Check for updates

Coping with interfraction time trends in tumor setup

Marta K. Giżyńska[a)]
Medical Physics Department, Maria Sklodowska-Curie Institute - Oncology Center, 02-781 Warsaw, Poland
Faculty of Physics, Department of Biomedical Physics, University of Warsaw, 02-093 Warsaw, Poland
Department of Radiation Oncology, Erasmus MC University Medical Center Rotterdam, 3015GD Rotterdam, Netherlands

Paweł F. Kukołowicz
Medical Physics Department, Maria Sklodowska-Curie Institute - Oncology Center, 02-781 Warsaw, Poland

Ben J. M. Heijmen
Department of Radiation Oncology, Erasmus MC University Medical Center Rotterdam, 3015GD Rotterdam, Netherlands

Purpose: Interfraction tumor setup variations in radiotherapy are often reduced with image guidance procedures. Clinical target volume (CTV)–planning target volume (PTV) margins are then used to deal with residual errors. We have investigated characterization of setup errors in patient populations with explicit modelling of occurring interfraction time trends.

Methods: The core of a "trendline characterization" of observed setup errors in a population is a distribution of trendlines, each obtained by fitting a straight line through a patient's daily setup errors. Random errors are defined as daily deviations from the trendline. Monte Carlo simulations were performed to predict the impact of offline setup correction protocols on residual setup errors in patient populations with time trends. A novel CTV-PTV margin recipe was derived that assumes that systematic underdosing of tumor edges in multiple consecutive fractions, as caused by trend motion, should preferentially be avoided. Similar to the well-known approach by van Herk et al. for conventional error characterization (no explicit modelling of trends), only a predefined percentage of patients (generally 10%) was allowed to have nonrandom (systematic + trend) setup errors outside the margin. Additionally, a method was proposed to avoid erroneous results in Monte Carlo simulations with setup errors, related to decoupling of error sources in characterizations. The investigations were based on a database of daily measured setup errors in 835 prostate cancer patients that were treated with 39 fractions, and on Monte Carlo–generated patient populations with time trends.

Results: With conventional characterization of setup errors in patient populations with time trends, predicted standard deviations of residual systematic errors (Σ_{res}) after application of an offline correction protocol could be underestimated by more than 50%, potentially resulting in application of too small margins. With the new trendline characterization this was avoided. With the novel CTV-PTV margin recipe with an allowed 10% of patients having nonrandom errors outside the margin, the observed percentage was $10.0\% \pm 0.2\%$. When using conventional characterization of errors and the van Herk margin recipe, on average $58.0\% \pm 24.3\%$ of patients had errors outside the margin, while 10% was prescribed. For populations with no time trends, the novel recipe simplifies to the generally applied $M = 2.5\Sigma + 0.7\sigma$ formula proposed by van Herk et al.

Conclusions: In populations with time trends in setup errors, the use of trendline characterizations in Monte Carlo simulations for establishment of residual errors after a setup correction protocol can avoid application of erroneous margins. The novel margin recipe can be used to accurately control the percentage of patients with nonrandom errors outside the margin. In case of daily image guidance of patients with multiple targets with differential motion, the recipe can be used to establish margins for the targets that are not the primary target for the image guidance (e.g., nodal regions). Probabilistic planning might be improved by using trendline characterization for modelling of setup errors. Population analyses of interfraction setup errors need to take into account potential time trends. © 2019 The Authors. Medical Physics published by Wiley Periodicals, Inc. on behalf of American Association of Physicists in Medicine. [https://doi.org/10.1002/mp.13919]

1. INTRODUCTION

In fractionated radiotherapy, tumor setup errors at the linac are often mitigated with image-guided corrections.[1–3] For planning, a Clinical target volume (CTV)–planning target volume (PTV) margin[4,5] is used to cope with residual errors. Both for estimating the expected impact of a setup correction protocol on treatment accuracy and for establishment or

validation of margin recipes, Monte Carlo (MC) simulations may be performed using a characterization of the setup errors in the patient population.

It is generally assumed that setup errors occurring in a fractionated treatment of a patient can be described with normal distributions and that they can be characterized by the mean (systematic) error during the treatment and day-to-day variations around the mean (random errors). Observed systematic and random errors are used to derive population parameters that characterize the error distributions along the principal directions, that is, anterior–posterior, superior–inferior, left–right. In this paper, this much applied characterization of setup errors in a patient population is designated as the "conventional characterization"[6] (see Appendix A for equations).

Daily tumor setup relative to the treatment unit isocenter may gradually change during a fractionated treatment, resulting in an interfraction time trend in setup.[7–14] Existence of interfraction time trends has been reported by several groups for different cancer sites. El Gayed et al.[7] reported trend motion of 4–11 mm for 2 rectal and 3 prostate patients out of 10 patients per cancer site. Hanley et al.[8] found statistically significant trends in 10 out of the 50 prostate patients with a range of 2–7 mm. Stroom et al.[9] compared 15 prostate patients treated in prone position with 15 treated in supine position and found time trends in rectum diameter and prostate translations. van der Heide et al.[10] investigated prostate treatment with fiducials for 453 patients receiving a 35 fraction treatment. They found total motion of 3.1 mm in AP and 1.7 mm in SI direction. Namysl-Kaletka et al.[11] analyzed 57 patients with gastric cancer treated with 25 or 28 fractions. They reported 1 and 1.6 mm total trend motion in LR and SI direction, respectively. Gangsaas et al.[12] showed caudal trend motion of up to 11 mm (average 3.2 mm) for 30 patients with laryngeal cancer. Penninkhof et al.[13] showed more than 5 mm total tumor bed trend motion in 20% of breast cancer patients treated with a simultaneously integrated boost technique.

Such time trends are not explicitly considered in the conventional characterization. Rather, they are implicitly treated as part of the random error. However, a time trend motion is clearly deterministic, with a gradual, cumulative shift of the tumor in the 3D dose distribution during the fractionated treatment. This deterministic motion may have an impact on the performance of offline setup correction protocols, with corrections based on setup measurements in the first fractions. Not explicitly accounting for time trends in the CTV-PTV margin may result in underdosage of tumor edges in substantial numbers of consecutive fractions. For example, for a patient with no systematic setup error, a time trend in LR direction can result in a systematic underdose in the left tumor edge in each of the first 50% of fractions, and an underdose in the right tumor edge in all subsequent fractions. Existing rather simple TCP models suggest that the order of fractions with underdose would not be important. However, to the best of our knowledge there is no evidence that systematically underdosing the same part of the tumor in many fractions at a row, and compensating it with adequate dose delivery in the other fractions, would be equivalent to a random ordering of fractions with underdose and adequate dose. There are many examples in the radiotherapy literature showing that time patterns in dose delivery can indeed matter.

In this paper, we investigated the explicit modelling of interfraction time trends in tumor setup errors, using so-called trendline characterizations of setup errors observed in patient populations. Trendline characterizations were compared to conventional characterizations regarding accuracy of Monte Carlo (MC) predicted residual setup errors in case a no-action-level (NAL) protocol[2] or an extended NAL (eNAL) protocol[3] was used for setup corrections. We developed a novel CTV-PTV margin recipe that assumes that deterministic underdosing of tumor edges due to time trends should preferentially be avoided. The approach was highly similar to van Herk's derivation of his well-known margin recipe,[4] aiming at equality of the recipes in the limit of no time trends in the population.

The investigations included synthetic patient populations with time trends and a database with daily setup errors measured in a large population of prostate cancer patients that experienced time trends ("Erasmus database").

2. MATERIALS AND METHODS

2.A. Trendline characterization of setup errors to model interfraction time trends in setup errors

In contrast to a conventional characterization of setup errors in a population (Appendix A), a trendline characterization explicitly models occurring time trends.[3] To this purpose, for each patient p the setup errors in the fractionated treatment along each of the principal axes are characterized with a linear trendline, fitted through the daily measured setup errors (see Fig. 1), and defined by the slope a_p (mm/fraction), and middle position m_p (mm). The latter is the setup error half-way the fractioned treatment according to the fitted trendline, that is, the mean of the trendline tumor positions in fraction 1 and the last fraction F. It can be easily proven that this m_p equals the mean setup error in the fractionated treatment as used in the conventional characterization. In the remainder of this paper, setup errors according to the trendline are designated "trendline errors." Daily deviations from the trendline are now defined as the random errors (see Fig. 1). This leads to the following parameters defining a trendline characterization, with M and Σ also used in a conventional characterization:

- the overall mean setup error in the population

$$M = \frac{\sum_p m_p}{N} \tag{1}$$

with m_p the mean setup error of patient p in the fractionated treatment
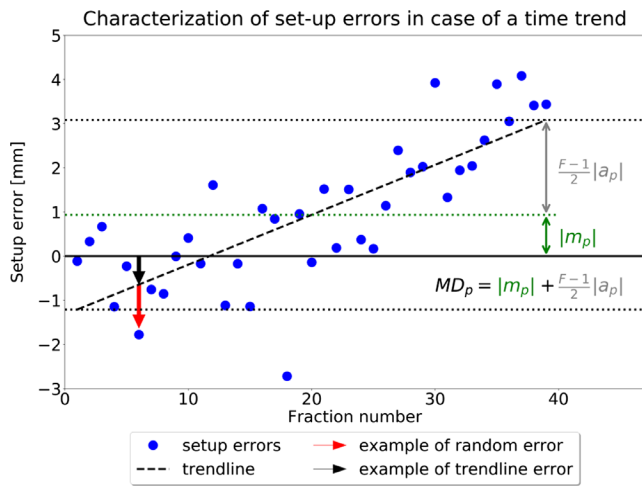
FIG. 1. Setup errors for an example patient with a time trend. The straight dashed line is the fitted trendline. Each fraction the total setup error (blue dot) is the sum of the error according to the dashed trendline (denoted as "trendline error," see black arrow as example) and the random error defined as the daily deviation from the setup according to the trendline (see red arrow). MD_p is the patient's maximum trendline setup error, used for calculation of clinical target volume (CTV)–planning target volume (PTV) margin for nonrandom errors (Section 2.F). m_p is the trendline error in the middle of the fractionated treatment which equals the systematic tumor setup. a_p is the trendline slope, F is the total number of fractions. [Color figure can be viewed at wileyonlinelibrary.com]

- the standard deviation describing the distribution of systematic (i.e., mean) setup errors in the population

$$\Sigma = \sqrt{\sum_p \frac{(m_p - M)^2}{N - 1}} \quad (2)$$

- the population mean of the trendline slopes:

$$M_a = \frac{\sum_p a_p}{N} \quad (3)$$

with a_p the trendline slope calculated for patient p, and N the number of patients in the population

- the standard deviation describing the interpatient variation in the trendline slopes:

$$\Sigma_a = \sqrt{\sum_p \frac{(a_p - M_a)^2}{N - 1}} \quad (4)$$

- the population standard deviation describing random errors relative to the trendlines (see also Fig. 1):

$$\sigma' = \sqrt{\sum_p \frac{SD_p'^2}{N}} \quad (5)$$

with SD_p' the standard deviation of random errors relative to the trendline observed for patient p

- the standard deviation describing the variation of SD_p' in the population[15]:

$$SD_{SD}' = \sqrt{\sum_p \frac{\left(\overline{SD_p'} - SD_p'\right)^2}{N - 1}} \quad (6)$$

2.B. The Erasmus database with prostate setup errors

Daily setup errors of 835 prostate cancer patients treated between November 2007 and May 2017 at the Erasmus MC Cancer Institute with 39 fractions of 2 Gy were used to build the database. Setup deviations in these patients were measured with kV/MV crossfire imaging of four implanted gold markers.[10,16–18] Setup errors were quantitated as marker center-of-mass displacements along the three principal directions, realizing that the mechanism for motion could in some cases also be rotations or deformations. The database was filled with setup errors that would have occurred in case no correction protocol would have been applied, that is, applied (*a priori*) setup corrections prior to imaging were subtracted from measured setup errors.

2.C. Synthetic patient populations with setup errors

Synthetic populations were created by choosing concrete values for the parameters in a trendline characterization and then using a MC approach to randomly create 39 fraction treatments for 10 000 patients (see Section 2.D for details). Parameters used for generation of synthetic setup errors along the three principal axes were: $M = 0$, $\Sigma \in \{1, 2, 3, 4\}$ mm, $\sigma' \in \{1, 2, 3\}$ mm, $SD_{SD}' = 0.75$ mm, $\Sigma_a \in \{0, 0.05, 0.1, 0.15\}$ mm/fraction, and $M_a \in \{0, -0.05, -0.1\}$ mm/fraction.

Simulations of the NAL and eNAL protocol (Sections 2.E and 3.B) were performed per principal direction, in line with the clinical application of these protocols.

Two types of synthetic populations were created for investigations on the CTV-PTV margin: isotropic (using the same distributions of trendline parameters for all three directions) and anisotropic (with different, randomly chosen distributions of trendline parameters for the three directions). By including all possible combinations of preselected parameters (above), a total number of 144 isotropic populations was generated. For generating the anisotropic populations, the same parameters were used as for isotropic populations but randomly chosen for each of principal directions (without replacements) resulting also in 144 anisotropic populations.

2.D. Monte Carlo (MC) generation of setup errors in a patient population

As described above, setup errors in a population are generally described by a conventional characterization. As discussed in this paper, alternatively, a trendline characterization can be used if time trends are (potentially) present. However, in both cases, when using the characterization parameters in a MC experiment for generating a

population of, for example, 10 000 new patients, the characterization parameters for these 10 000 patients will not be equal to the original parameters. The problem can be illustrated for a conventional characterization using a simple example: the random setup errors for a particular patient in the 39 fractions treatment are randomly drawn from a gaussian distribution $G(0, SD_p)$, with SD_p randomly drawn from the distribution $G(\sigma, SD_{SD})$. Due to the finite number of fractions (39), the mean of the drawn 'random' errors for the patient will in general not be equal to zero, that is, effectively the drawn errors are not completely random as they have a systematic component. Basically, this is caused by de-coupling of error sources in the characterization. Something very similar occurs with a trendline characterization: due to the finite number ($F = 39$) of drawn random errors, both the mean setup error of the patient and the slope of the trendline will in general be different from the drawn m_p and a_p, respectively. To avoid errors in the MC simulations, corrections described in Appendix B were always performed for drawn random errors.

2.E. NAL and eNAL offline protocols for correction of interfraction setup errors

The NAL[2] and eNAL[3] protocols are briefly summarized in Appendix C. In this study, we investigated for conventional and trendline characterizations the accuracy of MC simulated predictions of residual systematic setup errors for the NAL and eNAL protocols in patient populations with time trends. For both characterizations, Σ_{res}, the standard deviation describing the population residual systematic errors after NAL or eNAL was established.

2.F. A novel CTV-PTV margin recipe to account for time trends in setup errors

For the derivation of the margin recipe $M^{PTV} = 2.5\Sigma + 0.7\sigma$, based on the conventional characterization of setup errors, the margins for systematic and random setup errors ($M_{sys}^{PTV} = 2.5\Sigma$ and $M_{rand}^{PTV} = 0.7\sigma$, respectively) were independently established. Similarly, in the proposed margin recipe for a population with time trends, the margin contribution related to the trendlines, defined by a slope and a middle position (the latter equaling the patient's systematic setup error see Section 2.A), and the contribution from the random errors around the trendlines are treated separately (see Fig. 1). Actually, $M_{rand}^{PTV} = 0.7\sigma$ in the van Herk recipe is replaced by $M_{rand}^{PTV} = 0.7\sigma'$ [see Eq. (5)], while a new term, M_{trend}^{PTV}, is derived for coping with the trendline errors to replace the M_{sys}^{PTV} term. Equivalent to the work by van Herk et al.,[4] this new term is derived by requiring that for 90% of patients, the full CTV is within the PTV for 100% of the (nonrandom) trendline errors.

The margin component related to trendline errors, M_{trend}^{PTV}, is a 3D vector. In order to calculate its components, $M_{trend,i}^{PTV}$, for each principal direction, i, a procedure similar to that used

by van Herk et al.[4] for deriving $M_{sys,i}^{PTV} = 2.5 \cdot \Sigma_i$ is used, assuming for each direction a spherical 3D situation. In other words, for establishment of $M_{trend,i}^{PTV}$ it is assumed that the distributions of trendline errors for all three directions, k, are the same as for direction i. This procedure would be performed for each axis i based on the population parameters: M_i, Σ_i, $M_{a,i}$ and $\Sigma_{a,i}$ [see Eqs. (1), (2), ..., for definitions].

i.   randomly select for a large number of patients, p, the trendlines, that is, select $m_{p,k}$ and $a_{p,k}$ for the three principal directions, k, from the gaussian distributions $G(M_i, \Sigma_i)$ and $G(M_{a,i}, \Sigma_{a,i})$;

ii. determine for each patient, p, the maximum setup deviations following from the trendlines for the three principal directions as $MD_{p,k} = |m_{p,k}| + \frac{F-1}{2}|a_{p,k}|$ (see Fig. 1, note: the $|m_{p,k}|, and |a_{p,k}|$ distributions are folded-Gaussians);

iii. determine for each patient the length of the vector defined by the $MD_{p,k}$: $L_{p,i} = \sqrt{\sum_k (MD_{p,k})^2}$;

iv. establish $M_{trend,i}^{PTV}$ as the 90th percentile value of the distribution of $L_{p,i}$-values.

After calculation of $M_{trend,i}^{PTV}$ for the three principal directions, i, margins in any direction are derived from the three-dimensional ellipsoid defined by the $M_{trend,i}^{PTV}$.

2.G. Validation of the proposed margin recipe

The recipe for calculation of the margin component for coping with trendline errors as described in the previous section was validated for all 288 synthetic populations. For each of the populations we assessed for which percentage of the 10 000 patients all trendline errors were within the calculated margin. According to the design requirements (previous section) this should be 90%, so only 10% of patients can have one or more trendline errors outside the calculated margin. Mathematically, a trendline error with components $t_{p,i}$, of a patient p, is within the calculated margin if $\sum_{i=1}^{3} \left(\frac{t_{p,i}}{M_{trend,i}^{PTV}}\right)^2 \leq 1$.

For comparison, for all synthetic populations (most of them with time trends, see Section 2.A) we also established the percentage of patients with all trendline errors within the margin as calculated with the van Herk recipe. For a population with time trends, the true random errors are quantified by σ' [Eq. (5)], yielding a margin for random errors, $M_{rand}^{PTV} = 0.7\sigma'$ (Section 2.E). However, in the van Herk approach, trendline errors are treated as random errors, resulting in $M_{rand}^{PTV} = 0.7\sigma$, with σ defined in Eq. (A1). As $\sigma \geq \sigma'$, the prescribed margin for random errors in the van Herk approach is slightly larger than actually required for the true random errors. For this reason, for the van Herk approach, we established for each of the synthetic populations the percentage of patients with all trendline errors inside an ellipsoid defined by the margin components $2.5\Sigma_i + 0.7(\sigma_i - \sigma_i')$.

2.H. An analytical expression for the novel CTV-PTV margin

Section 2.F describes a numerical procedure for deriving the CTV-PTV margin, given the trendline characterization of the setup errors in the patient population. For populations with $M_i = 0$ and $M_{a,i} = 0$, that is, assuming that on average the patients' systematic setup errors and trendline slopes are zero, we also derived an analytical expression for the margin. To this purpose, margins calculated with the method presented in Section 2.E were fit to Eq. (7). The least square method as implemented in the SciPy package was used to establish the values for the fitting parameters α, δ, and γ.

$$M_{trend,i}^{PTV}(M_i = 0, \Sigma_i, M_{a,i} = 0, \Sigma_{a,i}, F)$$
$$= \alpha \cdot (\Sigma_i + \Sigma_{A,i}) + \delta \frac{\Sigma_i \cdot \Sigma_{A,i}}{\sqrt{\Sigma_i^2 + \gamma \cdot \Sigma_i \cdot \Sigma_{A,i} + \Sigma_{A,i}^2}} \quad (7)$$

with $\Sigma_{A,i} = (F - 1) \cdot \Sigma_{a,i}/2$, the standard deviation describing the distribution of trend motions during one-half of the fractions.

2.I. Origin of time trend errors

We propose a statistical method to determine whether observed trends in a population are caused by limited numbers of fractions or and by other (e.g., physiological) causes. First, for a large number of patients (10^6) setup errors are randomly generated for fractionated treatments, using M, Σ, σ, and SD_{SD} (i.e., ignoring trendline parameters). For all simulated patients, trendlines are then fitted. Next, the distribution of trendline slopes obtained from the simulations is compared to the distribution of slopes derived from the original (i.e., clinical) data using the Kolmogorov–Smirnov two-sample test.

3. RESULTS

3.A. Characterization of setup errors in the Erasmus database

Figure 2 shows for the three principal directions the distributions of total trend motion in the fractionated treatments. Absolute total trend motion for 10% of patients was larger than 2.6, 5.2, and 5.3 mm for left–right, superior–inferior, and anterior–posterior directions, respectively. Table I shows
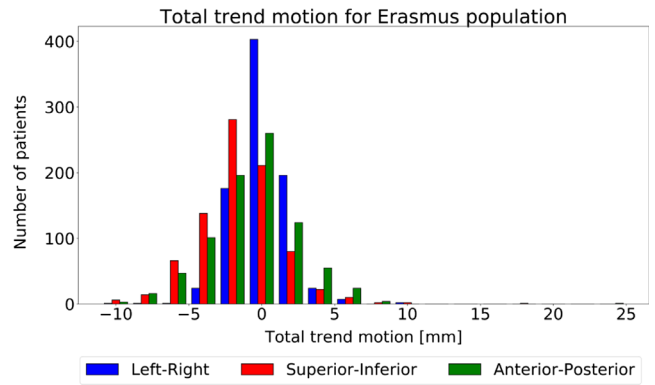

FIG. 2. Distributions of total trend motion, defined as the trendline error of a patient in the last fraction minus the trendline error in the first fraction, in the Erasmus database along the principal directions. [Color figure can be viewed at wileyonlinelibrary.com]

both the conventional characterization and the trendline characterization for the setup errors in the Erasmus database. Results of the test proposed in Section 2.I showed that observed trends in the Erasmus database are indeed larger than expected from the finite number of fractions ($P < 0.001$).

3.B. Monte Carlo simulations of residual setup errors for NAL and eNAL

For patient populations with time trends we investigated the impact of using a conventional characterization of setup errors for establishment of the distribution of residual systematic errors, Σ_{res}, instead of the more precise trendline characterization. Simulations were performed both for synthetic populations and for the measured errors in the Erasmus database. Always, the NAL/eNAL protocols were also simulated by directly using the fraction setup errors, that is, not using any characterization as an intermediate step. The latter simulations reflect the ground truth regarding the reduction of systematic setup errors with NAL/eNAL. For synthetic populations, a schematic overview of the investigations is provided in Fig. 3.

3.B.1. NAL for synthetic populations

For all synthetic populations, the investigations demonstrated that the simulations based on trendline

TABLE I. Conventional and trendline characterizations of uncorrected setup errors in the Erasmus database. See Eqs. (1)–(6) and (A1)–(A2) for definition of the parameters. M and Σ are part of both characterizations. Trendline slope parameters, M_a, Σ_a are given in mm/fraction. All other values are given in mm.

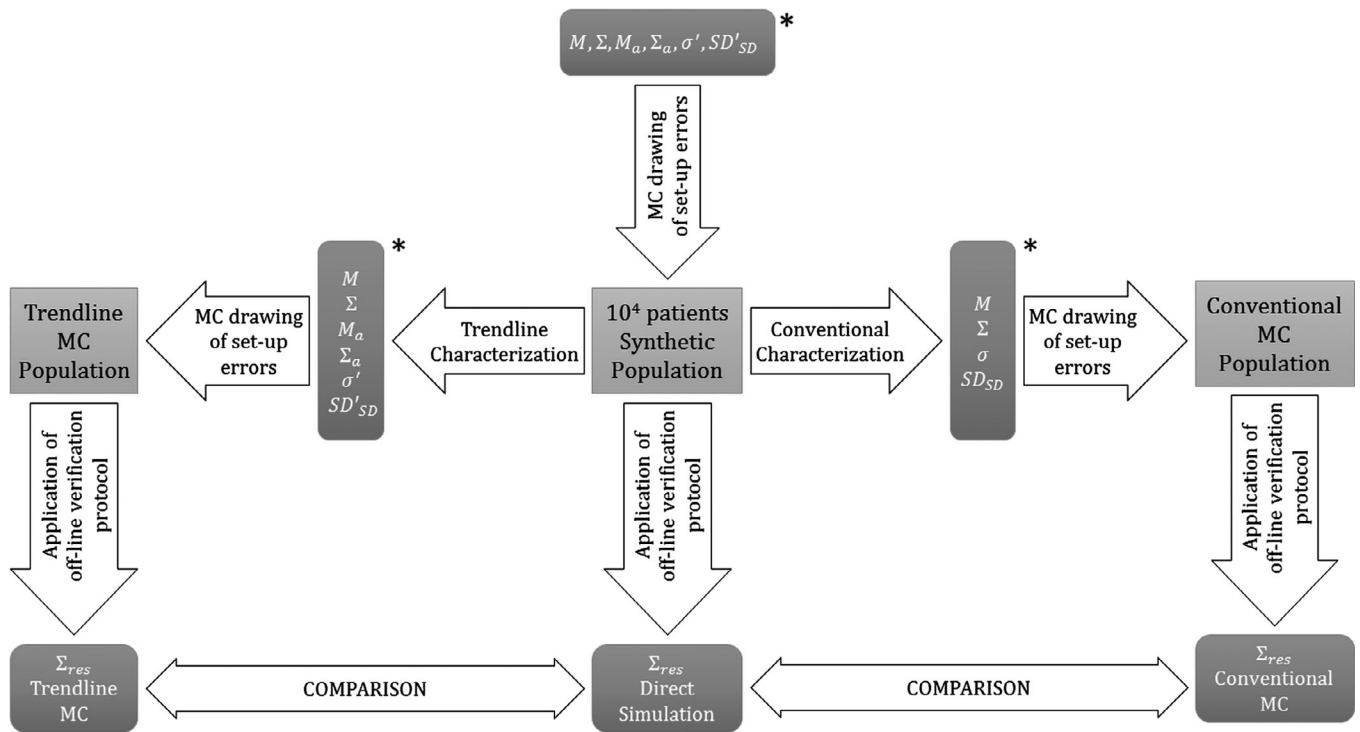|	Conventional characterization		Trendline characterization					
	σ	SD_{SD}	M	Σ	M_a	Σ_a	σ'	SD'_{SD}
Left–right	1.93	0.71	−0.32	2.50	0.002	0.046	1.86	0.69
Superior–inferior	2.64	0.68	−0.97	3.37	−0.042	0.075	2.45	0.60
Anterior–posterior	2.77	0.80	−0.54	3.47	−0.019	0.083	2.59	0.72

FIG. 3. Schematic overview of the investigations on Monte Carlo simulated residual setup errors for the no-action-level (NAL) and extended NAL off-line correction protocols for synthetic patient populations. * parameter values are different.

characterization did indeed accurately predict Σ_{res}, that is, the values were close to the ground truth values. In contrast, the use of conventional characterization did often result in significant deviations in estimated Σ_{res}. For all populations with $\Sigma_a > 0.05$, the simulation based on the conventional characterization overestimated the reductions in systematic setup errors with the NAL protocol. For $N = 3$, that is, imaging in the first three fractions, results for 12 out of 144 synthetic populations are summarized in Fig. 4. For these populations, the mean difference between the simulated Σ_{res} based on the trendline characterization and the ground truth value was -0.01 ± 0.02 mm. For simulated Σ_{res} based on conventional characterization of trendline errors the difference was -0.4 ± 0.5 mm.

3.B.2. eNAL for synthetic populations

Similar to the simulations for the NAL protocol, Σ_{res} estimated with the use of trendline characterization agreed very well with the ground truth (mean difference for $N = 3$: -0.01 mm ± 0.1 mm). Different from NAL, for eNAL simulations done with conventional characterization underestimated the positive impact of the protocol with a mean difference in Σ_{res} of 0.2 mm ± 0.1 mm. Figure 5 shows results for the 12 out of 144 synthetic populations. As explained in the M&M section, eNAL was developed to reduce residual errors in populations with time trends better than NAL. This is indeed observed when comparing the
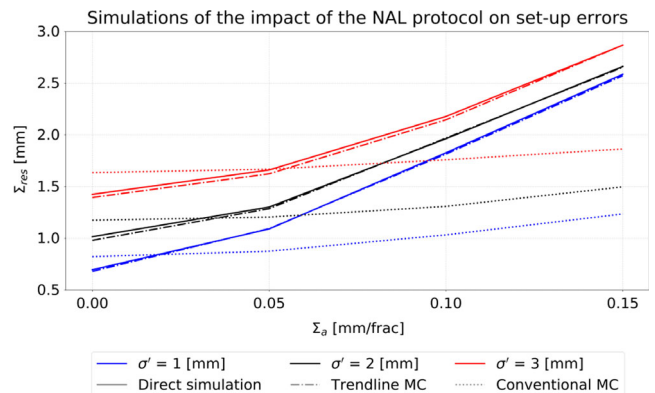

FIG. 4. Simulated residual systematic setup errors for the no-action-level protocol for 12 synthetic populations. In all simulations, imaging in only the first three fractions was assumed. For all populations: $M = 0$ mm, $\Sigma = 3$ mm and $M_a = 0$ mm/fraction. Σ_a are standard deviations describing distributions of trendline slopes. σ' are standard deviations describing distributions of random errors around trendlines. Direct simulation: no intermediate characterization used (ground truth), Conventional Monte Carlo (MC)/Trendline MC: MC simulation based on a conventional/trendline characterization of the population setup errors. [Color figure can be viewed at wileyonlinelibrary.com]

curves for Direct simulation/Trendline MC in Fig. 5. with those in Fig. 4.

3.B.3. NAL and eNAL for the Erasmus database

Also for the Erasmus database, the NAL simulations based on the trendline characterization clearly agreed best

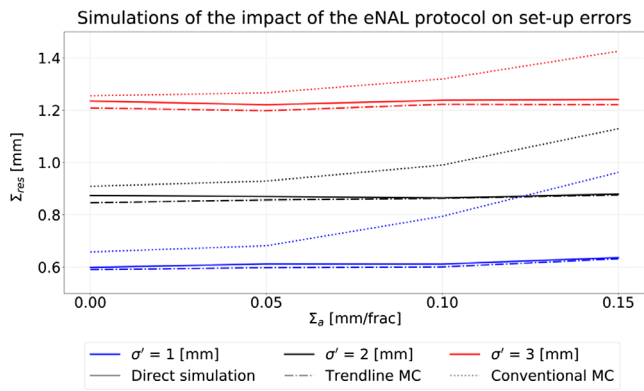## Simulations of the impact of the eNAL protocol on set-up errors

FIG. 5. Simulated residual systematic setup errors for the extended no-action-level protocol for 12 synthetic populations. In all simulations, imaging in the first three fractions was followed by image acquisition in the first fraction of each following week. For all populations: $M = 0$ mm, $\Sigma = 3$ mm and $M_a = 0$ mm/fraction. Σ_a are standard deviations describing distributions of trendline slopes. σ' are standard deviations describing distributions of random errors around trendlines. Direct simulation: no intermediate characterization used (ground truth), Conventional Monte Carlo (MC)/Trendline MC: MC simulation based on a conventional/trendline characterization of the population setup errors. [Color figure can be viewed at wileyonlinelibrary.com]

with the ground truth (Table II). However, the agreement was less good than observed for the synthetic populations (above). The eNAL simulations both for conventional and trendline characterization agreed well with the ground truth.

3.C. Required margins for trendline errors

Using the Python code provided in Data S1, margins $M^{PTV}_{trend,i}(M_i, \Sigma_i, M_{a,i}, \Sigma_{a,i}, F)$ were calculated for a wide range

of parameter values. The results are provided in Data S2 as look-up tables.

Results of fitting Eq. (7) to data with $M_i = 0$ and $M_{a,i} = 0$ are presented in Table III. Differences between $M^{PTV}_{trend,i}$ calculation with the Python code and Eq. (7) are negligible. In case of no time trends ($\Sigma_{A,i} = 0$), Eq. (7) reduces to the van Herk formula for systematic setup errors with equal α-values (compare columns 2 and 6 in Table III).

3.D. Validation of the proposed margin recipe for trendline errors

The simulations described in Section 2.G demonstrated that in the 288 synthetic populations (Section 2.C), $10.0 \pm 0.2\%$ of patients had one or more (nonrandom) trendline errors outside the margin calculated with the proposed novel recipe (y-axis Fig. 6), which is very close to the required 10.0%. In most populations, the van Herk margin for nonrandom errors $(2.5\Sigma_i + 0.7(\sigma_i - \sigma'_i)$, see Section 2.G) was clearly too small, with $58\% \pm 24\%$ of patients with one or more trendline errors outside (x-axis Fig. 6). The largest percentages with the van Herk approach were found for the largest population M_a and Σ_a values. The markers inside the depicted circle in Fig. 6 belong to the simulated isotropic synthetic populations that did not have time trends. They show that our simulations for the van Herk protocol (x-axis) are indeed correct, that is, they resulted in an expected 10% of errors outside the margin. Moreover, they demonstrate that for populations without time trends, the proposed novel recipe agrees with van Herk's recipe (compare x-axis with y-axis).

Table IV shows calculated margins for the Erasmus database and percentages of patients with trendline error(s) outside. For the novel margin recipe this was 9.6%, while the

TABLE II. Residual systematic setup errors, Σ_{res}, for the no-action-level (NAL) and extended NAL protocols applied to the Erasmus database. All values are given in mm. Direct simulation: no intermediate characterization used (ground truth), Conventional Monte Carlo (MC)/Trendline MC: MC simulation based on a conventional/trendline characterization of the population setup errors.

	NAL			eNAL		
	Direct simulation	Trendline MC	Conventional MC	Direct simulation	Trendline MC	Conventional MC
Left–right	1.4	1.2	1.1	0.8	1.1	1.0
Superior–inferior	2.0	1.7	1.3	0.8	1.1	1.0
Anterior–posterior	2.1	1.9	1.4	0.8	1.1	1.0

TABLE III. α, δ and γ: fit parameters for Eq. (7) as a function of the required percentage of patients (percentile) inside $M^{PTV}_{trend,i}$. $\Delta Margin$: for 441 combinations of $\Sigma_i \in [0, 5]$ mm and $\Sigma_{A,i} \in [0, 5]$ mm mean differences with ranges in $M^{PTV}_{trend,i}$, calculated with exact simulation with Python code and with Eq. (7). For comparison, the last column contains α-values given by van Herk et al.[4]

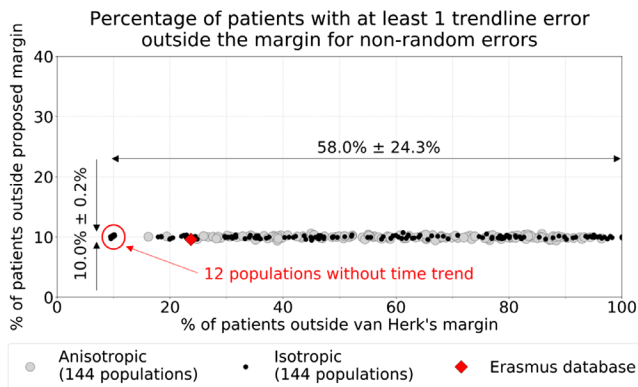| Percentile	α	δ	γ	$\Delta Margin$ (mm)	α by van Herk et al.
80	2.15	−0.94	0.86	2.65×10^{-5}; [−0.005, 0.004]	2.16
85	2.31	−1.08	0.67	4.33×10^{-5}; [−0.006, 0.005]	2.31
90	2.50	−1.27	0.52	8.19×10^{-5}; [−0.008, 0.007]	2.50
95	2.79	−1.54	0.38	1.11×10^{-4}; [−0.015, 0.012]	2.79
99	3.37	−2.09	0.26	1.75×10^{-4}; [−0.016, 0.019]	3.36

FIG. 6. For the 288 synthetic patient populations and the Erasmus database, percentages of patients with trendline error(s) outside the three-dimensional clinical target volume (CTV)–planning target volume (PTV) margin for non-random errors. X- axis: percentages for margins according to van Herk et al., Y-axis: percentages according to the proposed novel recipe. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE IV. Calculated margins and percentages of patients with one or more trendline errors outside, in case no setup protocols applied. To calculate total margins, $0.7\sigma'$ is added to the margin for nonrandom errors, see Section 2.G.

	Margin — van Herk recipe nonrandom/total	Margin — proposed recipe nonrandom/total
Left–right	6.3/7.6 mm	7.5/8.8 mm
Superior–inferior	8.6/10.3 mm	10.8/12.5 mm
Anterior–posterior	8.8/10.6 mm	11.0/12.8 mm
% patients outside 3D margin	23.7%	9.6%

FIG. 7. Population $\sigma(f)$ of setup errors for patients in the Erasmus database. For each patient, the standard deviation for each fraction, f, was established from the measured setup errors in the fractions f-2, f-1, and f. For each fraction, f, these patient-specific standard deviations, $SD_p(f)$, were then combined into a $\sigma(f)$ [see Eq. (A1)] as presented along the y-axis. [Color figure can be viewed at wileyonlinelibrary.com]

van Herk recipe resulted in 23.7%. Margins calculated with the van Herk recipe were up to 2.2 mm too small to guarantee that not more than 10% of patients would have trendline error(s) outside.

4. DISCUSSION

For the much applied NAL protocol for setup corrections it was demonstrated that MC simulations based on a conventional characterization overestimated the reduction in systematic setup errors. This is attributed to the fact that in a conventional characterization there is no explicit modelling of the time trends. These results demonstrate that the use of a conventional characterization for simulation of setup protocols in a patient population with trends may be potentially dangerous as it may point at required margins that are smaller than needed. As expected in populations with time trends, the eNAL protocol could better reduce setup errors than NAL. However, also for eNAL, predicted residual errors were inaccurate when simulations were based on conventional characterization.

Based on the proposed trendline characterization of setup errors, a novel CTV-PTV recipe for nonrandom errors was developed, ensuring that deterministic underdosing of tumor edges in multiple consecutive fractions as a result of trend
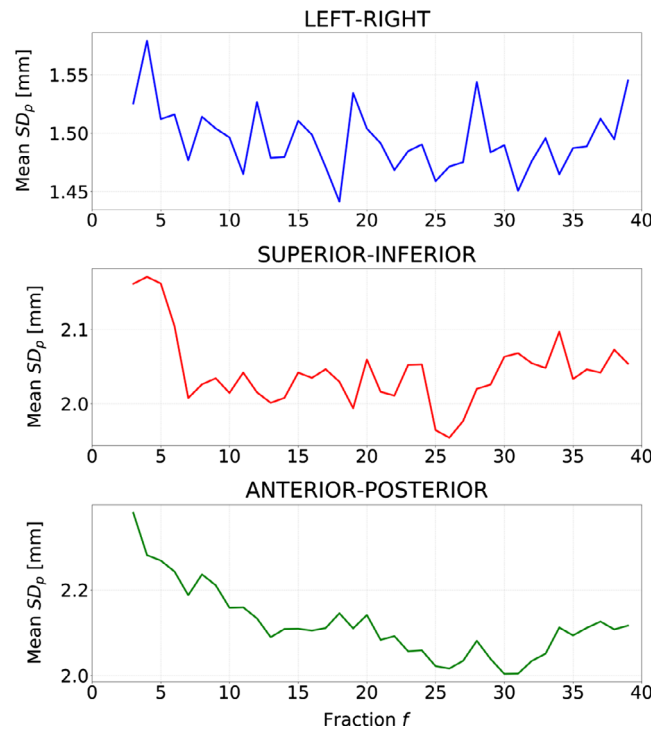
motion was avoided. Different from the approach proposed by van Herk et al.,[4] the nonrandom errors were not only characterized by mean setup errors in fractionated treatments but also by the slopes of fitted trendlines. Similar to the van Herk approach, the margin for nonrandom errors was defined by prescribing that only 10% of patients could have a non-random error outside the calculated margin. In the absence of time trends in the population ($M_a = 0$ and $\Sigma_a = 0$) the two margins are equal. The proposed recipe describes a numerical procedure for obtaining the margins. We derived an analytical margin formula [Eq. (7)] in case of zero mean population slopes and zero mean translational setup errors. We provided a Python code (Data S1) as well as look-up tables (Data S2) for if this is not the case.

For the proposed margin recipe we have adopted the generally applied approach of separating the total margin in two components, one for non-random errors (described by trendlines) and the other for random errors, that is, $M^{PTV} = M_{trend}^{PTV} + M_{rand}^{PTV}$, where M_{rand}^{PTV} is used to cope with the blurring of planned dose distributions due to random setup errors. In this paper we have used the well-known expression, $M_{rand}^{PTV} = 0.7\sigma'$.[4] On the other hand, we are aware of other recipes for calculating M_{rand}^{PTV}, which may be more appropriate, e.g. for lung tumors or SBRT[19–22]. If considered appropriate, the applied $M_{rand}^{PTV} = 0.7\sigma'$ can be substituted by any other recipe. It will not impact the proposed mechanism for calculation of M_{trend}^{PTV}.

To the best of our knowledge there is no radiobiological or clinical literature that confirms or denies a need for avoiding systematic underdosing of tumor edges in multiple consecutive fractions that can result from trend motion. Therefore, there is possibly no need, or not always a need, to (fully) apply the enlarged margins following from the proposed CTV-PTV margin recipe. On the other hand, as the recipe uses distributions of combined errors, MD_p (Section 2.F), resulting from systematic errors (m_p) and trends (a_p) (so no separate distributions of errors resulting from m_p and from a_p, i.e., no addition of margins), the margin increase compared to the van Herk approach may in practice be limited (see, e.g., Table IV; 7.6 mm compared to 8.8 mm). Nevertheless, as margin increases can result in increased OAR doses, there can be arguments to choose for the original van Herk margin recipe. Alternatively, margins could be calculated with the novel recipe while allowing some underdose in specific areas of the extended PTV that would otherwise result in unacceptable enhancement of OAR doses. In any way, systematic radiobiological studies are warranted to guide future clinical decisions regarding the margin in case of time trends. As described elsewhere in this section, also if enlarged margins would not be needed, there are still other arguments to apply trendline characterizations instead of the conventional fractionation in case time trends occur in the patient population.

Both in conventional and in trendline characterizations of setup errors there is a decoupling of random and nonrandom errors. This decoupling may result in erroneous conclusions from MC simulations. We have proposed correction schemes to avoid these issues (Section 2.D).

As explained in Section 2.D, even in populations that would not have systematic setup errors or time trends in case of an infinite number of treatment fractions, such errors and time trends are to be expected if the same patients would be treated with a (more realistic) finite number of fractions. A test has been proposed to find out whether observed time trends in a population are related to the finite number of fractions or whether there are other (physiologic) causes for observed trends (Section 2.I).

For the investigated synthetic patient populations, MC simulations using trendline characterization resulted in highly accurate residual setup errors after NAL corrections. For $N = 3$ the difference in Σ_{res} with the ground truth was -0.01 ± 0.02 mm. For the Erasmus database, ground truth Σ_{res} values were up to 0.3 mm larger than those obtained with MC simulations based on trendline characterization, depending on direction (see Table II). Part of the explanation for enhanced ground truth residual errors may be found in Fig. 7, showing $\sigma(f)$ values calculated with Eq. (A1) for patients' standard deviations in setup errors calculated over three fractions ($SD_p(f)$). It shows for the superior–inferior and anterior–posterior directions enhanced standard deviations at the start of treatment. In the ground truth simulations for NAL this is expected to result in increased Σ_{res} as the mean setup errors determined in the first three fractions, used for set-up corrections in the following fractions, may deviate more from the true mean set-up errors due to the larger random errors at the start of treatment. To investigate this further, we also performed direct simulations with set-up errors that were inverted in time, that is, fraction $F \rightarrow$ fraction 1, fraction $F - 1 \rightarrow$ fraction 2, etc. For the inverted fraction order, the achieved Σ_{res} indeed reduced: from 1.4 to 1.2 mm, from 2.0 to 1.6 mm, and from 2.1 to 1.8 mm for left–right, superior–inferior, and anterior–posterior directions, respectively. The new Σ_{res} do better agree with the MC predictions for trendline characterization, that is, 1.2, 1.7, and 1.9 mm, respectively (see Table II). The observed enhanced variation in setup errors at the start of treatment could be related to patients being more nervous at the start, rather than at later moments in the fractionated treatment.

Currently, many patients are treated with daily online setup corrections.[23,24] Obviously, this approach can reduce the occurrence of time trends. However, in case of differential motion between various targets this may not be the case. For example, time trends have been observed for the lumpectomy cavity in breast cancer patients[12] for primary larynx tumors[13] and for lung.[14] However, for these patient groups, the surrounding nodal targets do not move with the primary target. Therefore, correction of time trends in these targets based on daily setup corrections can induce effective time trends in the setup of the surrounding nodes. The proposed margin recipe can then be used for margin definition for the nodes. In that case, setup errors of nodes have to be described relative to the tumor center of mass. Such a procedure has to be further evaluated and verified prior to clinical implementation.

Probabilistic/robust planning does not need margins. However, it is based on distributions of geometric uncertainties in patient populations.[25–27] To the best of our knowledge, explicit inclusion of time trends in probabilistic planning has not yet been investigated. It would add an extra complexity to the plan generation. It would also need rethinking the probability requirements for CTV coverage in case the probability of deterministic underdosing parts of the CTV in multiple consecutive fractions needs to minimized. On the other hand, we hypothesize that using trendline characterizations of setup errors in patient populations that have time trends may improve the robustness of the generated plans. Further research is needed to investigate the full consequences of explicit inclusion of time trends in probabilistic planning.

5. CONCLUSIONS

The conventional characterization of tumor setup errors in patient populations, using only distributions of random and systematic errors, has limitations for populations with interfraction time trends. We have investigated the use of the trendline characterization that explicitly models these time trends. Trendline characterization resulted in more accurate simulated residual setup errors after the application of offline setup correction protocols, avoiding the application of erroneous margins. The proposed novel

CTV-PTV margin recipe for nonrandom errors can be used to avoid or reduce underdosing of tumor edges in multiple consecutive fractions caused by trend motions. In the limit of the absence of time trends in a population, the margin recipe reduces to the well-known van Herk recipe for systematic errors. In case of image guided therapy for patients with multiple targets with differential motion for daily correction of trend motion of one of the targets, the proposed recipe may be used to calculate margins for the other targets. In probabilistic planning, use of trendline characterization may result in more robust treatment plans. Population analyses of interfraction setup errors need to include the potential occurrence of time trends.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

Erasmus MC Cancer Institute has research collaborations with Elekta AB (Stockholm, Sweden) and Accuracy Inc (Sunnyvale, USA). This work was not part of these collaborations.

APPENDIX A

CONVENTIONAL CHARACTERIZATION OF SET-UP ERRORS IN A PATIENT POPULATION

Set-up errors occurring in a fractionated treatment of a patient can be described by the mean (systematic) error during the treatment, and day-to-day variations around the mean (random errors). Observed systematic and random errors are used to derive population parameters that characterize the error distributions along the principal directions i.e. anterior-posterior, superior-inferior, left-right. The following population parameters are established for each direction:

- the overall mean set-up error M in the population – see Eq. 1
- the standard deviation Σ describing the distribution of systematic (i.e. mean) set-up errors in the population – see Eq. 2
- the population standard deviation describing the random errors:

$$\sigma = \sqrt{\sum_p \frac{SD_p'^2}{N}} \tag{A1}$$

with SD_p the standard deviation of the random set-up errors observed for patient p

- the standard deviation describing the variation of SD_p in the population[15]:

$$SD_{SD} = \sqrt{\sum_p \frac{\left(\overline{SD_p} - SD_p\right)^2}{N-1}} \tag{A2}$$

In most publications SD_{SD} is not reported. This is related to the fact that in two well-known CTV-PTV margin recipes by van Herk et al.[4] ($M^{PTV} = 2.5\Sigma + 0.7\sigma$) and by Stroom[5] ($M^{PTV} = 2.0\Sigma + 0.7\sigma$) random errors are described only by σ as defined in Eq. A1.

APPENDIX B

CORRECTIONS DONE FOR MC SIMULATIONS BASED ON CONVENTIONAL CHARACTERIZATION

Corrected random set-up errors in the F fractions of the fractionated treatment of a patient p were established in a 3-step process:

i. randomly select the standard deviation of the patient's random set-up errors, SD_p, from the distribution $G(\sigma, SD_{SD})$;
ii. randomly select F random setup errors from the distribution $G(0, SD_p)$ and determine the mean of these errors, m_p^r, where r stands for random;
iii. subtract m_p^r from the random errors established in ii.

CORRECTIONS DONE FOR MC SIMULATIONS BASED ON TRENDLINE CHARACTERIZATION

Analogous to the conventional characterization, corrections were made to the MC generated random errors. To this purpose, for each patient, F random errors were first randomly selected from the distribution $G\left(0, SD_p'\right)$, with SD_p' drawn from $G(\sigma', SD_{SD}')$. Then, an auxiliary trendline was fitted through these errors. This line was then used to correct the initially selected random errors such that the overall mean and the time trend slope of the corrected random errors both became zero. In this way, a random selection of a finite number ($F = 39$) of random errors would not result in an undesired effective change in the patient's trendline drawn from $G(M, \Sigma)$ and $G(M_a, \Sigma_a)$.

APPENDIX C

DESCRIPTION OF THE NO ACTION LEVEL (NAL) AND EXTENDED NAL PROTOCOLS

The much applied NAL protocol[2] is based on off-line establishment of set-up corrections to be applied in future fractions. During the first N fractions (usually $N = 3$) of the patient's treatment, images are taken and no corrections are

applied. After fraction N, the mean set-up error during the first N fractions is calculated, which is considered as an estimate of the patient's mean (= systematic) set-up error in case no corrections would be applied. From fraction $N + 1$ till the last fraction, the patient is first set up on the original tattoos and then shifted by –(mean set-up error in first N fractions) using couch translations. The dose is then delivered without further imaging. Aim of this protocol is to reduce systematic set-up errors while random errors are not affected.

In 2007 an extended version of the NAL protocol (eNAL[3]) was proposed to also deal with set-up changes that can occur during a fractionated treatment, such as time trends. In the first week of treatment, eNAL is equal to NAL. In each subsequent week, the set-up correction vector is updated using set-up errors measured in its first fraction.

[a)]Author to whom correspondence should be addressed. Electronic mail: m.gizynska@erasmusmc.nl.

REFERENCES

1. Hurkmans CW, Remeijer P, Lebesque JV, Mijnheer BJ. Set-up verification using portal imaging; review of current clinical practice. *Radiother Oncol*. 2001;58:105–120.
2. de Boer HCJ, Heijmen BJM. A protocol for the reduction of systematic patient set-up errors with minimal portal imaging workload. *Int J Radiot Oncol Biol Phys*. 2001;50:1350–1365.
3. de Boer HCJ, Heijmen BJM. eNAL: an extension of the NAL setup correction protocol for effective use of weekly follow-up measurements. *Int J Radiot Oncol Biol Phys*. 2007;67:1586–1595.
4. van Herk M, Remeijer P, Rasch C, Lebesque JV. The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy. *Int J Radiot Oncol Biol Phys*. 2000;47:1121–1135.
5. Stroom JC, de Boer HCJ, Huizenga H, Visser AG. Inclusion of geometrical uncertainties in radiotherapy treatment planning by means of coverage probability. *Int J Radiot Oncol Biol Phys*. 1999;43:905–919.
6. Bijhold J, Lebesque JV, Hart AAM, Vijlbrief RE. Maximizing setup accuracy using portal images as applied to a conformal boost technique for prostatic cancer. *Radiother Oncol*. 1992;24:261–271.
7. El-Gayed AAH, Bel A, Vijlbrief R, Bartelink H, Lebesque JV. Time trends of patient setup deviations during pelvic irradiation using electronic portal imaging. *Radiother Oncol*. 1993;26:162–171.
8. Hanley J, Lumley MA, Mageras GS, et al. Measurement of patient positioning errors in three-dimensional conformal radiotherapy of the prostate. *Int J Radiot Oncol Biol Phys*. 1997;37:435–444.
9. Stroom JC, Koper PCM, Korevaar GA, et al. Internal organ motion in prostate patients treated in prone and supine treatment position. *Radiother Oncol*. 1999;51:237–248.
10. van der Heide UA, Kotte ANTJ, Dehnad H, Hofman P, Lagenijk JJW, van Vulpen M. Analysis of fiducial marker-based position verification in the external beam radiotherapy of patients with prostate cancer. *Radiother Oncol*. 2007;82:38–45.
11. Namysl-Kaletka A, Wydmanski J, Tukiendorf A, et al. Influence of interfraction motion on margins for radiotherapy of gastric cancer. *Br J Radiol*. 2015;88:20140610.
12. Penninkhof J, Quint S, Baaijens M, Heijmen B, Dirkx M. Practical use of the extended no action level (eNAL) correction protocol for breast cancer patients with implanted surgical clips. *Int J Radiot Oncol Biol Phys*. 2012;82:1031-1-37.
13. Gangsaas A, Astreinidou E, Yint S, Levendag PC, Heijmen B. Cone-Beam computed tomography – guided positioning of laryngeal cancer patients with large interfraction time trends in setup and nonrigid anatomy variations. *Int J Radiot Oncol Biol Phys*. 2013;87:401–406.
14. Weiss E, Robertson SP, Mukhopadhyay N, Hugo GD. Tumor, lymph node, and lymph node-to-tumor displacements over a radiotherapy series: analysis of interfraction and intrafractions variations using active breathing control (ABC) in lung cancer. *Int J Radiot Oncol Biol Phys*. 2011;82:e639–e645.
15. van Herk M, Witte M, Remeijer P.Performance of patient specific margins derived using a Bayesian statistical method. WC 2009, IFMBE Proceedings 2009;25:769–771.
16. Wu J, Haycocks T, Alasti H, et al. Positioning errors and prostate motion during conformal prostate radiotherapy using on-line isocentre set-up verification and implanted prostate markers. *Radiother Oncol*. 2001;61:127–133.
17. McNair HA, Hansen VN, Parker CC, et al. A comparison of the use of bony anatomy and internal markers for offline verification and an evaluation of the potential benefit of online and offline verification protocols for prostate radiotherapy. *Int J Radiat Oncol Biol Phys*. 2008;71:41–50.
18. van Haaren PMA, Bel A, Hofman P, van Vulpen M, Kotte ANTJ, van der Heide UA. Influence of daily set-up measurements and corrections on the estimated delivered dose during IMRT treatment of prostate cancer patients. *Radiother Oncol*. 2009;90:291–298.
19. Sonke JJ, Rossi M, Wolthaus J, van Herk M, Damen E, Belderbos J. Frameless stereotactic body radiotherapy for lung cancer using four-dimensional cone beam CT guidance. *Int J Radiat Oncol Biol Phys*. 2009;74:567–574.
20. van Herk M, Witte M, van der Geer J, Schneider C, Lebesque JV. Biologic and physical fractionation effects of random geometric errors. *Int J Radiat Oncol Biol Phys*. 2003;57:1460–1471.
21. Hoogeman MS, Nuyttens JJ, Levendag PC, Heijmen BJM. Time dependence of intrafraction patient motion assessed by repeat stereoscopic imaging. *Int J Radiat Oncol Biol Phys*. 2008;70:609–618.
22. Lujan AE, Larsen EW, Balter JM, Ten Haken RK. A method for incorporating organ motion due to breathing into 3D dose calculations. *Med Phys*. 1999;26:715–720.
23. Herman MG, Abrams RA, Mayer RR. Clinical use of on-line portal imaging for daily patient treatment verification. *Int J Radiot Oncol Biol Phys*. 1994;28:1017–1023.
24. van Herk M. Different styles of image-guided radiotherapy. *Semin Radiat Oncol*. 2007;17:258–267.
25. Moore JA, Gordon JJ, Anscher MS, Siebers JV. Comparisons of treatment optimization directly incorporating random patient set-up uncertainty with a margin-based approach. *Med Phys*. 2009;36:3880–3890.
26. Moore JA, Gordon JJ, Anscher MS, Siebers JV. Comparisons of treatment optimization directly incorporating systematic patient set-up uncertainty with a margin-based approach. *Med Phys*. 2012;39:1120–1111.
27. Chan TCY, Tsitsiklis JN, Bortfeld T. Optimal margin and edge-enhanced intensity maps in the presence of motion and uncertainty. *Phys Med Biol*. 2010;55:515–533.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Data S1. Python script for margin calculation.
Data S2. Tables containing margin sizes for different parameters.