UC Irvine
UC Irvine Previously Published Works

Title

Performance assessment of diffuse optical spectroscopic imaging instruments in a 2-year multicenter breast cancer trial

Permalink

Journal

ISSN

Authors

Leproux, Anaïs
O'Sullivan, Thomas D
Cerussi, Albert
et al.

Publication Date

DOI

Peer reviewed

Performance assessment of diffuse optical spectroscopic imaging instruments in a2-year multicenter breast cancer trial

Anaïs Leproux,[a] Thomas D. O'Sullivan,[a,b] Albert Cerussi,[a] Amanda Durkin,[a] Brian Hill,[a] Nola Hylton,[c] Arjun G. Yodh,[d] Stefan A. Carp,[e] David Boas,[e] Shudong Jiang,[f] Keith D. Paulsen,[f] Brian Pogue,[f] Darren Roblyer,[g] Wei Yang,[h] and Bruce J. Tromberg[a,*]

[a]University of California, Beckman Laser Institute and Medical Clinic, Irvine, California, United States
[b]University of Notre Dame, Department of Electrical Engineering, Notre Dame, Indiana, United States
[c]University of California, Department of Radiology, San Francisco, California, United States
[d]University of Pennsylvania, Department of Physics and Astronomy, Philadelphia, Pennsylvania, United States
[e]Massachusetts General Hospital, Harvard Medical School, Department of Radiology, Charlestown, Massachusetts, United States
[f]Dartmouth College, Thayer School of Engineering, Hanover, New Hampshire, United States
[g]Boston University, Department of Biomedical Engineering, Boston, Massachusetts, United States
[h]University of Texas MD Anderson Cancer Center, Department of Diagnostic Radiology, Houston, Texas, United States

Abstract. We present a framework for characterizing the performance of an experimental imaging technology, diffuse optical spectroscopic imaging (DOSI), in a 2-year multicenter American College of Radiology Imaging Network (ACRIN) breast cancer study (ACRIN-6691). DOSI instruments combine broadband frequency-domain photon migration with time-independent near-infrared (650 to 1000 nm) spectroscopy to measure tissue absorption and reduced scattering spectra and tissue hemoglobin, water, and lipid composition. The goal of ACRIN- 6691 was to test the effectiveness of optically derived imaging endpoints in predicting the final pathologic response of neoadjuvant chemotherapy (NAC). Sixty patients were enrolled over a 2-year period at participating sites and received multiple DOSI scans prior to and during 3- to 6-month NAC. The impact of three sources of error on accuracy and precision, including different operators, instruments, and calibration standards, was evaluated using a broadband reflectance standard and two different solid tissue-simulating optical phantoms. Instruments showed $<0.0010~\text{mm}^{-1}$ (10.3%) and $0.06~\text{mm}^{-1}$ (4.7%) deviation in broadband absorption and reduced scattering, respectively, over the 2-year duration of ACRIN-6691. These variations establish a useful performance criterion for assessing instrument stability. The proposed procedures and tests are not limited to DOSI; rather, they are intended to provide methods to characterize performance of any instrument used in translational optical imaging. © *The Authors. Published by SPIE under a Creative Commons Attribution 3.0 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.22.12.121604]*

1 Introduction

A variety of diffuse optical spectroscopic imaging (DOSI) methods for the detection, diagnosis, and clinical management of breast cancer are currently under investigation in clinical studies. DOSI, which employs near-infrared light, is a thick-tissue functional method capable of quantifying tissue oxy- and deoxy-hemoglobin, water, and bulk lipid concentration.[1] Recent applications of diffuse optical methods in breast cancer include monitoring chemotherapy[2–7] or hormonal therapy,[8] assessing breast density,[9–11] identifying tumors in dense breast tissue,[12] and differential diagnosis.[13–15]

In April 2011, we initiated the American College of Radiology Imaging Network (ACRIN)-6691 trial (ACRIN-6691) to test the effectiveness of DOSI to predict the final pathologic response of breast tumors in patients treated with neoadjuvant (i.e., presurgical) chemotherapy (NAC).[5] Sixty breast cancer patients receiving NAC were measured longitudinally using DOSI in six clinical research centers across the United States: Dartmouth College, Massachusetts General Hospital (MGH), MD Anderson Cancer Center (MDACC), University of California Irvine (UCI), University of California San Francisco (UCSF), and University of Pennsylvania (Upenn).

---

*Address all correspondence to: Bruce J. Tromberg, E-mail: bjtrombe@uci.edu

Prior to study activation, five identical DOSI platforms were distributed to all sites, excluding MDACC, which joined the study about 1.5 years later. An updated, miniaturized version of the DOSI instrument was delivered to MDACC. To prevent introducing additional variation, only data from the five identical instruments that were built simultaneously are compared in this study. Instrument calibration tools consisted of identical optical reflectance standards and breast tissue-simulating phantoms fabricated from the same batch of polysiloxane materials. Procedures and methods for instrument quality control, calibration operating procedures, and patient measurement protocols were adopted and revised from a previous multicenter breast cancer imaging study established by the National Cancer Institute (NCI) Network for Translational Research Optical Imaging.[16] These methods were designed to be simple to perform by any trained operator and incorporate redundant mechanisms to assure good data quality, such as multiple calibration steps and premeasurement system checks. The lead site (UCI) provided all operators with comprehensive training materials, including standard operating manuals and videos,[17] and several on-site measurement training sessions using human subjects. To ensure a consistently high level of operator performance, we offered support before, during, and after patient measurements and provided operator feedback in response to the observed data quality. Postcollection, data were sent to the lead site for quality control and processing using standardized methods.

To compare longitudinal measurements from multiple patients across sites, platforms, and operators, we developed a systematic approach for assessing the impact of three major potential sources of variability on DOSI accuracy and precision over the 2-year study duration: (1) calibration phantoms, (2) instrument performance, and (3) multiple operators. Starting with the test case of 1 operator running 1 instrument using 1 calibration phantom, we evaluate how each potential source of error contributes to overall DOSI performance. Results are presented across the five different sites and three relevant time frames: the duration of a patient measurement session (~1 h), the duration of the chemotherapy treatment (~4 months), and the entire study time frame (2 years).

2 Materials and Methods

2.1 *Diffuse Optical Spectroscopic Imaging* Instrumentation

A description of the basic components of the DOSI instrument has been previously presented.[18,19] Briefly, DOSI consists of broadband frequency-domain photon migration (FDPM) combined with broadband time-independent near-infrared spectroscopy (CW-NIRS). This provides absorption and reduced scattering spectra in centimeter-thick tissues from 650 to 1000 nm. The FDPM component uses six laser diodes (Blue Sky Research, Sanyo, Mitsubishi) at the wavelengths 660, 680, 785, 810, 830, and 850 nm that are coupled into a fiber bundle and delivered to a handpiece that contacts the tissue. The breast is illuminated sequentially by each laser diode, which is intensity modulated at 601 modulation frequencies from 50 to 600 MHz. Phase and amplitude are measured from three averaged frequency sweeps for a total of ~200 ms per laser diode. The diffusely reflected scattered light propagates through tissue and is detected by an avalanche photodiode (Hamamatsu module C5658) mounted inside a temperature-controlled hand-held probe. The CW-NIRS component uses a high-intensity tungsten–halogen source (Ocean Optics model HL-2000-HP-FHSA) to illuminate the tissue. The diffusely scattered broadband light that propagates through the tissue is detected by a grating-based spectrometer that collects light from 650 to 1000 nm (1024 pixels, BWTek model 611E). The FDPM and CW-NIRS sources are coupled by independent optical fibers mounted into the hand-held probe. The source–detector separation was 22 mm (phantom measurement) or 28 mm (breast measurement) for both FDPM and CW-NIRS components with fibers placed in an overlapping geometry.[20] A typical DOSI measurement for the Network Analyzer-based instrument used in this study requires a total of 2 to 5 s per scan location and is composed of three averaged FDPM sweeps of 601 modulation frequencies and time-integrated broadband CW reflectance spectra.

FDPM and CW-NIRS data are combined using a diffusion theory model to provide broadband absorption (μ_a) and reduced scattering (μ_s') spectra from 650 to 1000 nm. In the FDPM measurement, the phase and amplitude of the detected light are recorded as functions of source modulation frequency and fit to a diffusive model of light transport with semi-infinite boundary conditions to recover μ_a and μ_s' at each of the six laser wavelengths. CW-NIRS broadband reflectance spectra were converted into absolute absorption spectra using two steps. First, as per Mie scattering theory, the spectral shape of the reduced scattering spectrum is assumed to follow a power law of the form $\mu_s' = A\mu^{-sp}$, where A is the scatter amplitude and sp is the scatter power or the exponent of the reduced scattering spectrum. The power-law fit to the FDPM discrete laser diode spectrum provides a scatter correction for the CW-NIRS reflectance spectrum. We then fit the CW-NIRS reflectance intensity at each of the laser diode wavelengths to the reflectance calculated from the FDPM-measured absolute absorption values. Thus, the CW-

NIRS reflectance spectrum intensity is quantitatively scaled using the FDPM discrete μ_a measurements. The absolute absorption spectrum is then extracted by fitting the corrected reflectance spectrum to a diffusion reflectance model.[18]

A picture of the DOSI instrument used in this study is shown in Fig. 1(a). The calibration tools of the DOSI instrument are shown in Figs. 1(b) and 1(c). It consists of a solid tissue-simulating phantom [Fig. 1(b)], which is used to measure the FDPM instrument response function (IRF) (explicitly, phase offsets, and amplitude scale factors), and a reflectance standard for the broadband wavelength calibration [Fig. 1(c)]. The FDPM IRF measurement is similar to "blanking" performed in a conventional spectrophotometer; using a known attenuation (such as air), one can remove instrument artifacts such as wave-length-dependent detector responses. In the case of the FDPM calibration, the tissue-simulating phantoms have known μ_a and μ_s' values, such that the phase and amplitude changes due to tissue (or sample) can be separated from the IRF. In the case of the broadband wavelength calibration, a standard spectralon reflectance standard (Labsphere) is used to correct the wave-length-dependent variations caused by the light source, detector response, and optical fiber throughput.

2.2 ACRIN-6691 Trial—Protocol and Study Design

The ACRIN-6691 trial was designed to evaluate the sensitivity of DOSI to monitor and predict the pathologic response of breast tumors in patients treated with NAC.[5] The long-term goal is to provide oncologists with a simple, risk-free bedside tool that can be used to inform medical decisions. These include modifying or adapting the chemotherapy regimen, duration, and timing of surgery to maximize therapeutic response and minimize unnecessary toxicity.

A total of 60 breast cancer patients receiving NAC were measured longitudinally using the DOSI instrument in six different clinical sites. The five following sites, Dartmouth, MGH, UCI, UCSF, and Upenn used identical DOSI instruments. MDACC joined the study about 1.5 year after its activation and used a next generation instrument that was excluded from this analysis. The subjects were measured four times throughout NAC: base-line (pretreatment), early therapy (5 to 10 days after the initiation of the first cycle of the chemotherapy), midtherapy (halfway through the therapy protocol), and posttherapy (prior to surgery). For these patients, the typical chemotherapy length is 4 months.

Fig. 1 (a) DOSI instrument and (b) and (c) illustrate the probe placed on the calibration phantom (FDPM calibration) and on the reflectance standard (broadband calibration), respectively.

2.3 *Solid Phantom and Reflectance* Standard—*Calibration and Testing*

Each ACRIN site possessed one broadband reflectance standard and two solid tissue-simulating optical phantoms, Ph-A and Ph-B. The Ph-A phantoms, fabricated at UCI, were issued from the same fabrication batch specifically for the purpose of the ACRIN study. The specifics of these phantoms, including the fabrication process, have been previously published.[21] The Ph-B phantoms were developed by Institut National d'Optique, Inc. (Quebec, Canada). The phantoms within a Ph series are expected to have the same optical properties as they were fabricated from the same batch. All phantoms have similar optical properties as breast tissue. The Ph-A and Ph-B phantoms were arbitrarily chosen to calibrate the FDPM instrument response during patient measurements and used as an additional calibration phantom for instrument testing, respectively. Measuring the IRF each session corrects any possible IRF changes between measurements. To determine IRF, the optical properties of the reference phantom must be known, and these were derived from comprehensive series of multidistance, multifrequency, broadband measurements, a standard procedure that, assuming phantom homogeneity, utilizes the distance-dependent phase and amplitude behavior to calculate optical properties.[21,22]

The solid phantoms are contained in a custom-made case, as shown Fig. 1(b). Figure 1(c) shows the reflectance standard used to calibrate the broadband part of the instrument. A mask at the top of the phantom case and the reflectance standard is designed to limit placement error by helping the operator to repeatedly position the probe on the same location.[21]

Three sets of three calibration measurements using the two phantoms and the reflectance standard were performed for every patient measurement. These calibration measurements were acquired immediately before, midway through, and after each breast patient measurement. Iterating the three calibration measurements throughout the measurement session serves two purposes: (1) securing a robust, redundant calibration dataset and (2) assessing instrument performance throughout the session by monitoring the optical properties of a fixed phantom over time.

2.4 *Data Analysis*

Several phantom measurement procedures were designed to test instrument performance. These procedures consisted of repeat measurements either of the same phantom or of different phantoms and in various conditions. We report the results of these tests using the standard deviation (Stdev) of these series of measurements and the normalized standard deviation (NSD), which is defined as the Stdev/mean and reported in percent (%). In case of broadband data, we first obtain the Stdev and NSD at all wavelengths of the spectrum and report their spectral average, minimum, maximum, and median.

3 Results

3.1 *Methodology to Assess Uncertainties for* the ACRIN-6691 Study

Figure 2 presents the overview of the methodology used to assess the uncertainties in the context of the ACRIN-6691 study. Uncertainties originate from three main sources: phantom, instrument, and operator. Factors
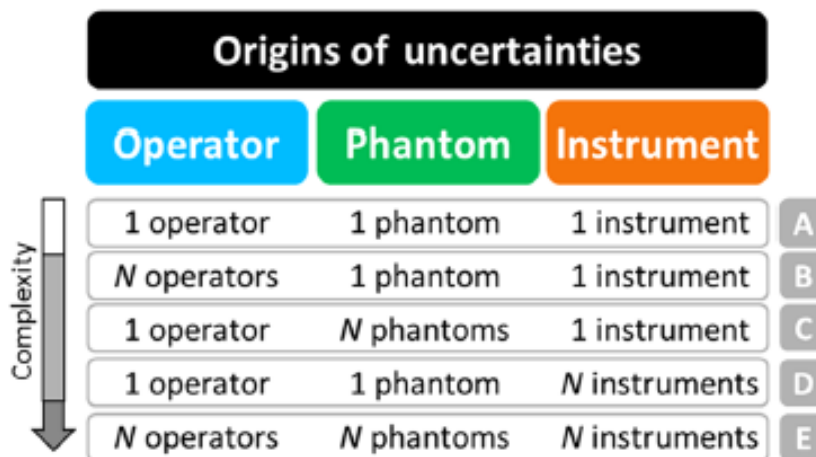

Fig. 2 Overview of methodology to assess uncertainties.

Conditions	Test description	Time frame	No. of scans
A	A1: Placement test over imaging session duration	1 h	9
	A2: Instrument stability over imaging session duration	1 h	60
	A3: Instrument stability over chemotherapy duration	4 months	36
B	B1: Instrument stability over study duration	2 years	150
C	C1: Phantom comparison	1 h	18
D	D1: Interinstrument variability on one phantom	1 week	15
E	E1: Interinstrument variability; multiple phantoms and operators	1 week	30
	E2: Overall study variability over study duration	2 years	480

Table 1 Description of the different tests performed for each condition in the context of the validation and standardization process of the ACRIN-6691 study. The number of scans refers to independent measurements of combined FDPM and CW-NIRS consisting of multiple temporal and spectral sweeps and lasting 2 to 5 s per scan (see Sec. 2.1).

contributing to each of these include, but are not limited to, (1) phantom material consistency, shelf life, and spatial uniformity, (2) instrument warm up, stability, reproducibility, and environmental effects, and (3) procedural consistency, probe operation, and positioning by different operators. To understand the relative contributions from each of these sources, we grouped the DOSI calibration data into five main conditions. The variability in calibration data across all conditions (A through E) in Fig. 2 is potentially a sum of all these factors.

The first, condition A, examines the DOSI platform performance with one operator, one phantom, and one instrument, which is the most favorable situation in terms of potential sources of uncertainty. Conditions, B, C, and D, consecutively, assess the study uncertainties by varying one of the three test parameters, operator, phantom, and instrument, respectively, and holding two parameters constant. For example, condition B examines multiple operators (N) with one instrument and one phantom, while C examines multiple phantoms (N) with one operator and one instrument, etc. Condition E combines the contribution of all sources to DOSI measurement error by examining the performance of all instruments, all phantoms and with several operators. Multiple measurements and time periods were used to characterize each of the five main test conditions. A brief summary and description of each test for the five main conditions is provided in Table 1.

3.2 *Condition A: One Operator, One Phantom, and* One Instrument

Placement and short-term instrument stability tests were conducted on the DOSI instrument at UCI using a Ph-A phantom. For the placement test, test A1, the phantom was measured for nine consecutive times, removing and replacing the probe between each measurement over 1 h, which is the typical patient measurement session duration. For the instrument stability test, test A2, a Ph-A phantom was measured continuously every minute over a 1-h period with the probe fixed onto the phantom. For both the placement and 1-h stability tests, tests A1 and A2, the broadband deviations were found to be lower than 0.0010 mm^{-1} for absorption and 0.01 mm^{-1} for reduced scattering coefficients (data not shown).

Instrument stability was also investigated over a longer time frame to determine the potential impact of variations over the course of chemotherapy, test A3. Measurements of a Ph-A phantom using the DOSI instrument at

UCI were acquired by the same operator at the four study time-points over 4 months (typical time of chemotherapy treatment at UCI). This corresponded to nine scans per time-point, resulting in a total of 36 scans. Figure 3 and Table 2 provide the resulting deviations in broadband optical properties. They ranged from 0.0002 to 0.0005 mm^{-1} for absorption and from 0.01 to 0.03 mm^{-1} for reduced scattering coefficients. As shown in Fig. 3(a) inset, the greatest deviation in absorption occurred at day 1 and did not seem to follow an increasing or decreasing pattern over time.

3.3 *Condition B: N Operators, One Phantom, and* One Instrument
The instrument stability was investigated on a Ph-A phantom over the 2-year duration of the study using the DOSI instrument at UCI, test B1. A total of 150 scans were performed by four different trained operators on 17 different days during these 2 years. The same Ph-B phantom was used to calibrate the instrument response of all measurements. The resulting deviations in broadband absorption and reduced scattering are presented in Table 3 and Fig. 4. The deviations ranged from 0.0002 to 0.0007 mm^{-1} in absorption and from 0.01 to 0.03 mm^{-1} in reduced scattering coefficients.

3.4 *Condition C: One Operator, N Phantoms, and* One Instrument
Using the DOSI instrument at UCI, one operator measured each phantom three times from the series of Ph-A phantoms to validate their uniformity, test C1. The phantoms were manufactured from the same batch and, thus, should have the same optical properties. Figure 5 provides the broadband optical properties from these compiled
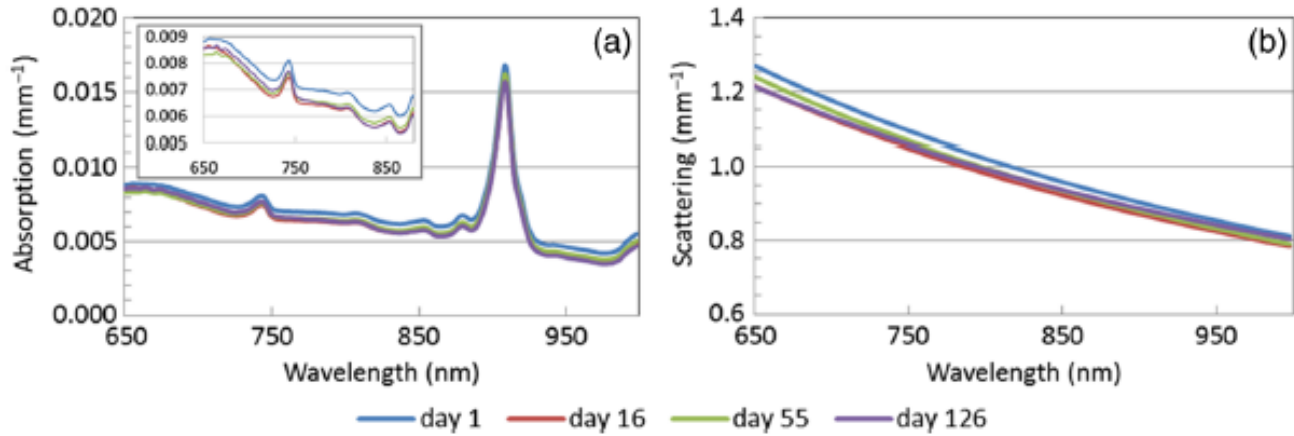

Fig. 3 Chemotherapy-long stability test of the DOSI instrument at UCI, test A3: broadband, (a) absorption and (b) reduced scattering, of a Ph-A phantom measured four times (day 1: presurgery, day 16: early therapy, day 55: midtherapy, and day 126: posttherapy) over a 4-month period. Inset: expanded scale (a).

Table 2 Deviations in broadband absorption and reduced scattering of a Ph-A phantom when measured four times over 4 months using the DOSI instrument at UCI.

Test A3	μ_a (mm^{-1})		μ_s' (mm^{-1})	
	Stdev	NSD (%)	Stdev	NSD (%)
Spectral average	0.0003	4.8	0.02	1.7
Spectral min	0.0002	2.4	0.01	1.5
Spectral max	0.0005	8.3	0.03	2.1
Spectral median	0.0003	4.3	0.02	1.6

Table 3 Deviations over 2 years in broadband absorption and reduced scattering of a Ph-A phantom measured at UCI by four different operators, test B1.

Test B1	μ_a (mm^{-1})		μ_s' (mm^{-1})	
	Stdev	NSD (%)	Stdev	NSD (%)
Spectral average	0.0004	5.3	0.02	1.8
Spectral min	0.0002	2.2	0.01	1.3
Spectral max	0.0007	8.9	0.03	2.5
Spectral median	0.0003	5.7	0.02	1.7

measurements, and Table 4 summarizes the Stdevs and coefficient of variation averaged over the whole spectral range (650 to 1000 nm). The broadband deviations within the series of Ph-A phantoms were found below 0.0003 and 0.01 mm−1 in absorption and in reduced scattering, respectively.

Overall, the absorption and reduced scattering coefficients were found to be similar among a series, which is consistent with the fact that they were fabricated in the same batch.

3.5 *Condition D: One Operator, One Phantom, and N Instruments*

The optical properties recovered from a Ph-A phantom measured using the different DOSI instruments across sites were compared. These measurements were performed by a single operator who traveled to the sites with a Ph-A phantom and an additional calibration phantom.

Figure 6 presents the recovered optical properties of Ph-A phantom from all sites. Similar absorption and reduced scattering coefficients were recovered by all instruments. The absorption spectra recovered at Upenn showed the largest disagreement compared to the other sites, specifically below 750 nm. Overall, the broadband deviations in absorption and in reduced scattering within the different instruments ranged from <0.0001 to 0.0010 mm−1 and from 0.01 to 0.02 mm−1, respectively (see Table 5).

3.6 *Condition E: N Operators, N Phantoms, and* N *Instruments*

The optical properties recovered from the Ph-A phantoms measured at their respective sites by the site-specific operator were compared, test E1. The deviations between the broadband- measured absorption and reduced scattering coefficients of the Ph-A phantoms range from 0.0001 to 0.0010 mm^{-1} and from to 0.02 mm^{-1}, respectively, see Table 6. The largest deviations in broadband absorption recovery originate from the lower wavelength range of the spectra, below 760 nm, see Fig. 7. As for condition D, this phenomenon is particularly visible for the data acquired at Upenn.
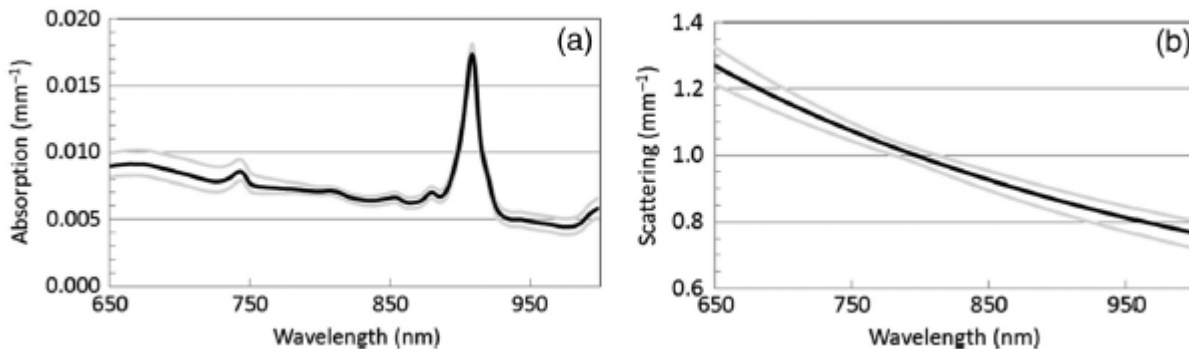

Fig. 4 Deviations in broadband (a) absorption and (b) reduced scattering of a Ph-A phantom measured 17 times over 2 years at UCI by four different operators, test B1. The black lines show the average spectra. The gray lines show the range (min/max) of spectra.
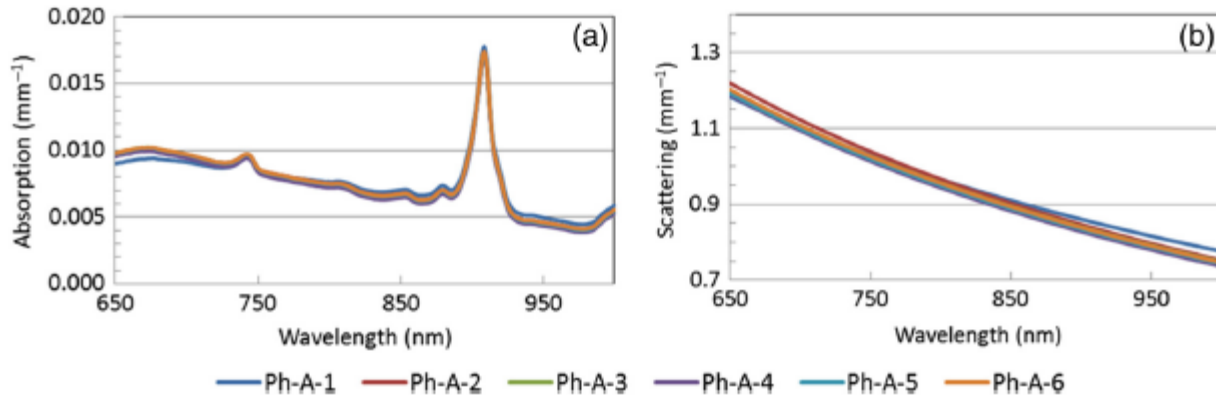
Fig. 5 (a) Absorption and (b) reduced scattering spectra of the Ph-A phantom series measured by one operator with a single DOSI instrument, test C1.

Table 4 Deviations in broadband absorption and reduced scattering within the Ph-A phantom series, test C1.

Test C1	μ_a (mm^{-1})		μ_s' (mm^{-1})	
	Stdev	NSD (%)	Stdev	NSD (%)
Spectral average	0.0002	2.8	0.01	1.2
Spectral min	0.0001	1.2	<0.01	0.8
Spectral max	0.0003	4.7	0.01	1.8
Spectral median	0.0002	2.7	0.01	1.0

Table 5 Deviations in broadband absorption and reduced scattering of a Ph-A phantom measured at five research sites by a single operator, test D1.

Test D1	μ_a (mm^{-1})		μ_s' (mm^{-1})	
	Stdev	NSD (%)	Stdev	NSD (%)
Spectral average	0.0004	6.3	0.01	1.4
Spectral min	<0.0001	0.6	0.01	1.2
Spectral max	0.0010	11.1	0.02	1.8
Spectral median	0.0003	6.8	0.01	1.4

Last, we compared optical properties recovered from all Ph-A phantoms measured at their respective sites by the on-site operators throughout the 2-year study duration, test E2. This corresponded to a total of 480 scans acquired on 58 different days by 18 different operators on the five different instruments; the resulting absorption and reduced scattering spectra are shown in Fig. 8. This situation represents the worst-case scenario in term of uncertainties, as it involved measurements on multiple Ph-A phantoms, by many different operators, using multiple instruments, over a 2-year period. Note that different Ph-B phantoms (the ones specific to the site) were also used to calibrate the instrument response of these measurements. The deviations between the broadband-measured absorption and reduced scattering coefficients of the Ph-A phantoms measured at their sites range from 0.0003 to 0.0010 mm^{-1} and from to 0.06 mm^{-1}, respectively, see Table 7.
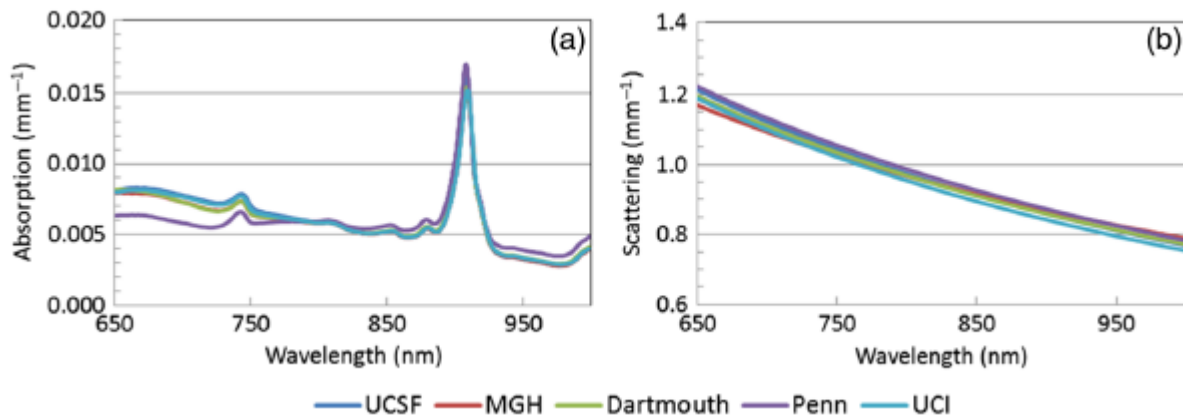
Fig. 6 (a) Broadband absorption and (b) reduced scattering spectra of a single Ph-A phantom measured at the five research sites by a single operator, test D1.

Table 6 Deviations in broadband absorption and reduced scattering of the Ph-A phantoms when measured at their corresponding ACRIN site, test E1.

Test E1	μ_a (mm^{-1})		μ_s' (mm^{-1})	
	Stdev	NSD (%)	Stdev	NSD (%)
Spectral average	0.0004	5.3	0.01	1.5
Spectral min	0.0001	0.8	0.01	1.4
Spectral max	0.0010	10.8	0.02	1.8
Spectral median	0.0002	5.2	0.01	1.5

Table 7 Deviations in broadband absorption and reduced scattering of 480 scans of the different Ph-A phantoms measured at their corresponding ACRIN site by 18 operators over 2 years, test E2.

Test E2	μ_a (mm^{-1})		μ_s' (mm^{-1})	
	Stdev	NSD (%)	Stdev	NSD (%)
Spectral average	0.0005	7.0	0.03	3.1
Spectral min	0.0003	4.1	0.02	2.6
Spectral max	0.0010	10.3	0.06	4.7
Spectral median	0.0004	7.1	0.03	2.9

4 Discussion

ACRIN-6691 is a multicenter trial of DOSI, an experimental imaging technology, used longitudinally to measure tumor functional changes during presurgical NAC. Because each patient measurement lasts up to ~4 months, and quantitative DOSI endpoints are used to predict the clinical outcome, it is essential that we develop new methods for assessing overall instrument accuracy and precision over the entire 2-year study period. As a result, we have devised a series of standardized tissue phantoms and procedures for evaluating the impact of different potential sources of uncertainly in DOSI measurements: variability between different instruments, operators, and calibration standards (tissue phantoms). These methods are designed to define the level of accuracy and precision required for

a multicenter study over several clinically relevant time frames: (1) 1 h, corresponding to the duration of a single patient measurement session, (2) several months, corresponding to the duration of the chemotherapy treatment, and (3) 2 years, corresponding to the duration of the study. The uniformity of a set of phantoms was validated, and the instruments were compared across all operators and sites. These studies are essential for determining the origins of DOSI performance variations and understanding how experimental methods such as DOSI compare to other radiologic approaches.

During a typical 1-h DOSI imaging session, three sets of three phantom measurements are acquired: before scanning the tumor breast, between scanning the tumor and normal breasts, and after scanning the normal breast. We simulated this in condition A by removing and replacing the DOSI probe on the Ph-A phantom nine times over a 1-h period, test A1. DOSI instrument stability over a typical patient measurement time frame was also evaluated by measuring a Ph-A phantom every minute for 1 h using one instrument and one operator, without removing the probe from the phantom, test A2. During these two different hour-long tests, the NSDs recovered for broadband absorption and reduced scattering were found to be lower than 1.4% and 1.0%, respectively. The DOSI instrument is thus highly stable over the duration of a typical patient measurement session, and removal/replacement of the probe on a test object has a negligible impact on optical property measurements of well-defined phantoms. This suggests highly reproducible probe placement on phantoms, good phantom homogeneity, and appropriate warm-up of the instrument. Similar results were found for the DOSI instruments at the other sites.[21]
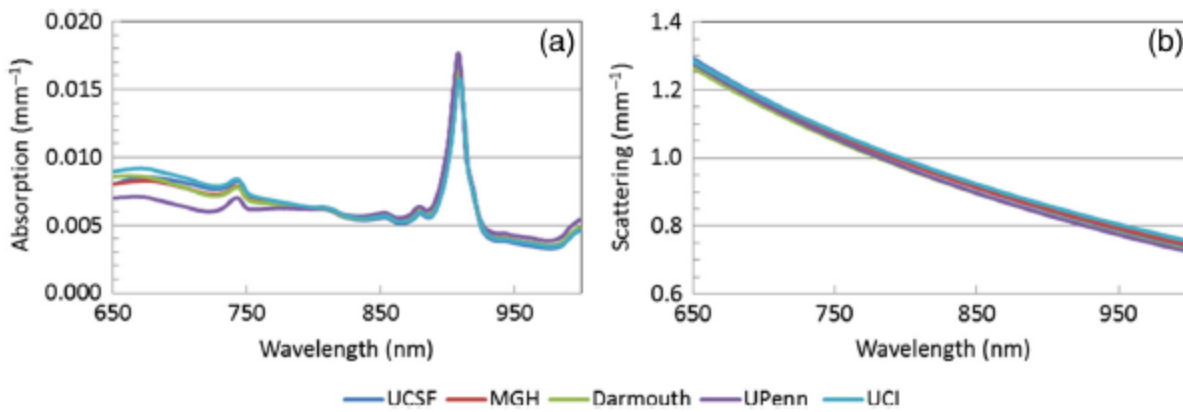

Fig. 7 (a) Broadband absorption and (b) reduced scattering spectra of the Ph-A phantoms measured at their sites, test E1.
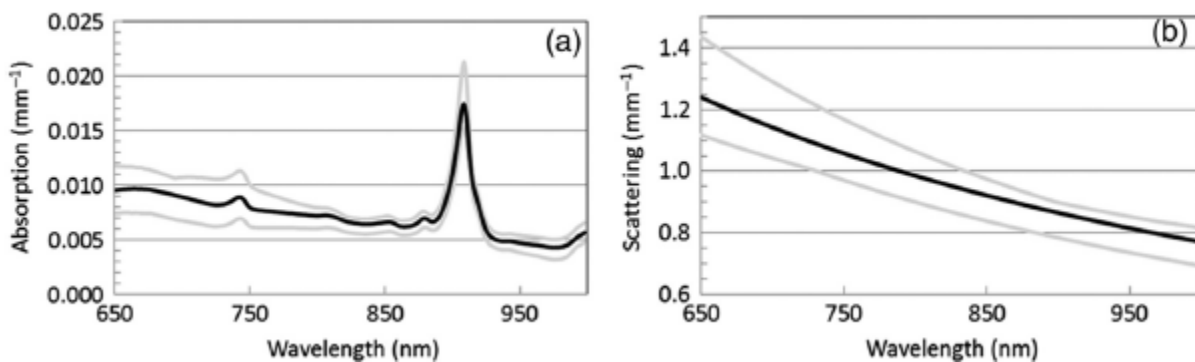

Fig. 8 Deviations in (a) broadband absorption and (b) reduced scattering of the Ph-A phantoms measured 480 times over 2 years across the five sites by 18 different operators, test E2. The black line shows the average spectra. The gray lines show the range (min–max) of spectra.

To assess single instrument performance over the chemotherapy treatment time frame with one operator and one phantom, DOSI instrument stability was tested by measuring a Ph-A phantom four times over 4 months, test A3. Nine measurements were acquired at each time-point for a total of 36 phantom measurements. Over these 4 months, the deviations in broadband absorption and reduced scattering were lower than 0.0005 mm^{-1} (NSD = 8.3%) and 0.03 mm^{-1} (NSD = 2.1%), respectively. There was no increasing or decreasing trend observed in these deviations over time. This suggests that the variations observed between these measurements are unlikely due to a

time-related drift of the phantom optical properties or of the instrument performance. In fact, as shown in the inset in Fig. 3(a), day 1 is the measurement that contributes most to the deviation. The origin of this disagreement in the optical properties measured at day 1 is unknown. It is possible that this is related to the warm-up of the DOSI instrument (typically about 15 min). Ambient conditions, such as room temperature and their impact on instrument noise, were likely responsible for the rest of the variations at days 16, 55, and 126.

We then explored the performance of the DOSI platform involving multiple operators, one phantom, and one instrument (condition B). DOSI instrument stability was tested over the study time frame by comparing the recovered optical properties of the calibration measurement on a Ph-A phantom obtained at all baseline patient measurements performed at UCI with four different operators. Over a period of 2 years and including 150 phantom scans, the deviations in broadband absorption and reduced scattering were found to be lower than 0.0007 mm^{-1} (NSD = 8.9%) and 0.03 mm^{-1} (NSD = 2.5%), respectively. These deviations are similar to those observed over 4 months with a single operator, suggesting that DOSI measurements are operator independent when performed on stable tissue phantoms. In addition to the causes of variation previously mentioned for the other tests, the shelf life of the phantom may contribute to these errors.

In condition C, we tested the performance of the DOSI platform involving one operator, multiple phantoms, and one instrument. The uniformity of the series of phantoms (Ph-A) was assessed at UCI by comparing the average of three measurements on each phantom. The phantom series is expected to have the same optical properties as it was fabricated from the same mixture. We obtained averaged deviations in recovered broadband absorption and reduced scattering under 0.0002 and 0.01 mm^{-1}, corresponding to NSD of 2.1% and 0.9%, respectively. Compared to the repeatability measurements done using only a single phantom (test A1), similar deviations were observed in condition C with respect to reduced scattering, and slightly greater deviations were observed in the absorption (NSD of 2.1% and 1.4%, for test C1 and A1, respectively). The silicon material used for fabricating phantoms is viscous before curing, which could potentially lead to some material inconsistency, resulting in some spatial heterogeneity among phantoms from the same batch. However, the low deviations observed within the Ph-A phantoms demonstrate uniformity between these phantoms fabricated from the same batch.

We then explored the performance of the DOSI platform involving one operator, one phantom, and all instruments (condition D). While the primary aim of ACRIN-6691 involves measuring relative changes in DOSI values during NAC, the secondary aim involves evaluating absolute values at specific time-points. It is thus important in the context of a multicenter study to be able to confidently compare quantitative tissue measurements across all subjects and all sites. The average optical properties recovered between the different instruments were compared on a single phantom measured three times at all sites by a unique operator over a period of 1 week. The observed deviations between sites on this single phantom were ranged from 0.0004 to 0.0010 mm^{-1} (NSD ranging from 0.6% to 11.1%) for the broadband absorption and 0.01 to 0.02 mm-1 (NSD ranging from 1.2% to 1.8%) for the broadband reduced scattering. These results suggest that all DOSI instruments are comparable in terms of their ability to recover optical properties. We then repeated this test by involving multiple operators, all phantoms, and all instruments (condition E). The interinstrument variability was assessed on the Ph-A phantoms at each site and measured by the on-site operator. The deviations between sites using their respective Ph-A phantoms were similar to the deviations obtained in condition D using a single phantom and operator: below 0.0010 mm^{-1} (NSD = 10.8%) for the broadband absorption and 0.02 mm^{-1} (NSD = 1.8%) for the broadband reduced scattering. These results suggest that all DOSI instruments are comparable in terms of optical property recovery even when different operators and phantoms are involved.

Last, we determined the performance of the DOSI platform involving 18 operators, all phantoms, and all instruments over the 2-year study duration (condition E). This test consisted of 480 measurements of any of the Ph-A phantoms across 58 different days over 2 years. The resulting deviations were below 0.0010 mm^{-1} (NSD = 10.3%) for the broadband absorption and 0.06 mm^{-1} (NSD = 4.7%) for the broadband reduced scattering. This represents the worst-case scenario in terms of uncertainties as many operators, all phantoms, and all instruments were involved over a 2-year time frame.

In conditions D and E, the greatest variation between instruments occurred below 760 nm and was particularly noticeable in the Upenn data. Later experiments suggested that this was due to a greater-than-expected variation in the spectrometer response in this spectral region caused by the reflectance calibration approach. A new spectrometer and calibration method has already been implemented in the next generation of instruments to mitigate this variation. We decided to keep the current spectrometers in the five instruments in use until the end of the study for consistency with already acquired subject data.

Although instrument performance was evaluated in the context of DOSI technology in ACRIN-6691, this general framework can be adapted to other experimental imaging technologies. As shown in Fig. 2, the first step is to identify the origin of uncertainties, which can vary depending on the instrumentation used. These sources of variation must then be isolated and assessed separately in the context of the study procedures. After the single sources of variation have been assessed, their combined effect can be addressed. Though the details of the tests will vary depending on the imaging technique and the study protocol, the proposed framework can be broadly used to characterize instrument performance in translational optical imaging studies as has been extensively reported in the positron emission tomography community.[23,24]

In conclusion, we have presented a methodology to assess performance uncertainties in an experimental imaging technology that was used in quantitative longitudinal measurements of patients over a 2-year period. The impact of multiple phantoms and operators, and intra- and interinstrument variability has been addressed separately and jointly. Using this approach, we have determined that despite 480 DOSI measurements acquired by 18 operators at five research sites over 2 years (58 measurement days), the resulting deviations were on average 0.0005 mm^{-1} (NSD = 7.0%) for the broadband absorption and 0.03 mm^{-1} (NSD = 3.1%) for the broadband reduced scattering. These data can be used to identify and minimize sources of error and potentially improve DOSI sensitivity to small changes in tumor physiology. In addition, the methodology we describe can be readily adapted and applied to other experimental imaging technologies undergoing evaluation in multicenter settings. However, additional analysis is required to assess the precise relationship between these sources of variation and relevant clinical decision-making thresholds.

References

1. T. D. O'Sullivan et al., "Diffuse optical imaging using spatially and temporally modulated light," *J. Biomed. Opt.* 17(7), 071311 (2012).
2. B. J. Tromberg and A. E. Cerussi, "Imaging breast cancer chemotherapy response with light," *Clin. Cancer Res.* 16(9), 2486–2488 (2010).
3. D. R. Busch et al., "Optical malignancy parameters for monitoring progression of breast cancer neoadjuvant chemotherapy," *Biomed. Opt. Express* 4(1), 105–121 (2013).
4. S. Jiang et al., "Predicting breast tumor response to neoadjuvant chemotherapy with diffuse optical spectroscopic tomography prior to treatment," *Clin. Cancer Res.* 20(23), 6006–6015 (2014).
5. B. J. Tromberg et al., "Predicting responses to neoadjuvant chemotherapy in breast cancer: ACRIN 6691 trial of diffuse optical spectroscopic imaging (DOSI)," *Cancer Res.* 76(20), 5933–5944 (2016).
6. A. Y. Sajjadi et al., "Normalization of compression-induced hemo-dynamics in patients responding to

neoadjuvant chemotherapy monitored by dynamic tomographic optical breast imaging (DTOBI)," *Biomed. Opt. Express* 8(2), 555–569 (2017).

7. B. E. Schaafsma et al., "Optical mammography using diffuse optical spectroscopy for monitoring tumor response to neoadjuvant chemotherapy in women with locally advanced breast cancer," *Clin. Cancer Res.* 21(3), 577–584 (2015).

8. T. D. O'Sullivan et al., "Abstract OT3-2-07: predicting hormonal therapy response in breast cancer using diffuse optical spectroscopic imaging (DOSI): ongoing clinical study," in *San Antonio Breast Cancer Symp.* (2015).

9. T. D. O'Sullivan et al., "Optical imaging correlates with magnetic resonance imaging breast density and reveals composition changes during neoadjuvant chemotherapy," *Breast Cancer Res.* 15(1), R14 (2013).

10. K. M. Blackmore et al., "Estimation of mammographic density on an interval scale by transillumination breast spectroscopy," *J. Biomed. Opt.* 13(6), 064030 (2008).

11. P. Taroni et al., "Breast density assessment by means of time domain optical mammography at 635–1060 nm," *Proc. SPIE* 8088, 80881E (2011).

12. A. Leproux et al., "Assessing tumor contrast in radiographically dense breast tissue using diffuse optical spectroscopic imaging (DOSI)," *Breast Cancer Res.* 15(5), R89 (2013).

13. R. Choe et al., "Differentiation of benign and malignant breast tumors by in-vivo three-dimensional parallel-plate diffuse optical tomography," *J. Biomed. Opt.* 14, 024020 (2009).

14. G. Quarto et al., "Estimate of tissue composition in malignant and benign breast lesions by time-domain optical mammography," *Biomed. Opt. Express* 5(10), 3684–3698 (2014).

15. A. Leproux et al., "Differential diagnosis of breast masses in South Korean premenopausal women using diffuse optical spectroscopic imaging," *J. Biomed. Opt.* 21(7), 074001 (2016).

16. A. Cerussi et al., "Quality control and assurance for validation of DOS/I measurements," *Proc. SPIE* 7567, 756703 (2010).

17. A. Durkin et al., "Diffuse optical spectroscopy & imaging," (16 July 2015).

18. F. Bevilacqua et al., "Broadband absorption spectroscopy in turbid media by combined frequency-domain and steady-state methods," *Appl. Opt.* 39, 6498–6507 (2000).

19. T. H. Pham et al., "Broad bandwidth frequency domain instrument for quantitative tissue optical spectroscopy," *Rev. Sci. Instrum.* 71(6), 2500–2513 (2000).

20. A. Cerussi et al., "In vivo absorption, scattering, and physiologic properties of 58 malignant breast tumors determined by broadband diffuse optical spectroscopy," *J. Biomed. Opt.* 11(4), 044005 (2006).

21. A. E. Cerussi et al., "Tissue phantoms in multicenter clinical trials for diffuse optical technologies," *Biomed. Opt. Express* 3(5), 966–971 (2012).

22. B. J. Tromberg et al., "Non-invasive measurements of breast tissue optical properties using frequency-domain photon migration," *Philos. Trans. R. Soc. London B* 352(1354), 661–668 (1997).

23. J. S. Scheuermann et al., "Qualification of PET scanners for use in multicenter cancer clinical trials: the American College of Radiology Imaging Network Experience," *J. Nucl. Med.* 50(7), 1187–1193 (2009).

24. D. W. Byrd et al., "Evaluation of cross-calibrated ^{68}Ge/^{68}Ge phantoms for assessing PET/CT measurement bias in oncology imaging for single- and multicenter trials," *Tomography* 2(4), 353–360 (2016).

Anaïs Leproux received her PhD in biomedical engineering from the University of Amsterdam, the Netherlands, in 2012. She is a postdoctoral scientist at the Beckman Laser Institute (BLI), the University of California, Irvine. Her BLI research in biomedical photonics is focused on using diffuse optical spectroscopic imaging technology in breast cancer applications, mainly for the detection and characterization of breast tumor response to therapy. She is a member of SPIE.

Thomas D. O'Sullivan is an assistant professor in the Department of Electrical Engineering at the University of Notre Dame. He received his PhD in electrical engineering from Stanford University in 2011, and then was a US Department of Defense Breast Cancer Research Program postdoctoral fellow at the University of California, Irvine Beckman Laser Institute (BLI). While at BLI, he also served as director of the Diffuse Optical Spectroscopy and Imaging Laboratory.

Arjun G. Yodh is professor of physics and astronomy and the James M. Skinner professor of science at the

University of Pennsylvania. He directs Penn's Laboratory for Research on the structure of matter, including its NSFMRSEC. His biomedical optics laboratory studies light transport, image reconstruction, and develops optical technology for clinical/preclinical applications. Current research spans functional imaging/monitoring of brain and breast, and monitoring of hemodynamics and metabolism during cancer therapy and in patients with vascular disease.

Stefan A. Carp received his BS degrees in chemistry and chemical engineering from MIT and a PhD from the University of California, Irvine, in biomedical optics. He is an assistant professor of radiology at Harvard Medical School and a member of Massachusetts General Hospital Martinos Center Optics Division. He focuses on the development of novel techniques for tissue hemodynamics and oxygen metabolism monitoring, as well as their translation to clinical use.

David Boas is the director of the Martinos Optics Division in the Department of Radiology at Massachusetts General Hospital, and a professor of radiology at Harvard Medical School. He received his BS degree in physics at RPI and his PhD degree in physics at the University of Pennsylvania. He is the founding president of the Society for Functional Near-Infrared Spectroscopy, founding editor-in-chief of Neurophotonics, and the recipient of the 2016 Britton Chance Award in biomedical optics.

Shudong Jiang received her PhD in opto-electronics, Tokyo Institute of Technology, Japan, in 1992. She was a research scientist in Tokyo Institute of Technology and Japan Science and Technology Corporation from 1992 to 1998. Since 1999, she has been involved in medical imaging research at Dartmouth College. Currently, she is working as a research associate professor at Dartmouth in the field of *in-vivo* spectroscopy and near-infrared tomography imaging of tissue.

Keith D. Paulsen is currently the Robert A. Pritzker professor of biomedical, engineering at Thayer School of Engineering at Dartmouth, professor of radiology at the Geisel School of Medicine, and director of the Advanced Imaging Center at Dartmouth Hitchcock Medical Center. His research has focused on the development and translation of advanced imaging technology, primarily for cancer detection, diagnosis, therapy monitoring, and surgical guidance. He has authored more than 350 archival publications with an active research program, continuously funded by the National Institutes of Health for 25 years.

Brian Pogue is professor of engineering, physics, and surgery at Dartmouth College, having his PhD from McMaster University Canada. He works on novel imaging systems for surgery and radiation therapy guidance. He has published >300 peer-reviewed and >400 conference papers, on work funded by NCI, NIBIB, and DoD. He is on editorial boards for physics in Medicine and Biology, Medical Physics, Journal of Biomedical Optics, and Breast Cancer Research and is an elected fellow of the OSA and AIMBE.

Darren Roblyer is an assistant professor in the Department of Biomedical Engineering at Boston University. After receiving his BS degree in biomedical engineering from Johns Hopkins University in 2004, and his PhD as an HHMI Med-into-Grad Fellow in 2009 from the Bioengineering Department at Rice University, where he studied under Rebecca Richards-Kortum. Prior to starting his faculty position in 2012, he was a postdoctoral fellow at Beckman Laser Institute studying under professor Bruce J. Tromberg.

Bruce J. Tromberg is a professor of biomedical engineering and surgery at the University of California, Irvine (UCI) and director of UCI's Beckman Laser Institute and Medical Clinic. His research interests are in the development of quantitative, broadband biophotonics technologies for characterizing and imaging tissue structure, function and composition across spatial scales.

Biographies for the other authors are not available.