do uncertainties outweigh potential benefits? Br J Radiol 2013;86:1–10. doi:10.1259/bjr.20130176.

[29] Liu W, Zhang X, Li Y, Mohan R. Robust optimization of intensity modulated proton therapy. Med Phys 2012;39:1079. doi:10.1118/1.3679340.

[30] 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.

[31] Langendijk J a, Lambin P, De Ruysscher D, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: The model-based approach. Radiother Oncol 2013;107:267–73. doi:10.1016/j.radonc.2013.05.007.

[32] Widder J, Van Der Schaaf A, Lambin P, Marijnen CAM, Pignol JP, Rasch CR, et al. The Quest for Evidence for Proton Therapy: Model-Based Approach and Precision Medicine. Int J Radiat Oncol Biol Phys 2016;95:30–6. doi:10.1016/j.ijrobp.2015.10.004.

[33] ICRU. Prescribing, Recording, and Reporting Proton-Beam. ICRU Report 78. vol. 7. Journal of the ICRU; 2007. doi:10.1093/jicru/ndi004.

[34] Tilly D, Ahnesjö A. Fast dose algorithm for generation of dose coverage probability for robustness analysis of fractionated radiotherapy. Phys Med Biol 2015;60:5439–54. doi:10.1088/0031-9155/60/14/5439.

[35] Goitein M. Nonstandard deviations. Med Phys 1983;10:709–11. doi:10.1118/1.595409.

[36] Grassberger C, Dowdell S, Lomax A, Sharp G, Shackleford J, Choi N, et al. Motion interplay as a function of patient parameters and spot size in spot scanning proton therapy for lung cancer. Int J Radiat Oncol Biol Phys 2013;86:380–6. doi:10.1016/j.ijrobp.2013.01.024.

[37] Harrington D, Liu W, Park P, Mohan R. SU-E-T-551: PTV Is the Worst-Case of CTV in Photon Therapy. Med Phys 2014;41:354–354.

[38] Lomax AJ, Pedroni E, Rutz H, Goitein G. The clinical potential of intensity modulated proton therapy. Z Med Phys 2004;14:147–52. doi:10.1078/0939-3889-00217.

[39] McGowan SE, Albertini F, Thomas SJ, Lomax a J. Defining robustness protocols: a method to include and evaluate robustness in clinical plans. Phys Med Biol 2015;60:2671–84. doi:10.1088/0031-9155/60/7/2671.

[40] Casiraghi M, Albertini F, Lomax a J. Advantages and limitations of the "worst case scenario" approach in IMPT treatment planning. Phys Med Biol 2013;58:1323–39. doi:10.1088/0031-9155/58/5/1323.

[41] McKenzie A, Van Herk M, Mijnheer B. Margins for geometric uncertainty around organs at risk in radiotherapy. Radiother Oncol 2002;62:299–307. doi:10.1016/S0167-8140(02)00015-4.

[42] Ribeiro CO, Meijers A, Korevaar EW, Both S, Langendijk JA, Knopf A-C. Comprehensive 4D robustness evaluation for pencil beam scanned proton plans. Radiother Oncol 2019.

[43] 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.

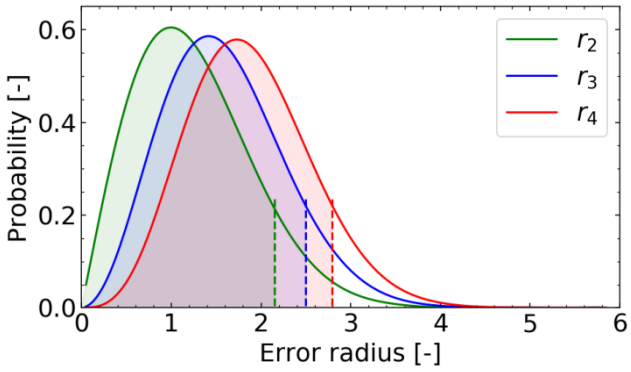
[44] Van Der Voort S, Van De Water S, Perkó Z, Heijmen B, Lathouwers D, Hoogeman M. Robustness Recipes for Minimax Robust Optimization in Intensity Modulated Proton Therapy for Oropharyngeal Cancer Patients. Int J Radiat Oncol Biol Phys 2016;95:163–70. doi:10.1016/j.ijrobp.2016.02.035.

Table 1. *Linear regression results of target coverage metrics (CTV V95 loss or D98) for evaluation dose distributions and scenario averages. See corresponding supplementary figures S1, S3 and S4)*

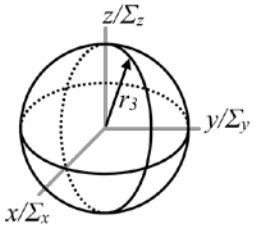
	Proton plans		Photon plans	
Evaluation dose metric	Slope	R^2	Slope	R^2
V95 loss of voxel-wise minimum	11.2	0.684	7.83	0.822
V95 loss of worst scenario	3.44	0.617	3.13	0.565
V95 loss of voxel-wise mean	0.249	0.27	0.229	0.34
D98 of voxel-wise minimum	0.976	0.998	0.977	0.999
D98 of worst scenario dose	0.99	0.999	0.985	0.999
D98 of voxel-wise mean	1.01	0.999	1.01	1.000

Abbreviations: CTV = clinical target volume; V95 loss = percentage volume of CTV below 95% of prescribed dose; D98 = minimum dose to 98% of the volume of the CTV; R = correlation coefficient

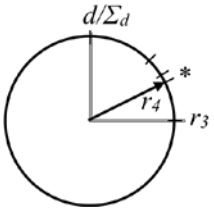
Figures Manuscript



a)



b)



c)

r_3	d/Σ_d
2.8	0.0
2.5*	1.24
2.36	1.5†
1.97	1.97
0.0	2.8

d)

*Value for the setup error from Van Herk's paper.

†Value for the range error from Paganetti's paper.

Figure 1a) Examples of probability distributions in 2D (green), 3D (blue) and 4D (red) simulated by taking error samples for each dimension from uncorrelated normal distributions with standard deviations of 1. On the x axis the error radius in 2D, 3D and 4D (r_2, r_3 and r_4, respectively), and on the y-axis the probability density. The vertical dashed lines indicate a 90% confidence level. It can be seen that increasing the number of dimensions widens the distribution such that a constant confidence level requires increase of the cut-off value; from 2.15 and 2.5 in 2D and 3D, to 2.8 in 4D, respectively. Figure 1b) Illustration of radius r_3, the length of the vector defined by x, y, and z setup errors relative to their standard deviations. Figure 1c) radius r_4 is defined by the r_3 setup error and the range error relative to its standard deviation (d/Σ_d). Marks on the circle correspond to the values given in figure 1d. Figure 1d) Example combinations of setup and range errors with 90% confidence level. Note that increase in the setup error confidence level allows reduction of the range error confidence while keeping the same overall confidence level.

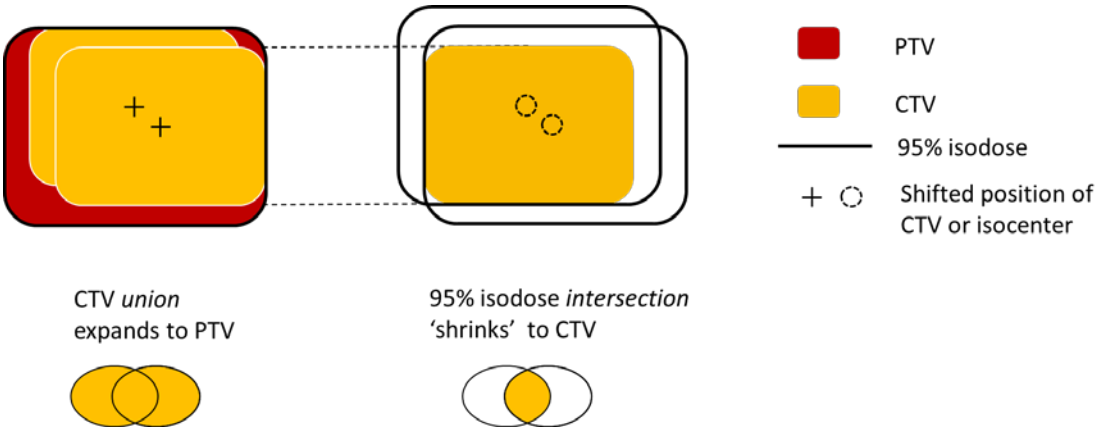


Figure 2. Under the static dose cloud approximation, PTV-based and scenario-based evaluation are equivalent in terms of CTV coverage under uncertainties. In PTV-based evaluation, the 95% isodose is static and covers the union of all translations of the CTV relative to the treatment isocentre (left image). In scenario-based evaluation, the CTV is static and is covered by the intersection of the 95% isodoses that are computed for all translations of the isocentre (right image).

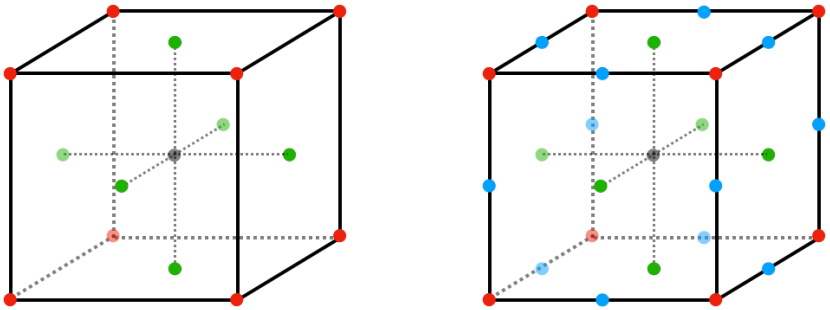
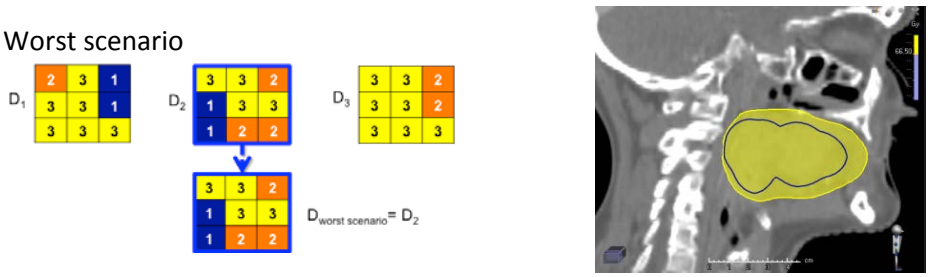


Figure 3. Examples of sampling setup error in a limited number of directions. Common choices include the 6 principal directions (6 green dots), from the centre to the vertices (8 red dots), to points in between (12 blue dots), and combinations of these that result in 14 or 26 directions. The magnitude of the shift is equal to the error radius $r3$ as described in figure 1. The sizes of the box in x, y, and z directions may be chosen anisotropic to reflect that in case of anisotropic setup uncertainties shifts in certain directions are more likely than in other directions.



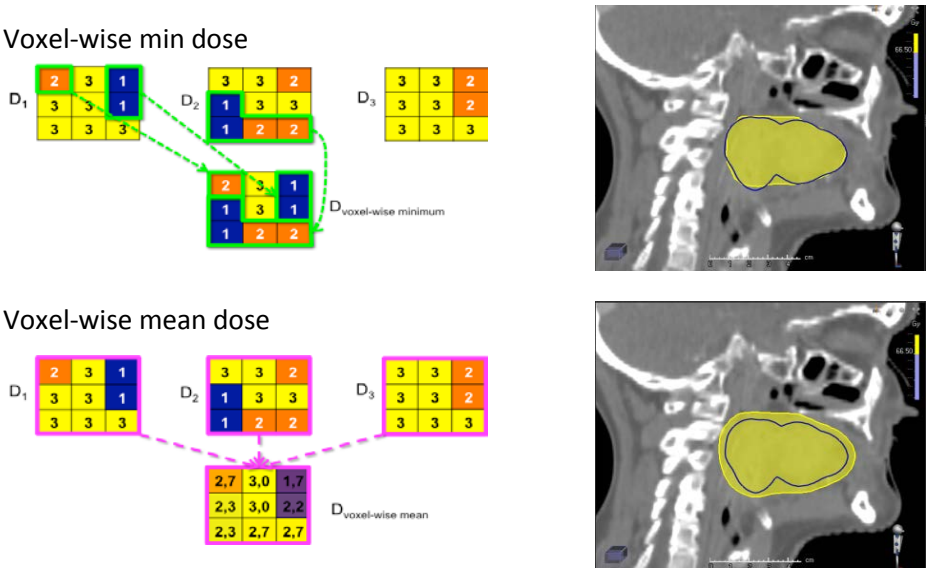
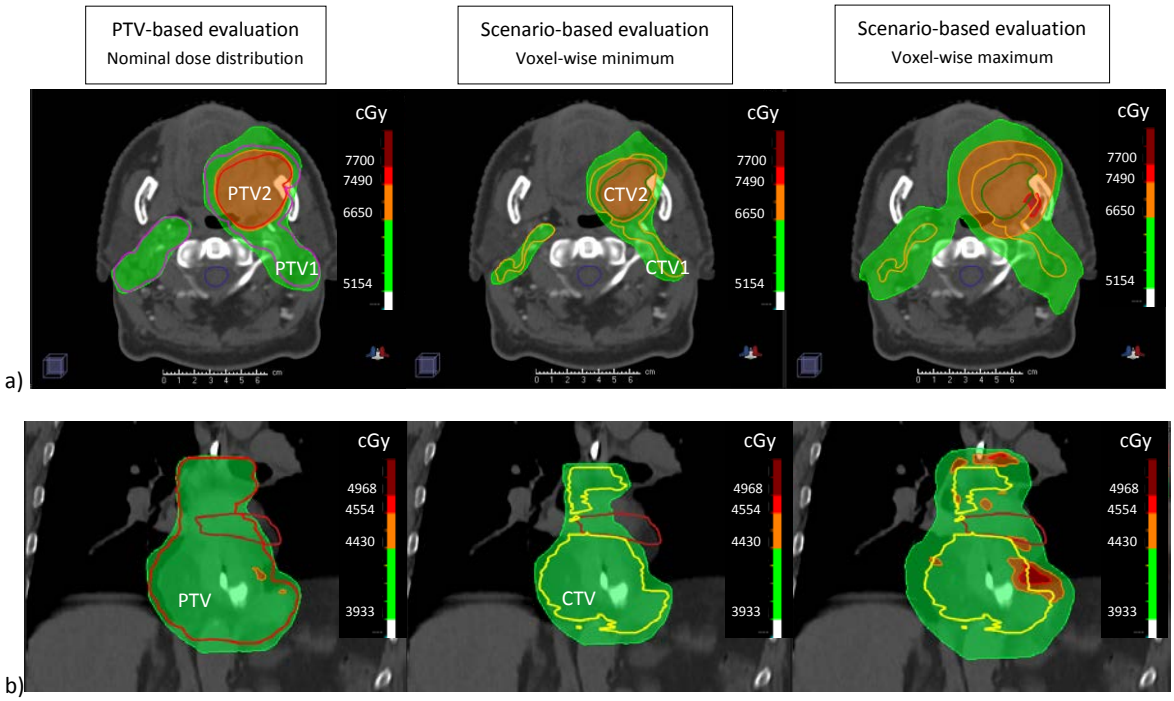
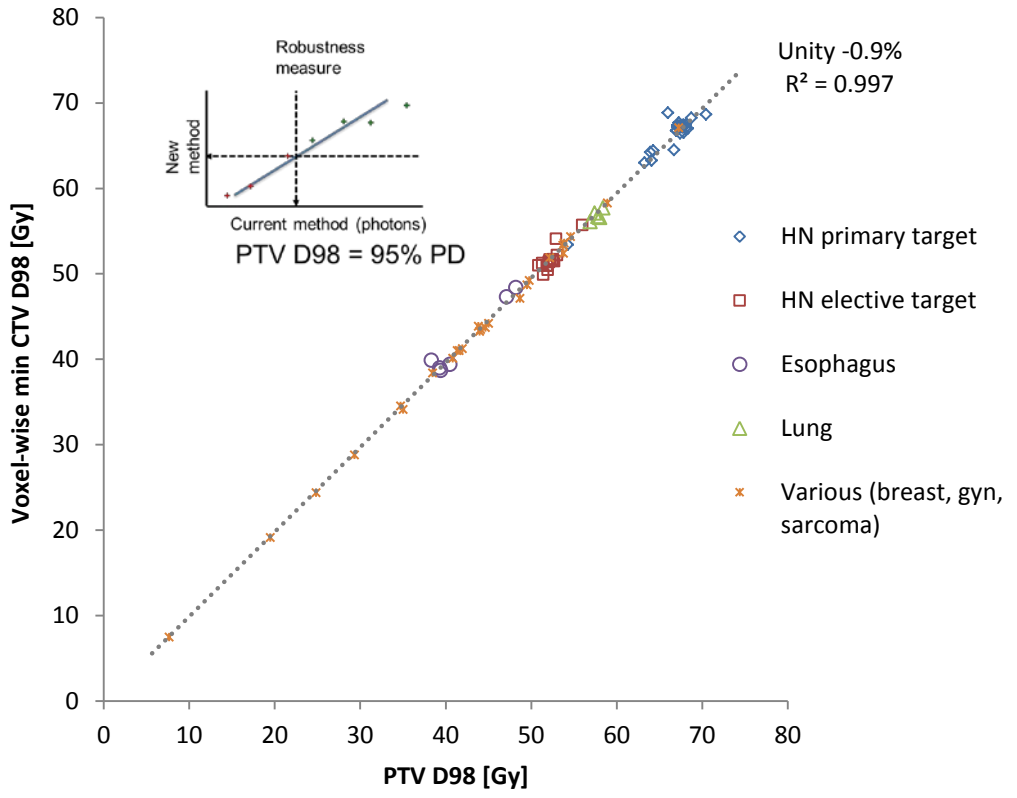


Figure 4. Example of the worst scenario dose, voxel wise minimum dose and voxel-wise mean dose for three scenarios (left column, higher values and brighter colours correspond to higher dose) and a photon treatment of an oropharynx patient (CTV in blue and 95% isodose in yellow). The worst scenario is a single scenario that is selected based on CTV dose metrics (e.g. lowest D98), the voxel-wise minimum is a composite of lowest dose per voxel in all scenarios, the voxel-wise mean dose is the scenario average dose per voxel. The worst scenario shows how tight the dose is only on one side of the target (in one scenario), the voxel-wise min dose shows a conformal dose whereas the voxel-wise mean dose does not correctly show how tightly the dose was planned.


PTV-based evaluation
Nominal dose distribution
Scenario-based evaluation
Voxel-wise minimum
Scenario-based evaluation
Voxel-wise maximum
a)
cGy
7700
7490
6650
5154
PTV2
PTV1
CTV2
CTV1
b)
4968
4554
4430
3933
PTV
CTV

c)


Robustness measure
New method
Current method (photons)
PTV D98 = 95% PD
Unity -0.9%
R² = 0.997
HN primary target
HN elective target
Esophagus
Lung
Various (breast, gyn, sarcoma)
Voxel-wise min CTV D98 [Gy]
PTV D98 [Gy]
0
10
20
30
40
50
60
70
80

d)

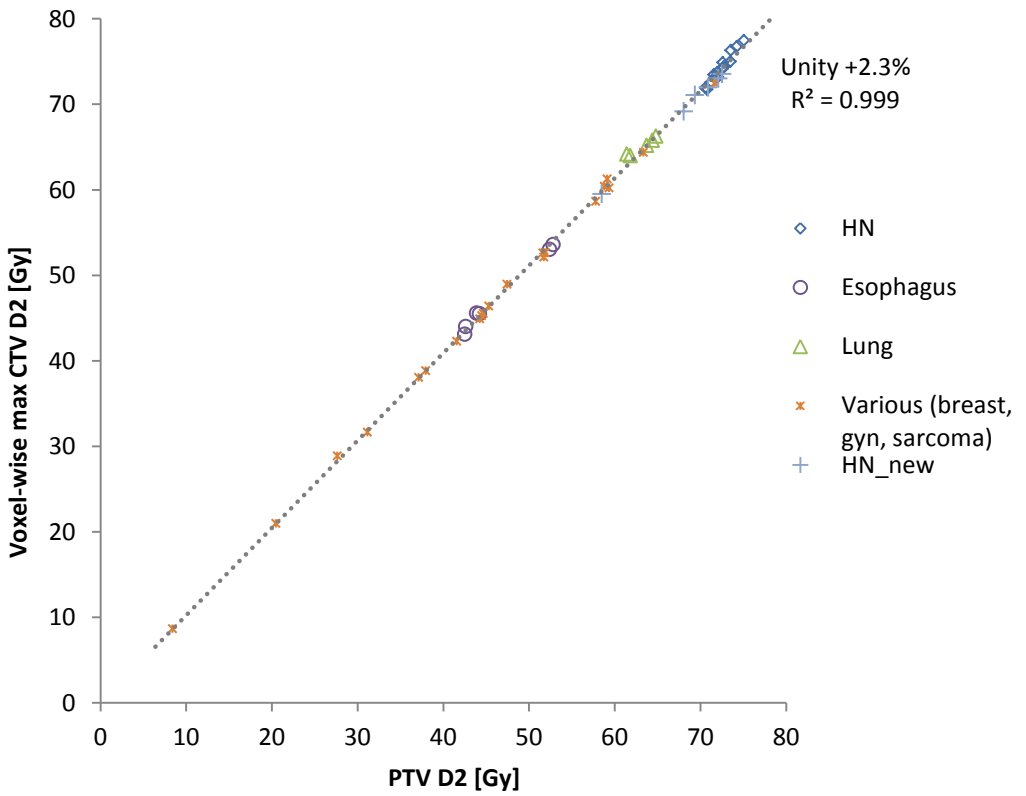


Figure 5. Comparison of PTV-based evaluation and robustness evaluation for various photon cases. a-b) examples of PTV-based evaluation (left) vs. scenario-based evaluation: voxel-wise min (middle) and voxel-wise max (right). In a) a head and neck VMAT (primary and elective PTV in red and pink, primary and elective CTV in orange and green and spinal cord in blue, 95% of primary and elective dose level in orange & green) is shown and treatment scenarios in the robustness evaluation included 5 mm systematic setup errors (including random errors converted to systematic as described in text). In b) an esophagus VMAT (PTV in red, ITV in yellow and heart in brown 95% dose level in green) is shown and treatment scenarios included 8 mm systematic setup errors. 95% dose conformity to CTV in voxel-wise minimum dose agrees with 95% dose conformity to PTV whereas voxel-wise maximum dose shows possible hot dose regions near density heterogeneities like bone (for head and neck) and lung/mediastinum. c-d) Scatter plots of voxel-wise minimum and maximum dose illustrates calibration of D98 and D2, respectively, with small correction factors and high correlation coefficients given in the figure. The inset in c) illustrates how criteria for the new (voxel-wise minimum) evaluation method can be derived from current PTV criteria and linear regression.

Figures Supplement

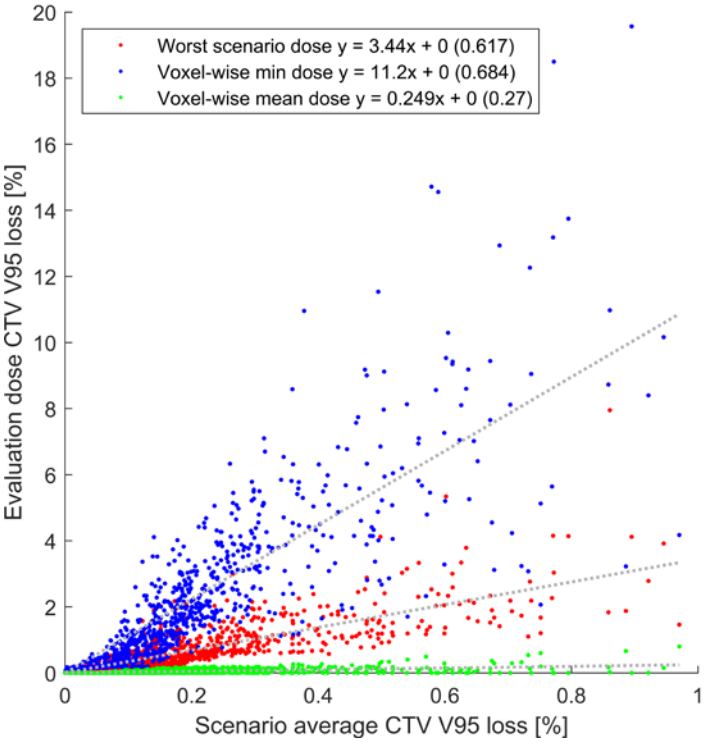


Figure S1. Robustness evaluations of CTV dose against setup and range errors of 842 proton plans in terms of V95 loss (defined as 100% – V95, i.e. the percentage of CTV not covered by the 95% isodose). The number of shifts were 14 or 26 and combined with two range errors resulted in 28 or 52 treatment scenarios per evaluation. V95 loss of evaluation doses on y-axis are plotted against scenario average values on the x-axis. Each marker corresponds to one treatment plan; different marker colours correspond to different evaluation doses (see legend). The dotted lines show linear regression lines for the three evaluation dose distribution types, with intercepts at zero and slopes given in the formulas (values before the 'x'). Correlation coefficients are given in brackets behind the formula as R^2 values.

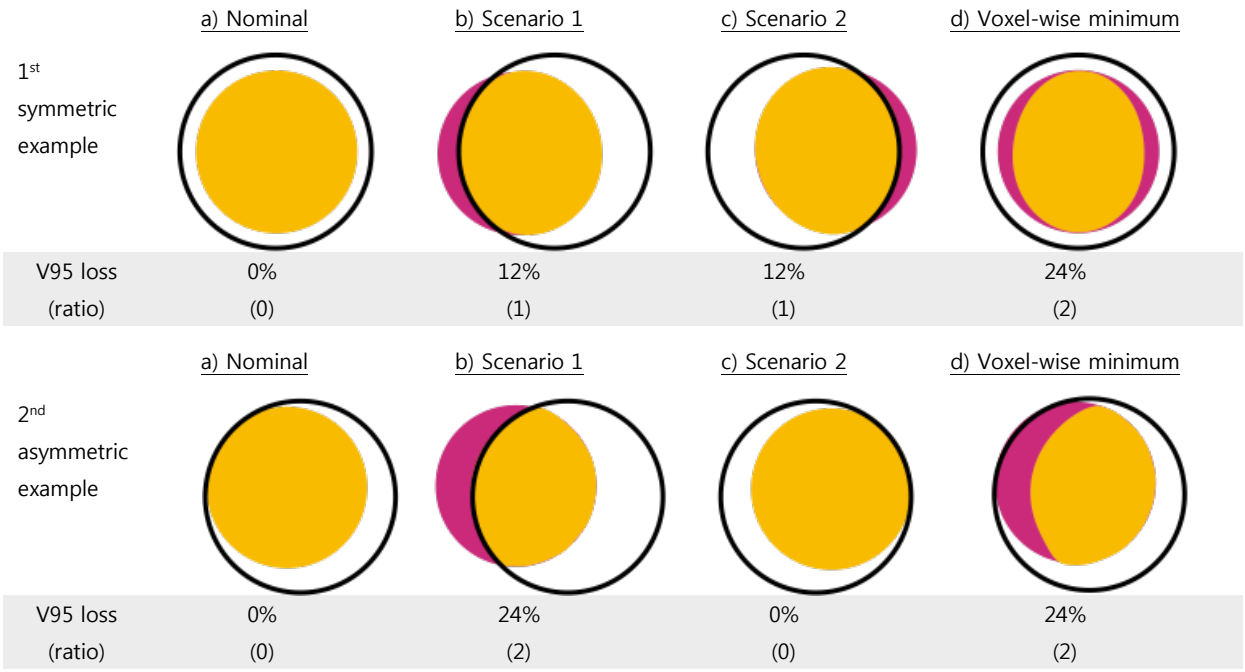


Figure S2. Two examples of a nominal 95% isodose line (black) that covers a CTV (yellow) symmetrically (1st row) or asymmetrically (2nd row). There are two scenarios with shifts in either left or right direction (2nd and 3rd column). The shifts cause parts of the CTV not being covered by the 95% isodose line (shown in purple). The relative size of this area (V95 loss) is given as percentage below the images, as well as the ratio to the scenario average value (in brackets below the percentage). In the *first example with symmetrical dose* the V95 loss is 12% in both scenarios, so the average V95 loss is 12% as well. The worst-case dose is found in both scenarios and is equal to 12%, with a ratio to the average value of 1. The voxel-wise minimum dose results in 24% V95 loss since the areas of both scenarios are included, resulting in a 24% area and ratio of 2 relative to the scenario average value. In the *asymmetric example* the scenario V95 loss values are 24% and 0% in scenario 1 and 2, respectively, resulting in an average of 12% (same as 1st example). The worst scenario dose V95 loss is found in scenario 1 and is 24% (ratio of 2 relative to the scenario average). The voxel-wise minimum dose V95 loss in the second example is the same as in the first example (24% or a ratio of 2 same as 1st example). These examples indicate that asymmetry of dose around the target may increase the relative V95 loss of worst-case scenario dose, whereas the number of (independent) scenarios may increase the relative V95 loss of the voxel-wise minimum dose.

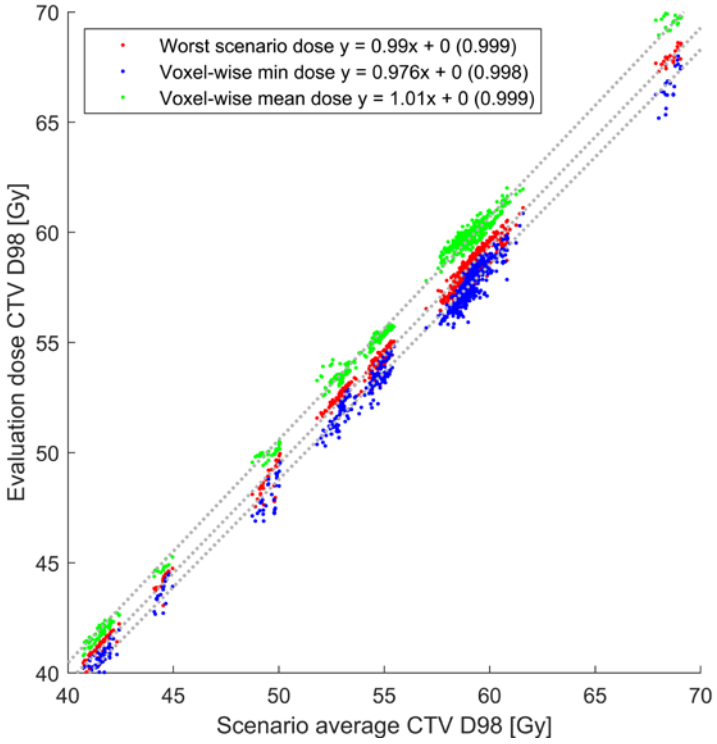
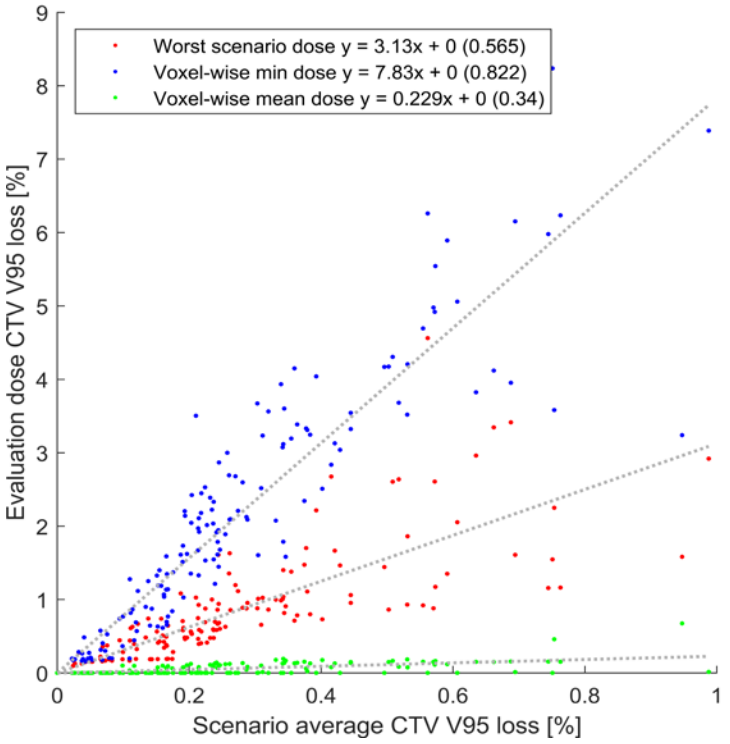


Figure S3. Robustness evaluations of CTV dose against setup and range errors of 842 proton plans (same plans as in figure S1) in terms of D98 of evaluation doses plotted against scenario average values. The linear regression lines (dotted lines) are shown with formulas in the legend (correlation coefficient R^2 given in brackets). Slopes are close to unity; slope of voxel-wise minimum dose is 2.4% below unity, worst scenario dose 1% below unity, and voxel-wise mean dose 1% above unity. Very strong correlations were found (R^2 close to 1.0) for all evaluation dose distributions.

a)



b)

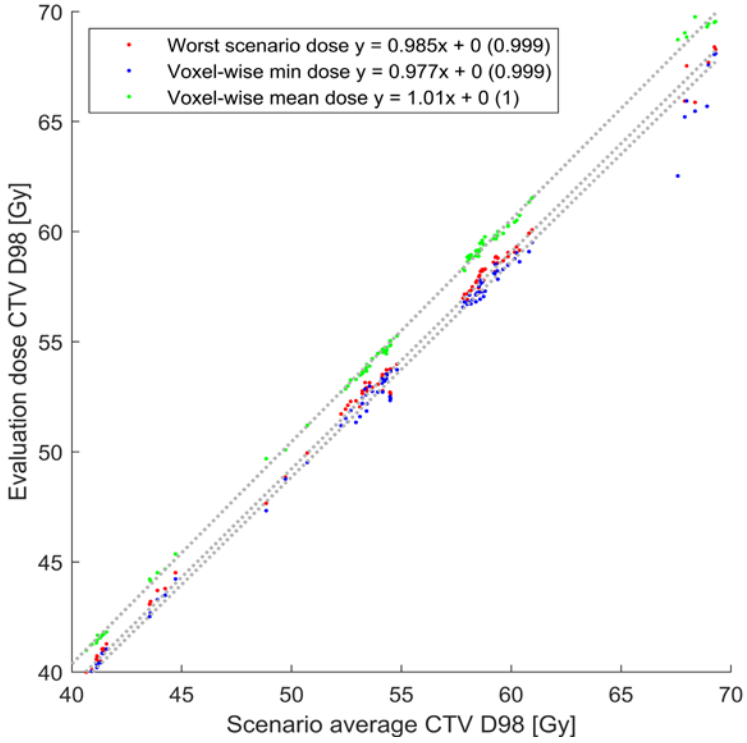


Figure S4. Robustness evaluations of CTV dose against setup errors of 150 photon plans in terms of V95 loss (a) and D98 (b) of evaluation doses plotted against scenario average values. The number of shifts was 14 or 26 per evaluation. The linear regression lines (dotted lines) are shown with formulas in the legend (correlation coefficient R^2 given in brackets). For V95 the variation in slopes is similar to the proton plans. For D98 the regression line slopes and correlation coefficients agree well with the results of the proton plans.

Supplement to manuscript: Practical robustness evaluation in radiotherapy – A photon and proton-proof alternative to PTV-based plan evaluation

V95 loss as target coverage metric

To investigate which evaluation dose distribution (e.g. the voxel-wise minimum dose or worst scenario dose distribution) is most suitable to assess target coverage, we introduced the 'V95 loss'. This metric is defined as 100% minus the V95 of the CTV, i.e. the volume percentage of the CTV that receives less than 95% of the prescribed dose. This metric is closely related to how MDs review target coverage by a slice-by-slice inspection of the three-dimensional (3D) dose distribution, and identify parts of the CTV that are not covered by 95% iso-dose line. The criterion of V95 of at least 98% of the CTV then corresponds to a V95 loss of 2% (100% - 98% = 2%).

Scenario average V95 loss vs. evaluation dose V95 loss

The V95 loss averaged over all scenarios is of interest as it shows how well the CTV is covered in case an error occurs (as defined by the error scenarios). If errors are sampled in a probabilistic way (i.e. smaller errors occur more frequently and larger errors occur less frequently, as defined by the probability distribution of errors), the scenario average is the expected value. Therefore, the scenario average V95 loss is expected to be a good measure of target coverage, but it lacks information in 3D. For plan evaluation in 3D, evaluation dose distributions are needed, but these have the disadvantage of possible bias. For instance, the V95 loss of the worst-case scenario is overestimated if the metric in the worst-case dose distribution is worse than in other, equally likely scenarios. A worst-case scenario only takes into account a single (bad) scenario whereas other, better scenarios are ignored.

Correlation: observations for protons and interpretation

The correlations between target coverage metrics of evaluation doses and scenario average values were investigated from 842 proton plans (figure S1). Note that the plans were partly made for research purpose, so may not be clinically acceptable regarding target coverage. Each regression line has a slope and a correlation coefficient. The first indicates the scale factor between target coverage metric of evaluation dose distributions and scenario average, and the latter the accuracy of the scale factor (a correlation coefficient of 1 shows perfect correlation and a value of 0 shows no correlation at all).

A slope of 11.2 was found for the V95 loss of the voxel-wise minimum dose, which means that the V95 loss in the voxel-wise minimum dose was a factor of 11.2 higher than the average of the scenarios (an average V95 loss of 1% corresponds to 11.2% V95 loss in the voxel-wise minimum dose). The high slope was expected from the way voxel-wise minimum dose is constructed; it is the composite of minimum dose values per voxel in various scenarios. If a part of the CTV is not covered by the 95% iso-dose in one scenario and another part is not covered in another scenario, both regions will be included in the voxel-wise minimum dose, thereby increasing the volume not covered as compared to that in individual error scenarios. This has been further clarified with a simple example of a circle shaped CTV and 95% iso-dose line (figure S2). For an example with only two error scenario's the ratio (or slope) between voxel-wise minimum and scenario average is two, so the high slope (>10) found in patient cases agrees with the large number of shifts in the error scenarios of at least 14.

The observation that the slope is less than the number of error scenarios can be explained by regions with dose below 95% in error scenarios that are not completely independent, i.e. partially overlap. Likewise, the under-covered regions caused by range errors in the proton evaluations probably often occur near the target edges, thereby overlapping the under-covered regions caused by shifts. It was also found that the slope was 10.7 and 11.7 if cases were filtered on number of error scenarios of 28 and 52, respectively (14 and 26 shifts and two range errors, respectively), which indicates that the results of 26 shifts largely overlap with 14 shifts.

The slope for the worst scenario dose was 3.4 (an average V95 loss of 1% corresponds to 3.4% V95 loss in the worst scenario). This number indicates the spread of V95 loss between scenarios; if in each scenario the V95 loss would be the same, then the worst is equal to the average and the slope is one. In the example in figure S2 it is shown how asymmetry in 95% dose around the CTV increases this slope.

Finally, the slope for the voxel-wise mean dose in proton patient cases was less than 0.3, so the V95 loss of this evaluation dose is more than a factor 3 below the scenario average. With regard to correlation coefficients, moderate correlations were found for the worst scenario and voxel-wise minimum dose distributions, whereas it was weak for the voxel-wise mean dose distribution. The correlation coefficients suggest that the voxel-wise minimum dose and the worst scenario dose are more suitable to estimate the scenario average V95 loss than the voxel-wise mean dose distribution.

Correlation: observations for D98 in proton plans

Because of the relatively large spread in data for V95 loss (figure S1), D98 was investigated as alternative metric for target coverage. Note that with a criterion of V95 ≥ 98% both V95 and D98 can be used. V95 might be sensitive in case large volumes exist with doses just above the 95% level in the nominal plan; small reductions in different error scenarios could result in large volumes below the 95% level. For the same 842 proton cases D98 of evaluation dose distributions were plotted against corresponding scenario average values and correlations were determined by linear regression (figure S3). Regression line slopes were close to unity; for the voxel-wise minimum dose it was 2.4% below unity, for the worst scenario dose 1% below unity, and for the voxel-wise mean dose 1% above unity. Furthermore, very strong correlations were found (R^2 close to 1.0) for all evaluation dose distributions. This suggests that D98 of all investigated evaluation dose distributions are suitable to estimate the scenario average D98 and require relatively small correction factors.

Correlation: observations for photons

The same robustness evaluations and analyses were performed for photon cases (figure S4). The slope of the regression line for V95 loss of the voxel-wise minimum dose was lower than for protons (7.8 vs. 11.2) which indicates that under-covered regions caused by range errors only partly overlap with those caused by setup errors (in case of 100% overlap, the slope would be the same for protons and photons). Similar slopes of the regression lines for V95 loss of the worst scenario and voxel-wise mean dose were found as well as similar correlation coefficients, compared to proton plan robustness evaluations. The regression lines for D98 agree well between photons and protons (slopes 2.3% vs. 2.4% below unity, 1.5% vs. 1.0% below unity and 1.0% vs. 1.0% above unity for voxel-wise minimum dose, worst scenario dose and voxel-wise mean dose for photons and protons, respectively). Correlation coefficients for D98 were approximately 1.0 for both modalities.