

TECHNICAL ADVANCE

Open Access

# Comparison of measurement methods with a mixed effects procedure accounting for replicated evaluations (COM<sub>3</sub>PARE): method comparison algorithm implementation for head and neck IGRT positional verification

Anuradha Roy<sup>1†</sup>, Clifton D. Fuller<sup>2\*†</sup>, David I. Rosenthal<sup>2</sup> and Charles R. Thomas Jr.<sup>3</sup>

## Abstract

**Purpose:** Comparison of imaging measurement devices in the absence of a gold-standard comparator remains a vexing problem; especially in scenarios where multiple, non-paired, replicated measurements occur, as in image-guided radiotherapy (IGRT). As the number of commercially available IGRT presents a challenge to determine whether different IGRT methods may be used interchangeably, an unmet need conceptually parsimonious and statistically robust method to evaluate the agreement between two methods with replicated observations. Consequently, we sought to determine, using an previously reported head and neck positional verification dataset, the feasibility and utility of a Comparison of Measurement Methods with the Mixed Effects Procedure Accounting for Replicated Evaluations (COM<sub>3</sub>PARE), a unified conceptual schema and analytic algorithm based upon Roy's linear mixed effects (LME) model with Kronecker product covariance structure in a doubly multivariate set-up, for IGRT method comparison.

**Methods:** An anonymized dataset consisting of 100 paired coordinate (X) measurements from a sequential series of head and neck cancer patients imaged near-simultaneously with cone beam CT (CBCT) and kilovoltage X-ray (KVX) imaging was used for model implementation. Software-suggested CBCT and KVX shifts for the lateral (X), vertical (Y) and longitudinal (Z) dimensions were evaluated for bias, inter-method (between-subject variation), intra-method (within-subject variation), and overall agreement using with a script implementing COM<sub>3</sub>PARE with the MIXED procedure of the statistical software package SAS (SAS Institute, Cary, NC, USA).

**Results:** COM<sub>3</sub>PARE showed statistically significant bias agreement and difference in inter-method between CBCT and KVX was observed in the Z-axis (both  $p$ -value < 0.01). Intra-method and overall agreement differences were noted as statistically significant for both the X- and Z-axes (all  $p$ -value < 0.01). Using pre-specified criteria, based on intra-method agreement, CBCT was deemed preferable for X-axis positional verification, with KVX preferred for superoinferior alignment.

**Conclusions:** The COM<sub>3</sub>PARE methodology was validated as feasible and useful in this pilot head and neck cancer positional verification dataset. COM<sub>3</sub>PARE represents a flexible and robust standardized analytic methodology for IGRT comparison. The implemented SAS script is included to encourage other groups to implement COM<sub>3</sub>PARE in other anatomic sites or IGRT platforms.

\*Correspondence: [cdfuller@mdanderson.org](mailto:cdfuller@mdanderson.org)

†Equal contributors

<sup>2</sup>Department of Radiation Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA

Full list of author information is available at the end of the article

## Background

Method comparison is a frequent problem encountered whenever different measurement devices/techniques are implemented in the absence of a gold standard [1–7]. Method comparison in radiological science is often a vexing issue [8–15], and is especially notable when competing imaging methodologies are used without establishment of the technical superiority in terms of accuracy of one platform. In a specific example, the explosion in applications of image-guided radiation therapy (IGRT), which necessitates repeated and exceedingly accurate spatial localization in order to carefully deliver conformal radiation dose, places a premium on both reproducibility and accuracy [14, 16–25]. Furthermore, the large number of divergent manufacturer-supported mechanisms for achieving image-guided target localization/positional verification (e.g., 2D- and 3D-ultrasound [26–34], 2-D radiography [35–39], megavoltage and kilovoltage 3-D tomography [40–44]) have arisen in the absence of a gold-standard, and thus have been the impetus for a large number of inter-modality comparative studies, which themselves often utilize a wide array of statistical methods to report between method measurement differences [22, 24, 25, 27, 33, 34, 44–47]. In an effort to more formally assess both inter- and intra-method bias, as well as to streamline comparatively time- and effort-intensive graphical and statistical analysis inherent in many method comparison statistical techniques, we sought to devise an algorithm to explore agreement between two methods of image-guided radiotherapy, using a novel linear mixed effects (LME) model with Kronecker product covariance structure in a doubly multivariate approach [48]. This integrated approach has great potential utility, formally evaluating inter-method bias, inter-subject variability and the intra-subject variability (i.e., agreement between the repeatability coefficients) of the two imaging methods/devices. Testing of all three aspects is crucial, as inter-subject variability is of import when estimating the difference between the two methods giving different measurements on the same subject, while intra-subject variability affords calculation of the random error among the replications taken by the same method on the same subject [49]. We use a doubly multivariate set-up (i.e., measurement data for each subject is considered at two levels, incorporating both the number of measurement methods and number of replicated measurements). This specific LME-based technique, which we shall refer to as COM<sub>3</sub>PARE (Comparison of Measurement Methods with Mixed Effects Procedure Accounting for Replicated Evaluations) is robust with regard to number of replicates, and is easily performed using SAS software (*vide infra*). LME models have improved fidelity in scenarios wherein observations are not fully independent and can more correctly model correlated errors, compared to

general linear models (GLM), which includes typical statistical analyses (e.g., t-test, ANOVA, linear regression). LME includes multiple random effect components, compared to single element evaluation in most GLM models, affording improved analysis of continuous variables where random effects, multiple hierarchical data levels, and replicated measurements are concerned. The number of replicated measurements on each patient or subject may not be equal, and also the number of replications of the two methods on the same subject may not be equal. The specific aims for this study included:

- First known application of LME-based COM<sub>3</sub>PARE hypothesis testing protocol for method comparison using imaging data.
- Demonstration of feasibility and utility of COM<sub>3</sub>PARE using an established head and neck positional verification dataset, previously presented with standard method comparison approaches.

## Methods

### Datasets

A previously presented dataset consisting of a series of 100 paired measures using two distinct positional verification techniques in a series of 28 sequential head and neck squamous cell carcinoma patients was utilized. As this manuscript is designed to specify a novel statistical methodology, interested readers are referred to the previous manuscript [50], wherein imaging parameters have been previously detailed. Briefly, CBCT and stereoscopic kV X-ray were acquired near-simultaneously at approximately biweekly intervals throughout a patient's course of treatment (dependent upon the scheduling exigencies in the department) for a series of patients with head and neck cancers. CBCT/kV X-ray analyses were performed using the attached on-board imager (Varian Medical Systems, Palo Alto, CA). Positional verification was performed with manufacturer-supplied software (Varian OBI 1.3/Varian Vision, Varian Medical Systems, Palo Alto, CA) for 3D-3D (CBCT-simulation CT) and 2D-2D (kV X-ray-DRR) automated matching using the aforementioned software. Recorded shifts represent the coregistration/allineation software derived values without physician/observer modification. For each paired-method positional acquisition, the origin was defined as the point in space identified by the initial isocenter position using immobilization-mask-based markers. Utilizing a three-dimensional Cartesian coordinate system, this spatial location was designated as a 'zero point' with X/Y/Z coordinates of (0, 0, 0). Software-derived shifts for each system were recorded in centimeters, specified as X-, Y- or Z-axis. Software-derived shifts were characterized as X- (lateral or left/right), Y- (vertical or anteroposterior) and Z- (longitudinal or superoinferior) axes, respectively, for both kV X-ray and CBCT IGRT techniques. For the purposes of clarity, we proposed the

following three conditions be met to verify whether two methods for measuring a variable (in this specific case, IGRT-suggested spatial shifts) can be considered interchangeable:

1. No significant bias (i.e., no difference between the means of the two methods under a pre-specified threshold nor a statistically significant difference between said means).
2. No statistically significant difference in the inter-subject (between-subject) variability of the two methods.
3. No statistically significant difference in the intra-subject (within-subject) variability (i.e., repeatability) of the two methods.

For this study, we pre-specified a bias threshold of an absolute value of  $< 0.1$  cm, with a statistically significant difference designated by  $\alpha < 0.05$ . To assess the aforementioned criteria, we implemented the LME methodology proposed by Roy<sup>48</sup>, referred to as COM<sub>3</sub>PARE (see Appendix A).

### Statistical analysis with COM<sub>3</sub>PARE

As mentioned in the introduction the number of replicated measurements on each patient or subject may not be equal, and also the number of replications of the two methods on the same subject may not be equal. Let  $p_i^{KVX}$  and  $p_i^{CBCT}$  be the number of replications on subject  $i$  by the established method (KVX), and a new method (CBCT) respectively. Let  $p_i = \max(p_i^{KVX}, p_i^{CBCT})$ , and  $n_i = 2p_i$ . Therefore, the number of observations on the  $i$ th subject is  $n_i$ , under the assumption that the  $i$ th subject has  $|p_i^{KVX} - p_i^{CBCT}|$  missing values.

Let  $y_{it}^{KVX}$  and  $y_{it}^{CBCT}$  be the responses by the established method and a new method of the  $i$ th subject at the  $t$ th replicate,  $i = 1, 2, \dots, N$ ,  $t = 1, 2, \dots, p_i$ . Let  $y_{it} = (y_{it}^{KVX}, y_{it}^{CBCT})'$  be the  $2 \times 1$  vector of measurements corresponding to the  $i$ th subject at the  $t$ th replicate. Let  $y_i = (y'_{i1}, y'_{i2}, \dots, y'_{ip_i})'$  be the  $(n_i \times 1)$ -dimensional random vector corresponding to the  $i$ th subject. That is, the vector  $y_i$  is obtained by stacking the responses of the KVX method, and the CBCT method at the first replication, then stacking the responses of the KVX method and the CBCT method at the second replication and so on. We write all responses ( $y_i$ ) of the  $i$ th subject in a matrix equation as

$$y_i = X_i\beta + Z_i b_i + \epsilon_i,$$

with  $b_i \sim N_m(\mathbf{0}, D)$ ,

and  $\epsilon_i \sim N_{n_i}(\mathbf{0}, R_i)$ ,

where  $b_1, b_2, \dots, b_N, \epsilon_1, \epsilon_2, \dots, \epsilon_N$  are independent, and  $y_1, y_2, \dots, y_N$  are also all independent. LME model

allows for the explicit analysis of between-subject ( $D$ ) and within-subject ( $R_i$ ) sources of variation of the two methods. We define the two methods by a vector variable Mvar; Mvar=1 for the KVX method and Mvar=2 for the CBCT method. We choose the intercept and the vector variable Mvar as fixed effects, thus the design matrix  $X_i$  has three columns, and consequently  $\beta = (\beta_0, \beta_1, \beta_2)'$  is a 3-dimensional vector containing the fixed effects. We also choose the vector variable Mvar as random effects, i.e., Mvar is random across individual subjects; thus the design matrix  $Z_i$  has two columns. Therefore,  $b_i = (b_{1i}, b_{2i})'$  is a 2-dimensional vector containing the random effects.

The solution for  $\beta$  gives the means of the two methods  $\mu_{KVX}$  and  $\mu_{CBCT}$ . The between-subject variance-covariance matrix  $D$  of the KVX method and the CBCT method is a general  $(2 \times 2)$ -dimensional matrix, and  $R_i$  is a  $(n_i \times n_i)$ -dimensional covariance matrix which depends on  $i$  only through its dimension  $n_i$ . The marginal density function of  $y_i \sim N_{n_i}(X_i\beta, Z_i D Z_i' + R_i)$ . Suppose the matrix  $\Sigma$  represents the within-subject variance-covariance matrix of the KVX method and the CBCT at any replicate; also, suppose  $V$  represents the  $p \times p$ -dimensional correlation matrix of the replicated measurements on a given method, where  $p = \max_i(p_i)$ . It is assumed that the  $2 \times 2$  within-subject variance-covariance matrix  $\Sigma$  is same for all replications, and the correlation matrix  $V$  is assumed to be the same for both the methods. We assume  $R_i = \dim(V \otimes \Sigma)$ , where  $V$  and  $\Sigma$  respectively are positive definite matrices as described above, and  $\otimes$  represents the Kronecker product structure. The notation  $\dim(V \otimes \Sigma)$ , represents a  $(n_i \times n_i)$ -dimensional sub- $n_i$  matrix obtained from the  $(2p \times 2p)$ -dimensional matrix  $(V \otimes \Sigma)$ , by appropriately keeping the columns and rows corresponding to the  $n_i$ -dimensional response vector  $y_i$ . Since the equicorrelated or compound symmetry (CS) structure assumes equal correlation among all replicated measurements, we assume that the correlation matrix  $V$  of the replicated measurements has equicorrelated correlation structure. For the above design matrix  $Z_i$  and between-subject  $D$  and within-subject  $R_i$  sources of variation, the observed  $(n_i \times n_i)$ -dimensional overall variance-covariance matrix  $\Omega_i$  for the  $i$ th individual is given by

$$\begin{aligned} \text{Cov}(y_i) &= \Omega_i = Z_i D Z_i' + R_i, \\ &= Z_i D Z_i' + \dim(V \otimes \Sigma)_{n_i}. \end{aligned}$$

Thus, the covariance matrix has the same structure for each subject, except that of the dimension. The  $2 \times 2$  block diagonals Block  $\Omega_i$  in the overall variance-covariance matrix  $\Omega_i$  represent the overall variance-covariance matrix between the two methods. Similarly, the  $2 \times 2$  block

diagonals in the overall correlation matrix  $\Omega_i$ . Correlation represent the overall correlation matrix between the two methods. Thus, the off-diagonal element in this  $2 \times 2$  overall correlation matrix gives the overall correlation between the two methods. It can be easily seen that the overall variability is the sum of between-subject variability and within-subject variability (see Roy<sup>48</sup> for detail). Thus, we see that if there is a disagreement in overall variabilities, then it may be due to the disagreement in either between-subject variabilities or within-subject variabilities or both.

### MIXED procedure of SAS

We use MIXED procedure (*PROC MIXED*) of SAS to get the maximum likelihood estimates (MLEs) of  $\beta$ ,  $D$ ,  $R_i$  and  $\Omega_i$ . *METHOD=ML* specifies MIXED procedure to calculate the maximum likelihood estimates of the parameters. The *COVTEST* option requests hypothesis tests for the random effects. *CLASS* statement specifies the categorical variables. *DDFM=KR* specifies the Kenward-Roger<sup>51</sup> correction for computing the denominator degrees of freedom for the fixed effects. Kenward-Roger correction is suggested whenever one has replicated or repeated measures data; also for missing data. The *SOLUTION (S)* option in the *MODEL* statement provides the estimate of the difference between the two mean readings (bias) of the two methods. *RANDOM* and *REPEATED* statements specify the structure of the covariance matrices  $D$  and  $R_i$ . See the sample program in Appendix A that demonstrates the use of *RANDOM* and *REPEATED* statements. *PROC MIXED* calculates the  $(n_i \times n_i)$ -dimensional submatrix  $R_i$  of the  $i$ th subject from the  $(2p \times 2p)$ -dimensional matrix  $(V \otimes \Sigma)$ , and eventually calculates  $(n_i \times n_i)$ -dimensional submatrix  $\Omega_i$ . When the number of replications on each subject by respective methods is unequal, *PROC MIXED* considers the case as missing value situation. Options *V=3* and *VCORR=3* in the *RANDOM* statement give the estimate of the overall variance-covariance matrix  $\Omega_3$  and the corresponding  $\Omega_3$  Correlation matrix, i.e., for the third subject. The option *G* in the *RANDOM* statement gives the estimate of the between-subject variance-covariance matrix  $D$ . Option *R* in the *REPEATED* statement gives the estimate of the variance-covariance matrix  $R_1$  for the first subject. One can get the  $\Omega_i$  variance-covariance matrix and the corresponding  $\Omega_i$  Correlation matrix for all subjects by specifying *V=1* to *N*, and *VCORR=1* to *N* in the *RANDOM* statement. When the correlation matrix  $V$  on the replicated measurements assumes equicorrelated structure and  $\Sigma$  as unstructured, we use the option *TYPE=UN* along with *SUBJECT=REPLICATE(PATIENT)* in the *REPEATED* statement. This gives the  $2 \times 2$  within-subject variance-covariance matrix  $\Sigma$ . See Appendix A.

### Related hypotheses testings to test the disagreement between KVX and CBCT

If there is a disagreement between the two methods, it is important to know whether it is due to the bias, due to the difference in between-subject variabilities or due to the difference in within-subject variabilities of the two methods. If it is due to the bias between the two methods, it is easy to correct. The output of *PROC MIXED* always gives the bias, its  $t$ -value and its  $p$ -value. Nonetheless, it is not straightforward to check the agreement or disagreement in between-subject variabilities and in within-subject variabilities of the two methods. We will accomplish these by the indirect use of *PROC MIXED* in two steps (described below) by using likelihood ratio tests.

#### Testing of hypothesis of difference between the means of KVX and CBCT

We are interested in testing the following hypothesis for bias:

$H_\mu$  : the two methods do not have the same mean,  
vs.  $K_\mu$  : the two methods have the same mean.

Output of *PROC MIXED* (Solution for Fixed Effects) gives the bias and the corresponding  $t$ -value and  $p$ -value.

#### Testing of hypothesis of difference in between-subject variabilities of KVX and CBCT

Here we are interested in testing the following hypothesis:

$H_d$  : the two methods do not have the same between-subject variabilities,  
vs.  $K_d$  : the two methods have the same between-subject variabilities.

We apply the likelihood ratio test for this hypothesis testing. To compute the test statistic  $-2 \ln \Lambda_d$ , where

$$-2 \ln \Lambda_d = \left[ -2 \ln \max_{K_d} L \right] - \left[ -2 \ln \max_{H_d} L \right].$$

The log likelihood function under both null hypothesis and alternating hypothesis must be maximized separately. We do this by setting the option *METHOD=ML* in *PROC MIXED* statement. The option *TYPE=UN* in the *RANDOM* statement, along with the option *TYPE=UN* in the *REPEATED* statement, is used to calculate the “-2 Log Likelihood” for the covariance structure under  $H_d$ . Similarly, the option *TYPE=CS* in the *RANDOM* statement, along with the option *TYPE=UN* in the *REPEATED* statement, is used to calculate the “-2 Log Likelihood” for the covariance structure under  $K_d$ .

*PROC MIXED* calculates “-2 Log Likelihood” under the heading of “Fit Statistics”, see Appendix B. The above

test statistic  $-2 \ln \Lambda_d$  under  $K_d$  follows a chi-square distribution with degrees of freedom (d.f.)  $\nu_d$ , where  $\nu_d$  is computed as

$$\nu_d = \text{LRT df (under } H_d) - \text{LRT df (under } K_d).$$

PROC MIXED calculates "LRT df" under the heading of "Null Model Likelihood Ratio Test", see Appendix B.

**Testing of hypothesis of difference in within-subject variabilities of KVX and CBCT**

We test the difference between the repeatability coefficients of the two methods by testing the following hypothesis:

- $H_\sigma$  : the two methods do not have the same within-subject variabilities,
- vs.  $K_\sigma$  : the two methods have the same within-subject variabilities,

As before here also we apply the likelihood ratio test for this hypothesis testing, and maximize the log likelihood function under both null hypothesis and alternating hypothesis separately to compute the test statistic  $-2 \ln \Lambda_\sigma$ , where

$$-2 \ln \Lambda_\sigma = \left[ -2 \ln \max_{K_\sigma} L \right] - \left[ -2 \ln \max_{H_\sigma} L \right].$$

The option *TYPE=UN* in the *RANDOM* statement, along with *TYPE=UN* in the *REPEATED* statement, is used to calculate the "-2 Log Likelihood" for the covariance structure under  $H_\sigma$ . *TYPE=UN* in the *RANDOM* statement, along with *TYPE=CS* in the *REPEATED* statement, is used to calculate the "-2 Log Likelihood" for the covariance structure under  $K_\sigma$ . The test statistic  $-2 \ln \Lambda_\sigma$  under  $K_\sigma$  follows a chi-square distribution with d.f.  $\nu_\sigma = \text{LRT df (under } H_\sigma) - \text{LRT df (under } K_\sigma)$ .

**Testing of hypothesis of difference in overall variabilities of KVX and CBCT**

We are interested in testing the following hypothesis:

- $H_\omega$  : the two methods do not have the same overall variabilities,
- vs.  $K_\omega$  : the two methods have the same overall variabilities,

As before here also we apply the likelihood ratio test to compute the test statistic  $-2 \ln \Lambda_\omega$ , where

$$-2 \ln \Lambda_\omega = \left[ -2 \ln \max_{K_\omega} L \right] - \left[ -2 \ln \max_{H_\omega} L \right].$$

The option *TYPE=UN* in the *RANDOM* statement, along with *TYPE=UN* in the *REPEATED* statement, is used to calculate the "-2 Log Likelihood" for the covariance structure under  $H_\omega$ . The option *TYPE=CS* in the *RANDOM*

statement, along with *TYPE=CS* in the *REPEATED* statement, is used to calculate the "-2 Log Likelihood" for the covariance structure under  $K_\omega$ . The test statistic  $-2 \ln \Lambda_\omega$  under  $K_\omega$  follows a chi-square distribution with d.f.  $\nu_\omega = \text{LRT df (under } H_\omega) - \text{LRT df (under } K_\omega)$ .

**Results**

Selected parts of the SAS output to test the within-subject variabilities are given in Appendix B. We present the sample SAS code (see Appendix A) to test within-subject variabilities by fitting the linear mixed effects model to our KVX and CBCT shifts for the lateral (X). We see that

$$-2 \ln \max_{K_\sigma} L = 273.7 \quad \text{and} \quad -2 \ln \max_{H_\sigma} L = 239.7,$$

with

$$\text{LRT df (under } H_\sigma) = 5 \quad \text{and} \quad \text{LRT df (under } K_\sigma) = 4.$$

Therefore,

$$\begin{aligned} -2 \ln \Lambda_\sigma &= \left[ -2 \ln \max_{K_\sigma} L \right] - \left[ -2 \ln \max_{H_\sigma} L \right] \\ &= 273.7 - 239.7 = 34.0, \end{aligned}$$

with

$$\nu_\sigma = \text{LRT df (under } H_\sigma) - \text{LRT df (under } K_\sigma) = 5 - 4 = 1.$$

The p-value for testing the within-subject variabilities of the two methods by using IML procedure of SAS is calculated at the third stage (see Appendix A). The p-value = 5.5112E-9 (see Appendix B).

Inter-method bias, inter-method agreement, intra-method agreement, overall agreement and correlation results from COM<sub>3</sub>PARE are presented in Tables 1, 2, 3, 4 and 5. Using COM<sub>3</sub>PARE, in this specific head and neck positional verification demonstration dataset, while inter-method bias was < 1 mm for all axes, a statistically significant between method bias was noted in the Z-axis (superoinferior axis). Also, was evidenced there was a statistically significant difference between CBCT and KVX inter-subject variation in the Z-axis (Table 1). Intra-subject variability was noted to be statistically significant for X- and Z-axes, as was overall variation. Correlation coefficient calculation estimation was performed using a mixed effects model (as per Roy [52]).

Using the aforementioned criteria, automated shifts from CBCT and kV X-ray, acquired and processed in the manner detailed are interchangeable only for measurements of the Y-axis (anteroposterior), and for example,

**Table 1** Between-method bias

	Bias (cm)	p-value
X	0.0335	0.6077
Y	-0.0428	0.2836
Z	-0.0942	0.0253

**Table 2** Inter-method agreement

	KVX (cm)	CBCT (cm)	<i>p</i> -value
X	0.0413	0.0670	0.4795
Y	0.0511	0.0464	0.7518
Z	0.0273	0.0848	0.0010

should not be used on alternating days in facilities with both systems in either X- or Z-axis. Additionally, our method suggests that, with lower intra-method variability in the X-axis (lateral), CBCT is the preferred measurement method, while in the Z-axis (superoinferior) kV X-ray measurement is preferable.

**Discussion**

The necessity for quantitative evaluation of competing measurement devices, in cases where one device has not been found to be superior, is a significant need in science generally [1, 2, 5–7], and particularly within the radiological sciences community. Specifically, this issue is encountered when comparing distinct positional verification methods for image-guided radiotherapy [34, 53–56]. The difficulty of assessing competing platforms is particularly vexing, as it impedes efforts at cross platform comparison. Our group [24, 50] and others have implemented several distinct methods for presenting such analysis [25, 27, 33, 34, 45, 57]. Our previous efforts have utilized several extant method comparison statistical presentations (including Bland-Altman<sup>7</sup>, Lin’s concordance [58], Deming orthogonal regression [59, 60]); however, what was gained in completeness was lacking in parsimony. To this end, we sought to define an improved algorithm for practical comparison of distinct imaging methodologies, with a non-fixed number of repeated measurements per patient, in the absence of a “gold standard”. Often, inappropriate statistical analyses are implemented in lieu of formal method comparison statistics. The analysis of different measurement devices is not as straightforward as the initial observer may suppose. Bland and Altman demonstrated that mean comparison and linear regression are insufficient for comparison of differing measurement techniques [1]. The Bland-Altman method is succinct and easily interpretable, making it a classic of medical literature. In a series of seminal papers [1–7], Bland and Altman defined the standard methodology for comparing differing measurements, as well as establishing effective techniques accounting for

**Table 3** Intra-method agreement

	KVX (cm)	CBCT (cm)	<i>p</i> -value
X	0.3396	0.1047	$5.5 \times 10^{-9}$
Y	0.1687	0.1693	1.0
Z	0.0485	0.0825	0.0034

**Table 4** Overall agreement

	KVX (cm)	CBCT (cm)	<i>p</i> -value
X	0.3809	0.1717	$3.7 \times 10^{-8}$
Y	0.2198	0.2157	0.9512
Z	0.0758	0.1673	$1.3 \times 10^{-5}$

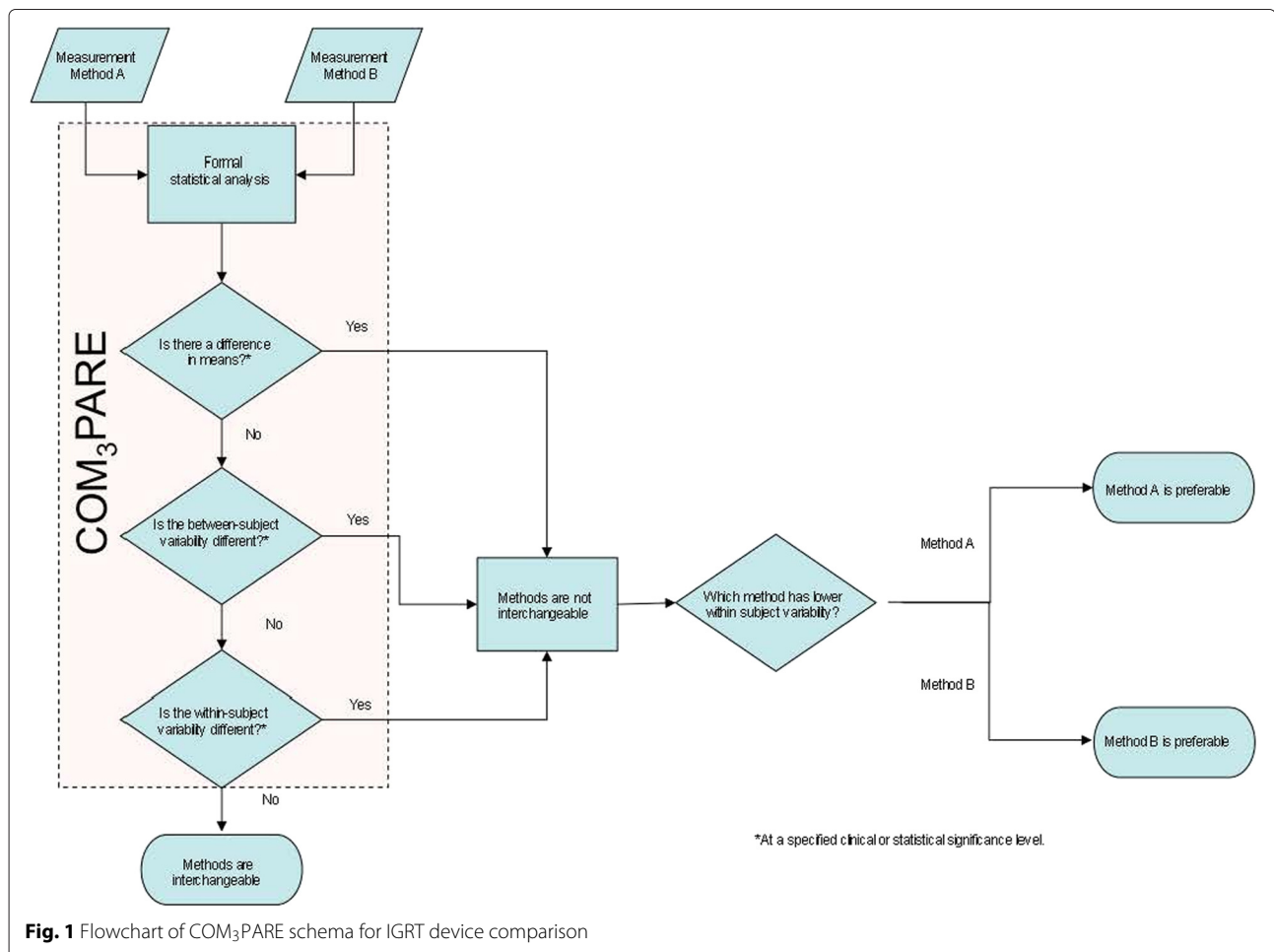
inter- and intra-method variability/repeatability. However, while the Bland-Altman methodology remains the current benchmark, it fails (by design, one should note) to include generation of a formalized *p*-value, instead recommending that a clinically meaningful difference between measures be utilized. Additionally, though repeatability estimation is a recommended component of accurate method comparison, the calculation for greater than two replicates is somewhat unwieldy using the methodology proposed by Bland and Altman. Since many IGRT datasets span > 30 repeated daily measures, the utility of a statistical methodology which can readily integrate large replicate numbers is desirable. The COM<sub>3</sub>PARE methodology presented herein represents an attempt to integrate several desirable methodological attributes into a unified, readily performed statistical process. COM<sub>3</sub>PARE has several advantages over existing method comparison statistical analyses. Specifically, compared to general linear model [61, 62] (GLM)-based approaches (such as the *t*-test, linear regression, and ANOVA [63]), which fail to account for multiple sources of random variance, the linear mixed effects (LME)-based COM<sub>3</sub>PARE platform integrates variation estimation at multiple hierarchical levels (i.e., between- and within- measurement methods/subjects) [48]. From a practical point of view, this allows factor-wise assessment of procedural or technical variability of each of the two methods rather than a combined assessment, so that there is the capacity to determine the exact source of disagreement. COM<sub>3</sub>PARE is also resilient with regard to uneven numbers of replicates per device, a feature of great practical utility in a clinical setting, such as daily IGRT recording, where the number of IGRT fractions received for each patient may differ based on fractionation regimen of clinical exigency. Additionally, since COM<sub>3</sub>PARE has the capacity to fit differences in said variability to a hypothesis testing-friendly Bonferroni-corrected *p*-value output, while still implementing clinician-determined thresholds

**Table 5** Mixed effects estimated correlation coefficient

	Correlation coefficient
X	0.5329
Y	0.8038
Z	0.7336

for agreement there is greater interpretability of statistical output, with no loss of clinical relevance. For instance, one could specify a priori that measurement differentials > 1 mm would represent a lack of interchangeability globally. Data presentation was performed in this study in an effort to illustrate potential applications of COM<sub>3</sub>PARE for replicated image-based measurements of the kind frequently encountered in radiation oncology. The specific dataset included have been previously presented using standard method approaches. By revisiting these data using compare we hope to illustrate implementation of what we perceive to be a more usable and parsimonious approach to conceptualizing method comparison for IGRT applications, expanding upon, rather than obviating the previous work. With regard to the specific dataset presented herein, our analysis points to the difficulties possible when comparing IGRT platforms. For instance, having set our criteria pre-analysis, we were surprised to note that differing measurement methods proved preferable in distinct axes (e.g., CBCT in X-axis, kV X-ray for the Z-axis), while appearing by

said criteria interchangeable in the Y-axis. A possible explanation of this phenomenon may appear as a feature of the imaging methodologies themselves. For CBCT, before three-dimensional reconstruction, data is acquired as axial slices (X-axis), while, previous to DRR referencing, the kV X-ray system uses orthogonal projections at oblique angles, parallel to the superoinferior plane (Z-axis). Consequently, method intra-subject repeatability may be tied to the reference plane of image acquisition, though this remains conjecture based on a single dataset. To our knowledge this technique represents the first formal hypothesis testing approach to integrate inter-method bias, inter-subject variability, and intra-subject variability of two methods with any number of replicated measurements for image-guided radiotherapy. As modeled on the aforementioned conceptual schema presented in the “Methods” section, we postulate that the following criteria be formally evaluated as feature of future image-guided radiotherapy measurement comparison studies comparing two imaging platforms, where multiple repeated observations



on the same subject is possible. To meet our criteria for interchangeability [48]:

1. The bias and overall agreement must fall within a pre-specified range (e.g., bias/agreement of < 0.1 cm between IGRT devices).
2. There should be no statistically significant, using a pre-specified threshold (e.g., < 0.05) difference in the inter-subject variability of the two methods.
3. There should be no statistically significant difference in the intra-subject variability (i.e., repeatability) of the two methods.
4. In cases where criteria 2 and 3 are NOT met, the preferred IGRT technique is the one exhibiting the lower intra-subject variability (i.e., greater repeatability).

These criteria are presented as a graphical schema (Fig. 1); notably analysis of criteria 1–3 is easily incorporated in a single step using the COM<sub>3</sub>PARE SAS Code (Appendix A). The a priori criteria set we specified for interchangeability represented what we considered a reasonable metrics for the given application (i.e. fractionated radiotherapy of 30+ fractions for head and neck cancer) with a standardized PTV margin. The COM<sub>3</sub>PARE methodology, however, allows specification of any specified difference/p-value combination. Consequently, if a scenario arose whereby either tighter tolerances are desirable (e.g. 3-fraction SBRT), such parameters can be easily defined as an acceptability criteria.

## Conclusion

COM<sub>3</sub>PARE represents an attempt at a unified conceptual schema and analytic algorithm for method comparison of IGRT platforms. Initial application in a head and neck positional verification dataset shows feasibility and utility.

## Appendix A

### SAS code

Below we provide the sample SAS code to test within-subject variabilities by fitting the linear mixed effects model to our KVX and CBCT shifts for the lateral (X). We first fit the linear mixed effects model for the null hypothesis, then we fit the linear mixed effects model for the alternating hypothesis, and then find the *p*-value for the test. Appropriate changes can be made to test between-subject variabilities and overall variabilities using the SAS commands as described in Sections Testing of hypothesis of difference in between-subject variabilities of KVX and CBCT, and Testing of hypothesis of difference in overall variabilities of KVX and CBCT. Appropriate changes can be also made for vertical

(Y) and longitudinal (Z) dimensions and for any other data sets.

```
/* Here we give the SAS code for fitting
the linear mixed effects model to the our
KVX and CBCT shifts data under the null
hypothesis that the two methods do not
have the same within-subject
variabilities.*/
```

```
options ls=64 ps=68 nocenter;
```

```
proc import datafile="c:\CBCTKVX.xls"
DBMS=EXCELCS
out=sasuser.max2 replace;
/*proc print data=sasuser.max2; run;*/
```

```
data KVXCBCTdataX; set sasuser.max2;
keep pat replicate xmlkvx xm2cbct;
```

```
data b; SET KVXCBCTdataX;
y=xmlkvx; mvar='1'; output;
y=xm2cbct; mvar='2'; output;
/*proc print data=b; run;*/
```

```
proc mixed data=b method=ml covtest;
classes pat mvar replicate;
model y = mvar/ s outpredm=pp ddfm=kr;
random mvar / type=un subject=pat g gcorr
v=3 vcorr=3;
repeated mvar/type= un
subject=replicate(pat) r rcorr;
run;
```

```
/* Here we give the SAS code for fitting
the linear mixed effects model to the same
data under the alternative hypothesis that
the two methods have the same
within-subject variabilities.*/
```

```
proc mixed data=b method=ml covtest;
classes pat mvar replicate;
model y=mvar /s ddfm=kr;
random mvar /type= un subject=pat v vcorr g;
repeated mvar/type= cs
subject=replicate (pat) r;
run;
```

```
/* Calculating the p-value to test the
within-subject variabilities of the two
methods */
```

```
proc iml;
pval=1- probchi(34.0, 1);
print pval; run;
```



**Appendix B****SAS output for covariance structure under the null and the alternating hypotheses**

Below we provide the selected portions of the output of the above program.

/\* The following output is for the null hypothesis \*/

Estimated V Matrix for Pat 3

Row	Col1	Col2	Col3	Col4	Col5	Col6
1	0.3809	0.1363	0.04126	0.03663	0.04126	0.03663
2	0.1363	0.1717	0.03663	0.06695	0.03663	0.06695
3	0.04126	0.03663	0.3809	0.1363	0.04126	0.03663
4	0.03663	0.06695	0.1363	0.1717	0.03663	0.06695
5	0.04126	0.03663	0.04126	0.03663	0.3809	0.1363
6	0.03663	0.06695	0.03663	0.06695	0.1363	0.1717

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
UN(1,1)	Pat	0.04126	0.03079	1.34	0.0902
UN(2,1)	Pat	0.03663	0.02384	1.54	0.1244
UN(2,2)	Pat	0.06695	0.02669	2.51	0.0061
UN(1,1)	Replicate(Pat)	0.3396	0.05238	6.48	<.0001
UN(2,1)	Replicate(Pat)	0.09963	0.02359	4.22	<.0001
UN(2,2)	Replicate(Pat)	0.1047	0.01688	6.20	<.0001

Fit Statistics

-2 Log Likelihood	239.7
AIC (smaller is better)	255.7
AICC (smaller is better)	256.5
BIC (smaller is better)	266.4

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
5	74.98	<.0001

Solution for Fixed Effects

Effect	mvar	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept		0.03022	0.06212	28.7	0.49	0.6304
mvar	1	0.03349	0.06447	26.5	0.52	0.6077
mvar	2	0	.	.	.	.

/\* The following output is for the alternating hypothesis \*/

Fit Statistics

-2 Log Likelihood	273.7
AIC (smaller is better)	287.7
AICC (smaller is better)	288.2
BIC (smaller is better)	297.0

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
4	41.03	<.0001

/\* The following output is for the p-value to test the within-subject variabilities of the two methods \*/

pval  
5.5112E-9

### Competing interests

CDF received/receives grant support from: the SWOG Hope Foundation Dr. Charles A. Coltman, Jr. Fellowship in Clinical Trials; the National Institutes of Health Paul Calabresi Clinical Oncology Award Program (K12 CA088084) and Clinician Scientist Loan Repayment Program (L30 CA136381-02); Elekta AB/MD Anderson Consortium; GE Medical Systems/MD Anderson Center for Advanced Biomedical Imaging In-Kind Award; the MD Anderson Center for Radiation Oncology Research, and an MD Anderson Institutional Research Grant Program Award. These listed funders/supporters played no role in the study design, collection, analysis, interpretation of data, manuscript writing, or decision to submit the report for publication.

### Authors' contributions

AR conceived of and performed the statistical methods described herein, and co-drafted the manuscript. CDF conceived of the study, and participated in its design and coordination, assisted with data collection and co-drafted the manuscript. DIR helped to draft the manuscript, provided scientific oversight and provided editorial assistance. CRT participated in the design of the study and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

Special thanks to Joseph Ting, PhD for dataset utilization permission.

### Conflict of interest disclosure

CDF was supported by the National Institutes of Health Clinical Research Loan Repayment Program (L30 CA136381), National Institute of Biomedical Imaging and Bioengineering (5T32EB000817-04), and the SWOG Hope Foundation Coltman Fellowship. The funder(s) played no role in study design, in the collection, analysis and interpretation of data, in the writing of the manuscript, nor in the decision to submit the manuscript.

### Author details

<sup>1</sup>Department of Management Science and Statistics, The University of Texas at San Antonio, One UTSA Circle, 78249 San Antonio, TX, USA. <sup>2</sup>Department of Radiation Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA. <sup>3</sup> Department of Radiation Medicine, Oregon Health & Science University, Portland, OR, USA.

Received: 17 July 2014 Accepted: 24 July 2015

Published online: 28 August 2015

### References

- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–10.
- Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet*. 1995;346:1085–7.
- Bland JM, Altman DG. Comparing two methods of clinical measurement: a personal history. *Int J Epidemiol*. 1995;24(Suppl 1):S7–14.
- Bland JM, Altman DG. Measurement error. *BMJ (Clinical research ed)*. 1996;312:1654.
- Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8:135–60.
- Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol*. 2003;22:85–93.
- Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat*. 2007;17:571–82.
- Bisdas S, Konstantinou G, Surlan-Popovic K, Khoshneviszadeh A, Baghi M, Vogl TJ, et al. Dynamic contrast-enhanced CT of head and neck tumors: comparison of first-pass and permeability perfusion measurements using two different commercially available tracer kinetics models. *Acad Radiol*. 2008;15:1580–9.
- Cronin P, Saab A, Kelly AM, Gross BH, Patel S, Kazerooni EA, Carlos RC. Measurements of pulmonary vein ostial diameter and distance to first bifurcation: A comparison of different measurement methods. *Eur J Radiol*. 2009;71(1):61–8.
- Kanza RE, Higashino H, Kido T, Kurata A, Saito M, Sugawara Y, Mochizuki T. Quantitative assessment of regional left ventricular wall thickness and thickening using 16 multidetector-row computed tomography: comparison with cine magnetic resonance imaging. *Radiat Med*. 2007;25:119–26.
- Kwee TC, Takahara T, Koh DM, Nieuwstein RA, Lijten PR. Comparison and reproducibility of ADC measurements in breathhold, respiratory triggered, and free-breathing diffusion-weighted MR imaging of the liver. *J Magn Reson Imaging*. 2008;28:1141–8.
- Mahnken AH, Spuentrup E, Niethammer M, Buecker A, Boese J, Wildberger JE Flohr T, et al. Quantitative and qualitative assessment of left ventricular volume with ECG-gated multislice spiral CT: value of different image reconstruction algorithms in comparison to MRI. *Acta Radiol*. 2003;44:604–11.
- Martin KE, Helvie MA, Zhou C, Roubidoux MA, Bailey JE, Paramagul C, et al. Mammographic density measured with quantitative computer-aided method: comparison with radiologists' estimates and BI-RADS categories. *Radiology*. 2006;240:656–65.
- Murakami R, Uozumi H, Hirai T, Nishimura R, Katsuragawa S, Shiraishi S, Toya R, Tashiro K, Kawanaka K, Oya N, Tomiguchi S, Yamashita Y. Impact of FDG-PET/CT fused imaging on tumor volume assessment of head-and-neck squamous cell carcinoma: intermethod and interobserver variations. *Acta Radiol*. 2008;49:693–9.
- Nieman K, Shapiro MD, Ferencik M, Nomura CH, Abbara S, Hoffmann U, et al. Reperfused myocardial infarction: contrast-enhanced 64-Section CT in comparison to MR imaging. *Radiology*. 2008;247:49–56.
- Agazaryan N, Tenn SE, Desalles AA, Selch MT. Image-guided radiosurgery for spinal tumors: methods, accuracy and patient intrafraction motion. *Phys Med Biol*. 2008;53:1715–27.
- Dawson LA, Jaffray DA. Advances in image-guided radiation therapy. *J Clin Oncol*. 2007;25:938–946.
- Grills IS, Hugo G, Kestin LL, Galerani AP, Chao KK, Wloch J, et al. Image-guided radiotherapy via daily online cone-beam CT substantially reduces margin requirements for stereotactic lung radiotherapy. *Int J Radiat Oncol Biol Phys*. 2008;70:1045–56.
- Ippolito E, Mertens I, Haustermans K, Gambacorta MA, Pasini D, Valentini V. IGRT in rectal cancer. *Acta oncologica (Stockholm, Sweden)*. 2008;47:1317–24.
- Kupelian PA, Langen KM, Willoughby TR, Zeidan OA, Meeks SL. Image-guided radiotherapy for localized prostate cancer: treating a moving target. *Semin Radiat Oncol*. 2008;18:58–66.
- Wertz H, Lohr F, Dobler B, Mai S, Wenz F. Dosimetric impact of image-guided translational isocenter correction for 3-D conformal radiotherapy of the prostate. *Strahlenther Onkol*. 2007;183:203–10.
- Wiersma RD, Mao W, Xing L. Combined kV and MV imaging for real-time tracking of implanted fiducial markers. *Med Phys*. 2008;35:1191–8.
- Yorke ED, Keall P, Verhaegen F. Anniversary paper: Role of medical physicists and the AAPM in improving geometric aspects of treatment accuracy and precision. *Med Phys*. 2008;35:828–39.
- Fuller CD, Thomas CR, Schwartz S, Golden N, Ting J, Wong A, et al. Method comparison of ultrasound and kilovoltage x-ray fiducial marker imaging for prostate radiotherapy targeting. *Phys Med Biol*. 2006;51:4981–93.
- Moseley DJ, White EA, Wiltshire KL, Rosewall T, Sharpe MB, Siewerdsen JH, et al. Comparison of localization performance with implanted fiducial markers and cone-beam computed tomography for on-line image-guided radiotherapy of the prostate. *Int J Radiat Oncol Biol Phys*. 2007;67:942–53.
- Artignan X, Smitsmans MH, Lebesque JV, Jaffray DA, van Her M, Bartelink H. Online ultrasound image guidance for radiotherapy of prostate cancer: impact of image acquisition on prostate displacement. *Int J Radiat Oncol Biol Phys*. 2004;59:595–601.
- Cury FL, Shenouda G, Souhami L, Duclos M, Faria SL, David M, et al. Ultrasound-based image guided radiotherapy for prostate cancer: comparison of cross-modality and intramodality methods for daily localization during external beam radiotherapy. *Int J Radiat Oncol Biol Phys*. 2006;66:1562–7.
- Fung AY, Ayyangar KM, Djajaputra D, Nehru RM, Enke CA. Ultrasound-based guidance of intensity-modulated radiation therapy. *NMed Dosim*. 2006;3:120–9.
- Fuss M, Salter BJ, Cavanaugh SX, Fuss C, Sadeghi A, Fuller CD, et al. Daily ultrasound-based image-guided targeting for radiotherapy of upper abdominal malignancies. *Int J Radiat Oncol Biol Phys*. 2004;59:1245–56.

30. Langen KM, Pouliot J, Anezinos C, Aubin M, Gottschalk AR, Hsu IC, et al. Evaluation of ultrasound-based prostate localization for image-guided radiotherapy. *Int J Radiat Oncol Biol Phys.* 2003;57:635–44.
31. Patel RR, Orton N, Tome WA, Chappell R, Ritter MA. Rectal dose sparing with a balloon catheter and ultrasound localization in conformal radiation therapy for prostate cancer. *Radiother Oncol.* 2003;67:285–94.
32. Choi M, Fuller CD, Wang SJ, Siddiqi A, Wong A, Thomas Jr CR, Fuss M. Effect of body mass index on shifts in ultrasound-based image-guided intensity-modulated radiation therapy for abdominal malignancies. *Radiother Oncol.* 2009;91(1):114–9.
33. Gayou O, Miften M. Comparison of mega-voltage cone-beam computed tomography prostate localization with online ultrasound and fiducial markers methods. *Med Phys.* 2008;35:531–8.
34. Johnston H, Hilts M, Beckham W, Berthelet E. 3D ultrasound for prostate localization in radiation therapy: a comparison with implanted fiducial markers. *Med Phys.* 2008;35:2403–13.
35. Aubry JF, Beaulieu L, Girouard LM, Aubin S, Tremblay D, Laverdiere J, et al. Measurements of intrafraction motion and interfraction and intrafraction rotation of prostate by three-dimensional analysis of daily portal imaging with radiopaque markers. *Int J Radiat Oncol Biol Phys.* 2004;60:30–9.
36. Balter JM, Sandler HM, Lam K, Bree RL, Lichter AS, Haken TENRK. Measurement of prostate movement over the course of routine radiotherapy using implanted markers. *Int J Radiat Oncol Biol Phys.* 1995;31:113–8.
37. Evans PM. Anatomical imaging for radiotherapy. *Phys Med Biol.* 2008;53:R151–191.
38. Lometti MW, Thurston D, Aubin M, Bock A, Verhey L, Lockhart JM, Bland R, Pouliot J, Roach 3rd M. Are lateral electronic portal images adequate for accurate on-line daily targeting of the prostate? Results of a prospective study. *Med Dosim.* 2008;33:22–9.
39. Sornsen de Kostevan JR, Cuijpers JP, Geest de FG, Lagerwaard FJ, Slotman BJ, Senan S. Verifying 4D gated radiotherapy using time-integrated electronic portal imaging: a phantom and clinical study. *adiat Oncol (London, England).* 2007;2:32.
40. Morin O, Gillis A, Chen J, Aubin M, Buccì MK, Roach 3rd M, et al. Megavoltage cone-beam CT: system description and clinical applications. *Med Dosim.* 2006;31:51–61.
41. Pang G, Bani-Hashemi A, Au P, O'Brien PF, Rowlands JA, Morton G, et al. Megavoltage cone beam digital tomosynthesis (MV-CBDT) for image-guided radiotherapy: a clinical investigational system. *Phys Med Biol.* 2008;53:999–1013.
42. Pisani L, Lockman D, Jaffray D, Yan D, Martinez A, Wong J. Setup error in radiotherapy: on-line correction using electronic kilovoltage and megavoltage radiographs. *Int J Radiat Oncol Biol Phys.* 2000;47:825–39.
43. Pouliot J, Bani-Hashemi A, Chen J, Svatos M, Ghelmansarai F, Mitschke M, et al. Low-dose megavoltage cone-beam CT for radiation therapy. *Int J Radiat Oncol Biol Phys.* 2005;61:552–60.
44. Serago CF, Buskirk SJ, Igel TC, Gale AA, Serago NE, Earle JD. Comparison of daily megavoltage electronic portal imaging or kilovoltage imaging with marker seeds to ultrasound imaging or skin marks for prostate localization and treatment positioning in patients with prostate cancer. *Int J Radiat Oncol Biol Phys.* 2006;65:1585–92.
45. Wu QJ, Godfrey DJ, Wang Z, Zhang J, Zhou S, Yoo S, Brizel DM, Yin FF. On-board patient positioning for head-and-neck IMRT: comparing digital tomosynthesis to kilovoltage radiography and cone-beam computed tomography. *Int J Radiat Oncol Biol Phys.* 2007;69:598–606.
46. Scarbrough TJ, Golden NM, Ting JY, Fuller CD, Wong A, Kupelian PA, Thomas Jr CR. Comparison of ultrasound and implanted seed marker prostate localization methods: Implications for image-guided radiotherapy. *Int J Radiat Oncol Biol Phys.* 2006;65:378–87.
47. Stutzel J, Oelfke U, Nill S. A quantitative image quality comparison of four different image guided radiotherapy devices. *Radiother Oncol.* 2008;86:20–4.
48. Roy A. An application of linear mixed effects model to assess the agreement between two methods with replicated observations. *J Biopharm Stat.* 2009;19:150–73.
49. Barnhart HX, Haber MJ, Lin LL. An overview on assessing agreement with continuous measurements. *J Biopharm Stat.* 2007;17:529–69.
50. Fuller CD, Scarbrough TJ, Sonke JJ, Rasch CR, Choi M, Ting JY, et al. Method comparison of automated matching software-assisted cone-beam CT and stereoscopic kilovoltage x-ray positional verification image-guided radiation therapy for head and neck cancer: a prospective analysis. *Phys Med Biol.* 2009;54:7401–15.
51. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics.* 1997;53:983–997.
52. Roy A. Estimating correlation coefficient between two variables with repeated observations using mixed effects model. *Biometrical J.* 2006;48:286–301.
53. Cui Y, Galvin JM, Straube WL, Bosch WR, Purdy JA, Li XA, Xiao Y. Multi-System Verification of Registrations for Image-Guided Radiotherapy in Clinical Trials. *Int J Radiat Oncol Biol Phys.* 2011;81(1):305–312.
54. Mohammed N, Kestin L, Grills I, Shah C, Glide-Hurst C, Yan D, et al. Comparison of IGRT Registration Strategies for Optimal Coverage of Primary Lung Tumors and Involved Nodes Based on Multiple Four-Dimensional CT Scans Obtained Throughout the Radiotherapy Course. *Int J Radiat Oncol Biol Phys.* 2012;82(4):1541–8.
55. Shi W, Li JG, Zlotecki RA, Yeung A, Newlin H, Palta J, et al. Evaluation of KV cone-beam ct performance for prostate IGRT: a comparison of automatic grey-value alignment to implanted fiducial-marker alignment. *Am J Clin Oncol.* 2011;34(1):16–21.
56. Thongphiew D, Wu QJ, Lee WR, Chankong V, Yoo S, McMahon R, Yin FF. Comparison of online IGRT techniques for prostate IMRT treatment: adaptive vs repositioning correction. *Med Phys.* 2009;36:1651–62.
57. Borst GR, Sonke JJ, Betgen A, Remeijer P, Herk van M, Lebesque JV. Kilo-voltage cone-beam computed tomography setup measurements for lung cancer patients; first clinical results and comparison with electronic portal-imaging device. *Int J Radiat Oncol Biol Phys.* 2007;68:555–61.
58. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics.* 1989;45:255–68.
59. Konings H. Use of Deming regression in method-comparison studies. *Surv Immunol Res.* 1982;1:371–74.
60. Martin RF. General deming regression for estimating systematic bias and its confidence interval in method-comparison studies. *Clinical Chem.* 2000;46:100–4.
61. Milliken G, Graybill F. Extensions of the General Linear Hypothesis Model. *J Am Stat Assoc.* 1970;65:797–807.
62. LaMotte L. A Canonical Form for the General Linear Model. *Ann Stat.* 1977;5:787–9.
63. Fennessey J. The General Linear Model: A New Perspective on Some Familiar Topics. *Am J Sociol.* 1968;74:1–27.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

