



Designation: ~~F2554 – 18~~ F2554 – 22

Standard Practice for Measurement of Positional Accuracy of Computer-Assisted Computer-Assisted Surgical Systems¹

This standard is issued under the fixed designation F2554; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reappraisal. A superscript epsilon (ϵ) indicates an editorial change since the last revision or reappraisal.

1. Scope

~~1.1 This practice addresses the techniques of document provides procedures for measurement and reporting of basic static performance (accuracy, repeatability, and so forth) of surgical navigation and/or robotic positioning devices under defined conditions. The scope covers the tracking subsystem, testing only in this practice the accuracy and repeatability. They can be performed on a subsystem (for example, tracking only) or a full computer-aided surgery system as would be used clinically. Testing a subsystem does not mean that the whole system has been tested. The functionality to be tested based on this practice is limited to the performance (accuracy in terms of bias and precision) of the system to locate individual points in space, regarding point localization in space by means of a pointer. A point in space has no orientation; only multi-dimensional/multidimensional objects have orientation. Therefore, orientation of objects is not within the scope of this practice. However, in localizing a point the different orientations of the localization tool/pointer can produce errors. These errors and the orientation of the localization tool/pointer orientation are within the scope of this practice. The aim is to provide a standardized measurement of performance variables by which end-users/end users can compare within a system (for example, with different fixed-reference frames/elements or stylus tools) – pointers) and between different systems (for example, different manufacturers) – different systems. – from different manufacturers). Parameters to be evaluated include (based upon the features of the system being evaluated):~~

- ~~(1) Location/Accuracy of a single point relative to a coordinate system.~~
- ~~(2) Relative point to point accuracy (linear). Sensitivity of tracking accuracy due to changes in pointer orientation.~~
- ~~(3) Repeatability of coordinates of a single point. Relative point-to-point accuracy.~~
- ~~(4) For an optically based system, the range of visible orientations of the reference frames or tools.~~
- ~~(5) This method covers all configurations of tool arrays in the system.~~

~~1.1.1 This method covers all configurations of the evaluated system as well as extreme placements across the measurement volume.~~

~~1.2 The system as defined in this practice includes only the tracking subsystem (optical, magnetic, mechanical, and so forth) stylus, computer, and necessary hardware and software. As such, this practice incorporates tests that can be applied to a prescribed phantom model in a laboratory or controlled setting.~~

1.2 This practice defines a standardized reporting format, which includes definition of the coordinate systems to be used for reporting the measurements, and statistical measures (for example, mean, standard deviation, RMS, and maximum error).

~~1.4 This practice will serve as the basis for subsequent standards for specific tasks (cutting, drilling, milling, reaming, biopsy needle placement, and so forth) and surgical applications.~~

¹ This practice is under the jurisdiction of ASTM Committee F04 on Medical and Surgical Materials and Devices and is the direct responsibility of Subcommittee F04.38 on Computer Assisted Orthopaedic Surgical Systems.

Current edition approved Nov. 1, 2018/Sept. 1, 2022. Published December 2018/September 2022. Originally approved in 2010. Last previous edition approved in 2010/2018 as F2554/F2554 – 18, –10. DOI 10.1520/F2554-18.10.1520/F2554-22.

1.3 The values stated in SI units are to be regarded as standard. No other units of measurement are included in this standard, except for angular measurements, which may be reported in terms of radians or degrees.

1.4 This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety, health, and environmental practices and determine the applicability of regulatory limitations prior to use.

1.5 This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.

2. Referenced Documents

2.1 ASTM Standards:²

E456 Terminology Relating to Quality and Statistics

E2281 Practice for Process Capability and Performance Measurement

2.2 Other References:³

ISO 10360 Geometrical Product Specifications (GPS)—Acceptance and Reverification Tests for Coordinate Measuring Machines (CMM)

3. Terminology

3.1 ~~Definition of Terms Specific to Accuracy Reporting~~

3.1.1 ~~accuracy, *n*—the closeness of agreement between a measurement result and an accepted reference value. **E456**~~

3.1.1.1 ~~Discussion—~~

~~The term accuracy, when applied to a set of measurement results, involves a combination of a random component and of a common systematic error or bias component.~~

3.1.2 ~~bias, *n*—the difference between the expectation of the measurement results and an accepted reference value. **E456**~~

3.1.2.1 ~~Discussion—~~

~~Bias is the total systematic error as contrasted to random error. There may be one or more systematic error components contributing to the bias. A larger systematic difference from the accepted reference value is reflected by a larger bias value.~~

3.1.3 ~~maximum error, *n*—the largest distance between any measured point and its corresponding reference position (for example, as measured by CMM) for any trial during a testing procedure.~~

3.1.4 ~~mean, *n*—the arithmetic mean (or simply the mean) of a list of numbers is the sum of all the members of the list divided by the number of items in the list. If one particular number occurs more times than others in the list, it is called a mode. The arithmetic mean is what students are taught very early to call the “average”. If the list is a statistical population, then the mean of that population is called a population mean. If the list is a statistical sample, we call the resulting statistic a sample mean.~~

3.1.5 ~~measurement range, *n*—see measurement volume.~~

3.1.6 ~~precision, *n*—the closeness of agreement between independent measurement results obtained under stipulated conditions. **E456**~~

3.1.6.1 ~~Discussion—~~

~~Precision depends on random errors and does not relate to the true value or the specified value. The measure of precision usually is expressed in terms of imprecision and computed as a standard deviation of the test results. The standard deviation is expressed as:~~

$$S = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N - 1}}$$

~~Less precision is reflected by a larger standard deviation. “Independent test results” means results obtained in a manner not~~

² For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard’s Document Summary page on the ASTM website.

³ Available from American National Standards Institute (ANSI), 25 W. 43rd St., 4th Floor, New York, NY 10036, <http://www.ansi.org>.

~~influenced by any previous result on the same or similar test object. Quantitative measure of precision depends critically on the stipulated conditions. Repeatability and reproducibility conditions are particular sets of extreme stipulated conditions.~~

~~3.1.7 *range, R, n*—the largest observation minus the smallest observation in a set of values or observations. **E456, E2281**~~

~~3.1.8 *repeatability, n*—precision under repeatability conditions. **E456**~~

~~3.1.8.1 *Discussion*—~~

~~Repeatability is one of the concepts or categories of the precision of a test method. Measures of repeatability defined in this compilation are repeatability, standard deviation, and repeatability limit.~~

~~3.1.9 *reproducibility, n*—precision under reproducibility conditions. **E456**~~

~~3.1.9.1 *Discussion*—~~

~~Ability of a test or experiment to be accurately reproduced, or replicated.~~

~~3.1.10 *resolution, n*—of a device/sensor, the smallest change the device or sensor can detect in the quantity that it is measuring. The resolution is related to the precision with which the measurement is made.~~

~~3.1.11 *standard deviation, n*—the most usual measure of the dispersion of observed values or results expressed as the positive square root of the variance. **E456**~~

~~3.1.12 *variance, n*—of a random variable, measure of its statistical dispersion, indicating how its possible values are spread around the expected value. Where the expected value shows the location of the distribution, the variance indicates the scale of the values. A more understandable measure is the square root of the variance, called the standard deviation.~~

~~3.1 *Definition of Terms Specific to Surgical Navigation and Robotic Positioning Systems: Definitions:*~~

~~3.1.1 *data integrity, accuracy, n*—condition in which data is identically maintained during any operation, such as transfer, storage, and retrieval; the closeness of agreement between a measurement result and an accepted reference value. **E456**~~

~~3.1.1.1 *Discussion*—~~

~~In the context of this standard, with the definitions of bias and precision (see below), it can be considered that the accuracy of a measurement of a point will be subject to some bias error and some precision error.~~

~~3.1.2 *bias, n*—the difference between the expectation of the measurement results and an accepted reference value. **E456**~~

~~3.1.2.1 *Discussion*—~~

~~In the context of this standard, bias represents the systematic error in a set of measurements of a target reference point making their average deviate from the actual reference point with a certain magnitude and direction.~~

~~3.1.3 *calibration, n*—the pre- or intraoperative registration of an item or device to its reference element.~~

~~3.1.4 *computer-assisted surgery (CAS), n*—the use of computers to facilitate or enhance surgical procedures via the use of three-dimensional space tracking of objects.~~

~~3.1.5 *coordinate measuring machine (CMM), n*—measuring system with the means to move a stylus and capability to determine spatial coordinates on a work piece surface. **ISO 10360-1**~~

~~3.1.6 *degree of freedom (DOF), n*—set of independent displacements that specify completely the displaced or deformed position of the body or system.~~

~~3.1.7 *dynamic reference base, n*—the coordinate system of a reference element that is intraoperatively attached to a therapeutic object and allows tracking that object. It defines the local coordinate system of the therapeutic object used for the tracking of other therapeutic objects.~~

~~3.1.8 *fiducial, n*—an artificial object/item (for example, a screw or sphere) that is implanted into, or a feature created on, a therapeutic object prior to virtual object acquisition to facilitate registration; a sphere) rigidly attached to a therapeutic object to facilitate its calibration.~~

3.1.9 *ground truth, n*—short name for the accepted reference value (see 3.1.1).

3.1.10 *marker, n*—a single 3-degree-of-freedom indicator on a reference element or dynamic reference base.

3.1.11 *maximum error, n*—the largest distance between any measured point and its ground truth for any trial during a testing procedure.

3.1.12 *mean, n*—of a population, μ , average or expected value of a characteristic in a population; of a sample, \bar{x} , sum of the observed values in the sample divided by the sample size. **E456**

3.1.13 *measurement range, n*—the interval of allowed values for a specific degree of freedom while performing the practice.

3.1.14 *measurement volume, n*—measuring range of a tracker, stated as simultaneous limits on all spatial coordinates measured by the tracker. **ISO 10360-1**

3.1.15 *navigation system, n*—~~a device set of devices consisting of a computer with~~computer, its associated software, and a localizer that tracks reference elements attached to surgical instruments or implants as well as one or more dynamic reference bases attached to the therapeutic object. ~~It tracker capturing the reference elements within the measurement volume. This system provides real-time feedback of the performed action by visualizing it within the virtual environment.~~state of the surgical scene under operation.

3.1.16 *phantom, n*—standardized measurement object. See **Appendix X1** for details regarding the design of the phantom used in this practice.

3.1.17 *pointer, n*—the device offered by the evaluated system to point and locate a position on any object including anatomical landmarks. The pointer is the whole device, including the stylus-like tip all the way to any reference element used to track it in space.

3.1.18 *precision, n*—the closeness of agreement between independent measurement results obtained under stipulated conditions. **E456**

3.1.18.1 *Discussion—*

In the context of this standard, precision represents scatter of a set of measurements of a point.

3.1.19 *range, R, n*—the largest observation minus the smallest observation in a set of values or observations. **E456, E2281**

3.1.20 *reference element, n*—~~a device attached to surgical instruments and implants and other devices that enables determination of an artificial item composed of rigidly bound markers in a unique and asymmetrical pattern recognizable by the tracker. While being rigidly attached to a therapeutic object, the position and orientation in 3d space (up to 6 degrees of freedom) of these by means of a tracker. It defines the local coordinate system of this instrument or implant.~~of the reference element can be used to determine those of the therapeutic object after its calibration.

3.2.9 *referencing, n*—tracking of a therapeutic object by means of a dynamic reference base.

3.1.21 *registration, n*—the determination of the ~~transformation~~spatial relationship between the coordinate spaces of the ~~therapeutic and virtual objects~~referential frames of two coordinate systems. This may occur between two reference elements or between the coordinate spaces of two virtual objects. Afiducials and a reference element of a therapeutic object. The registration is rigid if it consists only of ~~rotations, translations, and scaling~~; it is ~~rotations and translations (six degrees of freedom) and non-rigid if it also comprises scaling and/or local or global distortions~~distortions (seven degrees of freedom and more).

3.1.22 *repeatability, n*—precision under repeatability conditions. **E456**

3.1.23 *reproducibility, n*—precision under reproducibility conditions. **E456**

3.1.24 *robotic positioning system, n*—use of an active mechanical (mechatronic) device to position an instrument guide at a specified location in 3d3D space (up to 6six degrees of freedom).

3.1.25 *stylus, root mean square (RMS), n*—~~a mechanical device consisting of a stylus tip and a shaft. The stylus tip is the physical element that establishes the contact with the workpiece.~~ means of estimation of the scatter of a set of values, which consists of the square root of the average of the squared values. **ISO 10360-1**

3.1.26 ~~tool calibration, therapeutic object, n~~—the pre- or intraoperative determination of the location of points of interest on a navigated instrument (for example, its tip position, axis) in relation to a reference frame (for example, the attached reference element for a tracked instrument): a surgical item or a part of the patient.

3.1.27 *tracker, n*—a device that measures the spatial location and orientation of surgical instruments implants, or the therapeutic object that are instrumented with reference elements or a dynamic reference base respectively. A tracker may measure based on infrared light (see tracking, active and tracking, passive), ultrasound, electromagnetic fields, mechanical linkage, video streams, and so forth. detects and locates fiducials and markers in its measurement volume. This can be achieved by mechanical linkage or by analyzing signals of various types (visible or infrared light, electromagnetic field, or ultrasound).

3.2.15 *tracking, active, n*—a tracking technology that uses markers that emit energy (for example, an infrared light based tracking technology that uses pulsed LEDs as markers, ultrasound, electromagnetic fields, and so forth).

3.2.16 *tracking, passive, n*—a tracking technology that uses markers that absorb or reflect externally produced energy. (for example, an light based tracking technology that uses reflective spheres or similar objects as markers).

3.3 Others:

3.3.1 *computer assisted surgery (CAS), n*—the use of computers to facilitate or enhance Surgical Procedures via the use of three-dimensional space tracking of objects.

3.3.2 *coordinate measuring machine (CMM), n*—measuring system with the means to move a stylus and capability to determine spatial coordinates on a work piece surface. **ISO 10360-1**

3.3.3 *phantom, n*—standardized measurement object. See **Appendix X1** for recommendations regarding phantom design. Specific points referenced in this practice are with regards to the recommended phantom design in **Appendix X1**.

4. Summary of Practice

4.1 This practice provides recommendations for the collection, analysis, and presentation of data regarding the positional accuracy of (in terms of bias and precision) of surgical navigation and robotic positioning systems. systems under repeatable conditions.

4.2 Data to be provided include measured statistical distribution, maximum error, mean, and standard deviations, 5th and 95th percentiles of location and orientation accuracy. reported consists of all measurements, their corresponding errors if applicable, their statistical analysis, the test conditions, and the system conditions.

4.3 This practice provides protocols (Section 8) for measuring accuracy of the tracking system (optical, magnetic, mechanical, and so forth) made under repeatable conditions. Subsequent standards will address the system along with any necessary imaging modality (fluoroscopy, computed tomography, magnetic resonance imaging, ultrasound, and so forth) for image based systems, and the software for registering the images or the imageless data to the patient. Additional standards will also address task specific procedures and surgical applications (joint arthroplasty, osteotomy, tumor biopsy and/or resection, laproscopy, pedicle screw insertion, brain surgery, and so forth).

5. Significance and Use

5.1 The purpose of this practice is to provide data that can be used for evaluation of the accuracy of different CAS systems.

5.2 The use of surgical navigation and robotic positioning systems is becoming increasingly common and requires a degree of trust

by the user that the data provided by the system meets necessary accuracy requirements. In order to evaluate the potential use of these systems, and to make informed decisions about the suitability of a system for a given procedure, objective performance data of such systems are necessary. While the end user will ultimately want to know the accuracy parameters of a system under clinical application, the first their accuracy capability needs to be evaluated under clinical application and compared to the requirements. As the performance of a whole system is constrained by those of its subparts, a preliminary step must be to objectively characterize the digitization accuracy of the tracking subsystem in a controlled environment under controlled conditions.

5.3 In order to make comparisons within and between systems, a standardized way of measuring and reporting accuracy is needed. Parameters such as coordinate system, units of measurement, terminology, and operational conditions must be standardized.

6. Apparatus

6.1 The system under test is considered to have at least some tracking functionality, a pointer and associated hardware, and software. If the system is provided by the manufacturer with various combinations of parts, the evaluation must be performed at least with the combination known to present the worst-case scenario in terms of accuracy. For example, the tester may use the longest pointer with the smallest reference elements.

6.2 Standardized measurement object (phantom). This practice relies on a phantom. See ~~X1.1~~ Appendix X1 for design requirements. The phantom size and points have been designed to approximate a typical surgical site on the human body. All divots of the phantom shall be measured by a CMM (or another measurement system of similar performance traceable to NIST, FDA, EU, and ISO standards). These phantom measurements will constitute the ground truth used for the accuracy assessment. Therefore, the accuracy of the CMM must be better than that of the system being evaluated.

6.3 If the evaluated system relies on a dynamic reference base for its measurements, a reference element is attached to the phantom. This reference element and its attachment shall replicate as close as possible those used in a surgical setting.

6.4 System to be evaluated, including tracking system, stylus, and associated required hardware and software. While the software may be custom-written for If dedicated additional software functionality is used for assistance in performing the tasks outlined in this practice it should use the same algorithms and methodologies being implemented in the commercial/clinical practice, or for statistical analysis of the measurements, this addition must not alter the way the measurements are made by the system to be assessed used clinically.

7. Hazards

7.1 None.

8. Procedure

8.1 A standardized measurement object (phantom) measured with a CMM or similar measurement device calibrated to a traceable standard, will be used to evaluate the accuracy of the tracking subsystem. The resolution of this measurement device will determine the significant digits that can be reported in the results See Appendix X1 for phantom recommendations. The accuracy of the measurement device must be at least 4×, and preferably 10×, as accurate as the anticipated accuracy to be reported. The phantom shall be characterized by detailed measurements of the x,y,z measurements of all its points using the aforementioned device (see the example in Table X1.1.)

8.1 Rigid attachment of a dynamic reference base to the phantom is to be performed according to navigation system manufacturer's recommendations. Once testing has begun, the dynamic reference base should be set up the system to be evaluated, including the dynamic reference base, if used. The dynamic reference base shall not be repositioned relative to the phantom. Note that some systems, such as Robotic Positioning Systems, may not use a dynamic reference base. Measured points should phantom during the procedure. Measured points shall be expressed in the coordinate system defined by the dynamic reference base where applicable of the phantom, to enable them to be directly compared to those made by CMM.

8.2 Test Conditions—The conditions used for this procedure should be as close as possible to those typically found in the surgical setting for which the system is designed, and those conditions shall be reported. They include all known factors that may influence

the performance of the evaluated system, such as temperature, humidity, lighting, as well as potential sources of interference (for example, infrared noise and reflections for optical systems, large metal objects for electromagnetic systems). Some of these factors may not be replicated if their non-inclusion is reported and justified.

8.3 System Conditions—The system is composed of various parts and all their references and configuration shall be provided, including firmware and software versions. Any changes to the system beyond what is provided and configured by the manufacturer are to be reported and justified (for example, using third-party markers or pointer). Specific details of the phantom are also to be reported (for example, the divot dimensions). The measured points are to be acquired only through the firmware and software provided by the manufacturer.

8.4 Phantom Placement and Registration—In the first series of tests, place the system tracker nominally at the recommended distance from the phantom. At this location, a registration of the phantom to the dynamic reference base may be required for most systems. Any registration shall be performed as described by the manufacturer, simulating registration of a patient’s anatomy in the clinical environment. Registration can only be done once to cover the sequence of steps 8.6 – 8.9, but may be repeated in between.

8.5 Point Acquisition—In each trial, the tester locates the individual labeled points on the phantom and acquires its position using the pointer following the system manufacturer’s instructions for obtaining point data. This includes pointer orientation except for the rotation tests in 8.7.

8.6 Test 1: Single Point Accuracy—Treating the phantom as if it was part of the patient’s anatomy, this test requires the measurement of a designated point of the phantom multiple times, to compare the positions measured versus the actual position on the phantom relative to its local coordinate system. The single point measurement is then independently performed 20 times on the central divot (#20 in Fig. X1.1). Bias is estimated by the difference between the average of the measured points and the central divot. The result is a small error vector emanating from the target reference point. Calculate the average of all the error vectors by vectorial summation then dividing the length of the resulting vector by the number of vectors. Report the average error vector and the length of the longest error vector. For the determination of precision, start by calculating the average point of all measurements, which represents the system’s best estimate for the location of the target central divot. Calculate the distances of all the measurements from this average point, determine and report the RMS of these distances as well as their maximum.

8.7 Test Conditions: Pointer Rotation—

8.3.1 Test conditions (temperature, humidity, and so forth) should be recorded and reported (Section 9). Known potential sources of interference should be replicated by simulating as realistic an environment comparable to the operating room conditions (for example, infrared noise and reflections for optical systems, large metal objects for electromagnetic systems, and so forth). These steps quantify the variation in successive measurements of the same central divot (#20 Fig. X1.1) with various pointer orientations (tilt). The maximum physically possible angular range depends on the axis of rotation, but most systems have some limit to this range beyond which tracking is no longer provided. The actual range for which the system provides tracking data should be tested at no fewer than ten uniformly spaced intervals. The actual range, all angular increments, and the corresponding measurements shall be reported. The tester may use a protractor scale to measure the pointer angles or any other method. For each test in 8.7, the maximum distance between any two measurements shall be reported, as well as the RMS of the deviations between all measurements and the average point measured in 8.6 (Test 1).

8.3.2 System Conditions—Any changes to standard system (tool or stylus configuration, and so forth) configuration must be reported (Section 9). Specific details of the phantom to be used, with particular regard to the divot geometry (see Fig. X1.3).

8.3.3 Single point measurements to quantify accuracy of single points and effect of tool orientation. A single calibration point on the phantom will be digitized using the specifically designated and identified device (stylus, pointer, and so forth) specified by the Navigation and/or Robotic System manufacturer. Data will be collected by software on the Navigation and/or Robotic System. The measurement shall consist of an appropriate sample size, with a minimum of six independent trials performed at a single point (#20 Fig. X1.1) on the phantom. All measured tool tip positions shall be averaged to obtain the average tool tip position. The difference in distance between each measured tip position and this average shall be computed. The mean of the absolute error values, maximum error magnitude, and standard deviation of these calculated differences shall be reported. Optionally, this computation can also be performed for each of the 3 substeps below (that is, compute mean, maximum, and standard deviation of the calculated differences for just the data obtained in step 8.3.4, 8.3.7):

8.7.1 Test 2—The purpose of this step is to evaluate whether accuracy of the coordinates measured at the tip of the tool is affected