Transfer Accuracy and Precision Scoring in Planar Bone Cutting Validated With Ex Vivo Data

Federico Edgardo Milano,[1,2] Lucas Eduardo Ritacco,[3] Germán Luis Farfalli,[4] Luis Alberto Bahamonde,[5] Luis Alberto Aponte-Tinao,[4] Marcelo Risk[1,2]

[1]Department of Bioengineering, Instituto Tecnologico de Buenos Aires, Buenos Aires, Argentina, [2]CONICET (Consejo Nacional de Investigaciones Cientificas y Tecnicas), Buenos Aires, Argentina, [3]Department of Health Informatics, Italian Hospital of Buenos Aires, Buenos Aires, Argentina, [4]Institute of Orthopedics "Carlos E. Ottolenghi", Italian Hospital of Buenos Aires, Buenos Aires, Argentina, [5]Department of Orthopedics and Traumatology, University of Chile, Santiago de Chile, Chile

Received 2 September 2014; accepted 10 December 2014
Published online 25 February 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/jor.22813

ABSTRACT: The use of interactive surgical scenarios for virtual preoperative planning of osteotomies has increased in the last 5 years. As it has been reported by several authors, this technology has been used in tumor resection osteotomies, knee osteotomies, and spine surgery with good results. A digital three-dimensional preoperative plan makes possible to quantitatively evaluate the transfer process from the virtual plan to the anatomy of the patient. We introduce an exact definition of accuracy and precision of this transfer process for planar bone cutting. We present a method to compute these properties from ex vivo data. We also propose a clinical score to assess the goodness of a cut. A computer simulation is used to characterize the definitions and the data generated by the measurement method. The definitions and method are evaluated in 17 ex vivo planar cuts of tumor resection osteotomies. The results show that the proposed method and definitions are highly correlated with a previous definition of accuracy based in ISO 1101. The score is also evaluated by showing that it distinguishes among different transfer techniques based in its distribution location and shape. The introduced definitions produce acceptable results in cases where the ISO-based definition produce counter intuitive results. © 2015 Orthopaedic Research Society. Published by Wiley Periodicals, Inc. J Orthop Res 33:699–704, 2015.

Keywords: osteotomy; cut; accuracy; precision; clinical score

Clinical Background and Related Work

Computer-Assisted Preoperative Planning in bone surgery is becoming more common everyday. Its use has been reported in knee osteotomy,[1] mandibular reconstruction,[2] spine surgery,[3] foot and ankle surgery,[4] and tumor resection.[5] The existence of virtual osteotomy planning opens the possibility of measuring the discrepancies that occur in the transfer from the target to the executed osteotomies. These discrepancies are very common since many factors add random variations and biases to the surgical process: CT/MR imaging errors, three-dimensional modeling and segmentation omissions and the inherent inaccuracies of surgical tools and actions. Even though the evaluation of transfer accuracy (ACC) and precision (PRC) in bone cutting is a key application in orthopedics, there is no consensus about their definition in the field of planar osteotomies, neither there is an established method to measure them nor a score to assess them.[6]

In recent years, several research groups have approached the accuracy and precision estimation problem in bone cutting. Barrera et al.[7] introduced a way to evaluate planar osteotomies in total knee replacement using translational and rotational errors of the executed plane measured against the target plane. More recently, Cartiaux et al.[8] proposed a method based in the ISO 1101:2004 standard for geometrical tolerancing to evaluate differences between a cutting plane and a target plane. Their work shows that it is possible to express the most significant translational and rotational errors using only the location parameter (L) defined in the mentioned ISO standard. For experimental data gathering, a test bed with a block simulating bone tissue is used, and errors are estimated with a coordinate measuring machine set in the same frame of reference. Dobbe et al.[9] propose a method to measure and estimate the normal of an executed plane. This normal is used to compute the dihedral angle with the target plane, that is decomposed in sagittal and coronal plane angles. Then a distance error between the target and executed plane is computed taking the Euler distance between the centroids of the cross sections defined by target and executed planes. This method is validated using a cadaveric limb, with pre and postoperative computed tomography (CT) scans positioned in a common frame of reference using a registration algorithm.

The accuracy and precision measuring method proposed in this work borrows inspiration from the orthopedics oncology surgical pipeline where a resection specimen is preserved for histological evaluation. In our method, this specimen is also acquired by a medical imaging system and used to computationally find the executed planar osteotomies, which are then compared with the target osteotomies defined in a digital preoperative scenario. In particular, the specimens used in this work were CT scanned. CT imaging eases the identification and segmentation of the superficial cortical bone of the specimen, not requiring an extra peeling and preparation of the specimen. However, this computational method may also be used with any acquisition modality that allows mapping the surface of the specimen (e.g., a non-contact active three-dimensional surface scanner).

Grant sponsor: National Council for Scientific and Technical Research (CONICET); Grant sponsor: Ottolenghi Foundation.
Correspondence to: Marcelo Risk (T: +54 (11) 6393-4893; F: +54 (11) 6393-4821; E-mail: mrisk@itba.edu.ar)

This work also introduces the use of a well defined quality index, the process performance index (P_{pk}),[10] as a surgical score to assess both the accuracy and precision of a planar osteotomy. This index is widely used in the manufacturing industry for statistical production control. The use of this index in Computer-Assisted Surgery was proposed by Stiehl et al.[11] but, to the best of our knowledge, this index was never used for surgical performance evaluation.

In a previous publication by our group, we presented a series of Image Guided Surgery (IGS) cases, and used a previous version of the method presented here to evaluate differences in experimental groups defined by planar osteotomies configuration, affected bone, and type of tumor.[12] The previous method had several limitations in its efficacy to find good planar fits in noisy images. Also, there were no data of non-assisted surgery available at the time to compare with IGS data. Nevertheless, the obtained results were satisfactory, and brought forward a series of open-ended questions to be further investigated, relating to both the clinical and technical aspects of the method.

Hypothesis and Specific Aims

We hypothesize that the integration of the concepts, definitions, and methods that will be presented in this work may be a useful approach to a clinical method for measuring accuracy and precision in planar osteotomies. We pose a set of research aims that this study intends to address: (i) to define the transfer accuracy and precision for planar osteotomies and a method to measure them, (ii) to study the correlation of our score to the ISO 1101:2004 L parameter,[8] and (iii) to assess the capability of this score to distinguish between different techniques of preoperative planning transference to the operative field.

MATERIALS AND METHODS

Materials

The method presented in this paper was tested and validated using a set of CT images from seven subjects, comprising a total of 17 planar cuts used in the evaluation of the algorithm. Tumors were located in the pelvis in six subjects and in the scapula in one subject. All subjects were preoperatively CT scanned (Multislice 64, Aquilion; Toshiba Medical Systems, Otawara, Japan). Slices of 0.5 mm thickness were obtained using a soft tissue algorithm (matrix of 512×512 pixels). Magnetic resonance (MRI) images of tumoral regions were acquired using a 1.5 T (Magnetom Avanto; Siemens Medical Solutions, Erlangen, Germany). Slices of 1 mm thickness were obtained using T1-weighted or fat-suppressed sequences to optimize visualization of the signal intensity from the bone tumor (matrix of 256×256 pixels). MRI data are used during the virtual planning stage to establish resection margins. The resection procedure generates a surgical specimen for each subject. These specimens were CT scanned with the CT protocol described above. These images were semi-automatically segmented using Mimics version 14.1 (Materialise, Leuven, Belgium), and stored in the form of surface meshes. Surface models of corresponding subject and specimen pairs were aligned in a common frame of reference using landmark based registration as an initial transform estimate and a partial registration algorithm[13] for refining the alignment. The mean surface distance (in mm) was measured from the registered specimen surface to the subject surface.

The preoperative plan transfer was non-assisted for two subjects and assisted by an IGS (Navigation System II Cart with OrthoMap Hip Navigation Software, Stryker Leibinger GmbH, Freiburg, Germany) system in five subjects.

Transfer Accuracy, Precision, and Score Definitions

Let S denote the bounded surface in $\mathbb{R}^3$ generated by the cutting process. In theory, surface S should be planar since it is produced by following a target plane but, in practice, its shape may have a large amount of variations due to vibration, recuts, and unintended detachment of trabecular bone. We define the set of signed distances (in mm) from the surface S to the target plane T as $D = \{n \cdot (x - x_0) \forall_{x \in S}\}$ where n is the target plane unit normal vector, x is any point of surface S and x_0 is any point of T. Being Qk/n the kth n-quantile of set D, then we define the **transfer accuracy** as the median ($Q_{0.5}$), and the **transfer precision** as a measure of dispersion $| \; Q0.99865 - Q0.00135 \; |$.

Following an approach suggested by Stiehl et al.[11] we propose the use of the *process performance index* (P_{pk})[10] as a score to evaluate both of these properties in relation to two-sided tolerances: $P_{pk} = \min((UL - Q_{0.5})/(Q_{0.99865} - Q_{0.5}); (Q_{0.5} - LL)/(Q_{0.5} - Q_{0.00135}))$ where, UL and LL are the upper tolerance limit and lower tolerance limit respectively. These values were set to 10 and -10 mm for the evaluation of the ex vivo cases presented in this paper. This margin is the minimal wide margin in bone tumors as defined by Kawaguchi et al.[14] In other surgical domains these values should be adjusted according to the cutting margin defined for each particular setting.

The proposed score is unitless, since it is calculated as a ratio between the accuracy and precision, both measured in units of distance (mm). Different ranges of P_{pk} values have distinct meanings. A value of P_{pk} less than one signifies that the cutting process is outside tolerances, on the other hand a value of one means that the outcome of the process is just inside the bounds set by UL and LL limits. As P_{pk} increases, the process is better centered in the target value and has less dispersion.

In the context of these definitions, the location metric defined in the ISO 1101:2004 standard for geometrical tolerancing is $L = \max(|D|)$.[8] This parameter, being the maximum absolute distance from the cutting surface to the target plane, increases as the executed plane departs from the target.

Method for Estimating the Cutting Surface

The definitions of accuracy and precision given above are useful as long as there is a way to detect and segment the cutting surface S. Dobbe et al.[9] solve this problem by manually positioning spheres in the cutting surface region and using the local surface properties enclosed by the spheres to estimate the surface normal. The algorithm presented in this paper uses the target plane as a ray origin to automatically detect and segment the cutting surface S. The target plane is resampled at an interval of 0.5 mm. For each sampled point, a ray which orientation is perpendicular to the target plane is projected away from the plane in both senses. Most of the rays intersect the bone surface.[14] The intersection points are used to fit a planar model, using a locally optimized random sample consensus algorithm

(LO-RANSAC) for robustness.[15] The adjusted plane is then used to sample the bone surface keeping the inlier surface points, those that are to a distance less than 0.01 mm from the fitted plane, and discarding the rest. The largest inlier point set that fits the planar model is used as a point cloud estimation of surface S. Then, these inlier points are used to compute the signed distances that conforms the set D. A flowchart for this algorithm is shown in Figure 1.

Experimental Pipeline

The overall application of this method can be briefly described as follows. Based on the preoperative images of the patient, the surgeons who will intervene in the procedure decide the position and orientation of the target planes that compose the osteotomy, taking into account general medical factors affecting the patient, like the location and nature of the osteotomy and the particular approach to the surgery. The outcome of this activity is a virtual preoperative plan, that is, a set of target planes displayed together with the bone anatomy of the patient, in an interactive three-dimensional virtual scenario. The plan is then uploaded to an IGS system, if one is available, or to an operating room computer before the procedure is executed. During the surgery, if the team is using an IGS system, the system will show in its screen the position of a physical pointer in relation to the bone structures and the target planes. If there is no IGS system available, the surgeon has to mentally integrate the displayed images and the directly palpable and visually sensed anatomy to transfer the planned osteotomies to the patient. After the procedure, the resected surgical specimen is CT scanned (Fig. 2a). This surgical specimen is segmented and registered to the preoperative scenario. This registration step puts the target plane and surgical specimen in the same frame of reference (Fig. 2b). The error added by the segmentation and registration process is estimated computing the mean surface distance from the specimen surface to the preoperative bone surface. At this point, the executed plane parameters are unknown, but the information needed to estimate them is implicitly contained in the cutting surface S produced in the surgical specimen by the cutting tool. The measurement algorithm presented in this paper uses that information to estimate the plane parameters and to measure the discrepancies between the target and the executed planes (Fig. 2c). The mean surface error is added to the signed distances set D, so each measured value is penalized as if it were the carrier of the whole magnitude of the error. For each target plane that is part of an osteotomy, a score reflecting its execution accuracy and precision is calculated. A colorimetry depicting the signed distances for each sample point in the detected surface S is also generated. The surgeon could use this colorimetry as an error map to check the regions in need of a recut or adjustment (Fig. 2e). Supplementary Figure S2 shows a more detailed view of the pipeline.

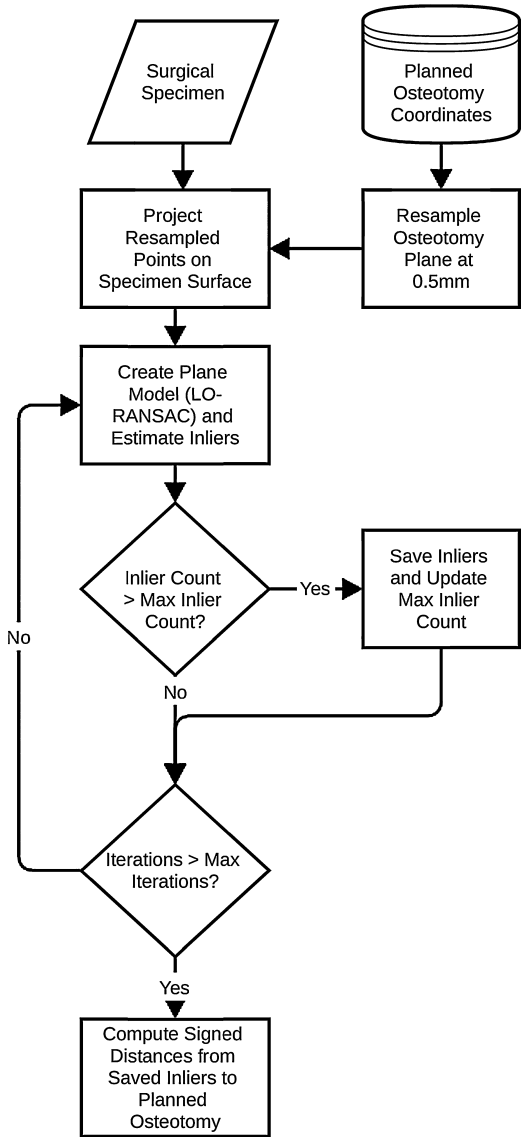


Figure 1. Surface S estimation algorithm. The inliers set is iteratively refined, producing as a result the surface S described as a point cloud. These inlier points are used to compute the set D of signed distances.

Simulation Model

A simulation model was created to characterize the mathematical definition of the score and evaluate its relationship with the location metric L. A schematic representation of the model is shown in Figure 3, following the convention introduced by Cartiaux et al.[8] of setting the xy plane as the target plane T. The model simplifies the cutting surface S to a plane. This plane is controlled by three parameters: h (the euclidean distance from T in the origin of coordinates, Fig. 3a), xz angle α (Fig. 3b), and yz angle β (Fig. 3c). A particular surface S is generated by uniformly sampling α and β in the range of $(-\pi/2,\pi/2)$ and h in the range of $(-3, 3\,\text{mm})$. Once these values are determined, the plane T is resampled with a sampling interval set to $\Delta = 0.5\,\text{mm}$ (Fig. 3c). The model then measures the signed distances from the sampled points to surface S and adds to each distance an uniformly distributed random value in the range of $(-0.5, 0.5\,\text{mm})$, simulating the noise introduced in the surface during the cutting process. The simulation is run in a volume of $40 \times 40 \times 70\,\text{mm}^3$. A total of 10,000 cutting surfaces were generated and the accuracy, precision, scores, and location metric L for each of these surfaces were calculated.

Statistical Analysis

The statistical analysis and simulation were performed using the R programming Language. Shapiro–Wilk test was used to evaluate the normality of the simulated and ex vivo data,

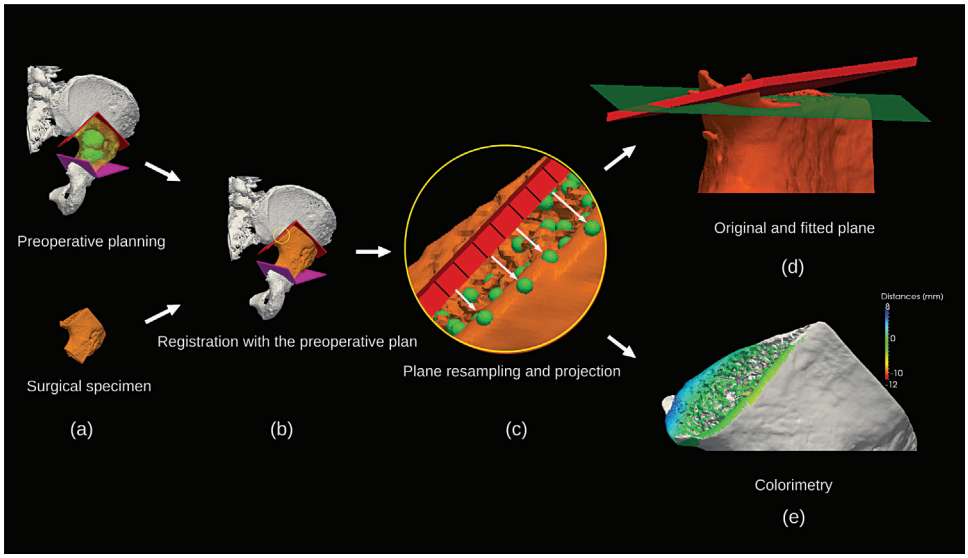
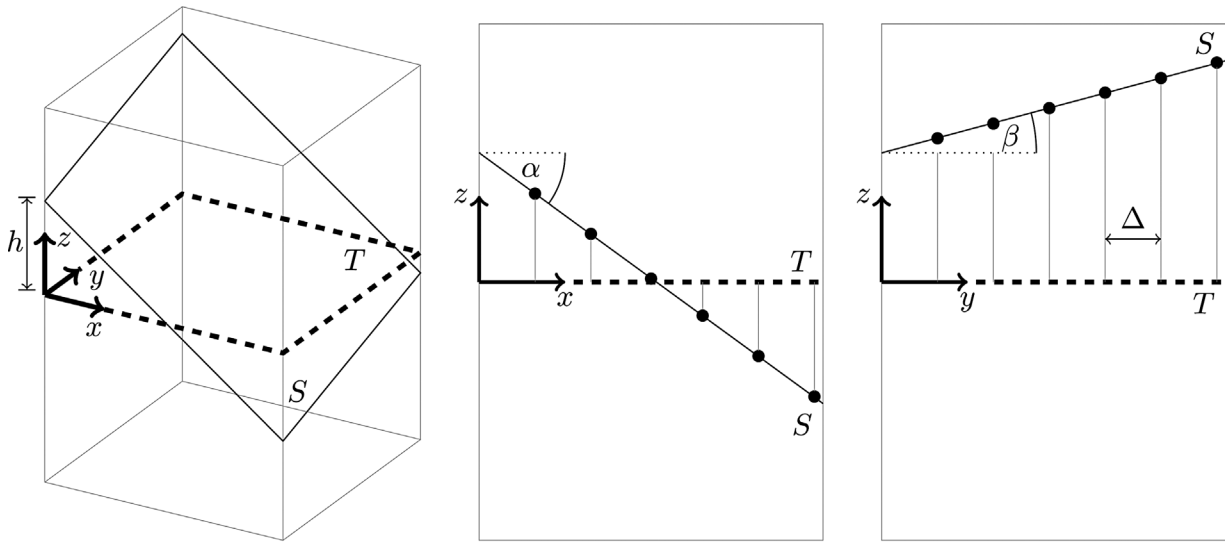


Figure 2. Processing pipeline. (a) The initial surgical plan and the digitized surgical specimen. (b) The preoperative plan is registered against the surgical specimen. (c) The osteotomy target plane is discretized and projected on the surgical specimen, generating a point cloud. A plane is robustly fitted to the point cloud. (d) The original plane is depicted in red, while the fitted plane is depicted in green. (e) An error map is generated using colorimetry; this visually shows whether the discrepancies in the cutting surface are larger than expected.



(a) 3D (b) xz plane (c) yz plane

Figure 3. (a) T is the reference or target plane, a finite region of xy plane. S is the surface generated by executing a cut; in the simulation it is defined by three parameters: a distance h and two angles, α and β. In (b) and (c) the spatial sampling of S is shown.

along with normal Q–Q plots and histograms. To study the relationship between the location metric L and the proposed score P_{pk}, the correlation between them was computed using Spearman rank correlation coefficient, for both the simulated and the ex vivo data. To address whether it is possible to discern between surgical transfer methodologies (non-assisted or assisted by an IGS system) using the proposed P_{pk} score, the Mann–Whitney U test was performed in the two groups of cases.

RESULTS

The simulation model shows that the signed distances in set D are not normally distributed in any of the cases. For a particular simulated plane, the Shapiro–Wilk normality test shields $W = 0.97$ ($p < 0.001$). The same results were found for the ex vivo series. We illustrate this by showing that the Shapiro–Wilk normality test for subject 6, plane b, yields $W = 0.95$ ($p < 0.001$).

The range of the error measured by the mean surface distances between the specimen and the preoperative bone surface is (0.35, 0.73 mm).

The simulated P_{pk} score has a very strong inverse correlation with the location metric L, with $\rho = -0.95$ ($p < 0.001$). The strong correlation found in the simulation matches the ex vivo measurements, showing also a strong inverse correlation with $\rho = -0.95$ at $p < 0.001$

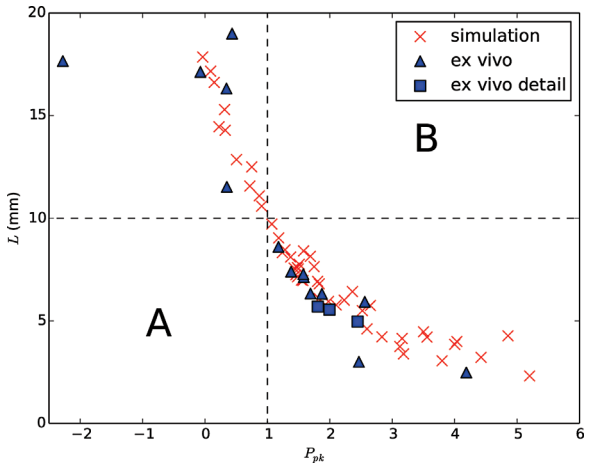


Figure 4. P_{pk} and L relationship. Ex vivo data and simulated data show a high inverse correlation between P_{pk} and L. Area A means an out of tolerance P_{pk} when the maximum value L is inside the tolerance range. Area B means an inside tolerance P_{pk} when L is out of tolerance. Since both areas are clear of points for ex vivo and simulated data, P_{pk} and L are consistent metrics. The square markers show the three ex vivo points analyzed in detail in Figure 5.

between the proposed score and the location metric L. The fact that the correlation is inverted agrees with the score interpretation. On the one hand a smaller P_{pk} means a less accurate and/or precise cut, therefore a large L parameter value is expected. On the other hand a larger P_{pk} means a more accurate and/or precise cut and a small L parameter value is anticipated. Figure 4 shows the relationship between P_{pk} and L, both for the simulated and ex vivo series.

Median P_{pk} score for non-assisted and assisted groups were 0.34 and 1.74, respectively; the distributions in the two groups differed significantly (Mann–Whitney $U = -11$, $n_1 = 5$, $n_2 = 12$, $p = 0.048$ two-tailed). A table with the mean surface distance, accuracy, precision, P_{pk}, and L values for assisted and non-assisted groups is shown in Supplementary Figure S5.

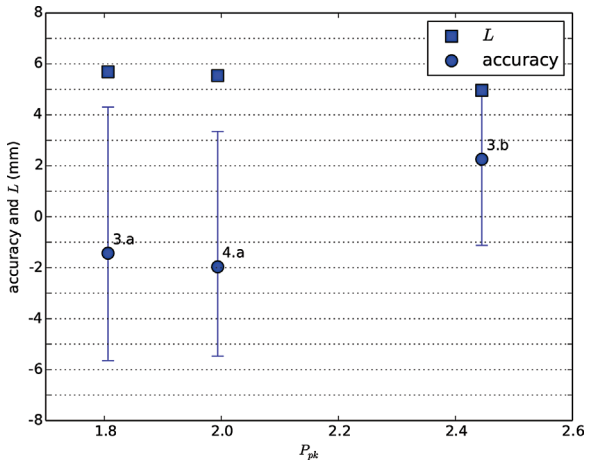


Figure 5. Ex vivo detail showing P_{pk}, accuracy, precision, and L for three ex vivo data points. These three data points show that it is possible to have a close L value and very different P_{pk} values, mainly due to different precision ranges. The precision is depicted as error bars around the accuracy value. Each case is labeled as subject plane and its details are shown in Supplementary Table S5.

DISCUSSION

Our main objective was to define the accuracy and precision of the transfer of a planar osteotomy described in a virtual scenario to the anatomy of the patient during a surgical process. We first defined both concepts based on an idealization in which we know the exact cutting surface and its bounds. Then we presented a method to automatically estimate this cutting surface using the target plane information. The mean surface distance between the specimen surface and the preoperative bone surface is considered the error of the cutting surface estimation process. The error never exceeds 0.73 mm, which is in the magnitude order of the CT scan resolution. This error is used to penalize the calculated signed distances propagating then to the accuracy, precision and score computations.

It is possible to estimate the cutting surface using other methods. The estimation method presented in Dobbe et al.[9] of enclosing the surface with spheres is compatible with our definition of accuracy and precision, since it could be used as a method for estimating the surface S; however, that approach presents some difficulties since its results would depend on the size of the spheres and in their manually chosen locations.

A secondary objective of this work was to apply a robust and industry proven score that could reflect both accuracy and precision in a single number. These are the main traits of the P_{pk}. It was shown that this score has a very good agreement with the ISO 1101:2004 location metric L proposed by Cartiaux et al.[8] for both simulated and ex vivo data. Figure 4 shows that P_{pk} is consistent with metric L. On the one hand, if L is above the tolerance limit of 10 mm, P_{pk} is below 1. On the other hand if L is below the limit of 10 mm, P_{pk} is above 1. Nevertheless, the location metric L does not always discern between different executed planes as it is shown by a local trend in the subset of ex vivo data plotted as square markers. The detail of this subset is illustrated in Figure 5, showing three ex vivo data points with close L values which are associated with very different P_{pk} values. This effect occurs because P_{pk} is computed from a central value and dispersion, while L is an extreme value of a set. This is graphically shown in the synthetic example in Figure 6. The figure displays two possible executions of an osteotomy based on the same target plane. Execution S_1 has a steeper angle than S_2. However, the metric L for both executed cuts yields the same value because this parameter only reflects the maximum of the distances between the executed surface and the target plane.

Another favorable point of the proposed score is that it is possible to set a different tolerance for each side of the osteotomy plane. This is useful in the field of orthopedic oncology since it is possible to penalize

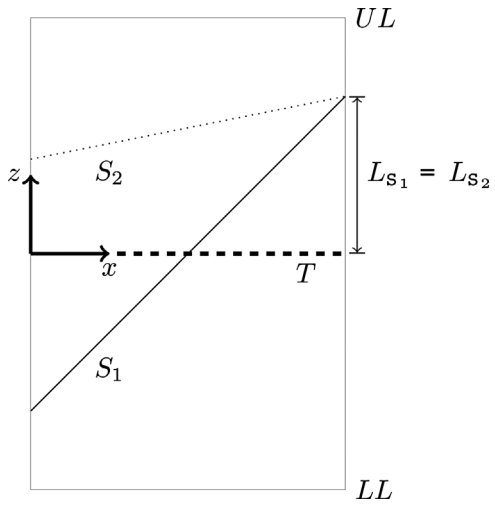


Figure 6. Location metric L ambiguity. T is the target plane. S_1 and S_2 are the cutting surfaces of two possibly executed planes. Their angles are $\alpha_{S1} = 45^\circ > \alpha_{S2} = 11^\circ$, while $\beta = 0^\circ$ for both planes. Their offsets are $|h_{S2}| < |h_{S1}|$. The location parameter L is the same for both planes. However, S_2 is closer to T in two parameters out of three. The proposed score makes explicit this difference since $P_{pk}(S_1) = 1.5 < P_{pk}(S_2) = 3.46$.

inaccuracies more on the tumor side than on the healthy side of the osteotomy. We have also shown that the proposed score is capable of distinguishing between different levels of assistance for surgical localization.

The method presented in this work has some limitations. The method relies in the availability of a surgical specimen. The given definitions and the method itself would still work with postoperative CT scans of the patient, but this was not evaluated in this work. This alternative workflow would also introduce new problems: the accuracy and precision information would not be available intraoperatively, since most operating rooms are not equipped with mobile CT scanners; the postoperative images may show artifacts produced by the osteosysnthesis or implants used for reconstruction; and last, the patient would be irradiated twice. It would also be possible to use surface scanning technologies intraoperatively; this approach may pose some challenges to the actual registration process and constitutes a future line of research.

The application of the concepts defined above is not constrained to the evaluation of the surgical process alone. These definitions and method could be used in the assessment of new surgical localization devices, in the preclinical evaluation of new cutting tools and technologies, and as a performance measuring tool for surgical training and education.

ACKNOWLEDGMENTS

This work was partly supported by a doctoral scholarship from the National Council for Scientific and Technical Research (CONICET) and from funds from the Ottolenghi Foundation for the Progress of Orthopaedics and Traumatology.

REFERENCES

1. Chao EY, Sim FH. 1995. Computer-aided preoperative planning in knee osteotomy. Iowa Orthop J 15:4–18.
2. Antony AK, Chen WF, Kolokythas A, et al. 2011. Use of virtual surgery and stereolithography-guided osteotomy for mandibular reconstruction with the free fibula. Plast Reconstr Surg 128:1280–1284.
3. Aurouer N, Obeid I, Gille O, et al. 2009. Computerized preoperative planning for correction of sagittal deformity of the spine. Surg Radiol Anat 31:781–792.
4. Sun SP, Chou YJ, Sue CC. 2009. Full-scale 3D preoperative planning system for calcaneal osteotomy with a multimedia system. J Foot Ankle Surg 48:528–539.
5. Wong K, Kumta S, Antonio G. 2008. Image fusion for computer-assisted bone tumor surgery. Clin Orthop Relat Res 466:2533–2541.
6. Abraham JA. 2011. Recent advances in navigation-assisted musculoskeletal tumor resection. Curr Orthopaed Pract 22:297–302.
7. Barrera OA, Haider H, Garvin KL. 2008. Towards a standard in assessment of bone cutting for total knee replacement. Proc Inst Mech Eng H 222:63–74.
8. Cartiaux O, Paul L. 2009. Accuracy in planar cutting of bones: an ISO-based evaluation. Int J Med Robot 5:77–84.
9. Dobbe JGG, Kievit AJ, Schafroth MU, et al. 2014. Evaluation of a CT-based technique to measure the transfer accuracy of a virtually planned osteotomy. Med Eng Phys 36:1081–1087.
10. Montgomery D. 2012. *Introduction to statistical quality control*, 7th ed. Hoboken: Wiley p 768.
11. Stiehl J, Bach J, Heck D. 2007. Validation and metrology in CAOS. In: Stiehl J, Konermann WH, Haaker RG, et al., editors. *Navigation and MIS in orthopedic surgery*. Heildelberg: Springer p 68–78.
12. Ritacco LE, Milano FE, Farfalli GL, et al. 2013. Accuracy of 3-d planning and navigation in bone tumor resection. Orthopedics 36:e942–e950.
13. Bronstein A, Bronstein M. 2008. Regularized partial matching of rigid shapes. In: Forsyth D, Torr P, Zisserman A, editors. *Computer vision—ECCV 2008*. Heildelberg: Springer p 143–154.
14. Kawaguchi N, Ahmed A, Matsumoto S, et al. 2004. The concept of curative margin in surgery for bone an soft tissue sarcoma. Clin Orthop Relat Res 419:165–172.
15. Milano F, Ritacco L, Farfalli G, et al. 2011. An algorithm for automatic surface labeling of planar surgical resections. J Phys 332:12037.
16. Chum O, Matas J, Kittler J. 2003. Locally optimized RANSAC. In: Michaelis B, Krell G, editors. *Pattern recognition*. Heildelberg: Springer p 236–243.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article.