

Review of consensus test methods in medical imaging and current practices in photoacoustic image quality assessment

Jorge Palma-Chavez,^{a,†} T. Joshua Pfefer^{b,†}, Anant Agrawal^b,
Jesse V. Jokerst,^{a,c,d} and William C. Vogt^{b,*}

^aUniversity of California San Diego, Department of NanoEngineering, La Jolla, California, United States

^bCenter for Devices and Radiological Health, U.S. Food and Drug Administration, Silver Spring, Maryland, United States

^cUniversity of California San Diego, Department of Radiology, La Jolla, California, United States

^dUniversity of California San Diego, Materials Science and Engineering Program, La Jolla, California, United States

Abstract

Significance: Photoacoustic imaging (PAI) is a powerful emerging technology with broad clinical applications, but consensus test methods are needed to standardize performance evaluation and accelerate translation.

Aim: To review consensus image quality test methods for mature imaging modalities [ultrasound, magnetic resonance imaging (MRI), x-ray CT, and x-ray mammography], identify best practices in phantom design and testing procedures, and compare against current practices in PAI phantom testing.

Approach: We reviewed scientific papers, international standards, clinical accreditation guidelines, and professional society recommendations describing medical image quality test methods. Observations are organized by image quality characteristics (IQCs), including spatial resolution, geometric accuracy, imaging depth, uniformity, sensitivity, low-contrast detectability, and artifacts.

Results: Consensus documents typically prescribed phantom geometry and material property requirements, as well as specific data acquisition and analysis protocols to optimize test consistency and reproducibility. While these documents considered a wide array of IQCs, reported PAI phantom testing focused heavily on in-plane resolution, depth of visualization, and sensitivity. Understudied IQCs that merit further consideration include out-of-plane resolution, geometric accuracy, uniformity, low-contrast detectability, and co-registration accuracy.

Conclusions: Available medical image quality standards provide a blueprint for establishing consensus best practices for photoacoustic image quality assessment and thus hastening PAI technology advancement, translation, and clinical adoption.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JBO.26.9.090901](https://doi.org/10.1117/1.JBO.26.9.090901)]

Keywords: phantoms; standardization; performance testing; photoacoustics; ultrasound; magnetic resonance imaging; x-ray computed tomography; quality assurance.

Paper 210176VSSR received Jun. 1, 2021; accepted for publication Aug. 17, 2021; published online Sep. 11, 2021.

*Address all correspondence to William C. Vogt, william.vogt@fda.hhs.gov

†Equal contribution

1 Introduction

Photoacoustic imaging (PAI) is a rapidly emerging modality that has been proposed for numerous clinical applications including cancer detection, mammography, vascular imaging, tissue oximetry, tumor margining, and biopsy/surgical guidance, among others.¹⁻⁵ This wide range of applications and the novelty of the field has resulted in a large variety in device designs. PAI device performance will generally vary with device design parameters (e.g., transducer geometry, optical source properties) as well as tissue parameters (e.g., properties and morphology). Quantitatively predicting how these parameters influence PAI device performance *in vivo* is challenging. Bench performance test methods can provide insight on design consequences, elucidate device working mechanisms, and help set performance expectations and limitations. Tissue-mimicking phantoms provide an invaluable approach for objective, quantitative evaluation of fundamental image quality characteristics (IQCs) as well as more technology-specific aspects of PAI system performance such as oximetry measurement accuracy, spectral recovery, or chromophore concentration accuracy.⁶⁻¹¹

However, no standardized phantom-based performance test methods have been established for PAI. This places a burden on researchers and device developers to design their own phantoms and test methods, thus increasing development time and cost while potentially causing redundancy of efforts across the community. Comparing device test results against those reported in the literature is also challenging given the variation in phantom design and testing methodology. Consensus PAI performance test methods are needed to facilitate consistent and scientifically rigorous, yet least burdensome evaluation of device performance. Such test methods can support many aspects of the medical product life cycle, including device development and optimization, benchmarking or inter-comparison, clinical trial standardization, quality management systems, regulatory evaluation, post-market studies, constancy testing, calibration, and accreditation. The US Food and Drug Administration (FDA) can formally “recognize” voluntary consensus standards as being suitable for regulatory purposes, which can potentially streamline regulatory decision-making.¹² Standards development is not only a key step in clinical translation and adoption of an imaging modality but may also improve device quality, increase device consistency across manufacturers, and serve as an indicator of technological maturity.

Standardized, phantom-based performance test methods have been developed for mature imaging modalities such as ultrasound, x-ray computed tomography (CT), x-ray mammography, and magnetic resonance imaging (MRI) through standards organizations such as the International Electrotechnical Commission (IEC), International Organization for Standardization (ISO), and National Electrical Manufacturers Association (NEMA). Additionally, consensus documents containing expert recommendations for image quality assessment have been developed by professional societies including the American Association of Physicists in Medicine (AAPM) and the American College of Radiology (ACR), as well as community-led working groups and consortia.¹³⁻¹⁶ These groups have designed accreditation programs that provide facilities performing medical imaging with recommendations on staff qualifications, equipment characteristics, phantom properties, quality control (QC) routines, and quality assurance (QA) tests. Some phantom manufacturers offer products that are specifically designed to meet the requirements of these standards removing the burden of fabrication and characterization from the developer or end user.¹⁷⁻¹⁹ Community interest in addressing these standardization needs is evidenced by the recent establishment of the International Photoacoustics Standardisation Consortium (IPASC), which aims to standardize PAI phantoms and performance test methods.²⁰ There is also a similar rise in standards development activities for other biophotonics technologies, including near-infrared cerebral oximeters²¹ and fluorescence-guided surgery.²²

Our overall goal is to support development of robust, consensus-based performance test methods for emerging PAI devices. We aimed to determine whether available medical imaging standards can be leveraged to inform and guide establishment of standardized test methods for PAI. To this end, we reviewed standards, consensus documents, and clinical accreditation guidelines describing image quality test methods for ultrasound, CT, x-ray mammography, and MRI. We also reviewed the PAI literature to capture the current state of the art in PAI phantom testing, compared findings against available image quality standards for mature modalities, and offered insights and recommendations for future standards development efforts in PAI.

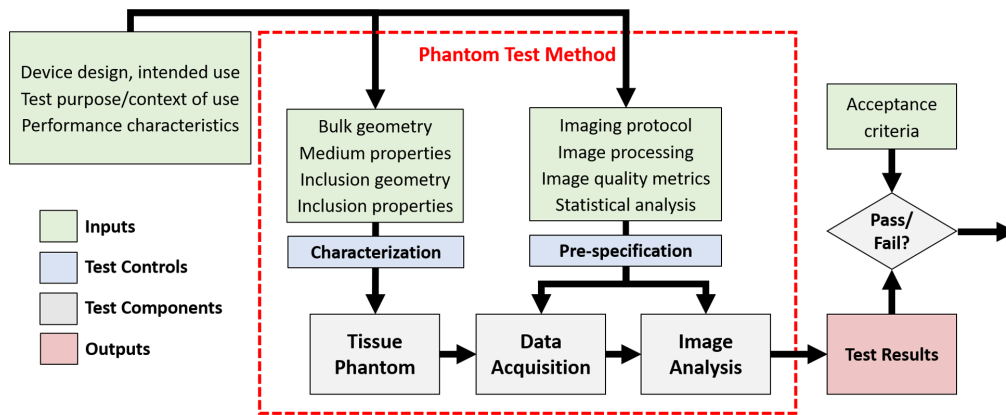


Fig. 1 Schematic of a phantom-based image quality test method.

2 Image Quality Test Methods for Established Modalities

The design of a standardized performance test method should begin with establishing the scope of device types the test applies to, the intended uses of those devices, the purpose of the test, key performance characteristics to be evaluated, and minimum acceptance criteria, if applicable (Fig. 1). Phantom test method specifications include phantom design requirements such as tissue-mimicking material (TMM) properties and the geometry of embedded targets. Phantoms should be rigorously characterized to ensure they meet desired specifications. In addition to phantom design, the methods for data acquisition and analysis also require careful consideration. The test method should provide a detailed protocol for taking phantom measurements, recommend best practices for image processing settings, and define appropriate image quality metrics. The test methodology should be “pre-specified,” meaning that the tester is not permitted to deviate from the specified protocol to produce more favorable outcomes (especially during execution of the test). Protocol modifications may be justifiable in certain situations (novel device configuration and inadequate phantom design), but in those cases the test should be repeated using the modified protocol.

Our review of image quality consensus test methods for ultrasound, CT, x-ray mammography, and MRI included research literature, standards, technical reports, consensus documents, and accreditation program requirements. We found that the scope and content of these documents varied widely. For instance, several clinical QA guidelines specified only high-level testing program requirements such as classes of image quality tests to perform (e.g., a generic requirement to evaluate spatial resolution using an unspecified test method).^{23–27} These documents also provided requirements for logistics of performance testing such as test report formats, recommended schedules for measurements in constancy testing, and “defect levels” that determine when system repair is needed to restore performance. In this review, we focused on documents that describe specific phantom-based image quality test methods because these fundamental details are of greatest interest for developing consensus test methods for PAI. Our review summarizes standardized test methods for evaluating IQCs most commonly used across all standards and most relevant to PAI including spatial resolution, geometric accuracy, image uniformity, depth of visualization, sensitivity, low-contrast detectability, and artifacts.

2.1 Spatial Resolution

Several standard test methods for evaluating in-plane spatial resolution were available for each of the three modalities, which is not surprising given the well-accepted importance of resolution in medical imaging. A key distinction was whether a test was based on qualitative (subjective) or quantitative (objective) image evaluation. Some ultrasound, CT, and MRI standards used a phantom containing various line or grid patterns with known target spacings [Figs. 2(b) and 2(c)], and resolution was determined as the spacing of the finest target in which the reader can distinguish the line pattern.^{14,15,28,29,31,32} However, this approach is subjective, depending on the individual

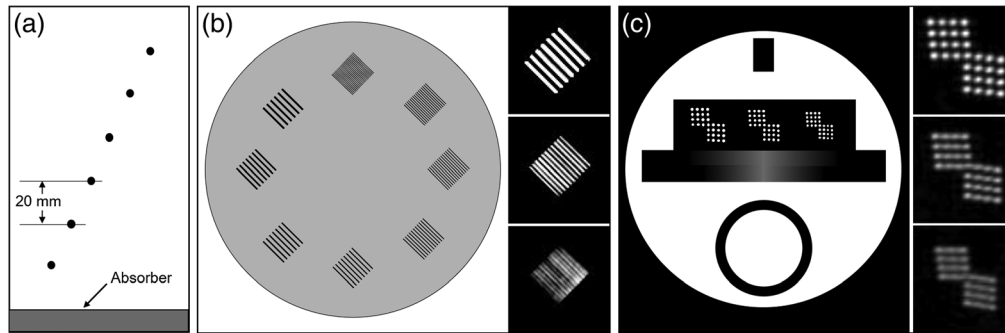


Fig. 2 (a) Schematic of an ultrasound PSF wire phantom. (b) Diagram and captured images of a CT resolution phantom containing aluminum bar targets. (c) Illustration and acquired images of MRI resolution phantom containing arrays of water-filled holes. Reproduced and adapted with permission from Refs. 28–30, respectively.

reader. Other standards describe objective, quantitative resolution tests, for instance by measuring the width of the point spread function (PSF) or line spread function (LSF) of a single sub-resolution target, usually specified as the full width at half maximum (FWHM), or less often, at tenth maximum (FWTM) [Fig. 2(a)].^{14,30} Placing several targets at various locations in the field-of-view also allows characterization of spatial variation in resolution. Another more comprehensive approach is to measure the modulation transfer function (MTF), a well-known approach used in optical imaging and endoscopy standards.^{33,34} A CT standard described computing MTF as the normalized Fourier transform of the PSF or LSF produced by a small, high-contrast wire, bead, or edge target embedded in a minimally attenuating background material. Spatial resolution was evaluated by reporting both the 10% and 50% points on the MTF curve.³⁵ It is worth noting that the common approach of measuring contrast, $C = (I_{\max} - I_{\min}) / (I_{\max} + I_{\min})$, versus spatial frequency in square-wave or bar patterns, such as the well-known 1951 USAF target, yields the contrast transfer function (CTF), which is not equal to the MTF.³⁴

Most resolution tests recommended use of high contrast targets at pre-specified positions. One ultrasound standard recommended using either (1) moderate-contrast nylon filaments in a “working liquid” with speed of sound 1540 ± 15 m/s, low acoustic attenuation (<0.1 dB/cm/MHz), and negligible scattering; or (2) high-contrast metal wires in a TMM with the same speed of sound, bio-relevant attenuation (0.5 ± 0.05 dB/cm/MHz), and an unspecified “moderate” level of scattering.³⁰ The first approach represents an engineering test under ideal conditions that may be useful for basic system characterization, and the latter represents a test closer to real-world conditions that may better predict *in vivo* performance. Accreditation programs often prescribed well-established, commercially available phantoms, some of which contained several “modules” for testing different IQCs.^{32,36} For example, the ACR CT phantom has an in-plane resolution module containing eight aluminum bar patterns ranging from 4 to 12 line pairs per centimeter embedded in a biologically relevant background [Fig. 2(b)]. The ACR MRI phantom contains a resolution module consisting of water-filled cylindrical cavities in various grid patterns [Fig. 2(c)].^{32,36}

Standards also specified tests for evaluating elevational (out-of-plane) resolution or section/slice thickness. These test methods typically used an angled object of known properties and dimensions slanted relative to the imaging plane.^{14,15,28–30,32,35,37} For example, an ultrasound test method describes scanning the transducer across a hyperechoic slab, angled at 75 deg relative to phantom surface, which appears in cross-sectional images as a rectangular object at variable depth [Figs. 3(a), 3(b)].³⁰ Elevational resolution, t , was determined as $t = x / \tan(75 \text{ deg})$, where x is the vertical height of the object. The ACR CT phantom contained two ramps of short wires positioned along out-of-plane inclines in opposite directions with elevational wire spacing of 0.5 mm [Figs. 3(c), 3(d)].²⁸ Slice thickness was computed as half the number of wires appearing at least 50% as bright as the central wires. MRI slice thickness has been determined by measuring FWHM of the signal intensity profile produced by a thin slab inclined at a 5 deg to 12 deg angle embedded in an MR-inactive material.³⁷ Slice thickness was calculated as the product of the FWHM of the trapezoidal profile and $\tan(\alpha)$. An alternative to imaging angled targets is to

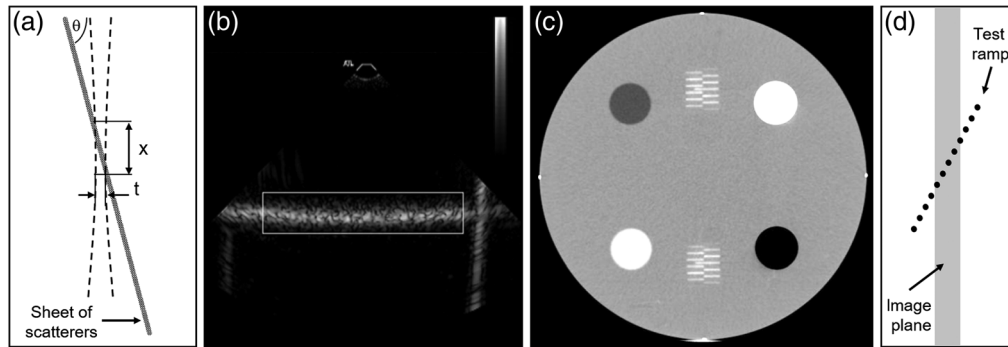


Fig. 3 (a) and (b) Diagram and captured image of an ultrasound slice thickness phantom using an angled plane ($\theta = 75$ deg) of scatterers, showing a typical ultrasound beam [dashed lines in (a)]. Reproduced and adapted with permission from Refs. 30 and 38, respectively. (c) and (d) Diagram of a CT slice thickness phantom using filament ramps. Reproduced and adapted with permission from Ref. 28.

scan a small point or line target in the elevational direction. For instance, an ultrasound standard described elevational resolution measurement by scanning a vertically oriented wire in a water bath,³⁰ whereas a CT standard characterized slice thickness by scanning a <0.1 mm-thick disk or bead.³⁵ Goodsitt et al.¹⁴ described a “less frequent” ultrasound test based on scanning an anechoic spherical object, although no explicit method for quantifying elevational resolution was provided.

2.2 Geometric Accuracy

Geometric accuracy, the ability of an imaging system to accurately represent tissue morphology, can be characterized by spatial measurement accuracy and image distortion. Assessment of tissue structure and geometry commonly involves the use of software-based image caliper tools in 1D (e.g., tissue layer thickness, distance between objects), 2D (e.g., vessel cross-sectional area), or 3D (e.g., tumor volume). In-plane spatial measurement accuracy test methods were available for ultrasound, CT, and MRI.^{14,15,28–30,32} These methods recommended imaging phantoms containing an array of high contrast targets [Fig. 4(a)] or a grid pattern [Fig. 4(c)] and comparing measured target distances in the image to known target distances. This approach can be used for linear, curvilinear, and circumferential measurements. Similarly, the accuracy of computed 2D cross-sectional areas and 3D inclusion volume can be evaluated by imaging a phantom containing 3D ovoid inclusions [Fig. 4(b)].³⁰

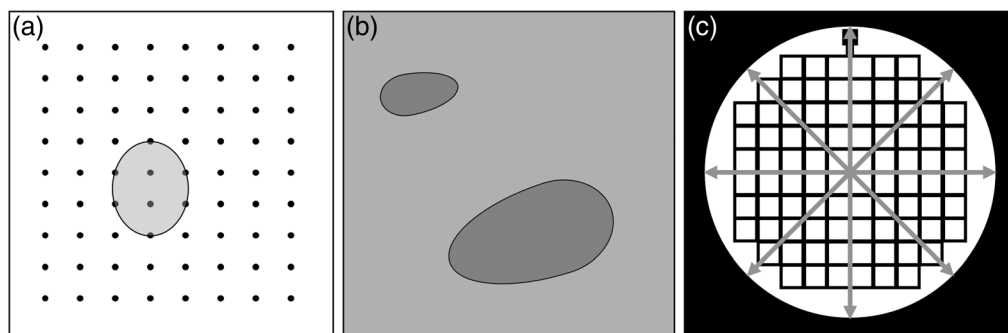


Fig. 4 Illustrations of (a) filament array for 1D distance and 2D area measurement accuracy (e.g., area of the drawn ellipse), and (b) ovoid inclusion phantoms for 1D, 2D, and 3D ultrasound spatial measurement accuracy testing (b). Reproduced and adapted with permission from Ref. 30. (c) Grid pattern phantom for MRI geometric accuracy evaluation. Reproduced and adapted with permission from Ref. 32.

Image distortion denotes spatial variation in magnification, such as well-known barrel or pincushion distortion effects in optical imaging. Distortion can also be asymmetric; for instance, incorrect ultrasound image reconstruction (e.g., poor speed of sound parameter) can cause significant distortion in the axial direction. An ultrasound consensus document described a qualitative distortion test by imaging a spherical or cylindrical phantom inclusion, which will appear as flattened or extended ovals if the image is distorted.¹⁴ Quantitative distortion tests often leveraged the same target grid phantoms used spatial resolution testing. One MRI distortion test recommended using a phantom containing a uniform grid or hole pattern to compute coefficient of variation of adjacent grid target spacings.³⁶ A different MRI approach involved imaging a phantom of known dimensions in all three orthogonal planes and computing the percent of geometric distortion (%GD) in each plane as

$$\%GD = 100 \times \frac{\Delta_{\text{actual}} - \Delta_{\text{measured}}}{\Delta_{\text{measured}}}. \quad (1)$$

where Δ_{actual} is the actual phantom dimension and Δ_{measured} is the dimension as measured on the image.³⁶

2.3 Uniformity and Depth of Visualization

Image uniformity describes spatial variation in sensitivity across an image field. Several documents recommended imaging a homogeneous, biologically relevant phantom and drawing several circular regions of interest (ROIs) to measure variations in image intensity across the field-of-view.^{15,28,29,31,32,35,37} In an ACR CT accreditation program, the mean CT number was computed for ROIs at the center and four edge positions [Fig. 5(c)], and uniformity was quantified as the absolute error between each edge ROI mean and the center ROI mean.²⁸ Similarly, an ACR MRI consensus document recommended drawing two small ROIs over regions having highest and lowest signal based on qualitative inspection.³² Mean signal intensity in these two ROIs (ROI_{high} , ROI_{low}) was measured to compute percent integral uniformity (PIU) as

$$PIU = 100\% \times \left[1 - \left\{ \frac{ROI_{\text{high}} - ROI_{\text{low}}}{ROI_{\text{high}} + ROI_{\text{low}}} \right\} \right]. \quad (2)$$

While CT and MRI systems can typically visualize signals within the entire field-of-view, ultrasound systems have finite imaging depth due to tissue attenuation and limited viewing angle. Thus, ultrasound documents considered depth of visualization or maximum penetration depth, the maximum depth to which the system provides useful information, but neglected lateral image uniformity.^{14,38-40} Maximum imaging depth was often assessed by imaging phantoms containing arrays of cylindrical inclusions that are anechoic or have specified contrast positioned at different depths [Fig. 5(b)], identifying the deepest visible inclusion by inspection or the depth

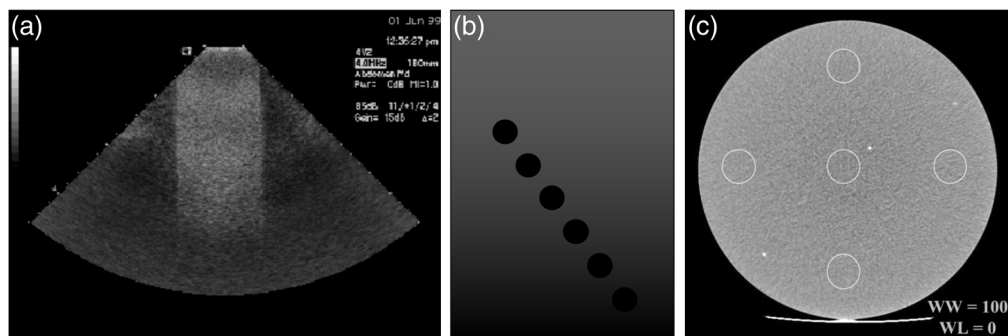


Fig. 5 (a) Ultrasound image of homogeneous phantom for evaluating depth of visualization. (b) Diagram for an ultrasound depth phantom containing anechoic inclusions in homogeneous background. (c) Captured image of the ACR CT uniformity phantom, showing circular ROIs. Reproduced and adapted with permission from Refs. 14, 28, and 39, respectively.

at which background texture “can barely be seen reliably.”¹⁴ A more quantitative approach computed the signal-to-noise ratio (SNR) of anechoic inclusions at various depths as

$$\text{SNR} = \frac{(m_{\text{target}} - m_{\text{background}})}{\sigma_{\text{background}}}, \quad (3)$$

where m_{target} is the mean target ROI intensity, $m_{\text{background}}$ is the mean background ROI intensity, and $\sigma_{\text{background}}$ is the standard deviation of the background ROI.¹⁴ However, since SNR values are only available at discrete depths where targets are placed, test results may depend on phantom design. Another standard described the use of a large, homogeneous phantom with specified acoustic attenuation and backscatter coefficient over 1 to 15 MHz [Fig. 5(a)].³⁹ Images were acquired in the phantom as well as with the transducer in air to measure electronic noise, and the maximum depth of penetration was defined as the axial location where the phantom signal decays to 1.4 times the noise signal, which corresponds to an SNR of 1 using the following definition:

$$\text{SNR}(j) = \sqrt{\frac{A(j)^2}{A'(j)^2} - 1}. \quad (4)$$

where $A(j)$ is the mean gray level of all pixels at a given depth, j , and $A'(j)$ is a similar measurement in the noise image.³⁹

2.4 Sensitivity and Low-Contrast Detectability

Sensitivity was most often used to describe the detection limit of an imaging system,^{14,39} but it may also describe the rate of change in image signal intensity versus target properties (e.g., target radioactivity, chromophore concentration).^{41,42} An ultrasound standard defined a closely related IQC, local dynamic range, as the difference in dB of echo amplitudes that produce minimum and maximum gray levels. Local dynamic range was evaluated using a phantom incorporating inclusions with different levels of relative contrast (e.g., -6 dB, -3 dB, +3 dB, and +6 dB) placed at the same depth within a biologically relevant echogenic background [Figs. 6(a) and 6(b)]. Local dynamic range was determined by finding the intercepts at 0 and 255 gray levels for a linear regression of ROI-averaged target amplitude versus known target contrast.³⁹ This standard also requires image processing settings to be reported for any local dynamic range measurement, as these controls will alter test results.

Low-contrast detectability denotes the ability to distinguish objects with similar brightness to the image background. Target size is typically varied in such tests to enable contrast-detail analysis, the combined evaluation of how object contrast and size impact object detectability. An ultrasound standard described an echogenic phantom containing arrays of 1- to 2-mm diameter anechoic spherical inclusions at various depths, where the smallest inclusion per depth was determined by inspection [Fig. 6(c)].⁴³ An alternative ultrasound approach used

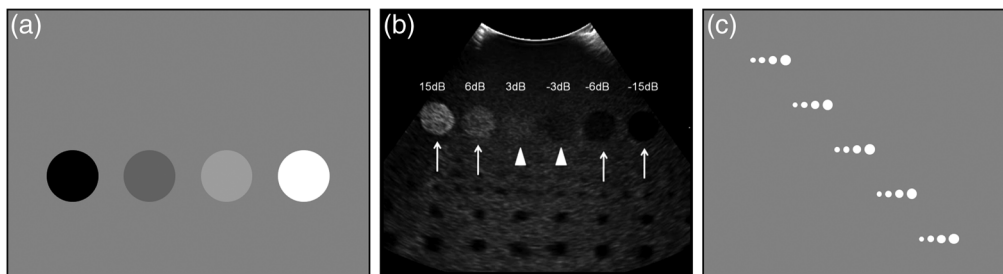


Fig. 6 (a) Diagram and (b) acquired image of an ultrasound phantom for local dynamic range measurements. (c) Diagram of an ultrasound low-contrast detectability phantom. Reproduced and adapted with permission from Refs. 39, 43, and 44, respectively.

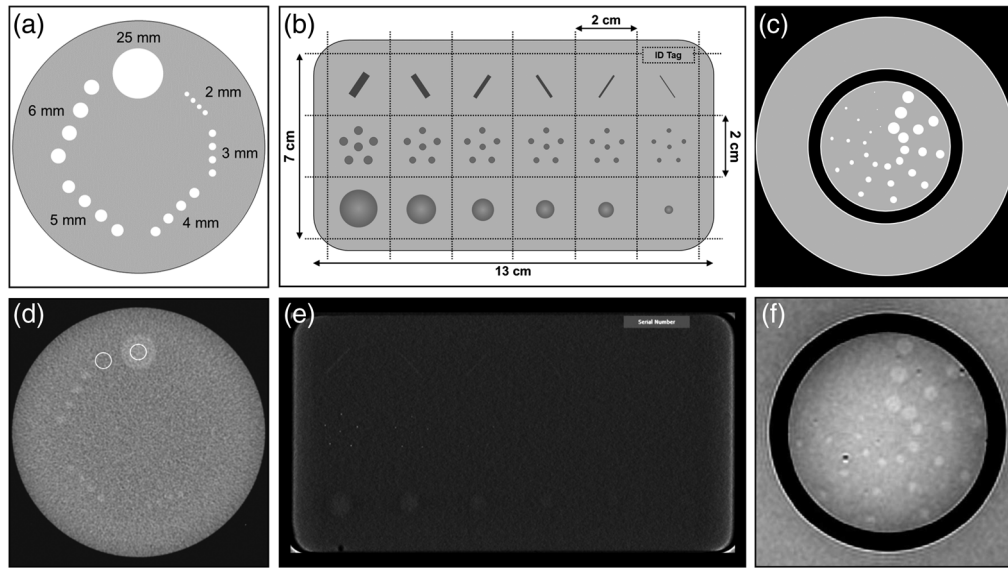


Fig. 7 (a) Diagram and (d) acquired image of a CT low-contrast detectability phantom. Reproduced and adapted with permissions from Refs. 28 and 31, respectively. (b) Diagram and (e) acquired image of the ACR Digital Mammography phantom. Reproduced and adapted with permission from Ref. 45. (c) Diagram and (f) acquired image of the ACR MRI low-contrast detectability spoke phantom. Reproduced and adapted with permission from Ref. 32.

a phantom containing $10\text{ cm} \times 20\text{ cm}$ conical inclusions with different contrast levels.⁴³ The transducer was scanned along the cone axis to change the in-plane cross-sectional area of the target cones, and the minimum detectable size for each contrast level was determined qualitatively.

Test methods for CT system low-contrast detectability involved a phantom containing arrays of cylindrical inclusions (2 to 10 mm in diameter) embedded in a biologically relevant medium [Figs. 7(a) and 7(d)].^{15,28,31,35} Detectability was either determined qualitatively by identifying the smallest set of “clearly delineated” inclusions or quantitatively by computing contrast-to-noise ratio (CNR):

$$\text{CNR} = \frac{m_{\text{target}} - m_{\text{background}}}{\sigma_{\text{background}}}, \quad (5)$$

where m_{target} is the mean signal of a target ROI, and $m_{\text{background}}$ and $\sigma_{\text{background}}$ are the mean signal and standard deviation of a local inclusion-specific background ROI.³¹ A similar MRI phantom contained radial “spokes” of 1.5- to 7-mm diameter cylindrical inclusions [Figs. 7(c) and 7(f)], as well as several elevational slices with inclusions at different contrast levels.^{32,36} Low-contrast detectability was determined as the number of spokes for which all three targets are distinguishable for each contrast level.

The ACR x-ray mammography QC manual prescribed an approach to evaluate low-contrast detectability using an approved ACR digital mammography phantom.⁴⁵ The phantom simulated a compressed breast of average density and contained a wax insert with groups of biomimetic inclusions relevant to breast cancer findings, such as tissue fibers (0.3 to 0.89 mm), specks representing calcifications (0.14 to 0.33 mm), and tumor-mimicking masses (0.2 to 1.0 mm) [Figs. 7(b) and 7(e)]. Minimum performance criteria were specified in terms of the smallest targets detected by a trained reader such as a radiologist. This approach differs significantly from other low-contrast detectability phantoms in that it uses three types of semi-idealized biological target features, as opposed to a more objective/quantitative but generalized evaluation using a single inclusion geometry. Both paradigms have merits and may be useful in device characterization and QC settings.

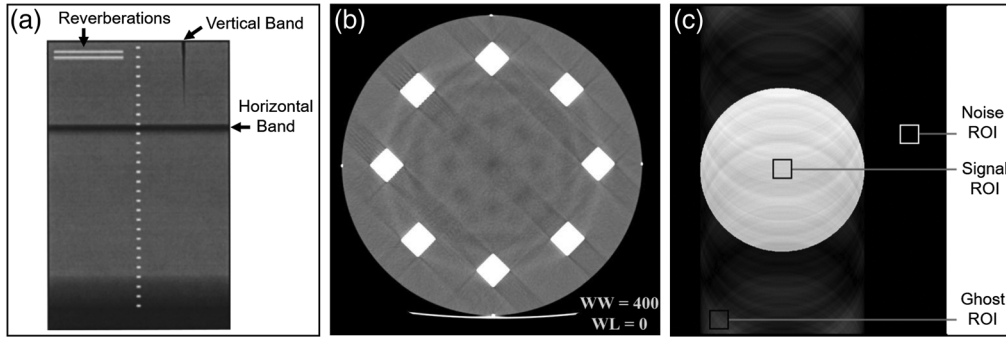


Fig. 8 (a) Phantom-based evaluation of ultrasound artifacts; (b) CT streak artifacts; and (c) MRI ghost artifacts. Reproduced and adapted with permission from Refs. 14, 28, and 37 respectively.

2.5 Artifacts

An image artifact is a visualized feature that is misrepresentative of the true object morphology and cannot be explained by random noise.³⁷ Artifactual shapes can either be reproductions of existing structures in the imaged object (e.g., ghosts, faint copies of an object superimposed on the image and displaced from its original location) or shapes unrelated to the imaged object. Artifacts can obscure true features of clinical interest, adversely affect diagnostic image interpretation, and corrupt phantom measurements of other performance characteristics. Test methods for artifacts tended to be less quantitative than those for other performance characteristics. AAPM QC procedures included evaluation of ultrasound image artifacts in a homogeneous tissue-mimicking phantom.¹⁴ Phantom images are inspected for streak artifacts not caused by beam coupling or phantom imperfections [Fig. 8(a)], and any deviations from the expected uniform image that rise to an action level (at which system repair should be made) or defect level (at which performance becomes severely affected) above the background are to be addressed. In the ACR CT accreditation program, artifact assessment relies on visual inspection of phantom images and manufacturer-specific corrective actions [Fig. 5(c)].^{15,23,28,31} These documents provided example images illustrating cupping, helical, ring, and streak/line artifacts [Fig. 8(b)]. MRI ghost artifacts, which are typically caused by patient motion or vibration and can be significant in low-contrast scenarios, can be evaluated using a homogeneous phantom as used for uniformity testing.³² A large primary ROI was drawn over the phantom as well as several background ROIs outside of the phantom, from which the ghosting ratio computed as

$$R_{\text{ghosting}} = \left| \frac{[S_{\text{top}} + S_{\text{bottom}}] - [S_{\text{left}} + S_{\text{right}}]}{2 \times S_{\text{large}}} \right|, \quad (6)$$

where S is the average pixel intensity in each ROI. A similar approach described in IEC 62464-1:2018 uses ROI measurements in a homogeneous phantom to compute ghost-to-signal ratio [Fig. 8(c)], ghost-to-noise ratio, and SNR:

$$\text{GSR} = \frac{I_G}{S}, \quad \text{GNR} = \frac{I_G}{I_N}, \quad \text{SNR} = \frac{S}{I_N}, \quad (7)$$

where I_G is the mean ghost ROI signal, S is mean phantom ROI signal, and $I_N = \sigma/0.655$ is the standard deviation of the background ROI, σ , corrected for image reconstruction effects.³⁷ The standard required reporting of all three metrics.

3 Current Image Quality Evaluation Practices in Photoacoustic Imaging

We used Web of Science to search for peer-reviewed journal articles published from 2010 to 2020 on PAI phantoms. This yielded 686 articles (search terms: [photoacoustic OR optoacoustic] AND imaging AND phantom). However, there was considerable variation in reported phantom

complexity, characterization, and context of use. To better align with our review of medical imaging standards, we excluded articles that (1) tested photoacoustic microscopy, elastography, non-imaging spectroscopy, flowmetry, or 1D depth profiling systems; (2) only used digital/numerical phantoms or *ex vivo* tissue; and (3) focused on non-image quality performance aspects such as oximetry measurement accuracy, fluence correction, or quantitative imaging. We focused our review on the 119 of the remaining 308 articles that used phantoms to quantify one or more specific IQCs, rather than only describing TMM development or qualitative performance testing. These articles included phantom studies of both 2D and 3D PAI systems.

A wide variety of background phantom materials was observed, including water,^{46,47} Intralipid,^{48–50} and various TMMs such as hydrogels (agar, gelatin, and polyvinyl alcohol),^{51–56} polyurethane,^{57–59} silicone,⁶⁰ gel wax,⁸ styrene-ethylene/butylene-styrene polymer,⁶¹ polydimethylsiloxane,^{62,63} and polyvinyl chloride plastisol (PVCP).^{48,64,65} Of the 119 studies of interest, 64 (54%) performed testing on targets immersed in non-turbid water baths or gels, rather than embedded in tissue-mimicking phantoms. This approach may be suitable in some cases to determine ideal performance (e.g., resolution testing) but may not be appropriate for IQCs that vary significantly with tissue attenuation (e.g., imaging depth). Only 36 (65%) and 8 (15%) of 55 studies using turbid phantoms characterized phantom optical and acoustic properties, respectively. In some cases, expected TMM properties were reported from previous literature, but many studies provided no discussion of phantom properties nor justification of their biological relevance. Phantom properties should be well-characterized to demonstrate biological relevance for an intended imaging application.

In-plane spatial resolution was by far the most commonly tested IQC, followed by depth of visualization and sensitivity (Fig. 9); other IQCs frequently encountered in medical imaging standards were significantly understudied. This may have been due to prioritization of IQCs that demonstrate the proposed advantages of PAI, namely, high-resolution imaging to detect deep, absorptive targets.⁶⁶ We also observed high variation in how IQCs were quantified, particularly for metrics related to target contrast and detectability. Reported image quality metrics included photoacoustic signal intensity (arbitrary units), SNR, signal-to-background ratio (SBR), contrast, contrast ratio (CR), and CNR. Adding to the confusion, these metrics have been defined many different ways (Table 1) or occasionally not explicitly defined. Note that the ratio of mean target image amplitude to mean background image amplitude (S/B) has been called SNR, SBR, CR, and CNR! The term SNR also requires careful interpretation as in some cases it referred to quality of raw, un-beamformed photoacoustic signals. To avoid ambiguity, image quality metrics and methods for their calculation should always be explicitly defined in a performance test method. It is important that both target contrast and background variation be considered when evaluating object detectability. One self-consistent set of metric definitions capturing both of these effects that we have employed is $SNR = S/\sigma_B$, $CR = SBR = S/B$, and $CNR = (S - B)/\sigma_B$, which also yields the relationship $CNR = SNR(1 - 1/SBR)$.⁹² One benefit of developing consensus documents is the establishment of standardized terms and definitions to enable reproducible data analysis and comparison of test results between systems.

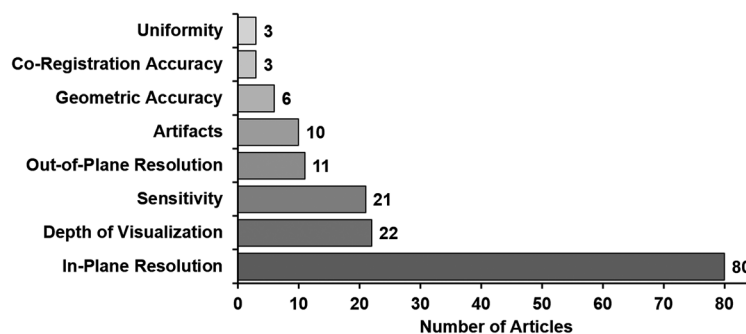


Fig. 9 Most commonly tested IQCs in reviewed PAI articles (some articles evaluated multiple IQCs).

Table 1 Reported definitions of image quality metrics in PAI studies, ranked in order of our descending preference (parentheses). S = mean target amplitude or power, B = mean background amplitude or power, σ_B = background standard deviation, “RMS” denotes root-mean-square, “max” and “min” denotes maximum and minimum values, subscript “2” denotes analysis of two-frame subtracted image, “pre-log” denotes using pre-log compression image amplitudes, and ‘global’ denotes analysis of the entire image (not ROIs, as for other definitions here).

IQ metric	Reported definitions		
SNR	(1) $\frac{S}{\sigma_B}$ ⁶⁷⁻⁷¹	(6) $\frac{S}{B}$ ⁵³	(11) $\frac{S_2}{\sigma_{B,2}} \sqrt{2}$ ⁷³
	(2) $\frac{S_{pre-log}}{\sigma_{B,pre-log}}$ ⁴⁸	(7) $20 \log_{10} \frac{S}{B}$ ^{72,73}	(12) $\frac{S_{max}}{B_{max}}$ ⁷⁴
	(3) $10 \log_{10} \frac{S}{\sigma_B}$ ⁷⁵	(8) $10 \log_{10} \frac{S}{B}$ ^{46,76}	(13) $\frac{S_{max}}{B_{RMS}}$ ⁷⁷
	(4) $\frac{S_{max}}{\sigma_B}$ ⁵⁵	(9) $20 \log_{10} \frac{S_{max}-S_{min}}{\sigma_S}$ ⁷⁸	(14) $\frac{S-B}{\sqrt{\sigma_S^2+\sigma_B^2}}$ ⁶²
	(5) $20 \log_{10} \frac{S_{max}}{\sigma_B}$ ^{79,80}	(10) $20 \log_{10} \frac{S_{RMS}}{B_{RMS}}$ ⁸¹	(15) $10 \log_{10} \frac{S-B}{\sigma_B}$ ⁸²
SBR	(1) $\frac{S}{B}$ ^{57,83}	(2) $\left(\frac{S_{max}}{B}\right)^2$ ⁸⁴	
Contrast or CR	(1) $\frac{S}{B}$ ⁸⁵	(3) $20 \log_{10} \frac{S_{pre-log}}{B_{pre-log}}$ ⁸⁰	(5) $\frac{S-B}{S+B}$ ⁸⁶
	(2) $20 \log_{10} \frac{S}{B}$ ^{68,71}	(4) $\frac{S-B}{B}$ ^{62,87}	
CNR	(1) $\frac{S-B}{\sigma_B}$ ⁴⁸	(5) $\frac{S-B}{\sqrt{\sigma_S^2+\sigma_B^2}}$ ⁶⁹	(9) $20 \log_{10} \frac{S}{\sigma_B}$ ⁸⁸
	(2) $20 \log_{10} \frac{S-B}{\sigma_B}$ ⁷³	(6) $\frac{ S-B }{\sqrt{\sigma_S^2+\sigma_B^2}}$ ^{71,79}	(10) $10 \log_{10} \frac{S_{RMS}-\sigma_B}{\sigma_B}$ ⁸⁹
	(3) $\frac{ S-B }{\sigma_B}$ ⁶⁸	(7) $\frac{S_{global}-B_{global}}{\sigma_{B,global}}$ ⁹⁰	(11) $\frac{S}{B}$ ⁵⁸
	(4) $20 \log_{10} \frac{ S-B }{\sigma_B}$ ^{72,91}	(8) $\frac{S}{\sigma_B}$ ⁹⁰	

3.1 Spatial Resolution

The most common approach for evaluating in-plane spatial resolution was to measure axial and/or lateral dimensions of the LSF produced by one or more line targets perpendicular to the image plane. It is worth noting that unlike some modalities described in Sec. 2, in-plane resolution is often anisotropic in PAI. This approach is essentially identical to resolution test methods described in ultrasound standards.³⁰ The ideal PAI resolution target should be much smaller than the resolution limit and produce high image contrast. Target size varied widely (6 μm to 1 mm) due to the broad range of minimum size requirements for PAI devices with different resolution limits. Line target materials included metal wires or filaments (tungsten, steel, copper, aluminum, or unspecified metal),^{48,60,67,78,79,93-99} carbon fibers,¹⁰⁰⁻¹⁰³ threads,^{51,104,105} sutures,^{48,89,106-108} graphite rods (pencil lead),^{50,109,110} or human/horse hairs.^{10,86,111-118} Some studies imaged inkjet-printed target patterns on paper or transparency film suspended in water or a tissue-mimicking medium.^{90,119} Almost all studies computed resolution as the FWHM (−6 dB width) of the measured PSF or LSF, although other metrics were observed including −3 dB width^{60,78} or half the FWTM.¹¹⁰ While targets were often aligned perpendicular to the image plane, some photoacoustic CT studies used line targets parallel to the plane.^{10,114} An alternative approach was to image spherical point targets such as 10- to 200- μm black polyethylene microspheres,^{11,51,55,56,119-123} 100- to 200- μm graphite particles,^{124,125} or 50- μm polyamide particles.¹²⁶ A few papers evaluated resolution using pairs of adjacent targets such as crossed threads, for instance using Sparrow’s resolution criterion.⁵¹ This method yielded somewhat larger results versus 50- μm microspheres (189 μm versus $129 \pm 16 \mu\text{m}$), which was attributed to out-of-plane absorber contributions. Another alternative approach for lateral resolution was to scan a 1951 United States Air Force (USAF) target immersed in water¹²⁷ or beneath a solid

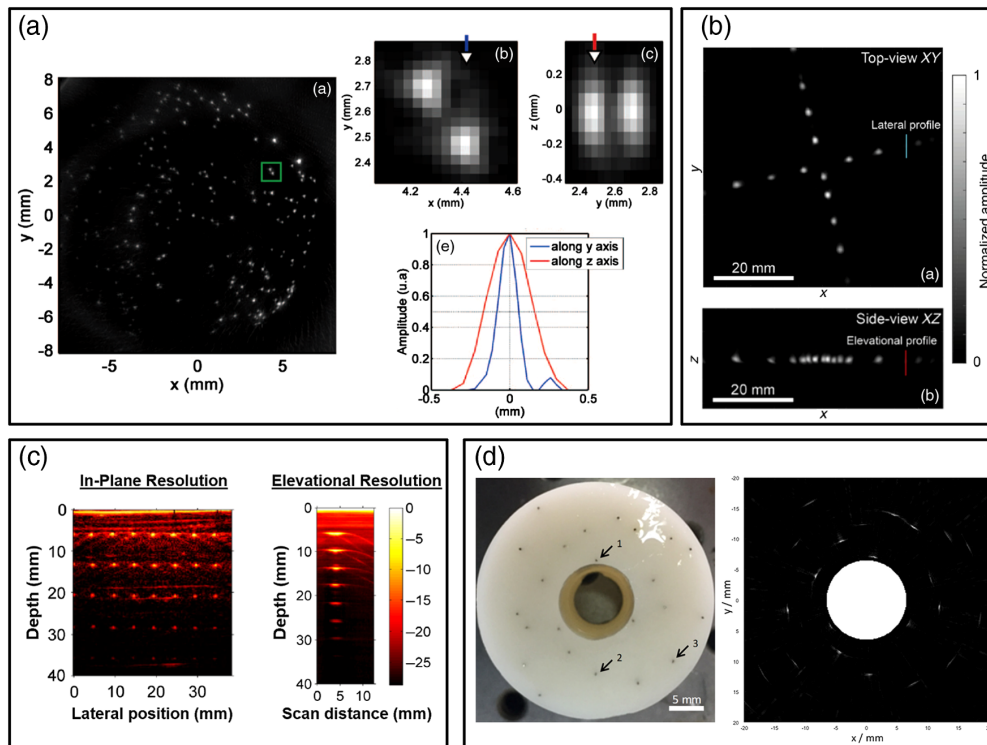


Fig. 10 Representative approaches to evaluate photoacoustic image in-plane and elevational resolution, including (a) black polyethylene spheres in agar; (b) black epoxy droplets in water; (c) steel wires in PVCP; or (d) metal wires in agar. Reproduced and adapted with permission from Refs. 46, 48, 51, and 99, respectively.

phantom^{88,128} and measure bar FWHM or contrast. However, it may be more appropriate to measure resolution with this target by computing the CTF or reporting line pairs per mm of the smallest discernable pattern by inspection. Also, this method requires vertical transducer scanning or different phantom layer thicknesses to characterize variations in resolution versus depth, whereas filament grids readily provide this information.

Unlike in-plane spatial resolution, elevational or out-of-plane resolution was less frequently considered. Medical imaging standards (Sec. 2) often used angled targets for elevational resolution testing, but these methods may not be acceptable for PAI due to light diffusion and limited elevational optical focusing. However, elevational resolution can often be measured using in-plane resolution phantoms—a concept that was seen in image quality standards (Sec. 2.1) (Fig. 10). We previously demonstrated this by scanning a column of steel wires in Intralipid or PVCP phantoms along the elevational direction to measure elevational FWHM versus target depth.⁴⁸ In addition to wire targets, spherical absorbers such as 50- to 100- μm black microspheres^{11,51,122,129} or 0.5- to 1.5-mm black epoxy drops,⁴⁶ have also been used for both in-plane and elevational resolution as the targets are sufficiently small in three dimensions. Another approach suited to photoacoustic CT was to measure the edge spread function of a small needle lowered into the image plane.⁷⁹

Resolution target size varied from ~ 1 to 10 times smaller than the measured FWHM, and it is unclear what size requirements are needed to ensure accurate resolution measurements. An MRI slice thickness test recommended feature size at least five times smaller than the FWHM,³⁷ whereas an ultrasound resolution test defined sub-resolution line targets such that a ten-fold reduction in diameter would not change apparent target size.³⁰ FWHM measurements should be interpreted carefully; if the FWHM is close to the actual target size, the target may not be sufficiently smaller than the resolution limit. PAI resolution should be assessed by measuring FWHM of high contrast, sub-resolution line or point targets positions placed at known locations throughout the field-of-view.

3.2 Geometric Accuracy, Distortion, and Co-Registration Accuracy

While geometric accuracy was a common consideration in standardized medical imaging test methods, few PAI articles reported specific geometric accuracy test methods (Fig. 9). Two of our group's studies leveraged spatial resolution phantoms for this purpose, in similar fashion to standardized test methods (Sec. 2.2). In one study, vertical and horizontal distances between steel filaments in a rectangular grid pattern in a turbid PVCP phantom were measured based on location of the brightest pixel.⁴⁸ These values were compared to nominal target spacing as well as distances measured in co-registered ultrasound images. Another study used a two-layer PVCP phantom with an irregular boundary representing breast fat-glandular tissue interfaces to study the impact of heterogeneity on axial position error of embedded steel wire targets.¹³⁰ Another study used a stacked-layer phantom to evaluate accuracy of PAI-measured layer thickness measurements for skin burn assessment.¹³¹ This phantom was comprised of thin inkjet-printed patterned polymer sheets containing red dye placed between slabs of turbid acrylic polymer. Similarly, one study evaluated accuracy of target localization (depth) measurements versus target blood content and size using turbid agarose phantoms containing blood-filled spherical gel lesions.¹³² PAI distortion was rarely tested or quantified, although it is well known that improper reconstruction parameters such as speed of sound can distort images, especially in the vertical direction. One study evaluated distortion by imaging a square loop target embedded in a brain-mimicking gelatin phantom beneath *ex vivo* ovine skull.¹³³ Distortion due to poor image acquisition settings may be corrected or calibrated, but tissue effects cannot always be avoided or completely mitigated. Especially in the latter scenario, distortion should be included in photoacoustic image quality testing. While no specific distortion test method was described in the literature, a filament grid phantom embedded in a phantom with well-characterized acoustic properties (Sec. 3.1) may be a reasonable approach.

Due to the nature of PAI technology, many PAI systems allow the collection of co-registered photoacoustic and ultrasound images. As with geometric accuracy, US-PAI co-registration accuracy is often not explicitly characterized but can be evaluated using spatial resolution phantoms to compare apparent target positions between US and PA images using either qualitative¹³⁴ or quantitative approaches.^{48,108} MRI-PAI co-registration has been calibrated using fiducial markers comprised of channels filled with gold nanoparticles and gadolinium solution in an Intralipid-agar phantom.¹³⁵ Additionally, one study characterized localization accuracy of tissue surface-generated photoacoustic signals as fiducial markers for co-registering ultrasound images and stereo camera video.¹³⁶ Co-registration was generally quantified using maximum or average target registration error (TRE), the Euclidean distance between matched points in different images. Co-registration accuracy should be tested in applications combining PAI with other imaging modalities.

3.3 Depth of Visualization and Uniformity

Depth of visualization was frequently evaluated in PAI phantom studies. The most common approach was to image a phantom containing an array of tubes placed at various depths, filled with relevant light-absorbing contrast media such as India ink, black dye, blood, or nanoparticles (Fig. 11).^{48,80,128,137-139} Alternative approaches included translating a single target to different depths in a liquid phantom¹⁰² or elevationally scanning the transducer over a phantom containing a vertically slanted tube¹³⁴ or graphite sheet.⁵⁰ Solid phantom inclusions were also used as imaging targets for depth testing such as black PVCP spheres in a PVCP background⁸³ or polyurethane cylinders within polyurethane background.⁵⁸ Some studies reported imaging depth based on detection of a target at one particular depth, which may underestimate maximum depth of visualization. While many studies focused on handheld epi-illumination PAI, one study tested depth of visualization for an endoscopic PAI device by placing 0.6-mm-diameter graphite rods at different radial positions in a cylindrical gelatin-milk phantom containing silica particles.¹⁴⁰ Similar studies of imaging depth were performed for PAI systems using interstitial light sources placed within the phantom or tissue.^{68,100,115} These approaches demonstrate how the common diagonal tube array phantom design can be modified to suit different imaging system configurations.

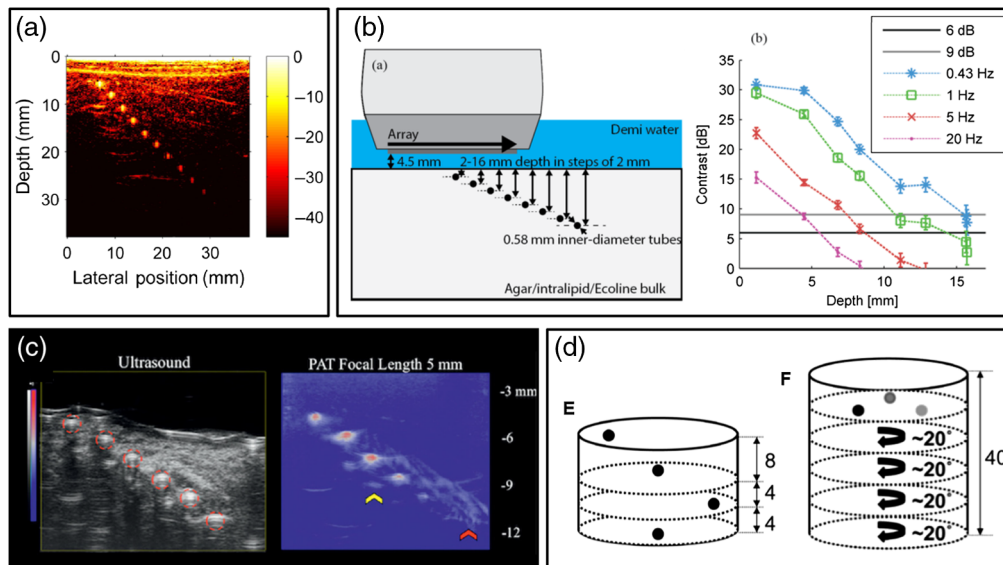


Fig. 11 Representative approaches to evaluate photoacoustic image maximum depth of visualization. (a) PA image of a PVCPhantom containing a diagonal array of India ink-filled tubes. Reproduced and adapted with permission from Ref. 48. (b) Schematic of an array of black ink-filled polyethylene tubes in an agarose phantom, and the plot of contrast versus depth and frame rate. Reproduced and adapted with permission from Ref. 137. (c) Ultrasound and PA images of a PVA phantom embedded with six PE-50 tubes. Reproduced and adapted with permission from Ref. 53. (d) Schematic of turbid PVCPhantoms containing PVCPhosphors with variable depth and absorption coefficient. Reproduced and adapted with permission from Ref. 83.

In most studies, all targets had the same absorption coefficient, isolating the impact of target depth on detectability from the effect of target absorption variation (see Sec. 3.4). This approach was similar to the ultrasound penetration depth phantom shown in Fig. 5(b).¹⁴ However, some PAI studies have also varied absorption coefficient of the target array, which is somewhat similar to low-contrast detectability phantoms described in Sec. 2.4.^{86,92} Because depth of visualization depends on target absorption coefficient, target absorption values should be relevant to the intended imaging application and should include low-contrast conditions.

In addition to phantom design, there was wide variation in how, if at all, maximum depth of visualization was quantified. The details of how such metrics were computed from image ROIs (ROI size, shape, and location using average versus maximum values) were not always provided. Also, specifying a maximum imaging depth requires selection of an appropriate signal threshold. Some studies interpolated an image quality metric versus depth to find the crossover with a pre-specified threshold (e.g., SNR = 2, or 6 dB), but others reported the depth of the deepest detectable target (even if the target appears well above the limit of detection). To determine maximum depth of visualization, there should be at least one target that is found to be undetectable such that maximum depth of visualization can be interpolated, as opposed to relying on extrapolation. To enable reproducibility, the methods of selecting ROIs and computing values from image data should always be comprehensively described.

Image uniformity was evaluated much less frequently than depth of visualization, despite the close relationship between these IQCs. While standards measured uniformity in terms of variation in large, positive-contrast homogeneous regions, photoacoustic images generally do not present such features, e.g., due to boundary buildup effects. Thus, photoacoustic image uniformity may be more appropriately described by how the apparent brightness of an absorbing target varies within the field-of-view. Several studies measured SNR or contrast of high-contrast targets such as wires to characterize imaging depth or target detectability versus depth,^{10,55,98,112,141,142} but few studies evaluated uniformity in other dimensions (most notably, lateral uniformity). One approach measured 2D image uniformity in a turbid PVCPhantom containing an array of metal wires, plotting average target amplitude versus target position [Fig. 12(a)].⁴⁸ Note that such wire or filament phantoms are often inappropriate for determining maximum depth of visualization

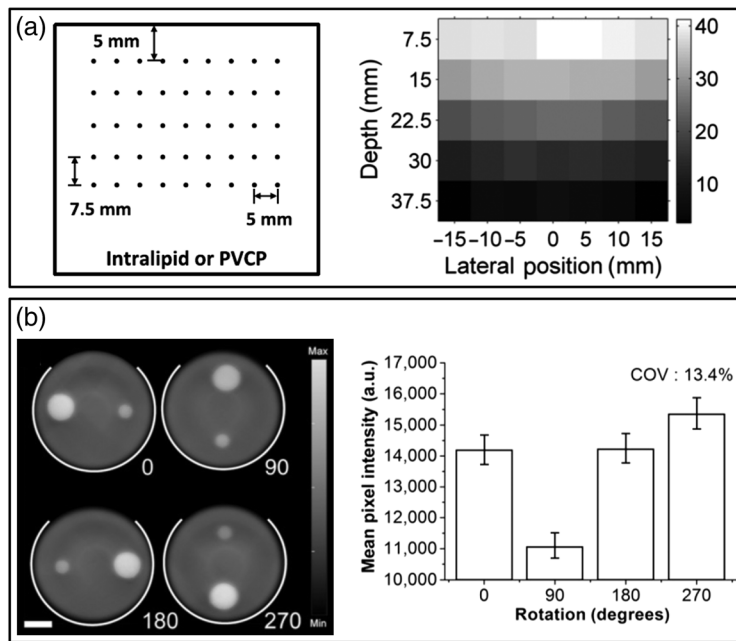


Fig. 12 Representative approaches for evaluating photoacoustic image uniformity. (a) Schematic and resultant uniformity map of a PVCPhantom containing a steel wire grid. (b) PA images and computed mean target intensities for a polyurethane phantom containing absorptive targets imaged at 4 different rotations (0 deg, 90 deg, 180 deg, and 270 deg). Reproduced and adapted with permission from Refs. 48 and 57, respectively.

owing to their high, non-biologically relevant absorption (unless the intended application involves detection of embedded manmade objects such as needles¹⁴³ or brachytherapy seeds⁶⁸). A few studies evaluated uniformity using larger inclusions with more moderate absorption levels, such as cylindrical absorptive inclusions in a turbid, acoustically attenuating polyurethane cylinder.⁵⁷ This phantom was scanned in different angular positions and uniformity was determined as the variation in average target intensity with location in the field-of-view [Fig. 12(b)]. Another study measured variation in image intensity of methylene blue-filled tubes both laterally and with depth using a 3D-printed housing to control tube alignment and positioning.¹³⁹

3.4 Sensitivity and Low-Contrast Detectability

Following medical imaging standards, we defined “sensitivity” testing as measurements of change in photoacoustic image amplitude versus target optical absorption or chromophore concentration to determine limits of detection. In some PAI articles, sensitivity referred to ultrasonic transducer sensitivity (e.g., responsivity in V/mPa or noise-equivalent pressure in Pa), rather than image sensitivity.^{123,144} Most sensitivity studies were performed to demonstrate detectability of exogenous contrast agents including dyes,^{85,112,145–147} encapsulated-ink microbubbles,¹⁴⁸ and nanoparticles,^{59,80,126,149–156} although other studies evaluated endogenous chromophores, such as melanoma cells^{11,157} or blood with varying hematocrit.¹²⁸ Some studies used generic absorptive targets such as embedded tubes^{48,102,114} or solid agar inclusions¹⁵⁸ containing colored inks. The common approach was to generate a linear fit of measured image signal/intensity (in arbitrary units) versus target concentration or absorption. Target depth varied considerably from 1- to 2-cm depths to entirely superficial/exposed targets. Some phantoms contained several targets with varying absorption, whereas others sequentially filled the same inclusion with different absorptive solutions. Several studies used a commercial cylindrical polyurethane phantom containing two cylindrical insertions/chambers [similar to Fig. 13(b)].^{145,146,149} Most studies did not implement or propose a limit of detection based on these test data.

This general approach, while commonly used, has several limitations: First, presenting PAI amplitude in terms of arbitrary units prevents direct comparisons between studies. Assessing

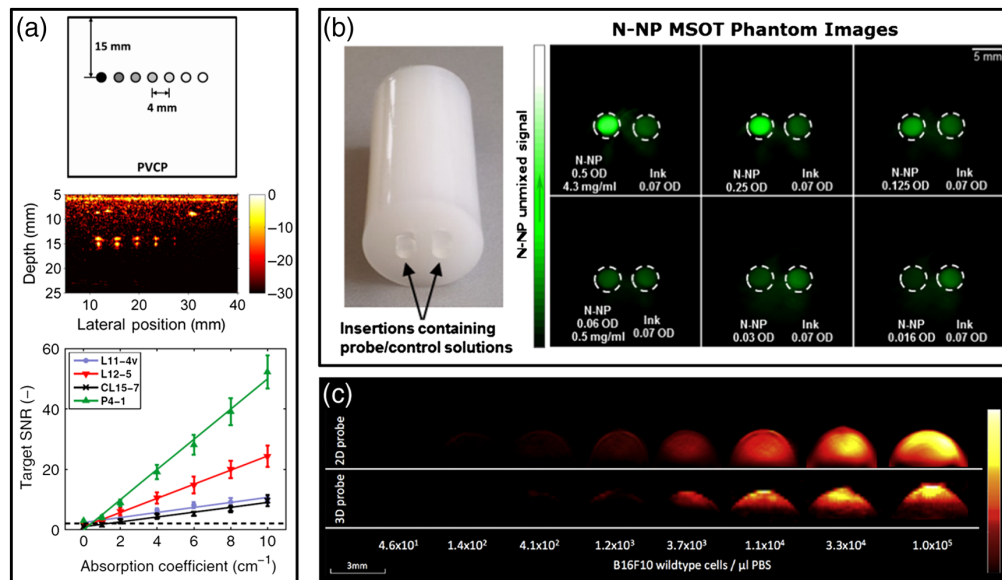


Fig. 13 Representative approaches to evaluate photoacoustic image sensitivity. (a) Schematic and PA image of a PVC phantom containing PTFE tubes filled with different concentrations of India ink, and plot of target pre-log compression SNR versus absorption coefficient for four transducers. (b) Photograph and PA images of an agar phantom with two cylindrical insertions filled with nanoparticles (P-NP) or a black ink solution. (c) PA images of agar plugs containing varying concentrations of B16F10 melanoma cells. Reproduced and adapted with permission from Refs. 11, 48, and 59, respectively.

sensitivity using image quality metrics such as target CR or SNR may better facilitate performance comparisons across PAI systems. Second, establishing quantitative detection thresholds that agree with limits determined by visual inspection may be more practical and reproducible. Third, test results expressed in terms of contrast agent concentration may have limited utility. A more universal approach would be to use phantoms containing stable, well-characterized chromophores at well-defined absorption coefficients.⁴⁸ It should then be possible to estimate results for different contrast agents if their molar extinction or absorption coefficients are known. Finally, most sensitivity phantoms contained targets of varying absorption strength but only at a fixed depth. The ideal phantom for testing sensitivity should have targets of various absorption coefficients located at several depths.^{92,128} It may also be appropriate to perform testing in phantoms with different background optical and/or acoustic properties to characterize how tissue background affects sensitivity and target detectability.^{139,153}

While we identified several PAI sensitivity test methods, we did not find any low-contrast detectability phantom studies using various target sizes. This was surprising given the prevalence of such testing in medical imaging standards (Sec. 2.4). Target size may be expected to affect detectability in PAI, for instance due to differences in intra-target fluence distribution and out-of-plane signal contributions, as well as boundary buildup effects in larger targets. This is a significant current gap in available phantom-based performance methods for PAI. Suitable phantom designs may build on sensitivity and imaging depth phantoms, such as turbid phantoms with arrays of targets of various absorption coefficient, placed at one or more depths.

3.5 Artifacts

Photoacoustic images are susceptible to several well-known image artifacts including image clutter,^{68,138} reflection artifacts,¹⁵⁹ out-of-plane artifacts,^{48,160} motion artifacts,¹⁶¹ scanning misalignment artifacts,¹⁰⁷ boundary buildup,¹⁶² laser-induced electromagnetic interference,¹⁶³ and limited view artifacts. Several studies used phantoms to evaluate performance of proposed correction techniques for specific types of artifacts. One study used a SMOFLipid-agar phantom containing 0.7-mm diameter graphite rods to evaluate reduction of x-shaped reconstruction

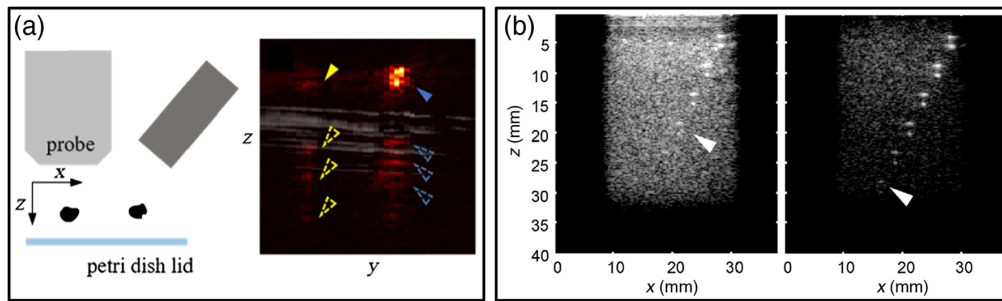


Fig. 14 Representative approaches to evaluate photoacoustic image artifacts. (a) Diagram of an agarose phantom containing two black absorbers, one inside and one 3 to 4 outside of the image plane. An overlaid ultrasound/PA image shows resultant in-plane and out-of-plane artifacts. (b) PA images of a gelatin phantom containing 2-mm absorptive gelatin cylinders, generated using either conventional image reconstruction (left) or clutter reduction methods (right). Reproduced and adapted with permission from Refs. 138 and 166, respectively.

artifacts using dynamic focusing and coherence weighting.¹²³ Another study evaluated a technique to remove reflection artifacts caused by acoustic heterogeneity using a clear gelatin phantom¹⁶⁴ or water bath¹⁶⁵ containing inclusions with different acoustic properties from the background medium. Artifact reduction was quantified using intensity reduction ratio, i.e., the ratio of original to corrected ROI intensity. Two articles by Nguyen and Steenbergen¹⁶⁰ and Nguyen et al.¹⁶⁷ described phantom-based evaluation of out-of-plane artifacts caused by photoacoustic signals from absorbers near the imaging plane [Fig. 14(a)]. These studies involved either transparent agarose phantoms or Intralipid solutions containing pairs of absorbers such as short lengths of sub-millimeter black threads or sutures. Phantoms either had inclusions at the same depth or positioned the out-of-plane absorber at a shallower depth in order to cause direct overlap of image artifacts with the in-plane target. One of these studies defined artifact-to-noise ratio, the mean artifact ROI amplitude divided by mean background ROI amplitude.¹⁶⁰ In another study, an acoustic radiation force technique for reducing photoacoustic image clutter was evaluated using gelatin phantoms doped with TiO₂, India ink, and cellulose, and containing an array of tubes at different depths [Fig. 14(b)].¹³⁸ Clutter reduction was evaluated in terms of improved SNR and maximum depth of visualization (see Sec. 3.3). A similar approach used a gelatin-cellulose phantom but quantified clutter reduction using target SBR.⁸⁴ While not all studies quantified artifact strength or reduction efficacy, most that did compared contrast-based image quality metrics, rather than noise-based metrics.

Due to the wide variation in PAI artifacts and how they impact performance, it may be difficult to develop a single phantom to quantitatively assess all possible artifacts. As with medical imaging standards, future consensus test methods may need to be tailored to individual artifacts. Still, we recommend establishment of general best practices for assessing PAI artifacts, such as use of biologically relevant phantoms that replicate artifacts of interest and establishment of well-defined metrics to quantify artifacts.

4 Discussion and Outlook

We reviewed 32 consensus documents and standards for established medical imaging modalities as well as nearly 120 PAI articles describing phantom-based image quality test methods. Our review of test methods for ultrasound, CT, x-ray mammography, and MRI revealed similarities and differences in terms of IQCs, phantom geometries, TMM properties, data acquisition and analysis procedures, and the level of prescribed detail for different aspects of testing. Insights gained from this review have the potential to facilitate standardization, clinical translation, and the maturation of PAI into a well-accepted medical imaging modality.

The most common IQCs used in medical imaging standards were in-plane spatial resolution, out-of-plane spatial resolution (slice thickness), geometric accuracy, image uniformity, depth of visualization, sensitivity, and low-contrast detectability. These IQCs should be considered in the

development of PAI standards, as well as others that address key aspects of image quality including distortion, artifacts, and co-registration accuracy. Unlike medical imaging standards, PAI literature focused on a smaller number of IQCs (e.g., in-plane resolution, depth of visualization, and sensitivity). It is possible that developers would elect to test more IQCs if the burden of developing and validating suitable test methods were reduced through phantom development and commercialization. Some of the understudied IQCs for PAI are linked to well-known device challenges: elevational resolution is often poor for linear array transducers and relates to out-of-plane artifacts; geometric accuracy, distortion, and co-registration accuracy relate to image reconstruction algorithm performance; and image uniformity and depth of visualization relate to fluence distribution. While it is important to ensure that a sufficient range of IQCs are tested to adequately characterize performance, PAI standards will need to balance this consideration against the potential for creating excessive burdens for developers and users. Achieving this balance could be accomplished, in part, by recommending the use of fewer IQCs and simpler test methods in roles such as post-market QC and constancy testing, whereas more extensive and rigorous testing would be reserved for device development, performance verification, and regulatory evaluation.

Tissue-simulating phantoms were critical components of nearly all image quality standards. These standards tended to implement relatively simple designs for objective, quantitative assessment of image quality, such as homogeneous regions with simple inclusions in repeating patterns. Phantom properties tended to be relevant to generic tissue, rather than matching a specific tissue type. While standards often specified required phantom material properties and geometry, they generally did not mandate a particular material for background regions or inclusions (although in some cases, suitable examples were mentioned). In principle, any TMM meeting test method requirements and relevant to the imaging application would thus be acceptable. But to maximize consistency in test results, future PAI standards may elect to identify a preferred TMM and allow other options if they are shown to generate identical test results. Also, most accreditation programs required use of specifically approved commercial phantoms that have been rigorously characterized by the manufacturer to ensure conformity to standards during acceptance testing, QC, and maintenance/repairs. Some of these phantoms are also traceable to gold standard metrology, such as those supported by the National Institute of Standards and Technology (NIST).¹⁶⁸ This may be an important future consideration for PAI standards, especially for quantitative imaging applications, and is an active area of development in biophotonics.^{169,170}

It should be stressed that while appropriate TMMs are essential for phantom-based test methods and the community is actively working toward addressing this need, careful design and consistent reproduction of phantom geometry, target inclusion sizes and patterns, and measurement/analysis protocols is equally important. Image quality standards often provided detailed, yet relatively simple, test protocols that specified ROI dimensions and locations, number of images to acquire, and explicit formulas for computing image quality metrics. Standards also often recommended using a fixed set of application-relevant image processing and display settings for a given test. While some variation in nomenclature and definition of image quality metrics was seen across medical imaging standards, we observed much broader variation in definitions for photoacoustic image quality metrics such as SNR, SBR, CR, and CNR. Future PAI standards should explicitly define recommended image quality metrics, and one self-consistent set of metric definitions would be $S = S/\sigma_B$, $CR = SBR = S/B$, and $CNR = (S - B)/\sigma_B$. Data acquisition procedures, image analysis methods, and image quality metrics should always be comprehensively described to ensure test reproducibility. It is notable that some test methods involved subjective image evaluation by a reader. While there is certainly value to such an approach as it mirrors how images will be used clinically, objective methods are typically preferred to maximize repeatability and reproducibility. Standards were often not accompanied by minimum acceptance criteria. While PAI studies generally have not attempted to establish minimum performance thresholds, such criteria may be useful for devices that focus on specific applications, such as breast cancer detection. In the development of PAI standards, it will be critical that procedures for data acquisition, image analysis, and metric calculation are comprehensively described, so as to optimize reliability of comparisons between tests performed by different groups. While this review has focused primarily on image quality standards,

additional standardized test methods will be needed for quantitative and functional PAI biomarkers such as blood oxygen saturation. These tests will likely require the use of specific materials such as blood or contrast agents incorporated within inclusions of a larger tissue-simulating phantom.^{6,171} Also, while not typically addressed in standards, future consensus test methods focusing on tissue-specific device applications may benefit from biomimetic, anthropomorphic phantoms to provide more clinically realistic, task-based image quality assessment approaches.^{172–174}

Many of the issues addressed in this review apply to the standardization of other existing and emerging biophotonic approaches. Some IQCs mentioned here have been addressed in endoscopy performance standards,³⁴ but may also be relevant to more advanced biophotonic modalities such as optical coherence tomography¹⁷⁵ or diffuse optical imaging.¹⁷⁶ Insights from this review on phantom design and test methodology may inform standards development in both sub-surface, cross-sectional optical imaging modalities (e.g., diffuse optical imaging/tomography, fluorescence tomography, and optical coherence tomography) and superficial, *en face* modalities (e.g., fluorescence, hyperspectral, and Raman imaging).

5 Conclusion

As the photoacoustics community and others within the field of biomedical optics work toward establishing consensus standards, available medical imaging standards should be consulted. These documents can facilitate and accelerate establishment of best practices for photoacoustic image quality assessment. The past decade has seen significant advances in TMM development for PAI, but more progress is needed on this topic and in development of standard image acquisition and data analysis protocols. Further work is also needed to expand and adapt existing phantom test methods into multiple variations that are useful for the broad range of PAI device configurations reported in the literature. These efforts should culminate in establishment of a PAI performance standard, which will mark a key milestone in the maturation of this technology. Such consensus documents have the potential to accelerate device development and optimization, minimize duplication of effort, and facilitate clinical translation.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

Acknowledgments

We gratefully acknowledge funding from the NSF/FDA Scholar-in-Residence Program (Award #1842387 and #1937674). The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services. This article reflects the views of the authors and should not be construed to represent FDA views or policies.

References

1. K. S. Valluru and J. K. Willmann, "Clinical photoacoustic imaging of cancer," *Ultrasonography* **35**(4), 267–280 (2016).
2. S. Manohar and M. Dantuma, "Current and future trends in photoacoustic breast imaging," *Photoacoustics* **16**, 100134 (2019).
3. N. Nyayapathi and J. Xia, "Photoacoustic imaging of breast cancer: a mini review of system design and image features," *J. Biomed. Opt.* **24**(12), 121911 (2019).
4. A. B. E. Attia et al., "A review of clinical photoacoustic imaging: current and future trends," *Photoacoustics* **16**, 100144 (2019).
5. S. Zackrisson, S. van de Ven, and S. S. Gambhir, "Light in and sound out: emerging translational strategies for photoacoustic imaging," *Cancer Res* **74**(4), 979–1004 (2014).

6. W. C. Vogt et al., "Photoacoustic oximetry imaging performance evaluation using dynamic blood flow phantoms with tunable oxygen saturation," *Biomed. Opt. Express* **10**(2), 449–464 (2019).
7. X. Zhou et al., "Evaluation of fluence correction algorithms in multispectral photoacoustic imaging," *Photoacoustics* **19**, 100181 (2020).
8. E. Maneas et al., "Gel wax-based tissue-mimicking phantoms for multispectral photoacoustic imaging," *Biomed. Opt. Express* **9**(3), 1151–1163 (2018).
9. J. Buchmann et al., "Quantitative PA tomography of high resolution 3-D images: experimental validation in a tissue phantom," *Photoacoustics* **17**, 100157 (2020).
10. P. K. Upputuri and M. Pramanik, "Performance characterization of low-cost, high-speed, portable pulsed laser diode photoacoustic tomography (PLD-PAT) system," *Biomed. Opt. Express* **6**(10), 4118–4129 (2015).
11. V. Neuschmelting et al., "Performance of a multispectral photoacoustic tomography (MSOT) system equipped with 2D vs. 3D handheld probes for potential clinical translation," *Photoacoustics* **4**(1), 1–10 (2016).
12. U.S. Food and Drug Administration, "Recognized consensus standards," 2019, <https://www.fda.gov/medical-devices/standards-and-conformity-assessment-program/federal-register-documents>.
13. G. J. Tearney et al., "Consensus standards for acquisition, measurement, and reporting of intravascular optical coherence tomography studies: a report from the International Working Group for Intravascular Optical Coherence Tomography Standardization and Validation," *J. Am. Coll. Cardiol.* **59**(12), 1058–1072 (2012).
14. M. M. Goodsitt et al., "Real-time B-mode ultrasound quality control test procedures. Report of AAPM Ultrasound Task Group No. 1," *Med. Phys.* **25**(8), 1385–1406 (1998).
15. C. H. McCollough et al., "The phantom portion of the American College of Radiology (ACR) computed tomography (CT) accreditation program: practical tips, artifact examples, and pitfalls to avoid," *Med. Phys.* **31**(9), 2423–2442 (2004).
16. A. Valladares et al., "Clinically valuable quality control for PET/MRI systems: consensus recommendation from the HYBRID consortium," *Front. Phys.* **7**, 136 (2019).
17. Computerized Imaging Reference Systems, Inc., <https://www.cirsinc.com/> (accessed May 2021).
18. Sun Nuclear Corporation, <https://www.sunnuclear.com/> (accessed May 2021).
19. JM Specialty Parts, Inc., <https://www.acraccreditation.org/Modalities/MRI> (accessed May 2021).
20. S. Bohndiek, "Addressing photoacoustics standards," *Nat. Photonics* **13**(5), 298–298 (2019).
21. International Organization for Standardization, ISO 80601-2-85:2021, "Medical electrical equipment – part 2-85: particular requirements for the basic safety and essential performance of cerebral tissue oximeter equipment," <https://www.iso.org/standard/72442.html> (2021).
22. B. W. Pogue et al., "Fluorescence-guided surgery and intervention – an AAPM emerging technology blue paper," *Med. Phys.* **45**(6), 2681–2688 (2018).
23. American College of Radiology, "CT Accreditation Program," <https://www.acraccreditation.org/modalities/ct> (2010).
24. American College of Radiology and American Association of Physicists in Medicine, "ACR–AAPM technical standard for diagnostic medical physics performance monitoring of magnetic resonance (MR) imaging equipment," <https://www.acr.org/-/media/ACR/Files/Practice-Parameters/mr-equip.pdf?la=en> (2019).
25. American College of Radiology, "Ultrasound Accreditation Program," <https://www.acraccreditation.org/modalities/ultrasound> (2014).
26. American College of Radiology and American Association of Physicists in Medicine, "ACR–AAPM technical standard for diagnostic medical physics performance monitoring of real time ultrasound equipment," <https://www.acr.org/-/media/ACR/Files/Practice-Parameters/US-Equip.pdf> (2016).
27. American Institute of Ultrasound in Medicine, "Routine quality assurance for diagnostic ultrasound equipment," <https://www.aium.org/resources/library.aspx> (2008).

28. American College of Radiology, Computed Tomography Accreditation Program, “Phantom testing: CT,” 2019, <https://accreditationsupport.acr.org/support/solutions/articles/11000056197-acr-ct-phantom-scanning-instructions>.
29. American College of Radiology, “Phantom test guidance for use of the small MRI phantom for the MRI accreditation program,” <https://www.acraccreditation.org/-/media/ACRAccreditation/Documents/MRI/SmallPhantomGuidance.pdf> (2018).
30. International Electrotechnical Commission, IEC 61391-1:2006+AMD1:2017, “Ultrasonics – pulse-echo scanners – part 1: techniques for calibrating spatial measurement systems and measurement of system point-spread function response,” <https://webstore.iec.ch/publication/61038> (2017).
31. American College of Radiology, “Computed tomography: quality control manual,” https://www.acr.org/-/media/ACR/NOINDEX/QC-Manuals/CT_QCManual.pdf (2017).
32. American College of Radiology, “Phantom test guidance for use of the large MRI phantom for the MRI Accreditation Program,” <https://www.acraccreditation.org/-/media/acraccreditation/documents/mri/largephantomguidance.pdf> (2018).
33. International Organization for Standardization, “Photography—electronic still picture imaging—resolution and spatial frequency responses,” <https://www.iso.org/standard/71696.html> (2017).
34. International Organization for Standardization, ISO 8600-5:2020, “Optics and photonics—medical endoscopes and endotherapy devices – part 5: determination of optical resolution of rigid endoscopes with optics,” ISO 8600-5:2020, <https://www.iso.org/standard/65019.html> (2020).
35. International Electrotechnical Commission, “Evaluation and routine testing in medical imaging departments – part 3–5: acceptance and constancy tests – imaging performance of computed tomography x-ray equipment,” IEC 61223-3-5:2019, <https://webstore.iec.ch/publication/59789> (2019).
36. American Association of Physicists in Medicine, “Acceptance testing and quality assurance procedures for magnetic resonance imaging facilities,” AAPM Report No. 100, https://www.aapm.org/pubs/reports/rpt_100.pdf (2010).
37. International Electrotechnical Commission, “Magnetic resonance equipment for medical imaging – part 1: determination of essential image quality parameters,” IEC 62464-1:2018, <https://webstore.iec.ch/publication/61163> (2018).
38. J. M. Thijssen, G. Weijers, and C.L. de Korte, “Objective performance testing and quality assurance of medical ultrasound equipment,” *Ultrasound Med. Biol.* **33**(3), 460–471 (2007).
39. International Electrotechnical Commission, “Ultrasonics – pulse-echo scanners – part 2: measurement of maximum depth of penetration and local dynamic range,” IEC 61391-2: 2010, <https://webstore.iec.ch/publication/5421> (2010).
40. American Institute of Ultrasound in Medicine, “AIUM – quality assurance manual for gray scale ultrasound scanners,” <https://www.aium.org/resources/library.aspx> (2014).
41. U. Kanniyappan et al., “Performance test methods for near-infrared fluorescence imaging,” *Med. Phys.* **47**(8), 3389–3401 (2020).
42. A. M. Grant et al., “NEMA NU 2-2012 performance studies for the SiPM-based ToF-PET component of the GE SIGNA PET/MR system,” *Med. Phys.* **43**(5), 2334 (2016).
43. International Electrotechnical Commission, IEC TS 61390:1996, “Ultrasonics – real-time pulse-echo systems – test procedures to determine performance specifications,” <https://webstore.iec.ch/publication/5419> (1996).
44. J. I. Choi et al., “Establishing cutoff values for a quality assurance test using an ultrasound phantom in screening ultrasound examinations for hepatocellular carcinoma: an initial report of a nationwide survey in Korea,” *J. Ultrasound Med.* **30**(9), 1221–1229 (2011).
45. American College of Radiology, “Digital mammography: quality control manual,” https://www.acr.org/-/media/ACR/NOINDEX/QC-Manuals/Mammo_QCManual.pdf (2018).
46. S. M. Schoustra et al., “Twente Photoacoustic Mammoscope 2: system overview and three-dimensional vascular network images in healthy breasts,” *J. Biomed. Opt.* **24**(12), 121909 (2019).
47. Y. Wang et al., “Slit-enabled linear-array photoacoustic tomography with near isotropic spatial resolution in three dimensions,” *Opt. Lett.* **41**(1), 127–130 (2015).

48. W. C. Vogt et al., “Phantom-based image quality test methods for photoacoustic imaging systems,” *J. Biomed. Opt.* **22**(9), 095002 (2017).
49. B. T. Cox and L. An, “Estimating relative chromophore concentrations from multiwavelength photoacoustic images using independent component analysis,” *J. Biomed. Opt.* **23**(7), 076007 (2018).
50. M. Kuriakose et al., “Optimizing irradiation geometry in LED-based photoacoustic imaging with 3D printed flexible and modular light delivery system,” *Sensors (Basel)* **20**(13) (2020).
51. J. Gateau et al., “Three-dimensional optoacoustic tomography using a conventional ultrasound linear detector array: whole-body tomographic system for small animals,” *Med. Phys.* **40**(1), 013302 (2013).
52. P. Hai et al., “Label-free high-throughput photoacoustic tomography of suspected circulating melanoma tumor cells in patients *in vivo*,” *J. Biomed. Opt.* **25**(3), 036002 (2020).
53. G. S. Sangha, N. J. Hale, and C. J. Goergen, “Adjustable photoacoustic tomography probe improves light delivery and image quality,” *Photoacoustics* **12**, 6–13 (2018).
54. H. Guo et al., “Co-registered photoacoustic and ultrasound imaging for tongue cancer detection,” *J. Innov. Opt. Health Sci.* **11**(3), 1850008 (2018).
55. H. He et al., “Importance of ultrawide bandwidth for optoacoustic esophagus imaging,” *IEEE Trans. Med. Imaging* **37**(5), 1162–1167 (2018).
56. G. Paltauf et al., “Piezoelectric line detector array for photoacoustic tomography,” *Photoacoustics* **8**, 28–36 (2017).
57. J. Joseph et al., “Evaluation of precision in optoacoustic tomography for preclinical imaging in living subjects,” *J. Nucl. Med.* **58**(5), 807–814 (2017).
58. Y. Asao et al., “Photoacoustic mammography capable of simultaneously acquiring photoacoustic and ultrasound images,” *J. Biomed. Opt.* **21**(11), 116009 (2016).
59. K. Liu et al., “Polymeric nanosystems for near-infrared multispectral photoacoustic imaging: synthesis, characterization and *in vivo* evaluation,” *Eur. Polym. J.* **88**, 713–723 (2017).
60. J. James, V. M. Murukeshan, and L. S. Woh, “Integrated photoacoustic, ultrasound and fluorescence platform for diagnostic medical imaging—proof of concept study with a tissue mimicking phantom,” *Biomed. Opt. Express* **5**(7), 2135–2144 (2014).
61. L. C. Cabrelli et al., “Stable phantom materials for ultrasound and optical imaging,” *Phys. Med. Biol.* **62**(2), 432–447 (2017).
62. C. Avigo et al., “Organosilicon phantom for photoacoustic imaging,” *J. Biomed. Opt.* **20**(4), 046008 (2015).
63. F. Ratto et al., “Hybrid organosilicon/polyol phantom for photoacoustic imaging,” *Biomed. Opt. Express* **10**(8), 3719–3730 (2019).
64. W. C. Vogt et al., “Biologically relevant photoacoustic imaging phantoms with tunable optical and acoustic properties,” *J. Biomed. Opt.* **21**(10), 101405 (2016).
65. J. Zalev et al., “Opto-acoustic imaging of relative blood oxygen saturation and total hemoglobin for breast cancer diagnosis,” *J. Biomed. Opt.* **24**(12), 121915 (2019).
66. L. V. Wang and S. Hu, “Photoacoustic tomography: *in vivo* imaging from organelles to organs,” *Science* **335**(6075), 1458–1462 (2012).
67. W. Xia et al., “Handheld real-time LED-based photoacoustic and ultrasound imaging system for accurate visualization of clinical metal needles and superficial vasculature to guide minimally invasive procedures,” *Sensors (Basel)* **18**(5), 1394 (2018).
68. M. A. Lediju Bell et al., “Short-lag spatial coherence beamforming of photoacoustic images for enhanced visualization of prostate brachytherapy seeds,” *Biomed. Opt. Express* **4**(10), 1964–1977 (2013).
69. T. Mitcham et al., “Photoacoustic imaging driven by an interstitial irradiation source,” *Photoacoustics* **3**(2), 45–54 (2015).
70. T. Guan et al., “A photoacoustic imaging system with variable gain at different depths,” *J. Innov. Opt. Health Sci.* **11**(5), 1850022 (2018).
71. K. M. Kempfski et al., “Application of the generalized contrast-to-noise ratio to assess photoacoustic image quality,” *Biomed. Opt. Express* **11**(7), 3684–3698 (2020).
72. M. Vallet et al., “Quantitative comparison of PZT and CMUT probes for photoacoustic imaging: experimental validation,” *Photoacoustics* **8**, 48–58 (2017).

73. N. Alijabbari et al., "Photoacoustic tomography with a ring ultrasound transducer: a comparison of different illumination strategies," *Appl. Sci. (Basel)* **9**(15), 3094 (2019).
74. M. Oeri et al., "Hybrid photoacoustic/ultrasound tomograph for real-time finger imaging," *Ultrasound Med. Biol.* **43**(10), 2200–2212 (2017).
75. J. P. Gray et al., "Multi-wavelength photoacoustic visualization of high intensity focused ultrasound lesions," *Ultrason Imaging* **38**(1), 96–112 (2016).
76. D. Das et al., "On-chip generation of microbubbles in photoacoustic contrast agents for dual modal ultrasound/photoacoustic in vivo animal imaging," *Sci. Rep.* **8**(1), 6401 (2018).
77. T. C. Hsiao et al., "Deep-penetration photoacoustic array imaging of calcifications," *J. Biomed. Opt.* **18**(6), 066002 (2013).
78. M. Mozaffarzadeh et al., "Double-stage delay multiply and sum beamforming algorithm: application to linear-array photoacoustic imaging," *IEEE Trans. Biomed. Eng.* **65**(1), 31–42 (2018).
79. Q. M. Barber and R. J. Zemp, "Photoacoustic-ultrasound tomography with S-sequence aperture encoding," *IEEE Trans. Ultrason Ferroelectr. Freq. Control* **64**(4), 688–693 (2017).
80. E. M. A. Anas et al., "Enabling fast and high quality LED photoacoustic imaging: a recurrent neural networks based approach," *Biomed. Opt. Express* **9**(8), 3852–3866 (2018).
81. A. Dima and V. Ntziachristos, "Non-invasive carotid imaging using optoacoustic tomography," *Opt. Express* **20**(22), 25044–25057 (2012).
82. M. A. Naser et al., "Improved photoacoustic-based oxygen saturation estimation with SNR-regularized local fluence correction," *IEEE Trans. Med. Imaging* **38**(2), 561–571 (2019).
83. S. E. Bohndiek et al., "Development and application of stable phantoms for the evaluation of photoacoustic imaging instruments," *PLoS One* **8**(9), e75533 (2013).
84. T. Petrosyan et al., "Rapid scanning wide-field clutter elimination in epi-optoacoustic imaging using comb LOVIT," *Photoacoustics* **10**, 20–30 (2018).
85. A. Hariri et al., "The characterization of an economic and portable LED-based photoacoustic imaging system to facilitate molecular imaging," *Photoacoustics* **9**, 10–20 (2018).
86. Z. Chen et al., "Performance of optoacoustic and fluorescence imaging in detecting deep-seated fluorescent agents," *Biomed. Opt. Express* **9**(5), 2229–2239 (2018).
87. V. P. Nguyen et al., "Feasibility study on photoacoustic guidance for high-intensity focused ultrasound-induced hemostasis," *J. Biomed. Opt.* **19**(10), 105010 (2014).
88. H. He et al., "Improving optoacoustic image quality via geometric pixel super-resolution approach," *IEEE Trans. Med. Imaging* **35**(3), 812–818 (2016).
89. A. Stylogiannis et al., "Continuous wave laser diodes enable fast optoacoustic imaging," *Photoacoustics* **9**, 31–38 (2018).
90. D. Van de Sompel et al., "Comparison of deconvolution filters for photoacoustic tomography," *PLoS One* **11**(3), e0152597 (2016).
91. E. Najafzadeh et al., "Photoacoustic image improvement based on a combination of sparse coding and filtering," *J. Biomed. Opt.* **25**(10), 106001 (2020).
92. A. Hariri et al., "Polyacrylamide hydrogel phantoms for performance evaluation of multi-spectral photoacoustic imaging systems," *Photoacoustics* **22**, 100245 (2021).
93. W. Wei et al., "Integrated ultrasound and photoacoustic probe for co-registered intravascular imaging," *J. Biomed. Opt.* **16**(10), 106001 (2011).
94. M. Wu et al., "Impact of device geometry on the imaging characteristics of an intravascular photoacoustic catheter," *Appl. Opt.* **53**(34), 8131–8139 (2014).
95. X. Wen et al., "High-robustness intravascular photoacoustic endoscope with a hermetically sealed opto-sono capsule," *Opt. Express* **28**(13), 19255–19269 (2020).
96. G. Wurzinger et al., "Simultaneous three-dimensional photoacoustic and laser-ultrasound tomography," *Biomed. Opt. Express* **4**(8), 1380–1389 (2013).
97. L. Xi et al., "Design and evaluation of a hybrid photoacoustic tomography and diffuse optical tomography system for breast cancer detection," *Med. Phys.* **39**(5), 2584–2594 (2012).
98. E. Filoux et al., "High-frequency annular array with coaxial illumination for dual-modality ultrasonic and photoacoustic imaging," *Rev. Sci. Instrum.* **84**(5), 053705 (2013).

99. B. Wang et al., "Modified back-projection method in acoustic resolution-based photoacoustic endoscopy for improved lateral resolution," *Med. Phys.* **45**(10), 4430–4438 (2018).
100. W. Xia et al., "Performance characteristics of an interventional multispectral photoacoustic imaging system for guiding minimally invasive procedures," *J. Biomed. Opt.* **20**(8), 086005 (2015).
101. N. Huang et al., "Curved-array-based multispectral photoacoustic imaging of human finger joints," *IEEE Trans. Biomed. Eng.* **65**(7), 1452–1459 (2018).
102. H. Leng et al., "Characterization of a fiber bundle-based real-time ultrasound/photoacoustic imaging system and its *in vivo* functional imaging applications," *Micromachines (Basel)* **10**(12), 820 (2019).
103. C. Miranda et al., "Side-viewing photoacoustic waveguide endoscopy," *Photoacoustics* **19**, 100167 (2020).
104. S. Jeon et al., "Real-time delay-multiply-and-sum beamforming with coherence factor for *in vivo* clinical photoacoustic imaging of humans," *Photoacoustics* **15**, 100136 (2019).
105. R. Ansari et al., "All-optical forward-viewing photoacoustic probe for high-resolution 3D endoscopy," *Light Sci. Appl.* **7**, 75 (2018).
106. H. He et al., "Hybrid optical and acoustic resolution optoacoustic endoscopy," *Opt. Lett.* **41**(12), 2708–2710 (2016).
107. M. Oeri et al., "Calibrated linear array-driven photoacoustic/ultrasound tomography," *Ultrasound Med. Biol.* **42**(11), 2697–2707 (2016).
108. I. Kosik et al., "Intraoperative photoacoustic screening of breast cancer: a new perspective on malignancy visualization and surgical guidance," *J. Biomed. Opt.* **24**(5), 056002 (2019).
109. K. J. Francis et al., "Multiview spatial compounding using lens-based photoacoustic imaging system," *Photoacoustics* **13**, 85–94 (2019).
110. S. Agrawal et al., "Design, development, and multi-characterization of an integrated clinical transrectal ultrasound and photoacoustic device for human prostate imaging," *Diagnostics (Basel)* **10**(8), 566 (2020).
111. X. Wang et al., "Photoacoustic imaging with a commercial ultrasound system and a custom probe," *Ultrasound Med. Biol.* **37**(3), 484–492 (2011).
112. S. R. Kothapalli et al., "Deep tissue photoacoustic imaging using a miniaturized 2-D capacitive micromachined ultrasonic transducer array," *IEEE Trans. Biomed. Eng.* **59**(5), 1199–1204 (2012).
113. S. A. Ermilov et al., "Three-dimensional optoacoustic and laser-induced ultrasound tomography system for preclinical research in mice: design and phantom validation," *Ultrason Imaging* **38**(1), 77–95 (2016).
114. J. Tang et al., "Wearable 3-D photoacoustic tomography for functional brain imaging in behaving rats," *Sci. Rep.* **6**, 25470 (2016).
115. M. Li et al., "Internal-illumination photoacoustic computed tomography," *J. Biomed. Opt.* **23**(3), 030506 (2018).
116. S. H. Pun et al., "Monolithic multiband CMUTs for photoacoustic computed tomography with *in vivo* biological tissue imaging," *IEEE Trans. Ultrason Ferroelectr. Freq. Control* **65**(3), 465–475 (2018).
117. G. Zhang et al., "Developing a photoacoustic whole-breast imaging system based on the synthetic matrix array," *Front. Phys.* **8**, 600589 (2020).
118. H. Ke et al., "Performance characterization of an integrated ultrasound, photoacoustic, and thermoacoustic imaging system," *J. Biomed. Opt.* **17**(5), 056010 (2012).
119. K. B. Chowdhury et al., "A synthetic total impulse response characterization method for correction of hand-held optoacoustic images," *IEEE Trans. Med. Imaging* **39**(10), 3218–3230 (2020).
120. A. Dima, N. C. Burton, and V. Ntziachristos, "Multispectral optoacoustic tomography at 64, 128, and 256 channels," *J. Biomed. Opt.* **19**(3), 036021 (2014).
121. A. Buehler et al., "Volumetric optoacoustic imaging with multi-bandwidth deconvolution," *IEEE Trans. Med. Imaging* **33**(4), 814–821 (2014).
122. X. L. Dean-Ben, "Hybrid-array-based optoacoustic and ultrasound (OPUS) imaging of biological tissues," *Appl. Phys. Lett.* **110**, 203703 (2017).

123. P. R. Torke, R. Nuster, and G. Paltauf, "Conical ring array detector for large depth of field photoacoustic macroscopy," *Biomed. Opt. Express* **11**(5), 2461–2475 (2020).
124. K. J. Francis et al., "Characterization of lens based photoacoustic imaging system," *Photoacoustics* **8**, 37–47 (2017).
125. B. Wang et al., "Photoacoustic tomography system for noninvasive real-time three-dimensional imaging of epilepsy," *Biomed. Opt. Express* **3**(6), 1427–1432 (2012).
126. R. Nagaoka et al., "Visualization of murine lymph vessels using photoacoustic imaging with contrast agents," *Photoacoustics* **9**, 39–48 (2018).
127. H. Zafar et al., "Linear-array-based photoacoustic imaging of human microcirculation with a range of high frequency transducer probes," *J. Biomed. Opt.* **20**(5), 051021 (2015).
128. L. J. Rich et al., "Performance characteristics of photoacoustic imaging probes with varying frequencies and light-delivery schemes," *Ultrason. Imaging* **41**(6), 319–335 (2019).
129. E. Mercep, X. L. Dean-Ben, and D. Razansky, "Combined pulse-echo ultrasound and multispectral photoacoustic tomography with a multi-segment detector array," *IEEE Trans. Med. Imaging* **36**(10), 2129–2137 (2017).
130. C. Jia et al., "Two-layer heterogeneous breast phantom for photoacoustic imaging," *J. Biomed. Opt.* **22**(10), 106011 (2017).
131. T. Ida et al., "Real-time photoacoustic imaging system for burn diagnosis," *J. Biomed. Opt.* **19**(8), 086013 (2014).
132. P. Ephrat et al., "Localization of spherical lesions in tumor-mimicking phantoms by 3D sparse array photoacoustic imaging," *Med. Phys.* **37**(4), 1619–1628 (2010).
133. R. Manwar, K. Kratkiewicz, and K. Avnani, "Investigation of the effect of the skull in transcranial photoacoustic imaging: a preliminary *ex vivo* study," *Sensors (Basel)* **20**(15) (2020).
134. A. A. Oraevsky et al., "Clinical photoacoustic imaging combined with ultrasound for coregistered functional and anatomical mapping of breast tumors," *Photoacoustics* **12**, 30–45 (2018).
135. W. Ren et al., "Automated registration of magnetic resonance imaging and photoacoustic tomography data for experimental studies," *Neurophotonics* **6**(2), 025001 (2019).
136. A. Cheng et al., "Direct three-dimensional ultrasound-to-video registration using photoacoustic markers," *J. Biomed. Opt.* **18**(6), 066013 (2013).
137. K. Daoudi et al., "Handheld probe integrating laser diode and ultrasound transducer array for ultrasound/photoacoustic dual modality imaging," *Opt. Express* **22**(21), 26365–26374 (2014).
138. M. Jaeger, J. C. Bamber, and M. Frenz, "Clutter elimination for deep clinical photoacoustic imaging using localised vibration tagging (LOVIT)," *Photoacoustics* **1**(2), 19–29 (2013).
139. S. J. Arconada-Alvarez et al., "The development and characterization of a novel yet simple 3D printed tool to facilitate phantom imaging of photoacoustic contrast agents," *Photoacoustics* **5**, 17–24 (2017).
140. A. B. Karpouk, B. Wang, and S. Y. Emelianov, "Development of a catheter for combined intravascular ultrasound and photoacoustic imaging," *Rev. Sci. Instrum.* **81**(1), 014901 (2010).
141. X. Bai et al., "Intravascular optical-resolution photoacoustic tomography with a 1.1 mm diameter catheter," *PLoS One* **9**(3), e92463 (2014).
142. U. Alqasbi et al., "Interlaced photoacoustic and ultrasound imaging system with real-time coregistration for ovarian tissue characterization," *J. Biomed. Opt.* **19**(7), 076020 (2014).
143. J. Su et al., "Photoacoustic imaging of clinical metal needles in tissue," *J. Biomed. Opt.* **15**(2), 021309 (2010).
144. W. Xia et al., "Design and evaluation of a laboratory prototype system for 3D photoacoustic full breast tomography," *Biomed. Opt. Express* **4**(11), 2555–2569 (2013).
145. C. J. Ho et al., "Multifunctional photosensitizer-based contrast agents for photoacoustic imaging," *Sci. Rep.* **4**, 5342 (2014).
146. A. B. Attia et al., "Phthalocyanine photosensitizers as contrast agents for *in vivo* photoacoustic tumor imaging," *Biomed. Opt. Express* **6**(2), 591–598 (2015).
147. C. Y. Lee et al., "Photoacoustic imaging to localize indeterminate pulmonary nodules: a preclinical study," *PLoS One* **15**(4), e0231488 (2020).

148. C. Kim et al., “Multifunctional microbubbles and nanobubbles for photoacoustic and ultrasound imaging,” *J. Biomed. Opt.* **15**(1), 010510 (2010).
149. Y. Shi et al., “Targeted Au-core-Ag-shell nanorods as a dual-functional contrast agent for photoacoustic imaging and photothermal therapy,” *Biomed. Opt. Express* **7**(5), 1830–1841 (2016).
150. S. J. Yoon et al., “Utility of biodegradable plasmonic nanoclusters in photoacoustic imaging,” *Opt. Lett.* **35**(22), 3751–3753 (2010).
151. A. Ray et al., “Targeted blue nanoparticles as photoacoustic contrast agent for brain tumor delineation,” *Nano Res.* **4**(11), 1163–1173 (2011).
152. Y. Li et al., “Hybrid polymeric nanoprobe for folate receptor-targeted photoacoustic imaging in vivo,” *Mater. Chem. Front.* **1**(5), 916–921 (2017).
153. L. Lim et al., “Can photoacoustic imaging quantify surface-localized J-aggregating nanoparticles?” *J. Biomed. Opt.* **22**(7), 076008 (2017).
154. E. Jung et al., “Molecularly engineered theranostic nanoparticles for thrombosed vessels: H₂O₂-activatable contrast-enhanced photoacoustic imaging and antithrombotic therapy,” *ACS Nano* **12**(1), 392–401 (2018).
155. C. Liu et al., “Switchable photoacoustic imaging of glutathione using MnO₂ nanotubes for cancer diagnosis,” *ACS Appl. Mater. Interfaces* **10**(51), 44231–44239 (2018).
156. Y. Yang et al., “A 1064 nm excitable semiconducting polymer nanoparticle for photoacoustic imaging of gliomas,” *Nanoscale* **11**(16), 7754–7760 (2019).
157. J. Lavaud et al., “Exploration of melanoma metastases in mice brains using endogenous contrast photoacoustic imaging,” *Int. J. Pharm.* **532**(2), 704–709 (2017).
158. H. Q. Wu et al., “Scanning photoacoustic imaging of submucosal gastric tumor based on a long focused transducer in phantom and *in vitro* experiments,” *J. Innov. Opt. Health Sci.* **12**(3), 1950011 (2019).
159. M. K. A. Singh and W. Steenbergen, “Photoacoustic-guided focused ultrasound (PAFUSion) for identifying reflection artifacts in photoacoustic imaging,” *Photoacoustics* **3**(4), 123–131 (2015).
160. H. N. Y. Nguyen and W. Steenbergen, “Three-dimensional view of out-of-plane artifacts in photoacoustic imaging using a laser-integrated linear-transducer-array probe,” *Photoacoustics* **19**, 100176 (2020).
161. X. L. Dean-Ben, E. Bay, and D. Razansky, “Functional optoacoustic imaging of moving objects using microsecond-delay acquisition of multispectral three-dimensional tomographic data,” *Sci. Rep.* **4**, 5878 (2014).
162. Z. Guo, L. Li, and L. V. Wang, “On the speckle-free nature of photoacoustic tomography,” *Med. Phys.* **36**(9), 4084–4088 (2009).
163. E. R. Hill et al., “Identification and removal of laser-induced noise in photoacoustic imaging using singular value decomposition,” *Biomed. Opt. Express* **8**(1), 68–77 (2017).
164. M. K. A. Singh et al., “Photoacoustic reflection artifact reduction using photoacoustic-guided focused ultrasound: comparison between plane-wave and element-by-element synthetic backpropagation approach,” *Biomed. Opt. Express* **8**(4), 2245–2260 (2017).
165. M. K. Singh et al., “*In vivo* demonstration of reflection artifact reduction in photoacoustic imaging using synthetic aperture photoacoustic-guided focused ultrasound (PAFUSion),” *Biomed. Opt. Express* **7**(8), 2955–2972 (2016).
166. H. N. Y. Nguyen and W. Steenbergen, “Reducing artifacts in photoacoustic imaging by using multi-wavelength excitation and transducer displacement,” *Biomed. Opt. Express* **10**(7), 3124–3138 (2019).
167. H. N. Y. Nguyen, A. Hussain, and W. Steenbergen, “Reflection artifact identification in photoacoustic imaging using multi-wavelength excitation,” *Biomed. Opt. Express* **9**(10), 4613–4630 (2018).
168. M. A. Boss et al., *Magnetic Resonance Imaging Biomarker Calibration Service: Proton Spin Relaxation Times*, National Institute of Standards and Technology (2018).
169. B. Zhu et al., “Determining the performance of fluorescence molecular imaging devices using traceable working standards with SI units of radiance,” *IEEE Trans. Med. Imaging* **35**(3), 802–811 (2016).

170. P. Lemaillot, J.-P. Bouchard, and D.W. Allen, "Development of traceable measurement of the diffuse optical properties of solid reference standards for biomedical optics at National Institute of Standards and Technology," *Appl. Opt.* **54**(19), 6118–6127 (2015).
171. M. Gehrung, S. E. Bohndiek, and J. Brunker, "Development of a blood oxygenation phantom for photoacoustic tomography combined with online pO₂ detection and flow spectrometry," *J. Biomed. Opt.* **24**(12), 121908 (2019).
172. M. Dantuma, R. van Dommelen, and S. Manohar, "Semi-anthropomorphic photoacoustic breast phantom," *Biomed. Opt. Express* **10**(11), 5921–5939 (2019).
173. Y. Liu et al., "Biomimetic 3D-printed neurovascular phantoms for near-infrared fluorescence imaging," *Biomed. Opt. Express* **9**(6), 2810–2824 (2018).
174. M. A. Gavrielides et al., "A resource for the assessment of lung nodule size estimation methods: database of thoracic CT scans of an anthropomorphic phantom," *Opt. Express* **18**(14), 15244–15255 (2010).
175. J. Pfefer and A. Agrawal, "A review of consensus test methods for established medical imaging modalities and their implications for optical coherence tomography," *Proc. SPIE* **8215**, 82150D (2012).
176. B. W. Pogue et al., "Image analysis methods for diffuse optical tomography," *J. Biomed. Opt.* **11**(3), 033001 (2006).

Jorge Palma-Chavez is a postdoctoral scholar in the in the Department of NanoEngineering at UC San Diego and an NSF Scholar-in-Residence fellow at FDA. He received his PhD in biomedical engineering from Texas A&M University, where he developed contrast agents for optical imaging techniques and designed vascular-targeted drug delivery platforms. He is currently developing phantom-based test methods for image quality assessment of photoacoustic imaging systems.

T. Joshua Pfefer received his PhD in biomedical engineering from the University of Texas at Austin and trained as a research fellow at the Wellman Laboratories of Photomedicine. He joined FDA in 2000 and currently leads the Optical Spectroscopy and Spectral Imaging Program. His research focuses on elucidating light-tissue interactions of emerging technologies and developing new methods to improve and standardize testing. In 2018, he was named a fellow of SPIE for contributions to biophotonics.

Anant Agrawal is a research scientist in the Division of Biomedical Physics at FDA's Center for Devices and Radiological Health, where he specializes in optical imaging and spectroscopy for medical device applications. Prior to joining FDA in 2003 he worked in the optical diagnostic medical device industry for five years. He received BS, MS, and PhD degrees in electrical engineering from the University of Virginia, the University of Texas, and Catholic University of America, respectively.

Jesse V. Jokerst is a professor in the Department of NanoEngineering at UC San Diego. He has received the NIH K99/R00 Pathway to Independence Award, the NIH New Innovator Award, the NSF CAREER Award, and Stanford Radiology Alumni of the Year Award.

William C. Vogt received his BS degree in mechanical engineering from the University of Massachusetts Amherst in 2009 and his PhD in biomedical engineering from Virginia Tech in 2013. Since joining the FDA in 2013, he has been conducting regulatory science to develop tools and test methods for evaluating the safety and effectiveness of photoacoustic imaging devices. His research interests include photoacoustic imaging, tissue phantoms, nanoparticles, standardization, and biophotonic medical device characterization and evaluation.