Image-guided surgery: what is the accuracy?
Robert F. Labadie, a Bryan M. Davis a and J. Michael Fitzpatrick b

Purpose of review
Use of image-guided surgery (IGS) systems in otolaryngology, particularly rhinology, has grown exponentially in recent years. Central to their use is the understanding of the accuracy of each system. The purpose of this review is to discuss the error inherent in all IGS systems. A standardized technique (currently used in the engineering literature) for understanding and reporting error in IGS systems is reviewed. Using this technique, the error of commercially available IGS systems is reviewed.

Recent findings
The most commonly used IGS systems depend on the conformation of the skin, as opposed to relying on bone-implanted devices. For these systems, mean accuracies 2 mm or less are routinely reported. This finding is independent of fiducial markers (e.g., proprietary headsets, skin-affixed markers, or laser scanning of skin surfaces). Techniques of fiducial localization and registration of CT scans to intraoperative anatomy are proprietary to each company. As such, there is great variability in reporting system specifications—particularly error of IGS systems. This lack of standardization makes comparison of one system to another difficult if not impossible.

Summary
Image-guided surgery systems commonly used in rhinology report mean accuracies of 2 mm or less. Surgeons must be aware that this value represents a mean of a distribution of errors. As such, 95% of the time error can be expected to be less than approximately 1.7 times its mean value. However, outliers (errors much larger and much smaller than the mean) may exist for each IGS intervention. As noted, IGS systems function to complement—not replace—knowledge of surgical anatomy.

Keywords
image-guided surgery, accuracy, fiducial localization error, fiducial registration error, target registration error

Introduction
Image-guided surgery (IGS) systems have found widespread use in rhinology. Analogous to global positioning systems (GPSs) but on a smaller scale, IGS systems allow real-time, intraoperative tracking of current location on preoperatively obtained tomograms. IGS systems are commonly used in revision sinus surgery and for complex dissections [1–5]. Some advocate their use for all sinus cases, because IGS theoretically provides an additional safeguard preventing damage to healthy collateral tissue. Critics of IGS argue that use of such systems builds dependence on them for anatomic identification, diminishing the importance of anatomic knowledge. Although the importance of anatomic knowledge cannot be overemphasized, IGS systems are becoming an integral part of the specialty. Because of this, it is important to understand both how they work and, more importantly, their limitations. Perhaps the biggest limitation of IGS systems is that they are inaccurate. Understanding how much error is associated with each system is important in deciding how much to trust an IGS system during surgery. The purpose of this article is to explain the concepts of error and accuracy as applied to IGS systems.

Error analysis for image-guided surgical systems
Central to all IGS systems are markers located on the patient that can be identified both in the preoperative tomograms, e.g., CT scan, and on the patient in the operating room. These markers are called fiducials and are used to superimpose, or register, the preoperative CT scan onto the intraoperative surgical anatomy. Fiducials are important because they are the basis of error in IGS systems. Numerous fiducial systems are used commercially. These include neurosurgical N-frames, bone-implanted markers, proprietary head frames, skin-affixed markers, and skin surface contours.

1 For the purposes of this discussion pertaining to rhinology, tomogram is assumed to mean CT scan. However, the same applies to MRI scans and flat-plate radiographs.
Fiducial localization error
To register the preoperative CT scan to the intraoperative surgical anatomy, the markers must be identified both in the CT scan and on the patient in the operating room. This process can never be accomplished perfectly, because error occurs when finding each fiducial marker. The error in identifying a fiducial marker is called fiducial localization error (FLE). FLE is a random, unknown error that occurs both when locating fiducials in the CT scans and when locating fiducials on the patient in the operating room. To localize fiducials in CT scans, most commercially applicable systems use pattern identification algorithms to find the center of the fiducial automatically. Factors that affect error in localizing the fiducials in the CT images include image distortion, signal-to-noise ratio, resolution, and human error in the computer algorithm development. To localize fiducials within the operating room, most commercially applicable systems use infrared tracking to determine the location of a probe placed on the fiducial marker. (At least one commercially available system uses deformity of a local electromagnetic field to determine location of the probe.) Factors that affect error in localizing the fiducials on the patient in the operating room include human error in placement of the probe on the fiducial and error associated with the tracking system.

Fiducial registration error
After the fiducial markers are localized both in the CT scan and on the patient within the operating room, they are used to register the CT scan to the patient. Registration involves superimposing the fiducials from the CT scan onto the fiducials identified in the operating room. Because the exact location of the fiducials is not known as a result of FLE, the alignment of the fiducials during registration is never perfect. Mathematically, the best fit is obtained by calculating the alignment that minimizes the differences between corresponding fiducial locations in the CT scan compared with the fiducial locations on the patient in the operating room. A gauge of this best fit is termed fiducial registration error (FRE). For most commercial systems, either FRE or a derivative of FRE is prominently displayed on the computer screen once the surgeon has registered a patient to the preoperative CT scan. FRE is calculated by taking the root mean square of the differences between corresponding fiducial marker locations in the CT scan and the operating room. Registration creates a transformation that superimposes each data point from the CT scan to the patient’s location in the operating room. It is important to note that individual anatomic points are aligned based on the localization of the fiducials. If a given system’s fiducial fit is consistently poor, the alignment of other points by this system is likely to be poor; conversely, if the fiducial fit is consistently good, the alignment of other points is likely to be good. It is important to point out, however, that a small FRE for a specific patient does not translate to more accuracy for that patient. In general, there is a threshold below which the FRE (registration) must be to ensure a necessary level of surgical accuracy. Most commercial systems will not let the process proceed unless this threshold has been met. However, having FRE dramatically below this threshold does not indicate that more accuracy is achieved.

Figure 1 demonstrates FLE and registration. For this demonstration, four skin-affixed fiducial markers are shown. The error in finding these fiducials on the patient and in the CT scan is FLE. Registration consists of aligning the fiducials such that a best fit is obtained. This error is termed FRE. It is important to keep in mind that FRE is only a general gauge of good registration. For example, if, for a given marker-based system, FRE is lower for one patient than for a second patient, there is no reason to expect more accurate surgical navigation for the first patient than for the second. However, if the average FRE over many patients is lower for one system than for a second system, using the same number of markers, the first system is likely to be more accurate. The number of markers is important. With more markers, one would correctly expect the system to be more accurate. However, surprisingly, the use of more markers usually results in larger FRE, because it is more difficult to align multiple markers. Thus, reducing the number of markers reduces FRE but generally makes surgical accuracy worse.

Target registration error
With this best fit registration complete, IGS begins real-time tracking. The location of targets of interest to the surgeon can be identified on the patient and compared with corresponding locations in the CT scan. Any difference in these positions is termed target registration error (TRE) and is the error of interest to the surgeon. TRE, like FRE, depends upon FLE. Target registration error is not the same for all points within the image. It varies depending on position relative to the fiducial markers and on the configuration of those markers. Guidelines for fiducial placement to minimize TRE have been previously reported [6]. In summary, avoid collinear placement of markers, use as many markers as possible, keep markers as far apart as possible, and place markers so that they surround the surgical target of interest.

The surgeon must be aware that TRE is a statistical measure. Recall that the registration is the best fit overall for the fiducial markers. Because the error in their localization is random, the error in target registration is random as well. Thus, target registration error is reported as a statistical distribution: TRE is \( x^2 \) distributed with 3 degrees of freedom [7]. The surgeon must be aware that TRE for any given point within the surgical field falls within this distribution. Ninety-five percent of the time, TRE can be expected to be less than approximately 1.7 times its mean value, but outliers may exist. For example, if the TRE of
a system is reported as 1 mm, the accuracy is, on average, 1 mm, but it may have considerably larger values (Fig. 2).

It is important to point out that for the surgeon the most critical error is TRE: the error associated with placing a tracked probe on the anatomic point of interest during surgical intervention. Akin to GPSs, TRE can be visualized in terms of a mountaineer climbing a tall peak (Fig. 3). In this illustration, the climber is checking his GPS to see whether it agrees with his assessment of reaching the zenith. The distance between where the GPS says he is and his actual location is the TRE of the system.

**Illustrative analogy**

An analogy to understanding FRE, FLE, and TRE can be made to fitting a knight with a suit of armor. The fiducial markers in this analogy are the key points necessary to align the suit to the knight. These fiducials would consist of neck size, arm length, waist size, hip size, chest size, and inseam. The FLE on the suit would be the error in the construction of the armor such that the sizes are not exact. The FLE on the suit would be the error in accurately measuring sizes. Dressing would be analogous to registration. The fit of the suit at the neck, arm, and other measured places would be analogous to FRE. The analogy of TRE would be the fit of the suit in areas of interest—such as the forearm or thigh—that were not precisely measured by the tailor. The suit fits better in some areas than in others, and, if the process is repeated with random measurement errors, some suits will fit better than others. Thus a distribution, both spatial and statistical, describes the fit better than a single number.

**Error in commercially available image-guided surgical skin-based systems**

With this background in place, error reporting for commercial IGS systems can be understood. This process is complicated by the fact that not all manufacturers abide by the aforementioned error analysis, which is the standard in the engineering community [8,9]. The sinus literature is replete with accuracy studies that include vague methodology but do not specify what type of error is being reported. This lack of standardization impairs a surgeon's ability to compare systems and to assess quickly the error of a system. Because TRE (the distance between an anatomic location of interest, e.g., the sphenoid ostium, and that location on the CT scan) is what the surgeon is interested in, we have attempted to extract TRE from the published literature. Recall that for each system, the number cited in Table 1 is only a mean value for a distribution of error. Thus, the true TRE, or accuracy, of the system will vary from one patient to the next, based on the statistical distribution.
of the skin as opposed to relying on bone-implanted devices. They are (in alphabetic order) the BrainLAB system (BrainLAB, Heimstetten, Germany), the InstaTrak System (GE Medical Systems, Lawrence, MA), the LandmarX system (Xomed, Jacksonville, FL), and the Stryker Image Guidance System (Stryker Leibinger, Kalamazoo, MI). For each of these systems, the literature was reviewed to find the most relevant article reporting accuracy of the system. In addition, each company was contacted to determine whether the cited study was the most up to date regarding error analysis.

For the BrainLAB system, no error studies could be identified specifically for sinus surgery, but such work has been done for neurosurgical applications, which use similar fiducials and registration methods. In a study performed at the University of Regensburg, Germany, error analysis for 36 patients undergoing intracranial surgery was performed [10]. FRE using skin-affixed fiducial markers was reported as 1.10 ± 0.53 mm and, using laser skin contouring, as 1.36 ± 0.34 mm. TRE was calculated by comparing a skin-affixed marker on the patient with that location in the corresponding radiographic image; for the skin-affixed marker registration, TRE was reported as 1.31 ± 0.87, and for laser skin contouring, TRE was reported as 2.77 ± 1.64 mm.

For the InstaTrak System, a multicenter, multisurgeon study was performed and published in *Laryngoscope* in 1997 [11]. In this study, registration was performed either with six skin-affixed markers or with a proprietary headset. TRE was measured using two skin-affixed targets placed on the lateral right and left supraorbital rims. Fiducial registration error was not reported. For the skin-affixed fiducial registration, a TRE of 1.97 mm with a 95% CI of 1.75 to 2.23 and a maximum value of 6.09 mm was reported. For the headset registration, a TRE of 2.28 mm with a 95% CI of 2.02 to 2.53 and a maximum value of 5.08 mm was reported.

For the LandmarX system, Metson *et al.* [2] performed a prospective study of 34 physicians performing 754 sinus cases over a 2½-year period [2]. Using five anatomic key points as fiducials (these points are undefined), they report, “mean accuracy of anatomical localization at the start of surgery was 1.69 ± 0.38 mm (range, 1.53 ± 0.41 mm to 1.79 ± 0.53 mm).” Although the specific methods are not given, this study appears to report FLE—the error associated with repeated identification of the fiducials—and not TRE—hence the question mark in Table 1.

The Stryker Image Guidance System was evaluated prospectively on 50 patients undergoing anterior cranial base surgery by Snyderman *et al.* [12]. Each patient had 10 skin-affixed fiducials placed over the scalp, lateral face, and mastoid. TRE was not measured in this study. Rather, the authors report the error within a zone of accuracy, which is an estimate of TRE based on FRE performed by an algorithm proprietary to Stryker. They report for their clinical applications that the Stryker system estimates that in a zone of accuracy, which it demarcates in image space, it has achieved a TRE less than 2 mm. The authors visually validated the registrations, but they report no independent error measurements to check these estimates.

### Table 1. TRE for commercially-available otolaryngologic IGS systems

<table>
<thead>
<tr>
<th>IGS System</th>
<th>Manufacturer</th>
<th>TRE (mm)</th>
<th>Fiducial</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>BrainLAB™</td>
<td>BrainLAB, Heimstetten, Germany</td>
<td>1.31 ± 0.87</td>
<td>Skin affixed markers</td>
<td>Schlaier, Warnat, and Brawanski, 2002 [10]</td>
</tr>
<tr>
<td>InstaTrak™</td>
<td>GE Medical Systems, Lawrence, MA</td>
<td>2.77 ± 1.64</td>
<td>Laser skin contour</td>
<td>Fried, Kleefield, Gopal et al., 1997 [11]</td>
</tr>
<tr>
<td>LandmarX™</td>
<td>Xomed, Inc., Jacksonville, FL</td>
<td>1.69 ± 0.38 (?)</td>
<td>Skin affixed markers</td>
<td>Metson, Cosenza, Dunningham, Randolph, 2000 [2]</td>
</tr>
</tbody>
</table>
Conclusion

There is a lack of standardization of error reporting for IGS systems. The engineering literature abides by a standard reporting FLE, FRE, and TRE for each set-up. This standard is not followed in clinical reports and makes comparisons among systems quite difficult.

Accuracy depends on the fiducial markers. Reducing FLE is crucial for reducing TRE. The surgeon must be an active participant in reducing FLE by accurately localizing markers on the patient before beginning surgery. In addition, the surgeon must keep in mind that more accurate TRE will be obtained by (1) using more stable markers to reduce FLE, (2) avoiding collinear placement of markers, (3) using more markers, (4) placing the markers far apart, and (5) placing the markers surrounding the surgical target of interest [6].

The numeric value an IGS system returns during registration (FRE or derivative) is only a guide for the accuracy of the system and must be interpreted correctly. Consistently small values for FRE indicate good registration and, thus, a highly accurate surgical navigation system. However, once FRE falls below a threshold, it gives no further information regarding accuracy. Thus, the surgeon should not be preoccupied with achieving extremely low FRE during registration.

Using skin-based fiducial systems, mean TRE for commercially available IGS systems for functional endoscopic sinus surgery approximates 2 mm. Although most studies report a mean value, the surgeon must keep in mind that this mean describes a distribution complete with outliers. Larger and smaller errors are expected to be encountered during each operation.

There is no substitute for anatomic knowledge! Although IGS systems are becoming more and more commonplace, the most commonly used systems are skin-based, and their error cannot be ignored. Thus, they function to complement—not replace—knowledge of surgical anatomy.

As the intent of this article was to describe a concept (accuracy in image-guided surgery) rather than to review the literature, articles are not highlighted for “special interest” or “outstanding interest.”

References