

# Experimental validation of predicted application accuracies for computer-assisted (CAS) intraoperative navigation with paired-point registration

Martina Perwög<sup>1</sup> · Zoltan Bardosi<sup>1</sup> · Wolfgang Freysinger<sup>1</sup>

Received: 11 March 2017 / Accepted: 24 July 2017 / Published online: 11 August 2017  
© The Author(s) 2017. This article is an open access publication

## Abstract

**Purpose** The target registration error (TRE) is a crucial parameter to estimate the potential usefulness of computer-assisted navigation intraoperatively. Both image-to-patient registration on base of rigid-body registration and TRE prediction methods are available for spatially isotropic and anisotropic data. This study presents a thorough validation of data obtained in an experimental operating room setting with CT images.

**Methods** Optical tracking was used to register a plastic skull, an anatomic specimen, and a volunteer to their respective CT images. Plastic skull and anatomic specimen had implanted bone fiducials for registration; the volunteer was registered with anatomic landmarks. Fiducial localization error, fiducial registration error, and total target error (TTE) were measured; the TTE was compared to isotropic and anisotropic error prediction models. Numerical simulations of the experiment were done additionally.

**Results** The user localization error and the TTE were measured and calculated using predictions, both leading to results as expected for anatomic landmarks and screws used as fiducials. TRE/TTE is submillimetric for the plastic skull and the anatomic specimen. In the experimental data a medium correlation was found between TRE and target localization error (TLE). Most of the predictions of the application accuracy (TRE) fall in the 68% confidence interval of the measured TTE. For the numerically simulated data, a prediction of TTE was not possible; TRE and TTE show a negligible correlation.

**Conclusion** Experimental application accuracy of computer-assisted navigation could be predicted satisfactorily with adequate models in an experimental setup with paired-point registration of CT images to a patient. The experimental findings suggest that it is possible to run navigation and prediction of navigation application accuracy basically defined by the spatial resolution/precision of the 3D tracker used.

**Keywords** Registration · Anisotropy · Surgical guidance · Error analysis · Accuracy · Navigation

## Introduction

Navigation is widely used in ENT surgery to support the surgeon. A crucial part of the whole navigation process is the registration of the patient to the preoperative CT/MRI images. Usually paired-point matching [1–3] or more recently surface registration [4,5] is used for registration. Homologous points on the patient and in the image (fiducials) are used to find the rigid transformation between them. Errors in localizing fiducials in image and patient space FLE lead to the FRE [6], which is the Euclidean distance between the corresponding fiducials after registration. Usually, fiducials on the surface of the patient are used for registration, but the operating area is inside the head. Tracking errors and errors in localizing fiducials on the patient or in the images prohibit perfect navigation. The TRE [6] allows surgeons estimating the accuracy of navigation inside the patient at the surgical target zone. This is thus a good measure for the theoretical clinical application accuracy of a navigation system. Knowing TRE before surgery is a key component for a reliable intraoperative use of information

✉ Martina Perwög  
martina.bickel@student.i-med.ac.at

<sup>1</sup> Medical University Innsbruck, Anichstr. 35, Innsbruck, Austria

guidance provided by navigation systems. The use of CAS systems might improve surgery, reduce peri- and postoperative complications, and thus might allow faster healing of patients [7]. Therefore, a prediction of the error in special regions inside the head, especially close to critical structures, is highly desirable. Different prediction methods for TRE were developed [6, 8–11]. From a clinical perspective good predictions should overestimate the real application error.

Theoretical comparisons [12], numerical simulations [6, 8–11], and clinical studies [13–15] tried proving the methods for predicting TRE. To the best of our knowledge, a comprehensive analysis of available prediction methods of application accuracy against experimental data in a surgical setup is not available yet.

The first raw analysis of the data presented in this paper has already been published in [16], where only isotropic registration with an isotropic FLE model was investigated. The present work extends [16] with a comprehensive analysis of the data by including isotropic and anisotropic errors of measurements, registrations, and prediction methods. The emphasis is on the most frequently used prediction method [6] or methods that fit the simulated surgery best: anisotropic prediction [8] and a general approach [10]. This investigation presents a critical appraisal of predictions and measurements for computer-assisted navigation, based on real data from experiments collected under realistic conditions.

Numerical simulations of the experiment that by definition fulfill all theoretical requirements served to compare purely theoretical predictions against predictions on base of experimental data. For both “experiments” statistical correlations between measured and predicted quantities (such as TRE) were calculated. For the simulated data also distributions of the measured and predicted errors were analyzed. The specific advantage of both experiments (numerical and real life) is that ALL positions in patient and image spaces, including target positions, are available and can be used for relevant calculations and measurements.

The next sections describe the data acquisition, all errors, measured and predicted, are defined, and the whole experiment is described. In the final sections, the results are presented and discussed.

## Materials and methods

### Data acquisition

For the experiments a plastic skull, an anatomic specimen, and a volunteer (“patient”), were registered with paired-point matching registration to their CT images [17, 18].

CT data for the plastic skull and the anatomic specimen were acquired with a Siemens Sensation 16 CT (Siemens, Erlangen, Germany). A Siemens Somatom Plus 4 Volume Zoom was used to acquire the imaging data for the vol-

unteer. The imaging parameters were: for the plastic skull: convolution kernel H60s, 120 kV, 74 mA, 1 mm slice thickness; for the anatomic specimen: convolution kernel H30s, 120 kV, 175 mA, 0.6 mm slice thickness; for the volunteer: reconstruction filter H30s, 140 kV, 150 mA, 1.25 mm slice thickness. Navigation was done with open4Dnav [19], an IGSTK-based application with optical tracking (active Polaris, first generation, NDI, Ontario, Canada) [20]. MATLAB R2012a (The Mathworks, Inc., USA) was used for analyzing the data.

Isotropic [17] and anisotropic [18] image-to-patient registration was executed with MATLAB to get the transformation between image and patient space. Fiducials and targets were defined before starting the registration process for each patient.

For image-to-patient registration 3, 5, 7, and 9 fiducial points were used. For the anatomic specimen (with Ti-screws) and the volunteer (with anatomic landmarks), 10 target points were used and 11 targets were used for the plastic skull (with Ti-screws).

To verify the registration, the surgeon used a probe to point on the fiducials in patient space (FRE). This is normal clinical practice and done prior to each surgical intervention to verify navigation. If the FRE was appropriate, the TTE was determined by measuring the difference between positions as displayed by the system and “real” target points in image space. The TRE was predicted for the real target (detailed definitions are presented in “Definition of the measured errors” section).

This process was repeated 10 times for each patient and each fiducial arrangement (i.e., 3, 5, 7, and 9 fiducials), yielding 240 registration points in total for each patient, 100 targets for the anatomic specimen and the volunteer, and 110 targets for the plastic skull.

For each target in image space, the mean value of the 10 repetitions of the localization data in image space was analyzed and set as reference target points [21].

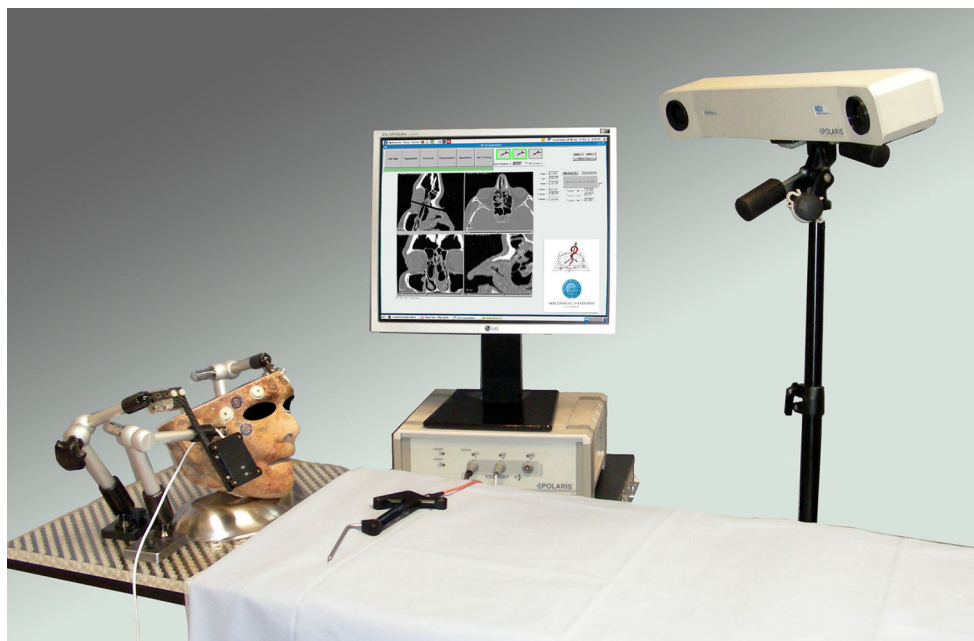
A detailed description of the experiment can be found in [16]; the setup can be seen in Fig. 1.

### Definition of the measured errors

Let  $x_{ij}$  and  $y_{ij}$  represent corresponding points (fiducials) in image and patient space, respectively, where  $i = 1, \dots, 10$  is the number of the registration and  $j = 1, \dots, m$  is the number of the fiducials;  $m = 3, 5, 7, 9$ .

Let  $r_{ik}$  and  $q_{ik}$  be the corresponding targets in image and patient space, respectively, where  $i = 1, \dots, 10$  is the number of registration,  $k = 1, \dots, n$  is the number of the target;  $n = 10$  for anatomic specimen and volunteer, and  $n = 11$  for the plastic skull.

Reference targets  $r_{k(m)}$  in image space are defined as  $r_{k(m)} = \sum_{i=1}^{10} \frac{r_{ik}}{n}$ , the mean of target  $k$  over all registra-



**Fig. 1** Experimental setup. The patient is fixed on the OR table. For all experiments the surgeon was using the same probe. The active NDI camera, the navigation system’s monitor, and the tracker unit are placed in optimal working distance. The DRF is attached near the patient

tions with  $m = 3, 5, 7,$  and  $9$  fiducials, respectively. For each experiment with  $m$  fiducials, the reference targets are calculated separately.

Isotropic registration: Image fiducials were registered to patient fiducials with the transformation that minimizes

$$FRE_{iso,i}^2 = \frac{1}{m} \sum_{j=1}^m (R_{iso,i}x_{ij} + t_{iso,i} - y_{ij})^2. \tag{1}$$

For registration  $i$ , the rotation matrix  $R_{iso,i}$ , the translation  $t_{iso,i}$ , and  $FRE_{iso,i}$  were saved.

The experimental TTE is the norm of the difference vectors of measured and navigated targets using

$$TTE_{exp,i} = \| r_k - (R_{iso,i}q_{jk} + t_{iso,i}) \|^2. \tag{2}$$

The values of  $TTE_{exp,i}$  were used as the reference TRE to be compared with the TRE of the different prediction methods.

$RMS(FLE_{i,img})$  and  $RMS(FLE_{i,pat})$  in image and patient space, respectively, were estimated as the traces of the covariance matrix of image and patient fiducials of the  $i$ -th repetition, respectively:

$$RMS(FLE_{i,img}) = \sqrt{\text{trace}(\text{cov}(x_{ij}))}, \tag{3}$$

and

$$RMS(FLE_{i,pat}) = \sqrt{\text{trace}(\text{cov}(y_{ij}))}. \tag{4}$$

The total FLE for registration  $i$ ,  $TFLE_i^2 = (RMS(FLE_{i,img}))^2 + (RMS(FLE_{i,pat}))^2$ , can be treated as a single random variable [22–24].

The target localization error for registration  $i$  (similar to FLE) is defined as

$$TLE_i^2 = TLE_{i,img}^2 + TLE_{i,pat}^2 \tag{5}$$

$TLE_{i,img}^2$  and  $TLE_{i,pat}^2$  are defined as  $RMS(TLE_{i,img}) = \sqrt{\text{trace}(\text{cov}(r_{ij}))}$  and  $RMS(TLE_{i,pat}) = \sqrt{\text{trace}(\text{cov}(q_{ij}))}$ , respectively. The TLE is equivalent to the FLE, but measured on targets, not on fiducials. Knowing all target positions in image and patient space allows to calculate the TLE, contrary to the definition in [25,26] where the target positions are unknown.

Anisotropic registration considers anisotropic noise in the measurement data, image data, etc., and FRE becomes

$$FRE_{aniso,i}^2 = \frac{1}{m} \sum_{j=1}^m W_{ij} |R_{aniso,i}x_{ij} + t_{aniso,i} - y_{ij}|^2 \tag{6}$$

has to be minimized.  $W_{ij} = V_{ij}^T \text{diag}(\sigma_{j1}^{-1}, \sigma_{j2}^{-1}, \sigma_{j3}^{-1}) V_{ij}$  is the weighting matrix, where  $I = V_j^T V_j$ , a  $3 \times 3$  identity matrix, and the columns of  $V_j$  are the principal axes of the FLE for fiducial  $j$ , and  $\sigma_{j\alpha}$ ,  $\alpha = 1, 2, 3$ , are the standard deviations of the FLE, resolved in three uncorrelated components along orthogonal principal axes [18].

## TRE prediction methods

For TRE prediction 6 different estimation methods were used and are described in this section:  $TRE_{F,FLE}$ ,  $TRE_{F,FRE}$ ,  $TTE_{F,FLE}$ ,  $TTE_{F,FRE}$ ,  $TRE_D$ , and  $TRE_W$ ;  $\langle \cdot \rangle$  denotes the expected value.

- (a) Fitzpatrick [6] derived an expression for the expected value of the TRE which is based on the linearization of the rigid point registration problem. It is a closed-form solution to estimate the  $TRE_F$ , where FLE follows an independent and identically distributed (iid) zero-mean Gaussian distribution. The expected  $TRE_F$  of a target  $r$  obtained on base of the FLE,  $TRE_{F,FLE}$ , is

$$\langle TRE_{F,FLE}^2(r) \rangle = \frac{\langle FLE^2 \rangle}{N} \left( 1 + \frac{1}{3} \sum_{k=1}^3 \frac{d_k^2}{f_k^2} \right) \quad (7)$$

where  $d_k$  is the distance of  $r$  from the principal axis  $k$  of the fiducial configuration and  $f_k$  is the RMS distance of the fiducials from that principal axis. For the prediction of  $TRE_{F,FLE}$  the measured  $TFLE_j$  was used as an approximation to  $\langle FLE^2 \rangle$ .

- (b) In addition,  $TRE_{F,FRE}$  was predicted with the expected value of  $\langle FLE_{iso,est,j} \rangle$  estimated from  $\langle FRE_{iso,j} \rangle$  of registration  $j$ ,  $\langle FLE_{iso,est,j}^2 \rangle = \frac{m}{m-2} \langle FRE_{iso,j}^2 \rangle$  [6];  $m$  is the number of fiducials used for registration.
- (c) The target localization error (TLE) is the error made in localizing the target (the probe is placed at target  $r$ , but the system reports target  $r'$  [26]). The system makes a TLE, which is uncorrelated to TRE and so  $\langle TTE_{F,x}^2 \rangle = \langle TRE_{F,x}^2 \rangle + \langle TLE^2 \rangle$ . TLE can be measured (see “Definition of the measured errors” section), and so the  $TTE_{F,x}$  can be reported. Here  $\langle TRE_{F,x}^2 \rangle$  and  $\langle TTE_{F,x}^2 \rangle$  indicate  $TRE_{F,FLE}$ ,  $TRE_{F,FRE}$ ,  $TTE_{F,FLE}$  and  $TTE_{F,FRE}$ , respectively.
- (d) Wiles [8] presented a closed-form solution estimation of  $TRE_W$  similar to [6], but with anisotropic normally distributed FLE. With this approach the  $\langle RMS(TRE_W)^2 \rangle$  and the covariance matrix of  $TRE_W$  can be obtained for predicting anisotropic application accuracy.
- (e) The generalized prediction of Danilchenko [10], given as  $\langle TRE_D \rangle$ , is valid for anisotropic and isotropic FLE and arbitrary weighting of the fiducials.

Isotropic registration was used for  $TTE_{exp}$ ,  $TRE_{F,x}$ , and  $TTE_{F,x}$ ; anisotropic registration was used for  $TTE_{exp,aniso}$ ,  $TRE_{F,x,aniso}$ ,  $TTE_{F,x,aniso}$ ,  $TRE_D$ , and  $TRE_W$ , with  $x = FLE$  or  $FRE$ . In [16] anisotropy of fiducials, measurements, and the setup was detected. Thus, it is clear that anisotropic registration had to be used for this analysis.

## ULE

The user localization error (ULE) is the pure user error of placing the probe on a fiducial and has already been defined and evaluated in [16]. Two different approaches were defined: Predict the  $ULE_F$  with  $TTE_{F,FRE}$  or calculate the ULE with measured errors ( $TFLE$ ,  $FLE_{img}$ ,  $FLE_{tracker}$ ,  $FLE_{probe\,calib}$ ):

$$\langle ULE_F^2 \rangle = \frac{1}{1 + \frac{1}{N} \left( 1 + \frac{1}{3} \sum_{k=1}^N \frac{d_k^2}{f_k^2} \right)} \times \left[ \langle TTE_{F,FRE}^2 \rangle - \frac{1}{N} \left( 1 + \frac{1}{3} \sum_{k=1}^N \frac{d_k^2}{f_k^2} \right) \left( \langle FLE_{tracker}^2 \rangle + \langle FLE_{probe\,calib}^2 \rangle \right) - \langle FLE_{img}^2 \rangle \right] \quad (8)$$

and

$$\langle ULE^2 \rangle = \langle TFLE^2 \rangle - \langle FLE_{img}^2 \rangle - \langle FLE_{tracker}^2 \rangle - \langle FLE_{probe\,calib}^2 \rangle. \quad (9)$$

Since measurement errors had occurred that we were not aware of in [16] (see “Data inspection and analysis” section), the results of the ULE are reported here correctly, calculated with formulae 8 and 9.

## Data inspection and analysis

While assessing anisotropic registration doubts arose about the validity of the raw data; re-analysis revealed that the experimental conditions must have changed during the measuring sessions. Unfortunately this was not discovered in our previous work [16]. All patients’ data were measured relative to a DRF; thus, changes that occurred due to temporal drift can be well observed. It was detected that the raw experimental data for the anatomic specimen, registration with three fiducials, repetitions 9 and 10, were off up to 2 mm in  $x$ -,  $y$ -, and  $z$ -direction in tracker space. This is not possible when bone-anchored fiducials are used that are still rigidly in place. Obviously the anatomic specimen, the patient tracker (dynamic reference frame, DRF), or the whole setup was slightly changed unnoticed. Therefore, these two repetitions were eliminated from further analysis.

The raw data for the volunteer for registrations with 7 and 9 points were not utilizable either: Every fiducial underwent a change of up to 10 mm in  $x$ -,  $y$ -, and  $z$ -directions. As this was an investigation with a volunteer, no general anesthesia was used and the volunteer’s head was fixed to the operating table with a tape, which is a common practice in our hospital [27]. Thus, it is probable that the causes for this discrepancy were unnoticed movements of

the head and/or a warming of the plastic material and a subsequent thermally induced mechanical deformation of the supporting material on the OR table. Certainly, user errors cannot be 100% excluded. Therefore, the experiments with 7 and 9 registration fiducials were eliminated from further analysis.

No problematic issues were detected for the data of the plastic skull.

All outliers that found were clearly detectable visually. These outliers were also detected by an outlier detection algorithm [28]. The algorithm found some more outliers, and clearly this is depending on the defined threshold. If one fiducial in one experiment was detected as an outlier, the whole repetition had to be removed. For the plastic skull repetitions 2, 2, 5, and 5 had to be removed for the 3, 5, 7, and 9 fiducial experiments, respectively. For the cadaver repetitions 2, 3, 1, and 5 for the 3, 5, 7, and 9 experiments, respectively, had to be removed.

For the volunteer parts of 7- and 9-point registrations had to be removed. For the 7-point registration only two repetitions and for the 9 points registration only 4 repetitions remained; thus, no statistically relevant result could be expected, and for this reason, the complete series was removed.

This generated a rather limited set of measurements, and so only the “obvious” outliers were excluded. The removal of outliers did not have significant influence on the data, see Tables 1 and 2.

In contrast to prior analysis [16], this work has used a reduced dataset without systematic errors. Therefore, a total data number of 240, 234, and 80 fiducials were available for the plastic skull, the anatomic specimen, and the volunteer, respectively.

Due to the rather small size of the dataset, a robust covariance matrix estimator for the FLE had to be used [29]; results were compared to the standardly implemented non-robust approach.

Contrary to [16] this work has used  $FRE_{iso,i}$  (for the current experiment  $i$ ) instead of the expected value of  $\langle FRE_{iso,i}^2 \rangle$  to estimate  $\langle FLE_{iso,est,i}^2 \rangle$ . Moreover, the FLE of probe calibration ( $FLE_{probe\_calib}$ ) was corrected from  $0.18^2 \text{ mm}^2$ , as used earlier [16], to  $0.36^2 \text{ mm}^2$ .

Summarizing, this is expected to provide a fairly comprehensive analysis of the application accuracy in rigid-body registration for computer-assisted surgery systems.

## Statistics

As from [30] the predicted TRE is influenced by the TLE, which leads to a larger (TTE) prediction error:

$$\langle TTE_{F,x}^2 \rangle = \langle TRE_{F,x}^2 \rangle + \langle TLE^2 \rangle. \quad (10)$$

Equation (10) assumes that  $TRE_{F,x}$  and TLE are uncorrelated, which is considered “likely” to be true in [30]. A significant correlation between  $TRE_{F,x}$  and TLE in “real” world would suggest that  $TTE_{F,x}$  as defined in (10) might not be useful for real intraoperative use/experiments. The correlation of the before-mentioned quantities was investigated with Kendall’s  $\tau$  and Spearman’s correlation coefficient.

Normality of the distributions of measured data was statistically tested with a one-sample Kolmogorov–Smirnov test. The equality (non-equality) of the different distribution pairs was tested with a two-sample Kolmogorov–Smirnov test.

The two-sided Wilcoxon signed rank test for zero median was applied to test for statistically significant differences between predictions and measurements. The null hypothesis for this test was  $H_0$ : the median of  $M - P = 0$ , and the alternative hypothesis was  $H_1$ : the median of  $M - P \neq 0$ , with  $M =$  measurement and  $P =$  prediction. From a clinical perspective it is senseful to provide an upper limit for the TRE, since overestimating the uncertainty in the application accuracy provides a larger safety margin to surgeons intraoperatively. The one-sided Wilcoxon signed rank test was used to test whether the predictor overestimates the “real” measured error (with  $H_0$ : the median of  $M - P = 0$ ;  $H_1$ : the median of  $M - P < 0$ ).

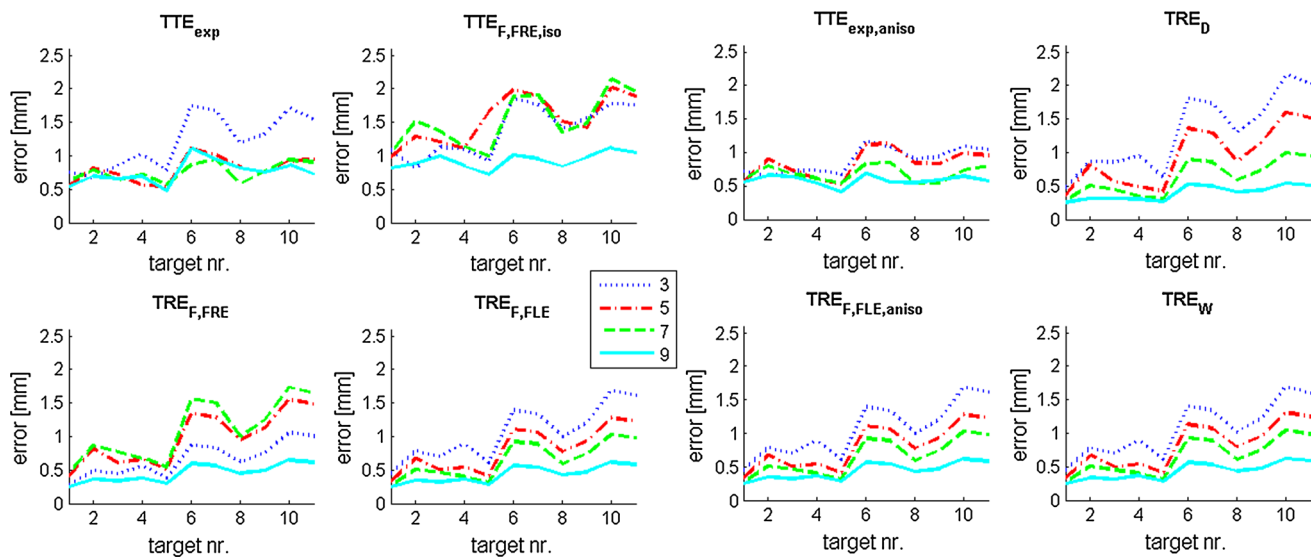
Throughout the analysis, the level of significance was 0.05.

## Numerical simulation of TRE prediction and measurements

A numerical simulation of the experiment might give information concerning eventual correlation of TRE and TLE random variables and how localization errors affect target errors. Two different experiments were made: an independent (unpaired) and a dependent (paired) one: Predicting the TRE with FLE led to an independent experiment, because all repetitions were used for the estimation of FLE (compared to the real experiments). On the other hand, the experiment is dependent if the FRE was used for the prediction of the TRE, because the very same samples were used both for measurement and for prediction.

The following simulation was repeated 100,000 times with 3, 5, 7, and 9 fiducials, respectively:

- (a) Creation of registration fiducials: Draw  $N$  3-dimensional random patient fiducial points  $X_i$ ,  $i = 1, \dots, N$ , in patient space inside a cube with an edge length of 200 mm. These are the true patient fiducials.  $N = 3, 5, 7, \text{ or } 9$ .
- (b) Create a random rotation matrix  $R_{\text{rand}}$  and apply to the  $X_i$  to yield  $N$  true image points  $Y_i$ .



**Fig. 2** Plastic skull. Mean TREs of anisotropic (right) and isotropic (left) registration. Three (blue dotted line), 5 (red chain line), 7 (green dashed line), and 9 (cyan solid line) fiducials were used for registration.  $TTE_{exp}$  was measured, and the different TREs were calculated. This was

- (c) Select a specific localization error in patient and image space,  $FLE_{sim,pat} = 1/3$  mm and  $FLE_{sim,img} = 0.0001$  mm, combined to  $FLE_{sim}^2 = FLE_{sim,img}^2 + FLE_{sim,pat}^2$ .
- (d) Perturb  $X_i$  with  $\Delta x$ , a zero-mean Gaussian noise with standard deviation  $FLE_{sim,pat}$  in all directions and  $Y_i$  with  $\Delta y$ , a zero-mean Gaussian noise with standard deviation  $FLE_{sim,img}$  in all directions, so that  $X'_i = X_i + \Delta x$ , and  $Y'_i = Y_i + \Delta y$ .
- (e) Register the  $X'_i$  to  $Y'_i$  to get rotation  $R_{sim}$ , translation  $t_{sim}$ , and  $FRE_{sim}$ .
- (f) Create one “true” random target patient point  $r$  inside the cube and transform it to image space with  $R_{rand} * r = q$ .
- (g) Generate  $M$  perturbed random target points  $r_j$  with mean( $r_j$ ) =  $r$  and std( $r_j$ ) =  $FLE_{sim,pat}$ ,  $j = 1, \dots, M$ ;  $M = 100,000$ . Transform  $r_j$  into image space ( $q_j = R_{sim} * r_j + t_{sim}$ ) and calculate the measured TRE for the  $M$  points,  $TTE_{sim}(q_j) = \|q - q_j\|$ .
- (h) Calculate  $TRE_{sim,F,FRE}$  for  $q$  and  $q_i$  (see Sect. 2.3b).

For the independent experiment, all steps (except h) are repeated again and  $TRE_{sim,F,FLE}$  is calculated (see Sect. 2.3a).

The simulation was repeated 10 times to calculate mean and standard deviation of all errors.

The distributions of  $TRE_{sim,F,FLE}$ ,  $TRE_{sim,F,FRE}$ , and  $TTE_{sim}$  were analyzed. Correlations between the errors were tested with Pearson’s correlation coefficient. Equality or difference of prediction and measurement was tested using a Wilcoxon signed rank test (in case of paired samples) and a

repeated 10 times. The mean of the 10 repetitions was calculated. For a clear view, no standard deviation is plotted. Using more fiducials for registration a decrease in  $TRE_{F,FLE}$ ,  $TRE_{F,FLE}$ ,  $TRE_D$ , and  $TRE_W$  can be observed

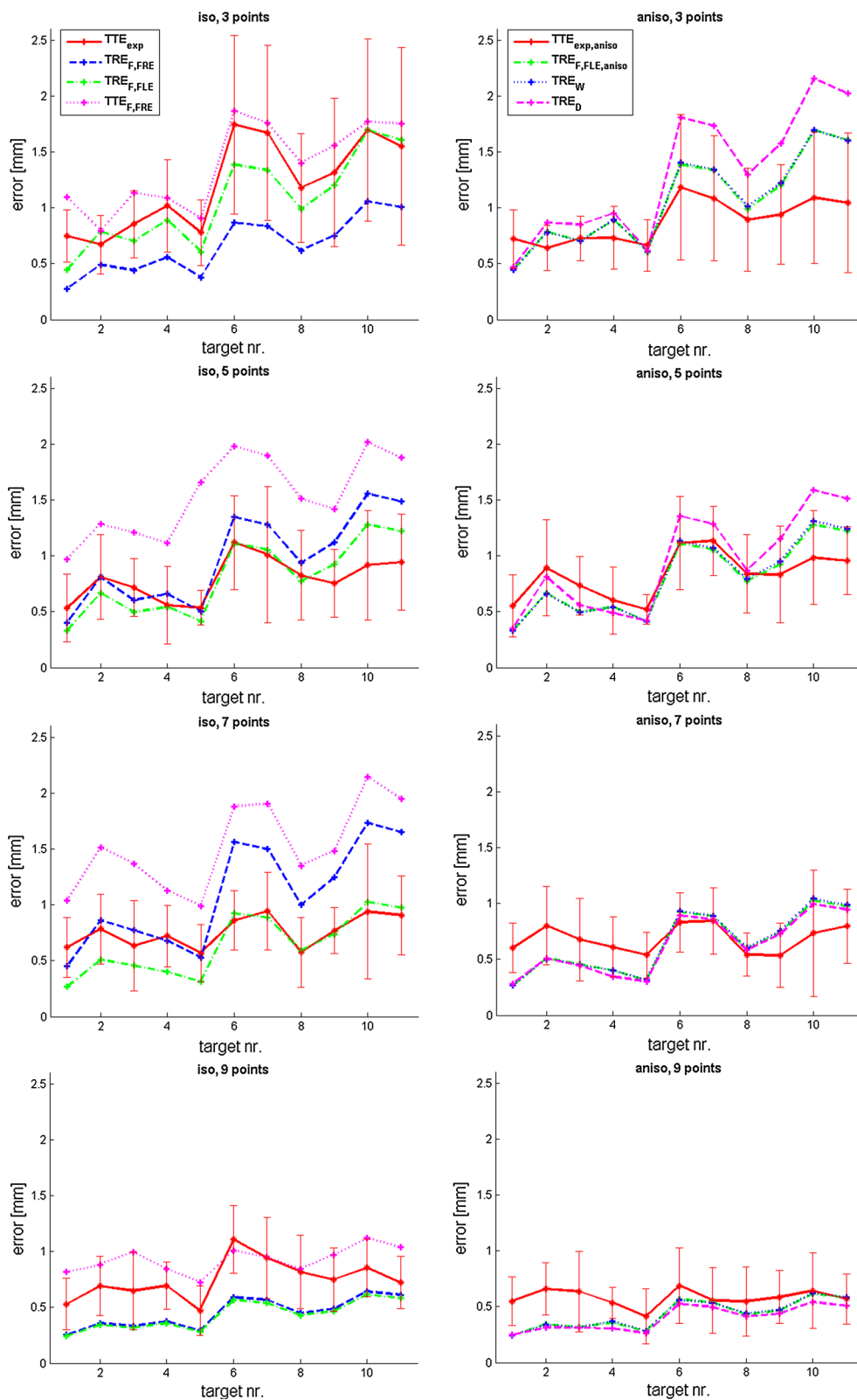
Wilcoxon rank sum test (in case of unpaired samples). Power and effect sizes of the experiments were evaluated.

## Results

The results for all patients with isotropic registration and prediction are presented on the left half of Figs. 2, 3, 4, 5, 6, and 7. The right half of Figs. 2, 3, 4, 5, 6, and 7 shows the results for anisotropic registration and predictions. For each target the mean measured and mean predicted error is visualized for the registrations with 3, 5, 7, and 9 fiducials. Tables 1, 2, 3, 4, 5, and 6 show the measured and predicted TREs and TTEs for all objects studied.

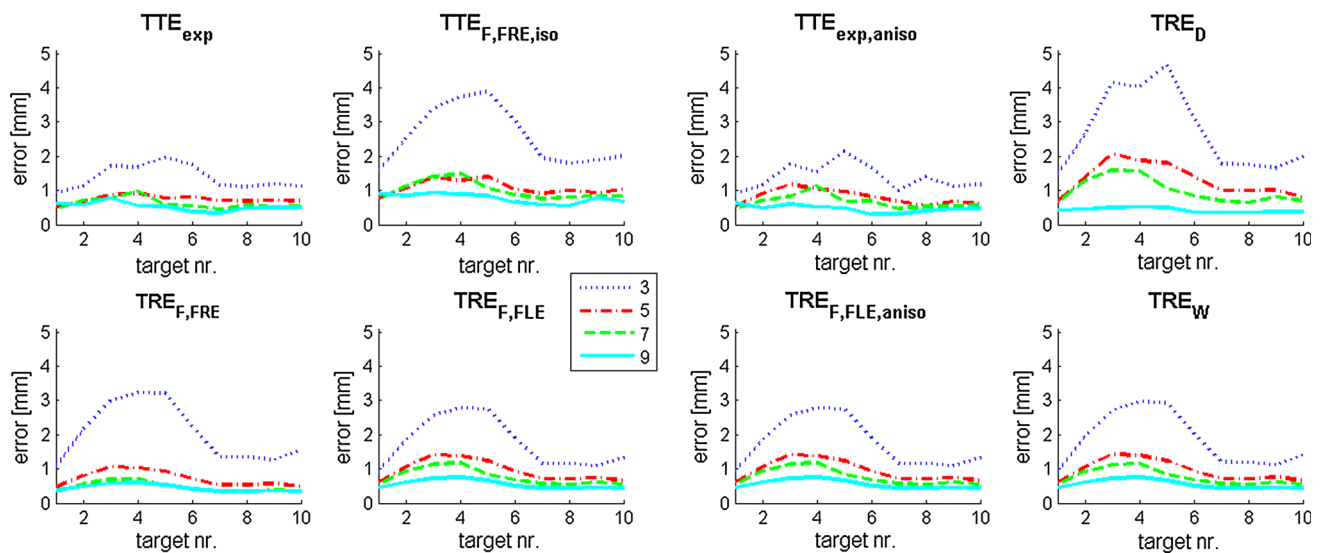
A robust estimation of the covariance matrix led to smaller TREs (Tables 4, 6). If the outliers are not included, and a robust estimation is used, the TRE gets larger again (Tables 1, 2). For the anatomic specimen robust approach gives an overall improvement, where more predictions equal the measurements in terms of statistical equivalence (see Table 8). We focus on the results of non-robust covariance matrix estimations. The detailed results with robust estimation can be seen in the mentioned tables. Table 10 shows  $ULE_{exp}$  and  $ULE_F$  as determined from the data. Table 11 shows the FLE as determined over all registrations and fiducials. Tables 7, 8, and 9 give the total number of targets, where equality or overestimation of the prediction can be statistically confirmed for all patients. The results of the numerical simulation are reported in Table 12.

Detailed remarks on the data:



**Fig. 3** Plastic skull. Mean TRE results of 3, 5, 7, and 9 fiducial arrangements (from top to bottom) and 10 repetitions. Standard deviation of  $TTE_{exp}$  (red solid line) is shown; it can be observed that most of the predicted TREs are lying within  $TTE_{exp} \pm$  standard deviation. Isotropic

registration on the left, anisotropic registration on the right side. Differences between anisotropic and isotropic registration can be observed, but also the similarity of predictions and measurements



**Fig. 4** Anatomic specimen. Mean TREs of anisotropic (*right*) and isotropic (*left*) registration. Three (*blue dotted line*), 5 (*red chain line*), 7 (*green dashed line*), and 9 (*cyan solid line*) fiducials were used for registration.  $TTE_{exp}$  was measured, and the different TREs were calculated. This was repeated 10 times. The mean of the 10 repetitions was calculated.

**Plastic skull** The lowest application error,  $TTE_{exp}$ , could be achieved with a registration with 9 fiducials; with anisotropic registration  $TTE_{exp}$  was improving.

Table 7 shows good correspondence of experiments and predictions. Regarding isotropic registration,  $TRE_{F,FLE}$  and  $TTE_{F,FRE}$  gave the most similar results for  $TTE_{exp}$  for 3 registration points. (with the statistical power  $\leq 0.61$  for 3-fiducial registration,  $\leq 0.94$  for 5 fiducials,  $\leq 0.24$  for 7 fiducials, and  $\leq 0.00001$  for 9 fiducials for  $TRE_{F,FLE}$ ). All targets were overestimated with  $TTE_{F,FRE}$  using 5 and 7 registration points. Regarding anisotropic registration  $TRE_{F,FLE,aniso}$  and  $TRE_W$  were predicting  $TTE_{exp,aniso}$  in about 70%.  $TTE_{F,FRE,aniso}$  and  $TRE_{F,FLE,aniso}$  were overestimating  $TTE_{exp}$  for all 11 targets in all experiments.

**Anatomic specimen** Predictions of target errors for different registration alternatives overestimated the measurements; this can clearly be seen in Figs. 4 and 5.

In case of isotropic registration  $TRE_{F,FLE}$  and  $TRE_{F,FRE}$  were predicting  $TTE_{exp}$  almost always (statistical power  $\leq 0.99$  for 3-fiducial registration,  $\leq 1$  for 5 fiducials,  $\leq 0.98$  for 7 fiducials, and  $\leq 0.99$  for 9 fiducials for  $TRE_{F,FLE}$ ).  $TTE_{F,FRE}$  was overestimating  $TTE_{exp}$  for most of the targets.

In the anisotropic case, for  $TRE_W$  and  $TRE_{F,FLE,aniso}$  equality to  $TTE_{exp,aniso}$  could be confirmed statistically in 82% of the targets. For all targets  $TRE_{F,FRE,aniso}$  overestimated  $TTE_{exp,aniso}$ .

The results for the *volunteer* (Figs. 6, 7) show that isotropic  $TRE_{F,FLE}$  was the best prediction method for  $TTE_{exp}$ ; most

overestimations were given by  $TRE_{F,FRE}$  and  $TTE_{F,FRE}$ . The statistical power for  $TRE_{F,FLE}$  is  $\leq 0.91$  for the 3-fiducial registration and  $\leq 0.99$  for the 5-fiducial registration.

With anisotropic registration  $TRE_{F,FRE,aniso}$  was equal to  $TTE_{exp,aniso}$  for 8 out of 10 targets, with 3 registration points.  $TTE_{F,FRE,aniso}$  overestimated  $TTE_{exp,aniso}$  with 3 fiducials only; using 5 fiducials no prediction method gave satisfying results.

The correlation between  $TRE_{F,FLE}$  and TLE was always larger than between  $TRE_{F,FRE}$  and TLE (Table 13), except in the case of the volunteer, where the correlation between  $TRE_{F,FRE}$  and TLE reached 0.47.

Almost all target errors, measured and predicted, were not normally distributed.

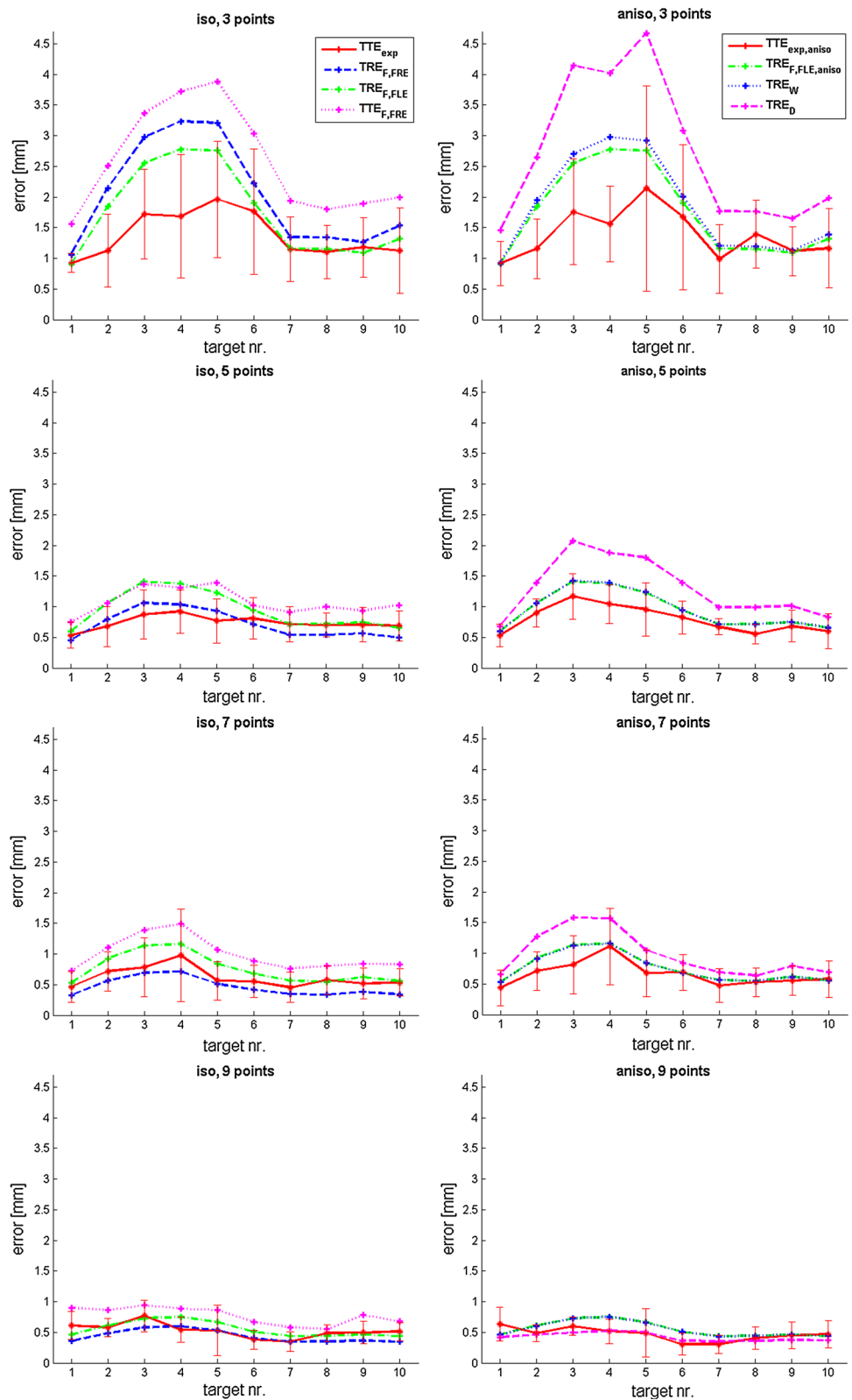
The results of  $ULE_F$  and  $ULE_{exp}$  are similar to each other for the anatomic specimen, but not for the plastic skull and the volunteer. The plastic skull with Ti-screws had the smallest ULE ( $ULE_{exp} = 0.4$  mm), while the volunteer, using anatomic landmarks only, had the largest ULE ( $ULE_{exp} = 1.6$  mm).

As a result of the numerical simulation it can be seen that the mean of the measured  $TTE_{sim}$  is always similar to the mean of the predicted  $TRE_{sim,F,FRE}$ ;  $TRE_{sim,F,FLE}$  is the largest error (Table 12). The smallest measured and predicted errors could be achieved using 9 fiducials, which is in agreement with earlier experiments and theory. The mean of  $TRE_{sim,F,FRE}$  of all perturbed targets equals  $TRE_{sim,F,FRE}$  on the true target (this is also valid for  $TRE_{sim,F,FLE}$ ).

No correlation could be found between  $FRE_{sim}$  and  $TRE_{rmsim,F,FLE}$  and between  $TTE_{sim}$  and  $TRE_{sim,F,FLE}$  and

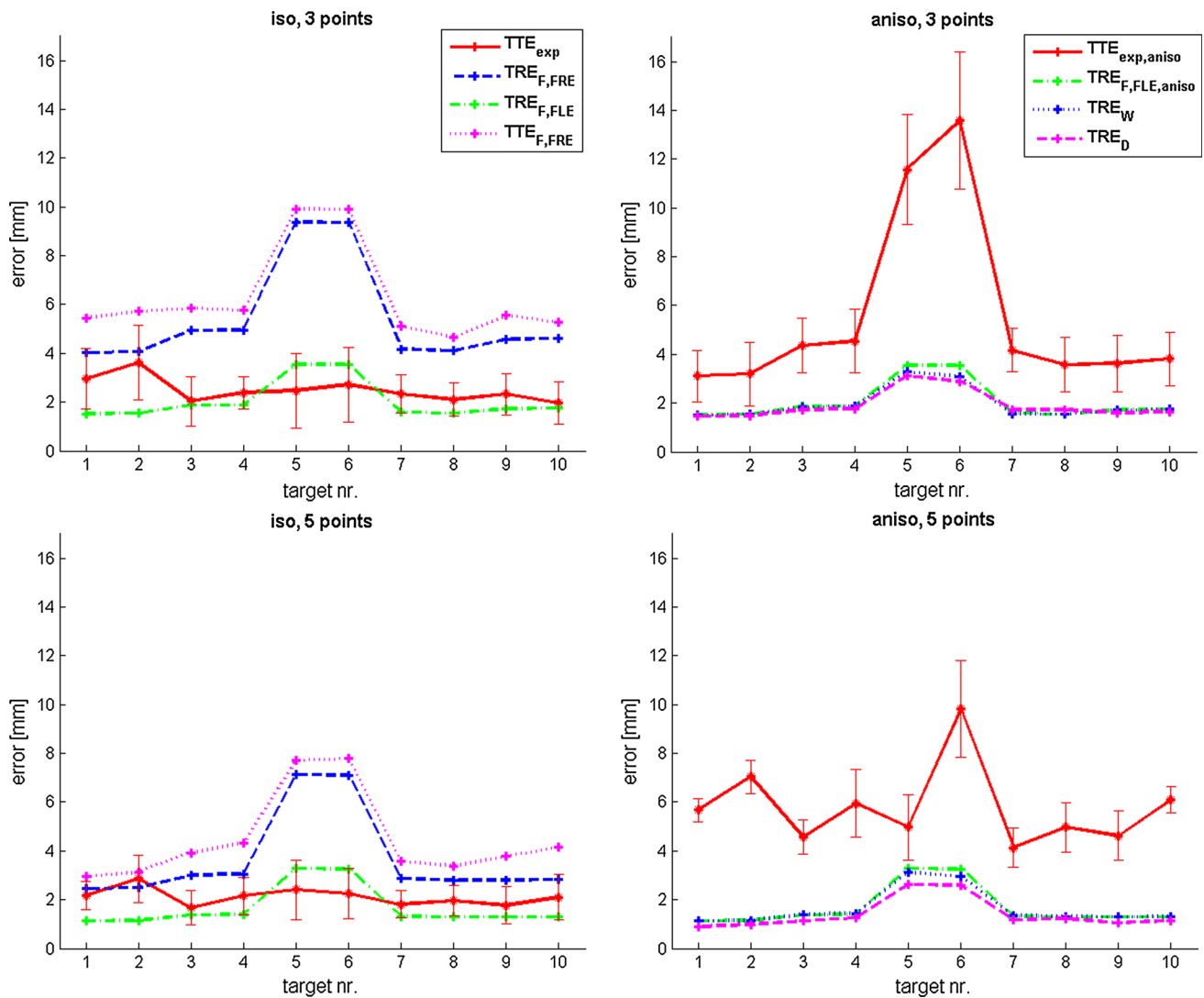


**Fig. 5** Anatomic specimen. Mean TRE results of 10 repetitions of 3, 5, 7, and 9 fiducial arrangements (from top to bottom). Standard deviation of  $TTE_{exp}$  (red solid line) is shown; it can be observed that most of the predicted TREs are lying within  $TTE_{exp} \pm$  standard deviation. Isotropic registration on the left, anisotropic registration on the right side. Differences between anisotropic and isotropic registration can be observed, but also the similarity of predictions and measurements



$TRE_{sim,F,FRE}$ , respectively (the mean correlation coefficient is always smaller  $\pm 0.02 \pm 0.00$ ). The correlation of  $TRE_{sim,F,FRE}$ , and  $FRE_{sim}$  was always 1.

Visual inspection showed that none of the errors were normally distributed. Statistical testing with the Kolmogorov–Smirnov method, the  $H_0$  hypothesis (that the error is normally



**Fig. 6** Volunteer. Mean TREs of anisotropic (*right*) and isotropic (*left*) registration. Three (*blue dotted line*) and 5 (*red chain line*) fiducials were used for registration.  $TTE_{exp}$  was measured, and the different TREs were calculated. This was repeated 10 times. The mean of the 10 repetitions was calculated. For a clear view, no standard deviation is plotted. With

a 5-point registration, the target error is smaller than with a 3 points registration for  $TRE_{F,FLE}$ ,  $TRE_{F,FLE}$ ,  $TRE_D$ , and  $TRE_W$ . The predictions lead to larger errors than the measurements with isotropic registration, different to anisotropic registration

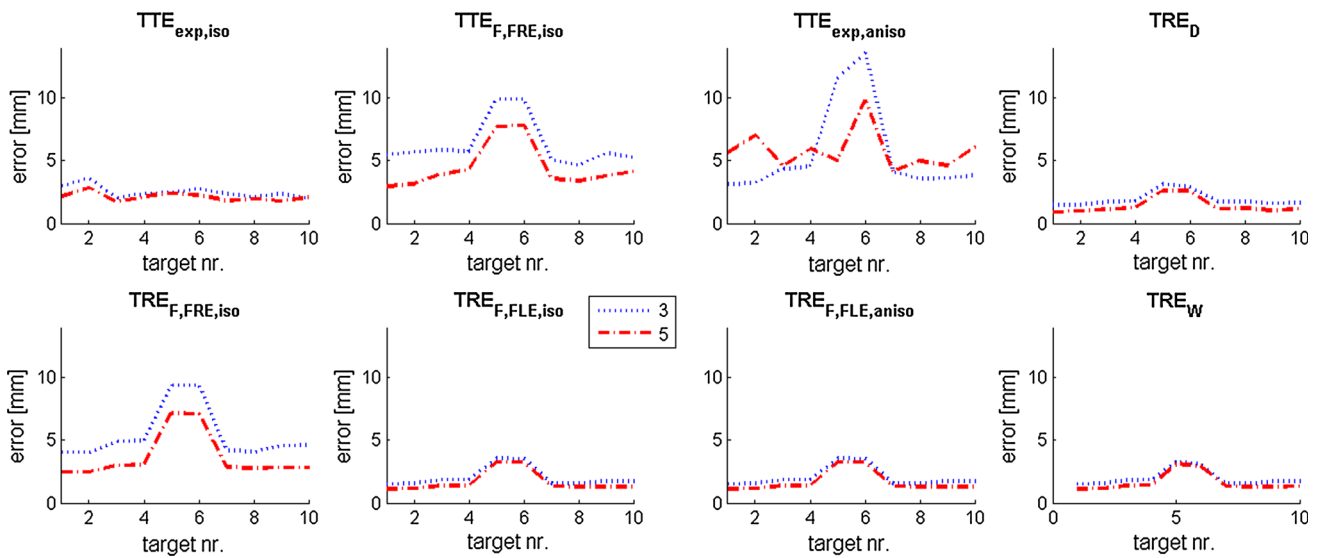
distributed) had to be rejected for all errors at the 5% significance level. Since the measured errors are Euclidean distances of normally distributed points, their distribution is expected to resemble the Maxwell distribution [31,32]. Figure 8, a plot of the pdf of the errors, shows this. The power of the numerical simulation is 1, with a small effect size. The difference of the distributions of measured and predicted errors could always be statistically confirmed. The distributions of the independent TREs were the same as in the dependent situation.

## Discussion

For rigid-body registration in clinical navigation, a complete and detailed analysis of anisotropic and isotropic prediction

methods for TRE is presented. Two major groups were distinguished based on isotropy and anisotropy. For the isotropic case, isotropic registration and isotropic prediction methods were used to measure and predict the TRE. The anisotropic case handled anisotropic registration with anisotropic prediction methods; the most widely used prediction method ( $TRE_{F,x,aniso}$ ) was added, though it is, strictly speaking, defined for isotropic FLE only [6].

The prediction methods studied gave a good estimation of the application error in the surgical environment for the plastic skull and the anatomic specimen. A two-sided test was used to statistically compare predicted and measured target errors. The one-sided test might be a better approach for predicting surgically relevant application accuracy (TRE),



**Fig. 7** Volunteer. Mean TRE results of 10 repetitions of the 3 and 5 fiducial arrangements (from top to bottom). Isotropic registration on the left, anisotropic registration on the right side. Standard deviation of  $TTE_{exp}$  (red solid line) is shown; it can be observed that only

$TRE_{F,FLE,iso}$  is lying within  $TTE_{exp} \pm$  standard deviation when using isotropic registration. For anisotropic registration  $TTE_{exp,aniso}$  is much higher than all the predicted TREs

**Table 1** Isotropic registration

Mean target error	$TTE_{exp}$ (mm)	$TRE_{F,FRE}$ (mm)	$TRE_{F,FLE}$ (mm)	$TTE_{F,FRE}$ (mm)	$TTE_{F,FLE}$ (mm)
Plastic skull	$0.90 \pm 0.31$	$0.63 \pm 0.24$	$0.56 \pm 0.29$	$1.24 \pm 0.35$	$1.21 \pm 0.39$
Anatomic specimen	$0.79 \pm 0.39$	$0.93 \pm 0.80$	$0.87 \pm 0.53$	$1.33 \pm 0.88$	$1.28 \pm 0.66$
Volunteer	$2.30 \pm 0.46$	$4.52 \pm 2.12$	$1.78 \pm 0.79$	$5.38 \pm 2.03$	$3.35 \pm 0.83$

All outliers found with the MCD algorithm [29] are removed (see Sect. 2.6). FLE is calculated via the robust estimation matrix. Results of experimental and predicted TREs. The mean value  $\pm$  standard deviation over all registration is given in mm

**Table 2** Anisotropic registration

Mean target error	$TTE_{exp,aniso}$ (mm)	$TRE_{F,FRE,aniso}$ (mm)	$TRE_{F,FLE,aniso}$ (mm)	$TRE_D$ (mm)	$TRE_W$ (mm)	$TTE_{F,FRE,aniso}$ (mm)	$TTE_{F,FLE,aniso}$ (mm)
Plastic skull	$0.84 \pm 0.24$	$2.58 \pm 1.27$	$0.56 \pm 0.29$	$0.75 \pm 0.47$	$0.59 \pm 0.31$	$2.84 \pm 1.19$	$1.21 \pm 0.40$
Anatomic specimen	$0.86 \pm 0.39$	$2.34 \pm 1.45$	$0.87 \pm 0.53$	$1.25 \pm 0.94$	$0.87 \pm 0.52$	$2.53 \pm 1.50$	$1.28 \pm 0.66$
Volunteer	$5.67 \pm 2.83$	$4.59 \pm 2.01$	$1.78 \pm 0.79$	$1.49 \pm 0.58$	$1.72 \pm 0.66$	$5.43 \pm 1.92$	$3.35 \pm 0.83$

All outliers found with the MCD algorithm [29] are removed (see Sect. 2.6). FLE is calculated via the robust estimation matrix. Results of experimental and predicted TREs. The mean value  $\pm$  standard deviation over all registration is given in mm

**Table 3** Isotropic registration

Mean target error	$TTE_{exp}$ (mm)	$TRE_{F,FRE}$ (mm)	$TRE_{F,FLE}$ (mm)	$TTE_{F,FRE}$ (mm)	$TTE_{F,FLE}$ (mm)
Plastic skull	$0.88 \pm 0.31$	$0.80 \pm 0.42$	$0.74 \pm 0.38$	$1.34 \pm 0.41$	$1.30 \pm 0.40$
Anatomic specimen	$0.81 \pm 0.40$	$0.91 \pm 0.79$	$0.99 \pm 0.61$	$1.35 \pm 0.85$	$1.40 \pm 0.72$
Volunteer	$2.30 \pm 0.46$	$4.52 \pm 2.12$	$1.86 \pm 0.82$	$5.38 \pm 2.03$	$3.40 \pm 0.85$

Results of experimental and predicted TREs. The mean value  $\pm$  standard deviation over all registrations is given in mm (3-, 5-, 7-, 9-point registration for plastic skull and anatomic specimen and 3- and 5-point registration for the volunteer). Only extreme and visible outliers are removed

because predictions should indicate a lower limit for the TRE on a specific target, rather than underestimate the real error. Underestimations can be very critical for patients in a real

intervention. For most of the targets, however, a good estimation or overestimation of the TRE was found; equality

**Table 4** Isotropic registration

Mean target error	TRE <sub>F,FLE</sub> (mm)	TTE <sub>F,FLE</sub> (mm)
Plastic skull	0.58 ± 0.30	1.21 ± 0.35
Anatomic specimen	0.81 ± 0.49	1.33 ± 0.73
Volunteer	1.78 ± 0.79	3.35 ± 0.83

FLE estimated with a robust covariance matrix. Results of TRE<sub>F,FLE</sub> and TTE<sub>FLE</sub>, all other TREs do not change, compared to Table 3. The mean value ± standard deviation over all registrations (3, 5, 7, 9 points for plastic skull and anatomic specimen and 3 and 5 points for the volunteer) is given in mm

**Table 5** Anisotropic registration

Mean target error	TTE <sub>exp,aniso</sub> (mm)	TRE <sub>F,FRE,aniso</sub> (mm)	TRE <sub>F,FLE,aniso</sub> (mm)	TRE <sub>D</sub> (mm)	TRE <sub>W</sub> (mm)	TTE <sub>F,FRE,aniso</sub> (mm)	TTE <sub>F,FLE,aniso</sub> (mm)
Plastic skull	0.75 ± 0.20	2.57 ± 1.15	0.74 ± 0.38	0.82 ± 0.52	0.74 ± 0.38	2.80 ± 1.12	1.30 ± 0.40
Anatomic specimen	0.83 ± 0.42	2.12 ± 1.55	0.99 ± 0.61	1.36 ± 1.07	1.02 ± 0.66	2.35 ± 1.58	1.40 ± 0.72
Volunteer	5.67 ± 2.83	4.60 ± 2.01	1.86 ± 0.82	1.66 ± 0.65	1.81 ± 0.70	5.44 ± 1.92	3.40 ± 0.85

Results of experimental and predicted TREs. The mean value ± standard deviation over all registrations is given in mm (3-, 5-, 7-, 9-point registration for plastic skull and anatomic specimen, 3- and 5-point registration for the volunteer). Only extreme and visible outliers are removed

**Table 6** Anisotropic registration

Mean target error	TRE <sub>F,FLE,aniso</sub> (mm)	TRE <sub>D</sub> (mm)	TRE <sub>W</sub> (mm)	TTE <sub>F,FLE,aniso</sub> (mm)
Plastic skull	0.58 ± 0.30	0.72 ± 0.45	0.60 ± 0.32	1.21 ± 0.35
Anatomic specimen	0.81 ± 0.49	0.88 ± 0.54	0.82 ± 0.48	1.26 ± 0.62
Volunteer	1.78 ± 0.79	1.49 ± 0.58	1.72 ± 0.66	3.35 ± 0.83

FLE estimated with a robust covariance matrix. Results of TRE<sub>F,FLE,aniso</sub>, TTE<sub>F,FLE,aniso</sub>, TRE<sub>W</sub>, and TRE<sub>D</sub>, all other TREs do not change, compared to Table 5. The mean value ± standard deviation over all registrations (3, 5, 7, 9 points for plastic skull and anatomic specimen and 3 and 5 points for the volunteer) is given in mm

**Table 7** Results of the statistical tests for the plastic skull

No. of fiducials	3		5		7		9		Total number of targets										
	M = P	M < P	M = P	M < P	M = P	M < P	M = P	M < P	Not robust		Robust								
	M = P	M < P	M = P	M < P	M = P	M < P	M = P	M < P	M = P	M < P	M = P	M < P							
<i>Plastic skull</i>																			
TRE <sub>F,FLE</sub>	9	4	0	8	2	0	7	6	0	1	0	0	25	2	18	0			
TRE <sub>F,FRE</sub>	1	0	8	7	3	5	6	1	0	15	9								
TTE <sub>F,FRE</sub>	9	2	0	11	0	11	4	8	13	32									
TTE <sub>F,FLE</sub>	7	8	7	5	0	11	0	1	11	4	7	8	6	11	37	16	33		
TRE <sub>F,FLE,aniso</sub>	8	2	0	8	5	2	0	7	6	2	7	5	0	30	6	24	2		
TRE <sub>F,FRE,aniso</sub>	0	11	0	11	0	11	0	11	0	11	0	11	0	0	44				
TTE <sub>F,FRE,aniso</sub>	0	11	0	11	0	11	0	11	0	11	0	11	0	0	44				
TTE <sub>F,FLE,aniso</sub>	0	11	0	2	11	10	0	11	0	1	11	0	11	0	44	3	43		
TRE <sub>D</sub>	3	8	8	3	7	6	4	3	7	6	2	0	7	5	0	24	14	25	6
TRE <sub>W</sub>	8	2	0	8	6	2	0	7	6	2	0	7	5	0	30	6	25	0	

Equality of measurement and prediction: H<sub>0</sub>: The median of M - P = 0, H<sub>1</sub>: M - P ≠ 0, where M is the measurement and P is the prediction. Overestimation of the measurement: H<sub>0</sub>: The median of M - P = 0, H<sub>1</sub>: M - P < 0. The number of targets where H<sub>0</sub> cannot be rejected (M = P) or has to be rejected (M < P) is reported. TRE<sub>D</sub> and TRE<sub>W</sub> were used with anisotropic registration only. The last column shows the total number of targets for 3-, 5-, 7-, and 9-point registration for each prediction method where M = P and M < P. In each column the numbers on the right show the results of the statistical testing, if a robust covariance matrix is used for calculating FLE, calculated with all fiducials defined in the data analysis section (240 for the plastic skull). Only the number of targets where the result has changed is reported

**Table 8** Results of the statistical testing for the anatomic specimen

No. of fiducials	3		5		7		9		Total number of targets											
	$M = P$		$M < P$		$M = P$		$M < P$		Not robust		Robust									
	$M = P$	$M < P$	$M = P$	$M < P$	$M = P$	$M < P$	$M = P$	$M < P$	$M = P$	$M < P$	$M = P$	$M < P$								
<i>Anatomic specimen</i>																				
TRE <sub>F,FLE</sub>	7	9	3	0	6	9	4	2	9	10	3	0	7	6	2	0	29	12	34	2
TRE <sub>F,FRE</sub>	6		5	4	9		0		8		0		6		0		29	5		4
TTE <sub>F,FRE</sub>	0	3	10	7	1	10	1	10	1	10	1	9	3	39	6	36				
TTE <sub>F,FLE</sub>	0		10		0	10	0	10	0	1	10	9	0	40	1	39				
TRE <sub>F,FLE,aniso</sub>	8	9	3	1	7	10	5	0	10	2	0	5	9	5	0	30	15	38	1	
TRE <sub>F,FRE,aniso</sub>	0		10		0	10	0	10	0	10	0	10	0	40		40				
TTE <sub>F,FRE,aniso</sub>	0		10		0	10	0	10	0	10	0	10	0	40		40				
TTE <sub>F,FLE,aniso</sub>	2	4	9	8	0	10	0	1	10	0	10	0	10	2	39	5	38			
TRE <sub>D</sub>	1	9	9	1	0	5	10	7	4	10	7	2	10	5	0	15	26	29	10	
TRE <sub>W</sub>	7	9	3	1	7	10	5	0	10	1	0	6	9	5	0	30	14	38	1	

Equality of measurement and prediction:  $H_0$ : The median of  $M - P = 0$ ,  $H_1$ :  $M - P \neq 0$ , where  $M$  is the measurement and  $P$  is the prediction. Overestimation of the measurement:  $H_0$ : The median of  $M - P = 0$ .  $H_1$ :  $M - P < 0$ . The number of targets where  $H_0$  cannot be rejected ( $M = P$ ) or has to be rejected ( $M < P$ ) is reported. TRE<sub>D</sub> and TRE<sub>W</sub> were used for anisotropic registration only. The last column shows the total number of targets, where  $M = P$  and  $M < P$ , for 3-, 5-, 7-, and 9-point registration for each prediction method. In each column the numbers on the right show the results of the statistical testing, if a robust covariance matrix is used, calculated with all fiducials defined in the data analysis section (234 for the anatomic specimen). Only the number of targets where the result has changed is reported

**Table 9** Results of the statistical testing for the volunteer

No. of fiducials	3		5		Total number of targets							
	$M = P$		$M < P$		Not robust		Robust					
	$M = P$	$M < P$	$M = P$	$M < P$	$M = P$	$M < P$	$M = P$	$M < P$				
<i>Volunteer</i>												
TRE <sub>F,FLE</sub>	6	5	1	1	9	3	2	1	9	7	3	1
TRE <sub>F,FRE</sub>	1		9		3		9		4		18	
TTE <sub>F,FRE</sub>	0		10		1		9		1		19	
TTE <sub>F,FLE</sub>	1		9		2		8		3		17	
TRE <sub>F,FLE,aniso</sub>	0		0		0		0		0		0	
TRE <sub>F,FRE,aniso</sub>	8		0		0		1		8		1	
TTE <sub>F,FRE,aniso</sub>	1		8		4		1		2		9	
TTE <sub>F,FLE,aniso</sub>	3		2		1		0		4		2	
TRE <sub>D</sub>	0		0		0		0		0		0	
TRE <sub>W</sub>	0		0		0		0		0		0	

Equality of measurement and prediction:  $H_0$ : The median of  $M - P = 0$ ,  $H_1$ :  $M - P \neq 0$ , where  $M$  is the measurement and  $P$  is the prediction. Overestimation of the measurement:  $H_0$ : The median of  $M - P = 0$ .  $H_1$ :  $M - P < 0$ . The number of targets where  $H_0$  cannot be rejected ( $M = P$ ) or has to be rejected ( $M < P$ ) is reported. TRE<sub>D</sub> and TRE<sub>W</sub> were used with anisotropic registration only. The last column shows the total number of targets, where  $M = P$  and  $M < P$ , for 3 and 5 fiducials registration for each prediction method. In each column the numbers on the right show the results, if a robust covariance matrix is used, calculated with all fiducials defined in the data analysis section, 80 for the volunteer. Differences between robust and not robust covariance matrix can only be observed calculating TRE<sub>F,FLE</sub>

**Table 10** Mean values of the ULEs of all patients, calculated with Eqs. (8) and (9)

Mean ULE	ULE <sub>exp</sub> (Eq. 8) (mm)	ULE <sub>F</sub> (Eq. 9) (mm)
Plastic skull	0.38	0.80 ± 0.23
Anatomic specimen	0.98	0.62 ± 0.21
Volunteer	1.61	3.92 ± 4.60

All values are given in mm

**Table 11** Experimental fiducial localization errors

Mean TFLE	TFLE (mm)	FLE <sub>img</sub> (mm)	FLE <sub>pat</sub> (mm)
Plastic skull	0.64	0.31	0.55
Anatomic specimen	0.85	0.61	0.55
Volunteer	2.23	1.16	1.85

Mean TFLE, FLE<sub>img</sub>, and FLE<sub>pat</sub> over all registrations and fiducials. The differences between anatomic landmarks (volunteer) and Ti-screws (plastic skull and anatomic specimen) can be well observed. All values are given in mm

**Table 12** Results of an example of a numerical simulation

Mean $\pm$ std	TTE <sub>sim</sub>	TRE <sub>sim,F,FRE</sub>	TRE <sub>sim,F,FLE</sub>
3 points	2.09 $\pm$ 1.96	2.26 $\pm$ 1.76	2.46 $\pm$ 1.48
5 points	0.73 $\pm$ 0.50	0.80 $\pm$ 0.37	0.83 $\pm$ 0.31
7 points	0.64 $\pm$ 0.37	0.70 $\pm$ 0.22	0.72 $\pm$ 0.18
9 points	0.44 $\pm$ 0.27	0.48 $\pm$ 0.19	0.49 $\pm$ 0.17

The mean of 100,000 error measurements, repeated 10 times, was calculated for 3-, 5-, 7-, and 9-point registration. The errors are getting smaller, the more the fiducials are used for registration, as expected

or that the prediction was an upper limit can be statistically confirmed.

Comparing true and reference targets (compared to [16]) showed that better agreement of measured and predicted

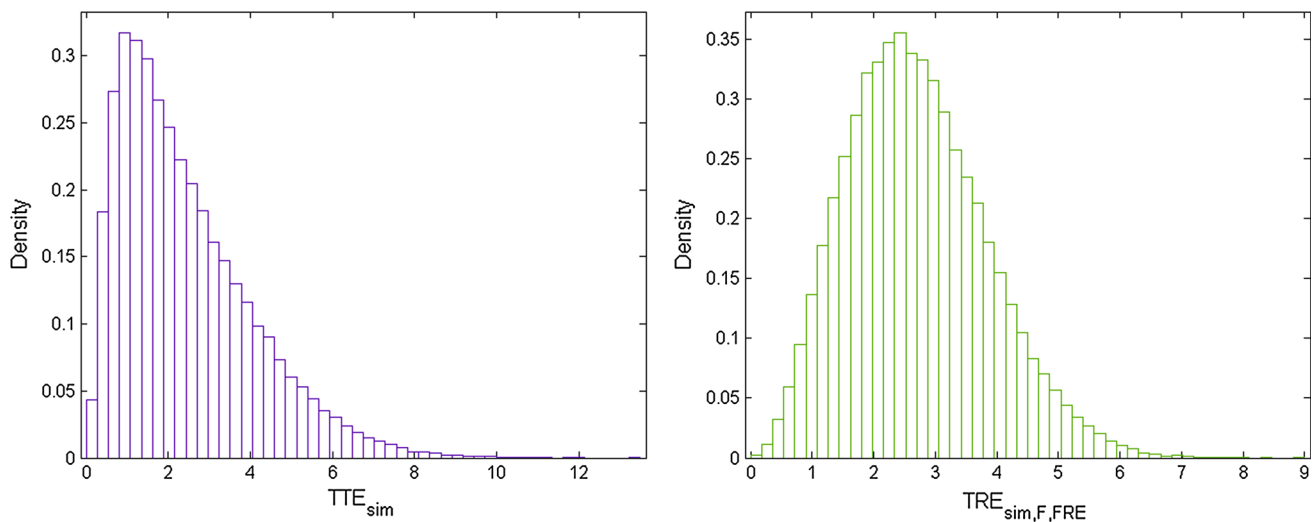
errors could be achieved if the mean of the targets of all repetitions was used for each experiment, because eventual biases were eliminated [21]. As a result, the TTEs were smaller and prediction approached the measured values.

Graphically it could be observed that for all patients the errors of measurements and predictions were getting smaller, when more fiducials were used for registration, as expected (Figs. 2, 4, 6). Using 5 fiducials or more leads to very similar results and shows that there is no need for using a large number of fiducials to improve accuracy of the navigation, as already investigated by many authors, e.g., [33]. An important result is that predictions were approaching measurements already when 3 or 5 fiducials were used for registration. Using 7 or 9 fiducials for registration TTE and

**Table 13** Mean experimental correlation coefficients

Correlation coefficient	Corr(TRE <sub>F,FLE</sub> , TLE)	Corr(TRE <sub>F,FRE</sub> , TLE)
Plastic skull	0.48	0.38
Anatomic specimen	0.65	0.64
Volunteer	0.41	0.47

The highest correlation can be found for the anatomic specimen for all variables studied



**Fig. 8** Example of a pdf of the measured TTE<sub>sim</sub> (left) and the predicted TRE<sub>F,FRE</sub> (right) in a numerical simulation. A 3 points registration was used to calculate 100,000 TTEs and TREs. TTE<sub>sim</sub> = 2.43  $\pm$  1.63 mm, TRE<sub>F,FRE</sub> = 2.70  $\pm$  1.13 mm. The difference between

the two functions can clearly be seen and is statistically significant. The Gamma distribution and the Nakagami distribution are the best fits for TTE<sub>sim</sub> and TRE<sub>F,FRE</sub>, respectively, for this experiment

TRE did not improve much, but less predicted TREs were similar to measurements.

Generally, the trend of predictions was always similar to measurements. Most of the predicted results were inside the measured  $TTE \pm$  one standard deviation.

Usually  $TTE_{F,FRE}$  was overestimating  $TTE_{exp}$ , but the overestimation was sometimes too large to be relevant (see Tables 7, 8, 9).

The best estimator for  $TTE_{exp}$  was  $TRE_{F,FLE}$ ; it predicted  $TTE_{exp}$  in 56.8, 72.5, and 45% of all targets of the plastic skull, the anatomic specimen, and the volunteer, respectively (the exact numbers of targets are shown in Tables 7, 8, 9).

In case of anisotropy, both  $TRE_{F,FLE,aniso}$  and  $TRE_W$  predicted 68.2 and 75% of all targets for the plastic skull and the anatomic specimen, respectively. For the volunteer  $TTE_{exp}$  could be predicted with  $TRE_{F,FRE}$  in 40% of all targets (the exact numbers of targets are shown in Tables 7, 8, 9).

It has to be mentioned that all prediction methods are grounded on the same theory and the predicted TREs (all but the  $TTE_{F,x}$ ) led to very similar results. This can be clearly seen in all figures and tables. However, the results of statistical testing showed that for these particular experiment settings different prediction methods did not lead to equal results; the number of targets where prediction was equal to measurement is different for all methods.

These findings confirm the importance of estimating FLE for the predictions; this was possible since the used navigation system provides access to all data. In a real surgical setup, it is difficult to estimate the FLE at the patient. It is always depending on how experienced the surgeon is with navigation, if the fiducials are in regions difficult to reach or if the probe at certain fiducials can be detected by the tracker.

$TTE_{exp}$  is submillimetric for plastic skull and anatomic specimen. For the volunteer a  $TTE_{exp} = 2.3$  mm is in a clinically acceptable range, with an  $ULE_{exp} = 1.61$  mm due to anatomic landmarks only. The difference in using screws and anatomic landmarks can be observed well. Screws lead to smaller errors due to the “exact” fiducial that can be located accurately, whereas locating anatomic landmarks accurately is difficult and leads to larger user errors.

In case of the real experiments medium correlation [34] could be found between  $TRE_{F,FLE}$  and TLE and between  $TRE_{F,FRE}$  and TLE (Table 13). Concerning this correlation considerable care must be taken using  $TTE_{F,x}$ , which can only be defined as is, if no correlation of TRE and TLE occurs. No correlations could be found for the results of numerical simulation, as required for the theoretical approach of TTE.

It could be observed that experiments with the navigation system and its simulation were leading to different results. A numerical simulation might be a good proof for theory, but it clearly differs from experiments in a real surgical situation. In reality, more complex error sources, like bias,

non-normality, and temporal variations of distributions, influence the experiments and make them difficult to generalize or predict.

Although the sample size of the real experiments is relatively small, especially for the volunteer, the results are providing a good insight into the possibilities of prediction methods for the TRE. The statistical power is large compared to the sample size; this is because the difference between prediction and measurement can be clearly seen for a lot of targets. Thus, the overestimation could be confirmed statistically with a one-sided test with a small alpha value. When the difference between the means of measurement and prediction was small, the power was getting smaller as well, especially when no overestimation of the measurement could be provided. For the numerical experiment the power was always 1 due to the very large sample size. Decreasing the sample size led to a smaller power also for the numerical simulations, as can be expected.

Due to the small sample size using a robust estimate for the covariance matrix of FLE is suggested, because the covariance matrix is sensitive to outliers. With a robust estimation  $TRE_W$  and  $TRE_D$  were changing and could predict more TTEs as with a non-robust calculation. Especially for the anatomic specimen results improved, the prediction got more “accurate.” Though the overall means did not change a lot, the equality of measurement and prediction could be confirmed more often when a robust method was used, and less overestimations occurred (see Tables 1, 2). For the other two patients no remarkable changes could be observed (see Tables 4, 6).

Taking into consideration that more fiducials were outliers and thus less registrations could be used for calculation, we had less variance of the experiments. This makes it more difficult for the prediction to be within the standard deviation of the experimental results.

The analysis of results is hardly affected whether outliers are removed or not. However, care should be taken when outliers are removed. For the experiments the FLE and TRE values decreased if the outliers were removed, but the whole measurement process itself is characterized as is and is not affected by this.

Only those points were excluded, where an obvious mistake in the measurement occurred, such that it was easy to confirm visually the impossibility to reach the position under consideration in the experimental setup. All other outliers found by the algorithm were a result of, i.e., systematic, temporally varying bias and non-static bias which are inherent to the measurement process (cf. “Data inspection and analysis” section). A detailed investigation of the effect of bias was already done in [35].

Registration had no influence on FLE; thus, there was no difference between  $TRE_{F,FLE}$  and  $TRE_{F,FLE,aniso}$  and furthermore between  $TTE_{F,FLE}$  and  $TTE_{F,FLE,aniso}$  (c.f. Tables 3, 5).

Surface registration might benefit of these results as well. Though the registration is different, it uses point correspondences too [4]. It is clear that only tracker and calibration error influence the FLE in surface registration; ULE and  $FLE_{\text{image}}$  are negligible. As we observed in our experiments eventual anisotropy did not influence the prediction much, Fitzpatrick's TRE is an adequate model to estimate the TRE in a surgical setting.

Previous numerical experiments were not analyzing the distributions of the predicted and measured errors ([6, 8–11]). Detailed analysis of the numerical simulation showed that prediction and measurement were coming from different distributions (an example is shown in Fig. 8). Thus, it is clear that the results of prediction and measurement could not be equal, both in real experiments and in simulation. In general, the distribution of a measurement of a vector length is similar to a Maxwell distribution [32]. In general, it is very challenging to know and to characterize the distribution of the experimental data, specifically for small sample sizes and when data acquisition is a lengthy and labor-intensive undertaking. For one repetition of the numerical experiment, the difference between the distributions of measurement and prediction could always be well observed and tested (Fig. 8). The means and standard deviations of measured and predicted TREs would suggest that it is possible to predict the measured TTE, because they were similar for prediction and measurement. However, predictions did not result in an upper limit. Whether overestimation of the measurement can be achieved with  $TRE_{F,FLE}$  is dependent on the FLE defined for the simulation. This indicates that the FLE is the important factor for the prediction and should be estimated well prior to the experiments. In the independent case, the difference between measurement and prediction was getting larger, because different fiducials and targets were used. Like real experiments, where measurement, tracking, and user errors influenced the prediction, numerical experiments showed that an improvement in the simplest TRE prediction ( $TRE_{F,FRE}$ ) might not be necessary.

## Conclusion

Experiments with a plastic skull, an anatomic specimen, and a volunteer were analyzed to predict and measure intraoperative application errors made before and during surgery. Isotropic and anisotropic registration and prediction methods were used, and prediction of the TRE was compared to the measured TRE. Best results for an upper limit of TRE were provided by  $TTE_{F,x}$ ; the most similar results were achieved with  $TRE_{F,FLE}$  and  $TRE_{F,FRE}$ . According to our experiments, using anisotropic registration and/or prediction methods did not significantly improve the results of the predictions. The smallest ULE was found for the plastic skull

with Ti-screws only; the largest ULE was found for the volunteer with anatomic landmarks only.

To our knowledge, this is the first investigation where the accuracy of navigation of simulated clinical experiments is compared to commonly used prediction methods, using anisotropic and isotropic registrations. A detailed error analysis of three patients in an experimental clinical setup was conducted, possibly due to a detailed data collection, that demonstrated the usefulness of an open navigation system.

**Acknowledgements** Open access funding provided by University of Innsbruck and Medical University of Innsbruck.

**Funding** This work was partly funded by the Austrian Science Foundation, Grant No. P-20604-B13, by the Jubilee Funds of the Austrian National Bank, Project No. 13003, and by the Austrian Research Promotion Agency (FFG) under Project Number 846056.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

**Human and animal rights** This article does not contain any studies with animals performed by any of the authors.

**Informed consent** This articles does not contain patient data.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Arun KS, Huang TS, Blostein SD (1987) Least-squares fitting of two 3-D point sets. *IEEE Trans Pattern Anal Mach Intell* 9(5):698–700
2. Maurer CR, Fitzpatrick JM (1993) A review of medical image registration. In: Maciunas (ed) *Interactive image-guided neurosurgery*. American Association of Neurological Surgeons, Park Ridge, pp 17–44
3. Maintz JB, Viergever MA (1998) A survey of medical image registration. *Med Image Anal* 2(1):1–36
4. Besl PJ, McKay HD (1992) A method for registration of 3-D shapes. *IEEE Trans Pattern Anal Mach Intell* 14(2):239–256
5. Zhang Z (1994) Iterative point matching for registration of free-form curves and surfaces. *Int J Comput Vis* 13(2):119–152
6. Fitzpatrick JM, West JB, Maurer CR (1998) Predicting error in rigid-body point-based registration. *IEEE Trans Med Imaging* 17(5):694–702
7. Kohan D, Jethanamest D (2012) Image-guided surgical navigation in otology. *Laryngoscope* 122(10):2291–2299



8. Wiles AD, Likholyot A, Frantz DD, Peters TM (2008) A statistical model for point-based target registration error with anisotropic fiducial localizer error. *IEEE Trans Med Imaging* 27(3):378–390
9. Balachandran R, Fitzpatrick JM (2008) The distribution of registration error of a fiducial marker in rigid-body point-based registration. In: *Medical imaging 2008 visualization image-guided procedures modeling, pts 1 2* 6918:M9180–M9180
10. Danilchenko A, Fitzpatrick JM (2011) General approach to first-order error prediction in rigid point registration. *IEEE Trans Med Imaging* 30(3):679–693
11. Moghari MH, Abolmaesumi P (2009) Distribution of target registration error for anisotropic and inhomogeneous fiducial localization error. *IEEE Trans Med Imaging* 28(6):799–813
12. Moghari MH, Ma B, Abolmaesumi P (2008) A theoretical comparison of different target registration error estimators. *Med Image Comput Assist Interv* 11(Pt 2):1032–1040
13. Shamir RR, Joskowicz L, Spektor S, Shoshan Y (2009) Localization and registration accuracy in image guided neurosurgery: a clinical study. *Int J Comput Assist Radiol Surg* 4(1):45–52
14. Fried MP, Kleefield J, Gopal H, Reardon E, Ho BT, Kuhn FA (1997) Image-guided endoscopic surgery: results of accuracy and performance in a multicenter clinical study using an electromagnetic tracking system. *Laryngoscope* 107(5):594–601
15. Shamir RR, Joskowicz L, Shoshan Y (2012) Fiducial optimization for minimal target registration error in image-guided neurosurgery. *IEEE Trans Med Imaging* 31(3):725–737
16. Güler Ö, Perwög M, Kral F, Schwarm F, Bárdosi ZR, Göbel G, Freysinger W (2013) Quantitative error analysis for computer assisted navigation: a feasibility study. *Med Phys* 40(2):21910
17. Horn BKP (1987) Closed-form solution of absolute orientation using unit quaternions. *J Opt Soc Am A* 4(4):629–642
18. Balachandran R, Fitzpatrick JM (2009) Iterative solution for rigid-body point-based registration with anisotropic weighting. In: *Proceedings of SPIE*, vol. 7261, p 72613D–72613D–10
19. Bickel M, Güler Ö, Kral F, Schwarm F, Freysinger W (2009) Evaluation of the application accuracy of 3D-navigation through measurements and prediction, In: *In World Congress on medical physics and biomedical engineering, September 7–12, 2009, Munich, Germany: Vol. 25/6 Surgery, Minimal Invasive Interventions, Endoscopy and Image Guided Therapy*, Dössel O, and Schlegel WC, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, pp 349–351
20. Güler Ö, Yaniv Z, Freysinger W (2011) Online temporal synchronization of pose and endoscopic video streams. In: *Proceedings of SPIE 7964, medical imaging 2011: visualization image-guided procedures and modeling*, pp 79641M–79641M
21. Bárdosi Z, Freysinger W (2016) Estimating FLE\_image distributions of manual fiducial localization in CT images. *Int J Comput Assist Radiol Surg* 11(6):1043–1049
22. Sibson R (1978) Studies in the robustness of multidimensional scaling: procrustes statistics. *J R Stat Soc Ser B* 40(2):234–238
23. Sibson R (1979) Studies in the robustness of multidimensional scaling: perturbational analysis of classical scaling. *J R Stat Soc Ser B* 41(2):217–229
24. Langron SP, Collins AJ (1985) Perturbation theory for generalized procrustes analysis. *J R Stat Soc Ser B* 47(2):277–284
25. Danilchenko A (2011) Fiducial-based registration with anisotropic localization error. *Vanderbilt University*
26. Fitzpatrick JM (2010) The role of registration in accurate surgical guidance. *Proc Inst Mech Eng Part H* 224(5):607–622
27. Caversaccio M, Freysinger W (2003) Computer assistance for intraoperative navigation in ENT surgery. *Minim Invasive Ther Allied Technol* 12(1):36–51
28. Verboven S, Hubert M (2004) LIBRA: a MATLAB library for robust analysis. *Chemom Intell Lab Syst* 75:127–136
29. Rousseeuw PJ, Van Driessen K (1999) A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41(3):212–223
30. Fitzpatrick JM (2009) Fiducial registration error and target registration error are uncorrelated. In: *Proceedings of SPIE*, vol. 7261, pp 726102–726102–12
31. Wiles AD, Thompson DG, Frantz DD (2004) Accuracy assessment and interpretation for optical tracking systems. *Proc SPIE* 5367:421–432
32. Hsu DY (1999) *Spatial error analysis: a unified application-oriented treatment*. Wiley, New York
33. Kim S, Kazanzides P (2017) Fiducial-based registration with a touchable region model. *Int J Comput Assist Radiol Surg* 12(2):277–289
34. Cohen J (1988) *Statistical power analysis for the behavioral sciences*. L. Erlbaum Associates, Hillsdale
35. Moghari MH, Abolmaesumi P (2010) Understanding the effect of bias in fiducial localization error on point-based rigid-body registration. *IEEE Trans Med Imaging* 29(10):1730–1738