Ultrasound Metrology in Mexico: a round robin test for medical diagnostics

R Amezola Luna, A L López Sánchez, A A Elías Juárez
Centro Nacional de Metrología (CENAM, México),
Carretera a Los Cués km.4.5, El Marqués, Querétaro, México. 76246.

E-mail: ramezola@cenam.mx

Abstract. This paper presents preliminary statistical results from an on-going imaging medical ultrasound study, of particular relevance for gynecology and obstetrics areas. Its scope is twofold, firstly to compile the medical ultrasound infrastructure available in cities of Querétaro-Mexico, and second to promote the use of traceable measurement standards as a key aspect to assure quality of ultrasound examinations performed by medical specialists. The experimental methodology is based on a round robin test using an ultrasound phantom for medical imaging. The physician, using its own ultrasound machine, couplant and facilities, measures the size and depth of a set of pre-defined reflecting and absorbing targets of the reference phantom, which simulate human illnesses. Measurements performed give the medical specialist an objective feedback regarding some performance characteristics of their ultrasound examination systems, such as measurement system accuracy, dead zone, axial resolution, depth of penetration and anechoic targets detection. By the end of March 2010, 66 entities with medical ultrasound facilities, from both public and private institutions, have performed measurements. A network of medical ultrasound calibration laboratories in Mexico, with traceability to The International System of Units via national measurement standards, may indeed contribute to reduce measurement deviations and thus attain better diagnostics.

1. Introduction
Ultrasound medical imaging examinations are a useful and reliable tool to perform medical diagnostics. In Mexico, close to 1 M of gyn-obstetric diagnostics by ultrasound are performed in the private sector [1], and it is estimated a similar amount in the public sector. Medical specialists can measure size and depth of human tissues and organs in order to diagnose and recommend a medical treatment. Therefore, the significance of detection and measurement accuracy is related to treatment effectiveness and thus patient’s health. Nevertheless, accurateness could be affected by factors such as capabilities of the equipment (ultrasound machine), medical expertise, patient biological characteristics and environmental conditions. Each factor may affect measurement reliability and decrease the quality to diagnose a patient. In order to establish if quality assurance programs, with
traceability to national standards, may indeed improve the quality of medical ultrasound imaging diagnostics, the Centro Nacional de Metrología (CENAM) started a field study focusing on gynaecology and obstetrics areas. The main goals were to compile the medical ultrasound infrastructure available in a relatively small Mexican province (State of Queretaro) with a population close to 1.6 M [2], in both public and private medical units. Information regarding location and amount of medical units and its specific ultrasound infrastructure for medical diagnostics, as well as how it is being managed and maintained were the starting issues for this study. In addition, considering that reference standards and calibration are metrological concepts not frequently found in the medical sector, it was set to gather information about the use of traceable measurement standards and how quality of ultrasound examinations are assured by medical specialists. In Mexico, usually, calibration of the ultrasound machine is not included during its periodic maintenance program; therefore users do not have enough information to assure its correct operation. This study, as far as the authors know, has not been done in Mexico before; it grew out as a concern of the government of the State of Queretaro about the actual operation of the ultrasound infrastructure available in the State, seeking to improve the quality of ultrasound medical services. To achieve these goals, a round robin test was setup among different medical units in cities belonging to the State of Queretaro. Partial results corresponding to on-site measurements from June 2009 to March 2010 are presented in the following sections.

2. Description of the round robin test
The experimental methodology is based on a round robin test using an ultrasound phantom for medical imaging. The phantom simulates acoustic properties of human tissue and is constructed of rubber-based tissue-mimicking material with nominal sound speed of 1450 m/s at 0.5 dB/cm/MHz at room temperature (23 °C) [3]. The phantom contains line targets and anechoic target structures simulating some biological tissues (see figure 1). During the test, staff from CENAM visited each medical unit and gave details about the measurement procedure. Each physician, using its own ultrasound machine, couplant and facilities, measured the depth, lateral distance or size of a set of pre-defined reflecting and anechoic targets of the reference phantom. Due to time restrictions, measurements taken were mainly limited to the use of one linear transducer and one convex transducer (average time for measuring: 20 min per each transducer).

Figure 1. Ultrasound phantom (artifact).
Measurements performed during the round robin test [4, 5, 6]:

a) **Measurement system accuracy.** The horizontal spacing between 2 consecutive line targets from the Horizontal Group was determined (nominal distance of 20 mm). Also, a vertical distance of 170 mm (for convex transducer) from the Vertical Group was measured.

b) **Dead zone.** The closest line target from the Dead Zone Group to the phantom entry surface was identified and its depth was measured. Closest line target available to the entry surface of the phantom is 2 mm, nominal. During the test, no supplementary accessory was used to perform measurements (e.g. delay line bag or couplers).

c) **Resolution.** Focusing on the Axial-Lateral Resolution Group, system capability to resolve between 2 contiguous targets in the axial axis was determined. The group of targets is form by 6 line targets with interval spacing of (5, 4, 3, 2 and 1) mm, nominal.

d) **Depth of penetration.** Maximum distance from the phantom entry surface to the farthest line target of the Vertical Group was measured. The depth of the farthest line target available in phantom is 170 mm, nominal.

e) **Anechoic target detection.** The diameter of three anechoic targets was determined. Nominal diameters of (10, 8 and 4) mm. Measurements were performed using linear transducers.

Every participant performed the same set of measurements referred above, however due to limitations of the ultrasound machines some measurements were not able.

3. **Results and discussion**

Measurements from 66 medical institutions (including both public and private health centers) are presented in here as preliminary results, corresponding to 82 individual ultrasound machines.

3.1. **Measurement system accuracy**

Measurement deviation was determined by:

\[ E_U = L_U - L_{ref} \]  \hspace{1cm} (1)

where

- \( E_U \) = Difference between the length measured by every individual ultrasound machine from the medical entity and the reference value, mm.
- \( L_U \) = Length measured with the ultrasound machine, mm.
- \( L_{ref} \) = Reference value taken as the average value from all measurements, mm.

Normalized error was calculated as [7],

\[ e_U = \frac{|L_U - L_{ref}|}{\sqrt{U_{L_U}^2 + U_{L_{ref}}^2}} \]  \hspace{1cm} (2)

where

- \( e_U \) = Normalized error. If >1, inconsistent results.
- \( U_{L_U} \) = Expanded uncertainty of the results obtained by the ultrasound machine from the medical entity participant, mm.
- \( U_{L_{ref}} \) = Expanded uncertainty of the reference value, mm.
Figure 2. Measurement system accuracy on (a) horizontal axis, and (b) vertical axis.

Results for Horizontal and Vertical Accuracy are shown in figure 2. Horizontal measurements were performed with linear and convex transducers. Vertical measurements include only convex transducers. Most of the participants have a deviation value within ±2 mm (±10 %) for the horizontal axis and ±5 mm (±3 %) for the vertical axis, respectively. Regarding horizontal accuracy 51 of 68 participants obtained consistent values (normalized error ≤ 1), while for vertical accuracy 74 of 76 participants obtained consistent values.

3.2. Dead zone
Most of the participants with linear transducers reached the closest line target available to the phantom’s entry surface, 2 mm. For convex transducers, widespread dispersion from 1 mm to 7 mm was obtained. This is shown in figure 3. As expected, in general, Dead Zone for linear transducers is smaller than for convex transducers. Some ultrasound machines had multi-frequency capabilities which allowed them to improve their measurements on targets corresponding to the dead zone.

3.3. Resolution
For convex transducers, 63.3 % of the individual ultrasound machines were able to resolve all the line targets of the Resolution Group, including the pair of targets separated 1 mm. The rest of the participants were able to resolve the line targets separated 2 mm.
For linear transducers, 79.3% of the medical entities were able to resolve up to 1 mm. Ultrasound machine capabilities for imaging and configuration had a considerable influence in the results. As expected, linear transducers showed much better resolution capabilities than convex transducers.

3.4. Depth of penetration

Most of the participants with convex transducers identified a line target at the maximum depth available in the phantom, 170 mm. For linear transducers, a widespread dispersion is shown from approximately 60 mm up to 170 mm, some of them only reached 80 mm due to maximum available range in the ultrasound machine (scale locked). On the other hand, two linear transducers reached the maximum available depth due to relatively low frequency operation (5 MHz). As expected, convex transducers reached a larger depth of penetration than linear transducers. See figure 4.

3.5. Anechoic target detection

In the case of anechoic targets measurements with a linear transducer, most of the participants are within data dispersion values of ±0.5 mm (±12%) for the 4 mm diameter target, ±1 mm (±12%) for the 8 mm diameter target and ±1.2 mm (±12%) for the 10 mm diameter target. See figure 5.

As items that may require further discussion and analysis, it is noted that measurement accuracy and equipment diagnostics capabilities may need to take into account the effect of factors, such as:

a) Equipment capabilities
   - Image quality
   - Spatial resolution
   - Beam divergence

b) Medical specialist expertise
   - Transducer alignment
   - Equipment operation
   - Measurement technique

c) Environmental conditions
   - Electrical voltage
   - Temperature
   - Humidity

d) Phantom
   - Geometry
   - Sound speed
   - Targets alignment

e) Protocol and logistic
   - Limited time to perform measurements

Above factors may have led to apparent equipment limitations and/or measurement inaccuracies. A line target should produce a snap point image; however an unfocussed beam produces a distorted elliptical form (see figure 6). On the other hand, misalignment of the transducer and measurement path (using the caliper) can give an overestimated length (see figure 7).

![Figure 5. Dispersion of data for anechoic target size.](image-url)
4. Conclusions

Results from this round robin test underline several aspects that need to be taken into account by a medical specialist performing ultrasound examinations or scannings, in order to improve the quality of the diagnostic. Namely, measurement capabilities may degrade as the equipment gets old, and metrological assurance is not the same as corrective maintenance. That is, ultrasound metrology and basic metrological practices may be overlooked. Measurements performed gave an objective feedback to medical specialists regarding dead zone, depth of penetration, resolution and anechoic targets detection of their ultrasound examination systems and their measurement accuracy. It is claimed that the use of an ultrasound phantom is an important tool to verify and monitor the correct operation of ultrasound machines; besides being a useful tool for training purposes.

As part of the future work, it will be convenient to find out if ultrasound machines are indeed periodically calibrated in other countries and how such calibrations are related to ultrasound measurement systems kept by the corresponding NMI. There seems to be a broken traceability chain that may need some attention. Among the remaining questions: 1) Which countries do widely have or enforce ultrasound medical diagnostics with traceability to SI units? and 2) Are there any benefits for a patient to receive ultrasound medical services with traceability to national standards?

5. References

[1] INEGI- 2003 Instituto Nacional de Estadística y Geografía, Mexico.
[2] INEGI- 2005 Instituto Nacional de Estadística y Geografía, Mexico.
[3] ATS Laboratories Phantom model 535. Operation manual.
[4] IEC 61390-96 Ultrasonics- Real time pulse-echo systems- Test procedures to determine performance specifications.
[5] AIUM Standard methods for measuring performance of pulse-echo ultrasound imaging equipment (July 13, 1990).
[6] Goodsitt, Carson, Witt, Hykes and Kofler. Real time B-mode ultrasound quality control test procedures. Report of AAPM Ultrasound task group No.1. Medical Physics, Vol. 25 No. 8, August 1998 (1385 to 1406).
[7] ISO/IEC Guide 43-1 (1997) Proficiency testing by interlaboratory comparisons-Part 1: Development and operation of proficiency testing schemes.

Acknowledgments

Authors thank the kind support of the Consejo Nacional de Ciencia y Tecnología (CONACYT) and the government of the State of Queretaro to perform this study (project FOMIX QRO-2008-C03-107938). Participation of all medical specialists and health institutions is also acknowledged.