Guidance for Performance Evaluation for Fluorescence Guided Surgery Systems

Brian W. Pogue[1,2], Timothy C. Zhu[3], Vasilis Ntziachristos[4], Brian C. Wilson[5], Keith D. Paulsen[2], Sylvain Gioux[6], Robert Nordstrom[7], T. Joshua Pfefer[8], Bruce J. Tromberg[9], Heidrun Wabnitz[10], Arjun Yodh[11], Yu Chen[12], Maritoni Litorja[13]

[1]University of Wisconsin-Madison, Madison WI USA

[2]Dartmouth College, Hanover NH USA

[3]University of Pennsylvania, Perelman School of Medicine, Philadelphia PA, USA

[4]Technical University of Munich, Helmholtz Zentrum Munich, Munich, Germany

[5]University of Toronto, University Health Network, Toronto ON, Canada

[6]University of Strasbourg, Department of Biomedical Engineering, Strasbourg, France

[7]National Cancer Institute, National Institutes of Health, Bethesda MD, USA

[8]US Food and Drug Administration, Center for Devices and Radiological Health, Silver Spring MD, USA

[9]National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda MD, USA

[10]Physikalisch-Technische Bundesanstalt, Berlin, Germany

[11]University of Pennsylvania, Department of Physics, Philadelphia PA, USA

[12]University of Maryland Baltimore County, Department of Biomedical Engineering, Baltimore MD, USA

[13]National Institute of Standards and Technology, Gaithersburg, MD USA

Abstract

The last decade has seen a large growth in fluorescence guided surgery (FGS) imaging and interventions. With the increasing number of clinical specialties implementing FGS, the range of systems with radically different physical designs, image processing approaches and performance requirements is expanding. This variety of systems makes it nearly impossible to specify uniform performance goals, yet at the same time, utilization of different devices in clinical trials indicates some need for common knowledge bases and a quality assessment paradigm to ensure that effective translation occurs. It is feasible to identify key fundamental image quality characteristics and corresponding objective test methods that should be determined such that there are consistent conventions across a variety of FGS devices. This report outlines test methods, tissue simulating phantoms and suggested guidelines, as well as personnel needs and professional knowledge bases that can be established. This report frames the issues with guidance and feedback from related societies and agencies having vested interest in the outcome, coming from an independent scientific group formed from academics and international federal agencies for the establishment of these professional guidelines.

Table of Contents

1. Introduction

1.1 General introduction on rationale for the report

1.2 Scope of the report

2. Background

2.1 Theory of light-tissue interaction as it affects images

2.2 System use and performance expectations

2.3 Phantoms to simulate human tissue

3. Protocol

3.1 Methods for performance testing

3.2 Performance testing paradigms

3.3 Considerations for clinical implementation of QA procedures

4. Recommendations for technical evaluation and guidance of new systems

4.1 Recommendations for technical evaluation

4.2 Recommendations for technical guidance for new systems

4.3 Recommendations for “qualified personnel” for using the FGS system

4.4 Suggestion for in service requirement for a staff to be qualified for using the FGS system.

5. Limitations of this report

6. Summary

7. Acknowledgements

8. Conflicts of Interest

9. References

1. Introduction

1.1 GENERAL INTRODUCTION ON RATIONALE FOR THE REPORT

The technological and logistical implementation of Fluorescence Guided Surgery (FGS) has evolved over several decades, in both investigational research and approved clinical indications [1]. However, just in the last half decade, increased clinical use has occurred with a spike in numbers of 510(k) pathway FDA cleared imaging systems for indocyanine green (ICG) [2-7]. With this growth in the industry, the range of research compounds being tested in humans has also expanded. Taken together, the increased range of systems and fluorescent reporters makes for a complex and evolving set of performance choices available for surgical work and surgical clinical trials. This report focuses on key performance issues that should be considered and quantified to facilitate scientific and medical decisions about trial design and system use for FGS hardware/software. The focus here is on macroscopic imaging systems used in surgical applications where the field of view was intentionally designed to allow scanning of surgical fields in a non-contact manner. The following paragraphs outline the rationale for addressing the system performance analysis of FGS systems.

Systems targeted for each surgical sub-specialty are different in features and their intended use and so it is implausible to establish universal standards that are highly specific. Even systems approved for the same indication are usually in competition with each other, and achieve this through design differentiation, cost reduction, strategic compatibilities or uniquely marketable performance metrics. Technological choices such as excitation/emission wavelengths, background filtering, illumination and image formation optics, each differentiate system specifications and performance. ICG emission occurs at near-infrared wavelengths with a peak in the range of 800 nm, which is not visible by the surgeon, so these systems are inherently tied to display-based guidance which augments traditional white light or x-ray views of the tissue [8-11]. The visual presentation of these images represents a developing paradigm in real-time diagnostics, so the performance metrics could inherently involve not only the hardware components but also the software processing and the real time display methodology [12]. These aspects are all critical parts of the integrated performance guidance and would benefit from standardization to enable consistent evaluation and quality control of new and existing systems.

Currently, most marketed clinical devices for FGS are designed and cleared for use with ICG [11, 13] to enable blood flow and tissue perfusion imaging applications. Other human use agents include fluorescein, methylene blue (MB), and aminolevulinic acid (ALA) to induce protoporphyrin IX (PpIX) in tissue, each of which have very different absorption and fluorescence spectra, and hence wavelength choices. Additionally, the development of new fluorescent probes to provide molecular information[14-

20] is a very active area of translational research. Fluorescence molecular imaging is being studied in investigator-initiated human trials to (a) track metabolism through PpIX production or protease activity, and (b) image immunologic targeting by antibodies and peptides[21]. It is common for trials with new agents to use FDA-cleared imaging systems, because of their commercial availability, ease of approvals with institutional review boards, and known safety profiles. However, sometimes custom-made systems are used in single-center or research-based studies. The growing divergence of device hardware and fluorescent molecular reporters has set up a complex landscape, with little authoritative guidance from professional societies involved in this field, and no clear consensus on how to evaluate system performance and effectiveness.

The responsibility of training users typically rests with the manufacturer, yet in the current direct-to-surgeon market, technically trained support staff are not commonly involved. The Medical Physics or Biomedical Engineering communities can help fill this gap, especially as devices become more complex and risks of misuse grow, such as using a new the fluorescent agent with a non-ideal device or carrying out multi-center clinical trials with systems that are not comparable in performance. Just as the sheer variety of systems makes it difficult to specify exact performance requirements, this situation will also require a variety of approaches to define expert users and their methods for system performance evaluation and calibration. However, the goal of this initiative was to identify key scientific image quality characteristics and corresponding objective test methods that could be important, as is reasonable for the specified use cases, and to point towards those individuals who are optimally situated for this work.

1.2 SCOPE OF THE REPORT

This report was produced and charged with addressing three issues related to the clinical implementation of FGS systems, including: 1) Provide recommendations on how to select FGS systems for clinical use and how to use them clinically; identify specific requirements and performance goals necessary for their clinical implementation; 2) Provide recommendations on how to calibrate these systems and other appropriate aids, such as targets and phantoms that test technical functionality in planned use; and 3) Provide recommendations on risk-based approaches to quality management for Fluorescence guided surgery systems. This report covers all three, although a bit more on the latter two points, as it was determined that the clinical use (point 1) was a bit outside of the scope of the technical working group. Details of the specific requirements, performance goals, calibration, targets & phantoms, and risk-based management needs are each outlined in the sections below.

Part of the process of this work was to frame the issues with guidance and feedback from related societies and agencies with vested interest. This independent committee of scientists regularly work on fluorescence in clinical trials or have been involved in optical device clinical trials and/or regulatory evaluation. Interaction has included discussion with members of the Optical Navigation Workgroup of the World Molecular Imaging Society (WMIS), and several groups that meet regularly at the International Society for Optics and Photonics (SPIE) Biomedical Optics (BiOS) conference. These groups focus on the range of needs for clinical trials and reporter agents specifically, as well as aspects of system performance. There has been iterative feedback from participants at the meetings while dissemination of ideas has been achieved through presentations at these venues, and the meetings provided a cost-effective and time-efficient way for the members to geographically meet[22]. In addition to the majority participation by academic investigators involved in research on fluorescence guided surgery, there has been participation of scientific staff from the US Food and Drug Administration, the NIH National Cancer Institute, the US National Institute of Standards and Technology and the German counterpart, Physikalisch-Technische Bundesanstalt (PTB). This has been a part of the planning to ensure that the correct balance of information and guidance is reached. Additionally, outreach to industry has occurred through public forums via presentation, such as at SPIE BiOS and WMIS meetings.

TABLE 1. Symbols used in this report.

Symbol	Name – (Conventional Units)
Φ	Radiant energy fluence rate - (W/m^2)
H	Radiant exposure - (J/m^2)
μ_a	Absorption coefficient – (mm^{-1})
μ_s	Scattering coefficient – (mm^{-1})
μ'_s	Reduced or Transport scattering coefficient - (mm^{-1})
g	Average cosine of the scattering angle – (unitless)
λ	Wavelength, (nm)
A	Area, (m^2)
E_x	Irradiance excitation light - (W/m^2)
E_m	Irradiance emission light - (W/m^2)

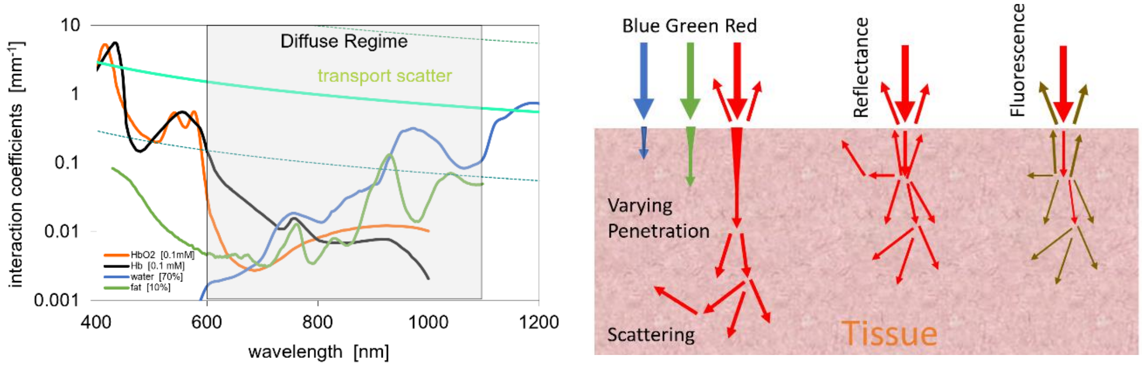
2. Background

2.1 THEORY OF LIGHT-TISSUE INTERACTION AS IT AFFECTS IMAGING

2.1.1. Light transport in tissue

Perhaps the most important factor in understanding the unusual needs for optical system performance is that light interaction with tissue is complex, affected by both the tissue surfaces and the interior tissue optical properties.[23] The primary light-tissue interactions present inside tissue are elastic scattering and absorption, each of which can be characterized by macroscopic interaction coefficients: $\mu_s(\lambda)$ is the probability per unit distance of an elastic scattering event, and $\mu_a(\lambda)$ is the probability per unit length of absorption, each at wavelength λ. There can be a strong spectral dependence to these parameters, as illustrated in Figure 1, and there is potential for re-emission of light by fluorescence or phosphorescence from specific molecules within the tissue. To make this even more complex, in the near field of a scattering event – typically hundreds of microns – light propagation is highly anisotropic with the average cosine of the scattering angle, g, typically being higher than 0.7 and often higher than 0.9, depending upon the tissue and wavelengths. Thus, light entering and exiting tissue can have spatial patterns that are highly directional and the intensity can vary by orders of magnitude across millimeters in depth. This exponential attenuation of light in tissue makes the measured or observed light exiting tissue very surface weighted in FGS, and the interaction of absorption and scattering can distort the remitted colors from white light illumination or alter fluorescence signals from deeper layers of tissue.

Figure 1. The spectral components of major chromophores and scatterers present in soft tissue (a) with the



scattering range of biological values shown (dotted lines). An illustration of how the depth of penetration varies with wavelength (b) as well as reflectance and fluorescence light propagation at the photon level.

Measurements that span source to detection distances, d_{SD}, greater than a few millimeters, or those in the longer wavelengths beyond 600 nm can appear fully diffuse, with a transport or reduced scattering coefficient, $\mu_s'(\lambda)$, that describes the level of scattering magnitude under the assumption that each event was isotropic [24]. This assumption provides for simpler diffusion theory modeling of the interactions but must be interpreted with the limitations inherent in applying the diffusion approximation to this situation. The key condition of validity for the diffusion approximation is that the reduced scattering coefficient is much larger than the absorption coefficient (i.e. $\mu_s'(\lambda) >> \mu_a(\lambda)$), and that the source to detection distance is larger than the average distance between scatterers (i.e. $d_{SD} >> 1/\mu_s'(\lambda)$)[25]. Diffusion modeling of large area reflectance is often used as an approximation to interpret the light signals, although more precision is achieved with discrete particle stochastic simulations such as Monte Carlo modeling. [26]

The wavelength dependence of these parameters is a function of the concentration of individual tissue constituents[23, 27-30]. The major chromophores observed in tissue are hemoglobin and oxy-hemoglobin present in all red blood cells, as well as melanin in the upper layer of the skin. In addition to this, in the NIR wavelengths, water, lipids and collagen all have absorbing features as well, and in the blue/UV ranges, water, hemoglobin and other proteins are the major absorbing features. The importance of these issues is significant to this report due to their impact on the performance of FGS systems. Differences in wavelength, optical design, or filtering can all alter the detected signal in ways that are affected by the tissue optical properties. Additionally, some systems are designed for optimal performance in the face of the type of optical properties present in specific organs.

2.1.2. Optical penetration depth, absorption and fluorescence image information

Each of these tissue factors affects the depth into tissue that light signals sample in an FGS system, as illustrated in Figure 2(a), where the wavelengths of light have different attenuation levels, with red and near infrared wavelengths having the most penetration and UV/blue wavelengths having the least[23]. The magnitude of the attenuation and the resulting depth of sampling depends upon the wavelengths of light used, the design features of the system such as the geometry of the light source and imaging sensor relative to the tissue surface[31]. An example of how fluorescence imaging results may be non-intuitive in a tumor is illustrated [32] in Fig 2(c), where the signal is observed to decrease even though there are greater fluorophore levels in the tumor than the surrounding normal tissue, which also agrees with the reflected light image, in Fig 2(b). This effect is most severe at blue or green wavelengths, where light absorption by hemoglobin is very strong (two orders of magnitude greater than in the NIR). In such cases, the impact of increased blood volume due to angiogenesis may dominate over simultaneous increases in fluorophore concentration due to probe binding. Normalization can remove some of this effect in red/near-infrared wavelengths, Fig 2(d) [32]. This is one example of the complex interplay between fluorescence, absorption and scattering of tissue, as well as the geometry of the optical measurement and other design considerations such as data processing algorithms. This issue is especially relevant in oncology malignancy, which commonly have increased capillaries and hence higher blood volume in lesions. The result is that fluorescence measured is not always a linear reporter of the contributions of fluorophore concentration. Because of this well-studied effect, reflectance has been shown as a surrogate measure of the light penetration or remittance intensity and is sometimes used to normalize or process the fluorescence signal for variations in absorption or scattering. While doing this type of correction requires knowledge of the full absorption and scattering coefficients and anisotropy patterns detected by the system, in practice often empirical ratios or weighted ratios are used to provide a more heuristic or empirical correction for light interaction with the tissue[33-35].

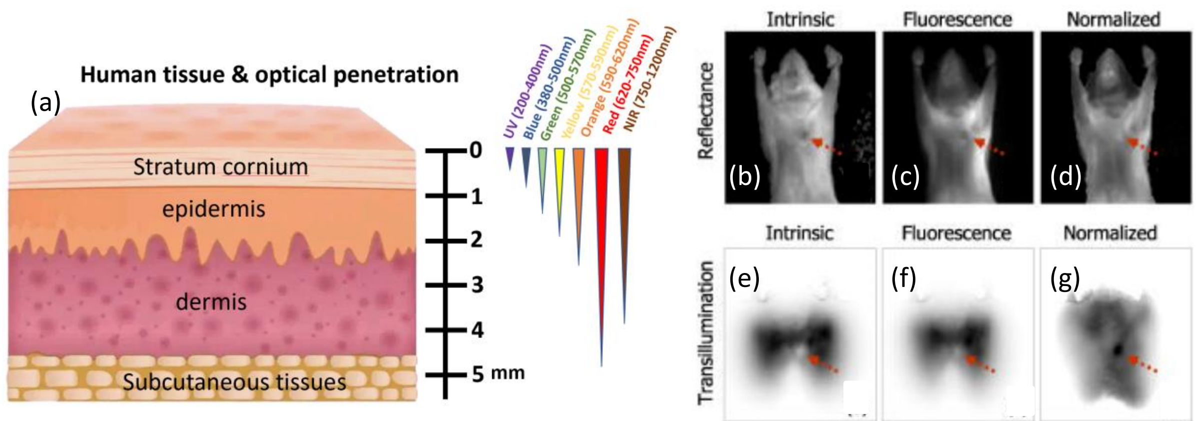


Figure 2. The attenuation of light in tissue is exponential with depth, and varies considerably with wavelength (a), with blue/green being much more highly attenuated than red and near-infrared. A visual example of the effect that absorption can have on the detection of fluorescence in epi-illumination or reflectance mode, is shown where the reflectance image of a tumor (arrow) in (b) with the fluorescence image (c), and the normalized fluorescence to reflectance image (d) showing the contrast of the tumor shifts from negative to positive (white is more signal, while black is less signal in these images). In (e)-(g) the native data from transillumination geometry are shown. [32]

2.2 SYSTEM USE AND PERFORMANCE SPECIFICATIONS

2.2.1 Fluorophores currently approved and under development

Fluorophores used in current clinical practice are relatively few[11, 36, 37]. While there are some endogenously present in tissue such as collagen, NADH, FAD, and porphyrins, exogenously administered agents include largely only ICG and fluorescein[38]. Others such as methylene blue, isosulfan blue, and proflavine are used but in research trials of fluorescence[39]. Additionally, the fluorophore precursor aminolevulinic acid is now commonly used in neurosurgery [40] and bladder imaging [41], as it induces production of PpIX and a collection of associated porphyrins[42]. Approved photodynamic therapeutic agents also happen to fluoresce and are used in locally approved institutional trials, with a large range of approved porphyrins, phthalocyanines and chlorins[43-45]. Perhaps most important to recognize from this issue is that each agent has different excitation and emission wavelengths that are optimal, and these choices can even vary between manufacturers[45].

2.2.2 Fluorescence basics

Fluorescence imaging consists of exciting a contrast agent with appropriate wavelengths of light and detecting the resulting emissions (at different wavelengths of light) by means of a camera and filters. The subtleties of effective system design can require much more complexity[46-50]. The most straightforward configuration is a continuous-wave (CW) system where the source intensity is constant in time. The excitation light, typically from a laser diode, a filtered white light source, or a light emitting diode, excites fluorescent molecules from the ground state to a higher energy level. In return, the molecules relax back to the ground state by means of two processes – either non-radiative vibrational transition producing mainly heat, or via radiative transition with emission of a fluorescent photon. Because of partial non-radiative relaxation processes, the energy of each emitted photon is lower than the energy of the original excitation photon, and therefore, the re-emission occurs at a longer wavelength, energy-shifted from the excitation photon by an amount called the Stokes shift. The main challenge in performing efficient fluorescence detection is therefore filtration, i.e., isolating the fluorescence emission of interest from other sources of light, in particular, the excitation light that is typically several orders of magnitude greater than the Figure 3 shows a generic system and its key parameters influencing the measured fluorescence intensity. The field-of-view is illuminated with excitation light that has been filtered to reduce wavelengths that overlap the range of the fluorescence emission. This light reaches the tissue where it gets absorbed and scattered. Fluorescent contrast agents absorb a portion of this light and re-emit the signal isotropically as fluorescent photons. This emission light is then captured by an objective lens equipped with emission filters that isolate the fluorescence photons from the excitation photons, with the resulting image captured by a camera.

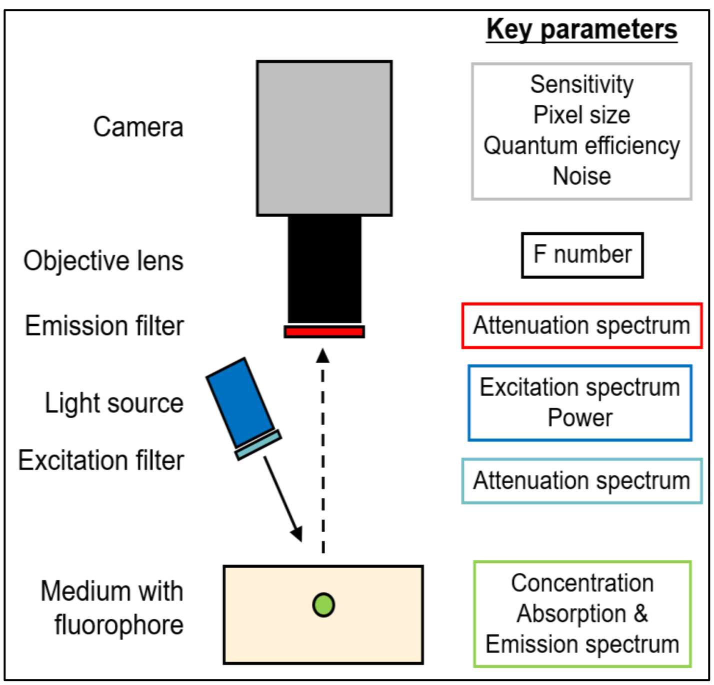


Figure 3. Generic Fluorescence Imaging System: Filtered excitation light illuminates the medium where the fluorescent contrast agent is located. Fluorescence is emitted and captured on a camera using an objective lens

equipped with an emission filter. Key parameters for each component are listed to the right, in the corresponding color-coded box.

Table 2 Components of the fluorescence signal and imaging system & factors that can affect performance.

Components affecting the image signal	Example factors that can alter the signal or image
Fluorophore	concentration, localization
Excitation Light Source	wavelengths, intensity, homogeneity, modulation
Light Filtration	wavelengths, suppression optical density, actual vs specified performance, bleaching
Background Signal(s)	excitation leakage, room light, filter performance
Biological Background Signals	autofluorescence
Biological kinetics & compartmentalization	localized by vascular, cellular, or albumin binding
Camera lens	f-number, depth of field, aperture, focus
Camera system	sensor type, intensified, noise level, frame rate
Image Pixelation & Digitization	spatial resolution, contrast resolution, dynamic range
Image Processing & Corrections	image alteration, autogain, background removal

Light source parameters: Light source technology typically employed in the fluorescence imaging system consists of either laser diodes, white light sources or LEDs (Light Emitting Diodes). Controlling the source power and spectral distribution is key to providing the right amount of light to the fluorophore. To this extent, narrower bandwidth sources, such as laser diodes and LEDs, confine the excitation power spectrally in order to match the absorption spectrum of the contrast agent. In addition, the small source size of laser diodes and LEDs are easily manipulated to provide the desired illumination characteristics. In particular, the illumination should be designed to cover the field of view and, in most cases, be as homogeneous as possible to minimize fluorescence intensity variation resulting across the imaging plane.

Temporal modulation of the light source is used by several systems to overcome limitations of CW fluorescence imaging. One method consists of pulsing the light source and performing lock-in detection of the fluorescence intensity to isolate contributions from fluorophores only, in case a steady background signal exists [51]. An advanced embodiment of this method sends very short and intense pulses of light to increase significantly the apparent fluence, and therefore, the sensitivity of the imaging system. Another very different method captures the time-dependent fluorescent signal of the dye flowing into the tissue and being cleared over time, in order to analyze the raw fluorescence intensity according to its derivative or the slope [52]. This approach has been valuable in vascular surgeries where the "time-to-peak" fluorescence can be indicative of vascular defects, or to assess tissue function such as kidney or liver [53-55]. Finally, fluorescence lifetime can be measured using short-pulsed or rapidly modulated sources to distinguish contrast agents having similar wavelengths, or to quantify environmental conditions (e.g., pH) or detect molecular binding with specialized agents [56].

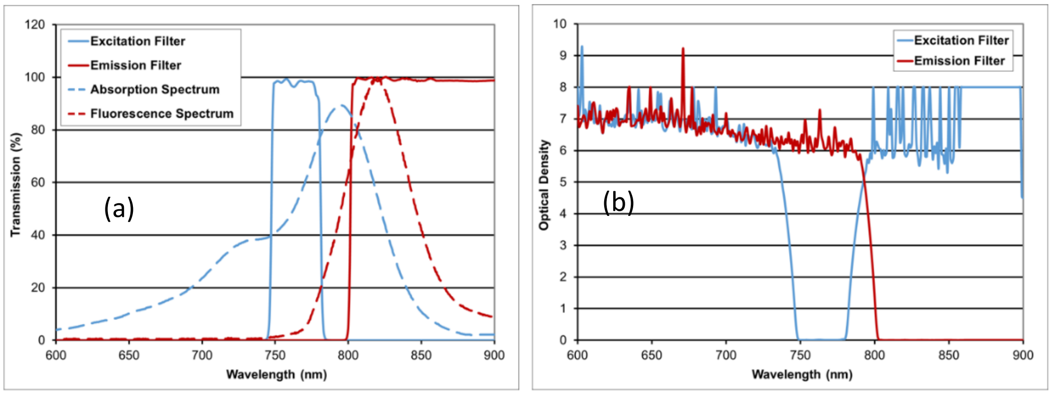


Figure 4. Typical ICG Filtration Scheme: left the transmission plots of the excitation and emission filters on top of the ICG absorption and emission spectra (dash); right the optical density plots of the excitation and emission filters (note the crossing point above OD=5).

Filtration parameters and effects: Optical filtration is one of the most critical elements in design of a fluorescence imaging system[57, 58]. The number of detectable fluorescent photons from a sample is considerably smaller than the large amount of excitation photons reflected or scattered back to the detector. This difference can be several orders of magnitude, so proper elimination of the returned excitation signal is needed to isolate the desired fluorescent signal. The optical isolation of these fluorescence photons is central to system performance. Thus, excitation and emission filters should be analyzed in terms of both transmittance and optical density (OD). As depicted in Figure 4, these quantities can be seen to differ, and close attention should be paid to both. The transmission plot, presented in the context of ICG detection, indicates locations for the passing bands of each filter, on the light source and camera sides. One can appreciate in this example the confinement of the excitation light (745-780nm) that is necessary due to the low, but significant, amount of excitation photons in the fluorescence emission side, with a full range which is most fully appreciated on the OD plot (b). Beyond 800nm, the excitation is cut down by 6 orders of magnitude and there is high transmission for the ICG fluorescence. Both filters have near 100% transmission in the wavelengths where they are designed to pass, but the edges which look sharp in the linear transmission graph (a) are somewhat less sharp when viewed on the logarithmic OD graph (b). The separation between these two filters is essential to the performance of the system. As a rule of thumb, the two filters should have their longest wavelength crossing point above a vertical blocking value of OD=5 relative to each other's transmission values, in to offer satisfactory rejection of excitation photons.

Importantly, when considering filtration, the design is highly dependent upon the objective lens and the F-number of the objective. The F-number controls the solid angle of the collected photons, and therefore, has a large influence on the photon angles that will pass through the emission filter. Two different filter technologies exist, one based on absorption and the other based on interferences. Because the interference filter characteristics are strongly angle-dependent, the F-number should be relatively high compared to absorption filters that are not angle-dependent. However, because of their impressive characteristics (high transmission, high OD, fast spectral response), interference filters offer better performance. Thus, the combination of the filter technology with the objective lens represents a compromise and certainly a significant challenge in the design of a fluorescence imaging system to ensure high quality performance, without unintended leakage or performance loss from high-angle light signals.

Background Signals: One very important example where non-linearity can enter into these types of systems is the contaminating effect that background signals can have on the measured intensity and images. Background can come from several causes in imaging systems, including:

a. Excitation light leaking through the emission filter(s).
b. Room light leakage into the emission band of the system.
c. Sub-optimal filter performance

Each of these are described briefly below.

The first listed source of background signals is from excitation photons, and this is because the number of excitation photons is orders of magnitude higher, fluorescent photons should be very well isolated through proper filtration, as already mentioned. Typically, the excitation light source should be filtered since a fraction of photons from the source can still be detected through the emission filter. While contamination occurs with a small fraction of excitation photons, they can be of comparable intensity to the fluorescence signal. Not filtering the source will result in unnecessary amounts of excitation background. A good analogy is in fluorescence microscopy where there are three layers of filtering in most systems; where the source is filtered to reduce emission band signals, a dichroic filter is placed between the excitation and emission paths, and then the emission band is filtered again as a third stage to further remove excitation light.

The second major cause of background signals arises from ambient room white light sources, either provided locally by the imaging system or globally in the room. These sources are typically not filtered and can contain significant amounts of light in the same wavelength range as the fluorescence emissions. If this light has avoided the excitation path filter, which is commonly the case in open area imaging, then it will pass through the emission filter. NIR emission imaging system often try to passively use the 800nm band, where room light levels are significantly suppressed, and the use of LED lighting instead of incandescent lighting further lowers this background light. This issue is much less relevant for endoscopic, laparoscopic or intra-cavitary imaging systems where the presence of unwanted light inside the body is much lower.

The third listed cause of background signal is from filtering that may not be ideally designed for an application, given that excitation light can be many orders of magnitude higher than the emission light intensity. In particular when using interference filters, special attention should be paid to the specified range for transmission and OD properties since the filter may be designed for limited filtering capacity. As a result, small amounts of light above 900nm, for instance, can be captured within an emission filter that has not been designed to reject light outside of its working range. Similarly, some systems are designed to allow in some room or excitation light as a visual aid to the user, and this can significantly affect the limit of fluorescence detection.

Biological background signals: The causes of background can also be biological in nature and not optical system issues, and these confounding problems can lead to misinterpretation of images and so are briefly mentioned here[59-61] For systems that image in visible wavelengths there are a range of background fluorescence signals that lead to high background, even in the absence of fluorophore[62, 63]. These signals, such as those from NADH in the blue and porphyrins in the red are transient and can change with body site and physiology[38]. These are often the features that limit detection from the biology of the tissue.

Biological kinetics & compartmentalization signal effects: Many contrast agents are injected intravenously and are distributed to the entire organism through systemic circulation. This phase of activity is typically called biodistribution and is followed by a clearance phase in which the agent is excreted, either filtered by the liver or the kidneys. During these two phases, contrast agents may bind preferentially, with different proteins, cells and structures. Because the binding is a probabilistic phenomenon, with a dissociation constant describing the likelihood of the binding, contrast agents not only bind to their targets, but also to surrounding structures producing a background signal. This undesired retention degrades the ability to identify targets of interests, in particular, tumor margins. Additionally, even simple non-binding agents can have quenching issues, where the fluorescence can be suppressed due to excessively high concentrations or due to microenvironmental effects that alter the energy structures of the chemical species. The design of a contrast agent, in particular its nature (small molecule, antibody, peptide) and physical-chemical properties, is responsible for its fate in the organism. Several strategies have been used to design improved contrast agents, mainly to augment the signal but also more recently to reduce background effects. For instance, using small amphiphilic molecules with a hydrodynamic diameter under 5nm, results in rapid clearance leaving behind only the highest affinity interactions consisting of the contrast agent binding to its target. The design of the contrast agent, therefore, obeys certain rules to ensure proper behavior, or has high affinity to its target and low affinity to surrounding tissues. However, no "recipe" exists for creating the perfect contrast agent, and a balance must be found between all desired properties. Background effects can sometimes be avoided by using different delivery strategies such a topical application of sprays. A class of contrast agents of great interest relies on being activated when binding to the target, in which case the bound agent can be detected since the unbound fraction is not fluorescent. The knowledge of performance of all these parameters are typically worked out during the system manufacturing design process and application testing, however these factors can all be important in how the system performs in human use.

Camera lens parameters: In addition to the choice of F-number, which is critical for filtration, the objective lens plays an important role in the ergonomics and user experience of the imaging system. While the objective lens should match the sensor size of the camera and be designed for the wavelength range of interest, the strategy chosen for focal length and F-number will have an impact on ergonomics. While a wide-open aperture (low F-number) would allow many fluorescence photons to be detected, this strategy strongly impacts the depth-of-field of the imaging system and the ability to optically filter the fluorescence signal. The depth-of-field plays an important practical role as a narrow depth-of-field limits the range at which the system will be used, leaving structures blurry above and below a narrow depth on the field-of-view. Large depth-of-field is typically preferred in order to observe all structures in the field-of-view more easily. It is obtained at high F-numbers and is compatible with use of interference filters. The focal length also affects the working distance and the depth of field, and so careful choice of each of these parameters is needed for appropriate system design for the intended use.

Camera parameters: A large array of possible camera technologies exist for imaging fluorescence. Most commercial systems use regular CMOS cameras, which are produced at low cost, have high pixel density and fast readout. However, some use CCD cameras with higher dynamic range and linearity. Thermoelectrically cooled devices are used for lower noise, and higher bit readouts are used for higher dynamic range. Pixel density counts and sensor size vary dramatically between cameras, and this choice can alter sensitivity by a large amount. Often, high electronic gain is used to increase sensitivity at the expense of slightly higher noise, but because real-time image feedback is important to usage, this approach can provide increased frame-rate at lower limit of concentration because of the amplified signal. In almost all functional systems, video rate imaging is the desired performance goal for instantaneous feedback to the surgeon in both white light and fluorescence modes. Systems have also been produced with cameras that use intensifiers in front of the imaging sensor [64] [65], to allow for time-gating of the detection and for fast acquisition of low signal levels. Additionally, there are a large class of single photon detection technologies [66] that have time of flight capability as well [67], that may become more relevant for lifetime based imaging [68] or for distance and/or depth ranging into tissue.

Image pixelation & digitization: The pixelation of an image from a CMOS or CCD camera is inherent in the image capture process, with common cameras now being HD size or above. However, the number of pixels is not synonymous with the spatial resolution of the system, as the optical lens design commonly has the limiting effect upon the spatial resolution. Spatial resolution measurements are described later. Additionally, the light interactions with tissue and the scattering present can alter the effective resolution performance of a system in any given application as well, and so the user should be aware of the tradeoff between spatial resolution and contrast resolution of their imaging system. Measurements of this are described below.

The digitization level of a CMOS camera is typically the major factor that limits the dynamic range with the lowest performing cameras having 8 bits of depth to each pixel, and more modern cameras having 12, 14, or 16 bits at video rate output (>25 frames per second). Even when the system has a digitization level, it is common that the noise level on the bottom makes several of these bits not useful, and they are often deleted right out of the hardware pipeline, producing output video with images that have lower digitization per pixel. The most basic systems can work off of 6 bit effective depth, whereas more advanced ones use a full 14 bit dynamic range. This difference can be stark when appreciating the difference in gray scale levels (64 levels for 6 bit, 256 levels for 8 bit, 1024 levels for 10 bit, 4096 levels for 12 bit and 16384 levels for 14 bit, etc). Since display systems can only encode 8-bit output, some camera systems synthesize a high dynamic range output though the use of this higher compression or multiple exposures mapped together to provide very high bit depth to the user in a logarithmically compressed image intensity. This is common in commercial cameras and is likely to enter fluorescence surgical instruments as capabilities grow [69]. It is not common to calibrate these instruments to absolute units like photons/mm^2 or photons/str/mm^2 because the variable geometry between the camera and tissue will always alter this value. Most imaging is done with simple readout of intensity with a variable gain value, dynamic to the intensity being imaged. Further discussion of this calibration is below.

Image frame rate & display latency: The frame rates of systems are typically designed to be video rate (≈30 frames per second) if possible, however at times the signal levels can be low and there are situations where some systems might use substantially lower frame rates. However, for ICG imaging the concentrations are typically high and so the frame rates even with most CMOS cameras can be video rate. However, it is relatively obvious that integrating for longer periods of time, and/or using post processing algorithms on sequences of images can improve image quality. Related to this, there can be a temporal latency of display or a slowed display rate below video rate if the camera or pipeline of images are delayed relative to real time. Performance assessment of a system should occur in the intended use frame rate which would be utilized by the surgeon, incorporating the normally used frame rate and image display latency, rather than on images which have been optimized for acquisition but might not reflect normal video rate usage.

Image processing & corrections: All modern cameras and imaging systems have imperfections in their performance which are corrected for through firmware or software processing. The most extensive of these, such as defect pixels or readout irregularities are done by the manufacturer through online firmware processes inherent to the camera that get applied prior to readout. However, some of the more system specific effects such as lens distortions or background removal or noise suppression are applied in software after the readout or during the readout process. These methods are specific to each system, and get folded into the performance of how the entire system performs. The measures of these are described below.

2.3 PHANTOMS TO SIMULATE HUMAN TISSUE

Performance evaluation of imaging systems often involve the simulation of the signal in a stable test object, made of materials that represents the pertinent features of the tissue of relevance to the indicated use. To evaluate certain image quality characteristics, the test object can be relatively straight forward. If the purpose of the measurement is to simulate the signal effects that might be present in the intended use, a phantom that mimics the pertinent properties of tissue is utilized[70]. In this case, the interaction between tissue absorption and scattering and the fluorophore being imaged is commonly required. The effects of varying depth into tissue, concentration of fluorophore, layers and wavelengths used are common features to consider, especially when comparing performance of systems with different optical components.

2.3.1 Common phantom choices – strengths and weaknesses

The most important qualities of a tissue simulating phantom are to:

1) provide sufficient ***simulation of optical properties*** such that results generated are relevant to FGS.
2) exhibit characteristics to ***enable task specific metrics*** (e.g. fluorophore distribution within the FOV).
3) be ***manufacturable*** in a way which is repeatable to the desired accuracy.
4) be ***stable*** over the lifetime of the test needs.

Phantoms for FGS generally consist of four types of components:

1) *base* ***matrix material***, typically a solid or liquid that is relatively non-absorbing and non-scattering; in a solid phantom its primary function is to provide mechanical rigidity
2) ***scatterer***, having similar Mie-like scattering to tissue, (i.e., TiO_2, AlO_2, lipids), which is anisotropic elastic scattering which is strongly dependent upon the particle size, but largely having the macroscopic appearance of broadband 'white light' scattering.
3) ***absorber***, matching the spectral distribution and magnitude of biological molecules such as hemoglobin, water, melanin.
4) ***fluorophore***, includes relevant clinical dyes used for excitation/emission spectra or dyes that mimic them sufficiently.

Constituent materials can be combined in different ways to create tissue-mimicking materials or phantoms, which are incorporated as bulk structures (e.g., layers) or inclusions (e.g., spheres, cylinders) in a phantom. The importance of optical phantoms for spectroscopy, imaging and dosimetry has been reviewed and a list of materials that can simulate tissue scatter, absorption and fluorescence is well established[71, 72].

One key decision issue for fluorescence phantoms is if the phantom needs to be highly stable and exactly the same every time of use, or if it needs to be adapted and modified over time. Highly stable and consistent phantoms have typically been produced from solid matrix materials, but the establishment of intralipid as a standard for liquid organic matrix has also been reasonably robust when biologically preserved and allows use of organic dyes that are used in humans. Thus, there are two distinct paths which have been followed and each are reviewed here and summarized briefly in Table 3 below.

Table 3. Listing of types of tissue simulating phantom types and their components

Characteristics	Liquid (temporary) Phantom	Solid (permanent) Phantom
Matrix	Water, Gel	Polyurethane, Silicone, Plastics
Scatterer	Lipids (Intralipid)	TiO2, Al_2O_3
Absorbers	Blood, Melanin, Water	Inks
Fluorophores	All organic & inorganic	Some organic & inorganic
Strengths	Easily mixed Biologically compatible Widespread research use Mixture process is easy	Stable over years Transportable Manufacturable One time construction

Weaknesses	Single use construction Not easily transportable Subject to human error Prescription supplied	Not biological/organic Rigorous creation process

2.3.2 Phantom materials

Liquid phantoms: The most dominant choice in the field of biomedical optics for a turbid phantom has been the various forms of commercially available lipid emulsions, used for intravenous feeding of patients. The leading version of this is called Intralipid®[73, 74], shown in Figure 5, however other trade names from other companies are also used, such as Liposyn II®. These come in various concentrations of lipids (10%, 20%, 30%) and are commonly diluted in water to near 1% to mimic the scattering of tissue. The lipid component in these is highly regulated by health agencies, which produces the scattering nature of the liquid. There are smaller amounts of egg phospholipids as an emulsifier as well as glycerin. Because this is regulated to tight manufacturing criteria, it can serve as a stable matrix with scatterer embedded. Since it is based on water, it is inherently biologically compatible with hydrophilic organic dyes and other biomolecules or cells common in the human body. However, since it is comprised of lipid molecules, it is not stable unrefrigerated and must be re-established each time of use, limiting the time over which a sample may be used continuously. The detraction of this approach is that the mixing process is then subject to human errors in the process each time and requires a person who is knowledgeable about this process to prepare. Additionally, since this is a pharmaceutical product, most laboratories must order this as a prescription compound with medical authority to supply it. Nonetheless, this is likely the most widely utilized tissue phantom matrix material in scientific laboratories, and there is widespread literature on its optical characteristics and use [73-77].

The use of blood as an absorber is widely utilized since it perfectly mimics the blood and water absorption which dominates soft tissue [78] and is widely commercially available from non-human sources. Also, most fluorescent agents used in humans can then be directly dissolved in the phantom, allowing for a good match to the *in vivo* situation with nearly identical spectral characteristics and calibration approaches. This parallel to human tissue is the dominant attraction for this approach, although there is still some potential for microenvironmental effects of the fluorophore with the intralipid in a manner which is not representative of the in vivo use, and so the user must be aware of this risk when using agents which have unknown behavior in a high aqueous lipid environment.

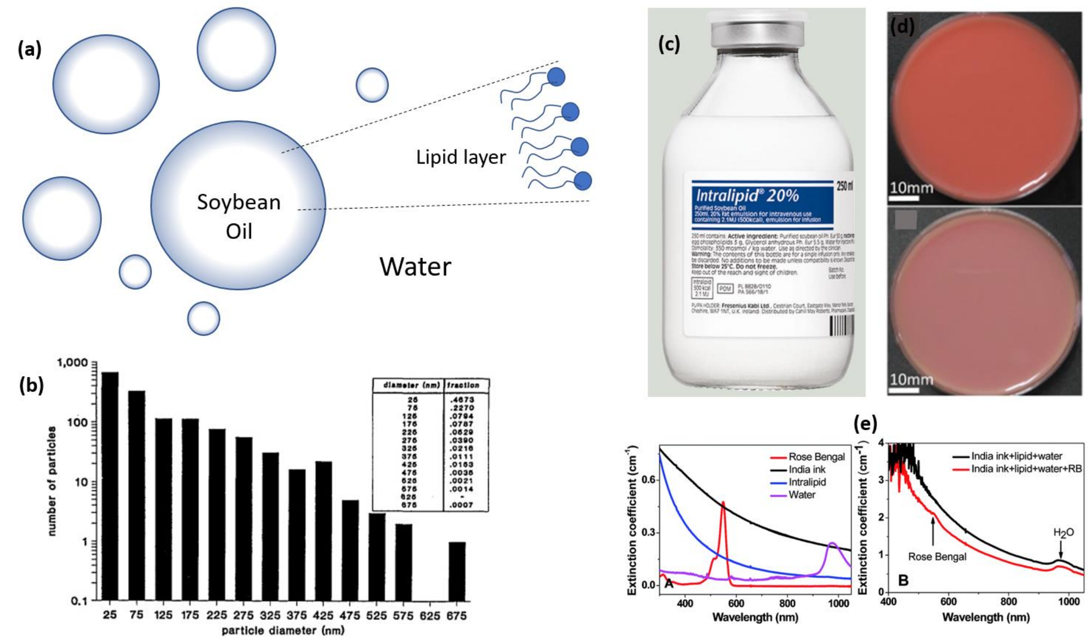


Figure 5. A schematic of Intralipid composed largely of soybean oil droplets in water is illustrated (a) with a histogram of particle sizes measured by electron microscopy (b) [74]. A vial of Intralipid is shown as supplied by one manufacturer (c) with example Intralipid-blood phantoms [79] (d) with an aqueous mix of 1% intralipid and 1% blood, at full oxygenation (top) and deoxygenated (bottom), and extinction spectra of Intralipid phantoms with added constituents for fluorescence measurement (e). [80]

Solid Phantoms: Despite the deep historical use of Intralipid-based phantoms, the detraction of not being able to have a stable, easily-used phantom that does not require any knowledge of the mixing process, has been an issue. In terms of supplying a manufactured product or test object, the use of Intralipid seems less reasonable. Additionally, in parallel to other radiological imaging systems, there is a need for permanent phantoms that can be used for quality audit over many years, and so are therefore manufactured by plastics or resins, with stable, long-term form factors. Several companies and research groups in the FGS field have focused on solid phantoms for test objects, due to their superior consistency both in terms of mechanical and optical properties over time, and their ability to be independently manufactured and quality controlled outside of the point of use, and shipped to any site in the world[81].

Within this context, several different phantom designs and corresponding uses have been suggested. One commonly used matrix has been polyurethane due to its stability and machineability [81-84]. An additional choice which has a mechanical flexibility closer to human tissue is a silicone matrix [85-87], however this has less machinability than the resin-based ones. The choice of this type of matrix necessitates a compatible non-organic scatterer such as titanium dioxide (TiO_2) particles, which can mimic tissue scattering spectra. One caveat is that these powder particles tend to be smaller and have a higher index of refraction than lipids, and so the scattering spectrum can tend more towards Rayleigh shape than Mie shape. This is important because it affects the scattering spectrum and the anisotropy phase function, but still the tradeoff of having a permanent matrix phantoms has been thought to be reasonable in several applications. While powders have been used extensively, their level of aggregation is high, so strategies for incorporation as either a pre-mixed liquid form (common in white paint) have been demonstrated, as have rigorous mixing protocols to ensure maximum homogenization is reached before solidifying. Examples of these are shown in Figure 6. Methods for controlling scattering in solid phantoms have been adapted over many years, with one more comprehensive report showing how to titrate the scattering spectrum more precisely[88]. According to this work, epoxy-resin is an ideal material for the phantom matrix, while a combination of TiO_2 and aluminum oxide is suggested for scattering anisotropy and phase function control.

Fluorophores: Varying quantities of NIR fluorescent agents can be also incorporated in the polyurethane hardener. The initial idea for solid phantoms came from Firbank et al [82, 83] who demonstrated this use for optical tomography; subsequently, this approach was widely adopted by many groups [34, 60, 89-91], and even commercially supplied by a few companies, notably INO in Quebec and PerkinElmer in Hopkinton MA. The use of polyurethane-based solid phantoms for FGS came more recently as a stable standard test phantom [92-94]. While organic dyes such as ICG or protoporphyrin IX (PpIX) are desired to be used in vivo, they are not always stable in a matrix such as polyurethane, and so it has been a challenge to find ways to directly sample ICG as a fluorophore in a permanent test phantom. Inorganic particles can be incorporated, and many versions have high stability and high quantum yield of emission. Some laser dyes that are manufactured for high stability are able to be embedded into resin with high stability, and IR125 has been found to both match the ICG spectrum as well as be stable in this application[95]. One of the most stable options are nanoparticles (quantum dots), 2-6 nm in size, fabricated from semiconductor materials such as silicon and germanium with specific examples of cadmium selenide or indium arsenide. These commonly have broad UV/blue/green absorption spectra, and a large variety of emission characteristics can be chosen. Published photostability tests revealed that these phantoms exhibited less than 1.0% variation in fluorescent intensity over 50 days, thus indicating that quantum dots may be suitable for FGS phantoms. Rigorous photostability tests must be made at an ongoing frequency to establish the optimal choice of agents for these solid phantoms [92-94]. The one major caveat with quantum dots is that for high concentrations, the cost of purchase can be a practical limitation, and so from this standpoint more attention is focused on lower cost solutions such as laser dyes.

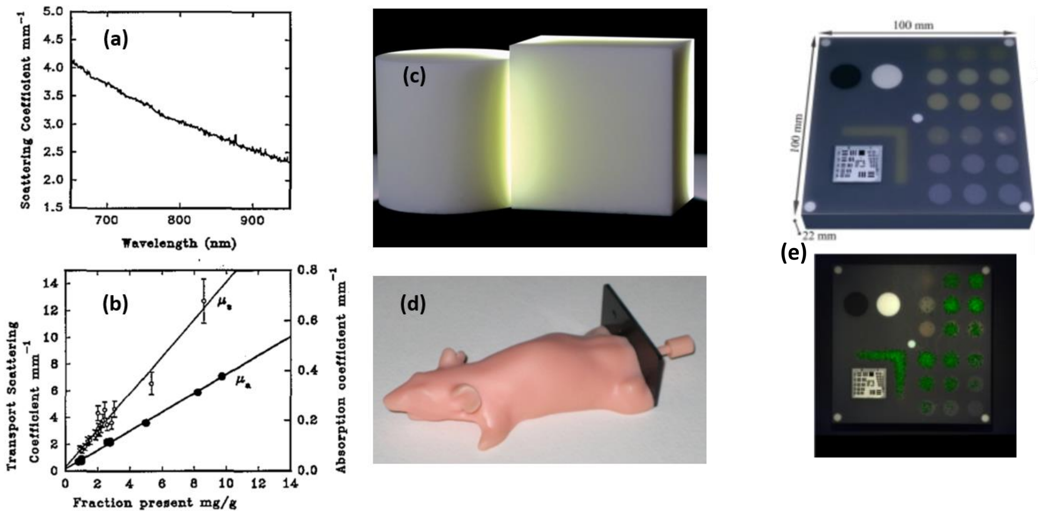


Figure 6. Scattering spectra (a) from the original work of Firbank et al [83, 96] based upon polyurethane resin, with values of reduced or transport scattering as a function of concentration (b) of scatterer from TiO_2 and Al_2O_3 concentration, and absorption from black inkjet dye. These types of phantoms are now available in calibrated custom machined forms, as shown in (c) (INO, Quebec Canada) and in an anthropomorphic mouse shapes (Xfm-2 phantom, PerkinElmer, Hopkinton MA) (d). In (e) this type of material was combined with nanoparticles and cast into a test phantom for sampling a range of imaging properties, as a prototype design for comprehensive system assessment with a single phantom. [93]

Absorbers: Absorption can come from a range of pigments (ie. India Ink, Nigrosin, Phthalo Green, Phthalo Blue Royal, Cinnabar, Haematite, Cobalt Blue, Cobalt Blue Turquoise, and Cobalt Violet), while most efforts have simply used a single such pigment with absorption in the spectral region of importance for the test. A common flat spectral agent widely chosen is either India Ink or Nigrosin dyes. Organic dyes which are highly stable can be added, although their emission spectrum is well known to shift when embedded in a resin matrix. The fluorophores are dissolved either in the resin directly or first pre-dissolved in organic solvent and then mixed into the resin[97, 98]. For Rhodamine dye, fluorescence emission has been shown to remain stable for up to 3 months in one case [6].

Heterogeneous phantoms: A modular phantom can be used to target depth variations of the fluorescent layers [99-101], adjustable layers, allowing tests of fluorescence imaging as a function of depth. A similar concept was also adopted by Leh *et al.* to propose phantoms with variable properties and geometries [102]. The inclusion of background fluorescence signals mimicking the autofluorescence of human tissues can be especially important in the visible wavelengths, and so Rodamine B or fluorescein (FITC) can be employed. Rodamine B presents emission peak at 580 nm similar to lipopigments, while FITC emission at 515 nm, is similar to flavins. Excitation in the blue wavelengths especially is dominated by fluorescence from these agents. Phantoms consisting of multiple diffuse reflectance targets has been shown [103], with reflectance values used to assess increased excitation light leakage through the fluorescence detection path. This is an important specification of systems which is commonly forgotten, especially by untrained users. Critically employing this strategy will allow assessment of the crosstalk signal which will limit the lower detection level, or in some cases could be used to correct for this baseline offset.

Besides regular flat shaped test phantoms, organ or tissue shaped anthropomorphic phantoms have been used for task-specific assessments or training [104, 105]. These are commonly developed in four possible ways, via gelatin, via polyurethane, via silicone or via 3D printing methods. The benefits of gelatin and silicone are that they can be poured from liquid into a mold, whereas polyurethane can be machined.

3D printing solid phantoms: Fabrication from 3D printing has been a topic of research development [106], which could potentially simplify and standardize the process of performance testing. Future assessment paradigms may involve 3D-printed calibration targets or phantoms with biomimetic morphologies containing fluorophore-doped inclusions [107]. This field is still emerging, however undoubtedly the widespread penetration of 3D printers and the low cost reductions associated with their materials, and ease of use of the software for design could likely make this pathway more and more attractive, as long as the reproducibility and batch-to-batch and system-to-system consistency is sufficiently stable.

2.3.4 Phantom validation: optical properties and morphology

Validation of phantom properties breaks down into the three major functions of 1) scattering, 2) absorption, and 3) fluorescence. Most manufacturers focus on fluorescence intensity recovery, however each parameter can significantly affect the signal; so, the performance and value of a phantom depends upon these properties mimicking tissue reasonably well. Characterization methods for tissue properties can largely be microscopic or macroscopic in nature. Macroscopic methods have generally been preferred, because in the end the scattering and absorption coefficients are defined as 'bulk' values. [108]

Bulk tissue optical property estimation requires a measurement methodology which can separate out the dominant effect of multiple scattering from the absorption and fluorescence signals. As such, a light transport model is also routinely required to fit the measurements to deconvolve this out. Commonly either diffusion theory or Monte Carlo models are applied to measurement data to fit for independent absorption and scattering coefficients [109]. Bulk measurements can be taken with invasive insertion of fibers [74] or on the surfaces for solid phantoms with measurements which are one of: 1) time-resolved with sub-nanosecond resolution [110, 111], 2) frequency domain in the 100's of MHz [112], 3) spatially resolved to better than 1mm resolution [76]; 4) spectrally resolved with constraints on the fitting spectra [113]. Commercial systems for these are not widely available nor used, but still some versions are available, such as devices based on frequency domain (ISS, Inc., Champaign, IL) or spatial Fourier domain (Modulated Imaging, Irvine, CA) approaches. Numerous custom-made instruments are used in research laboratories. The utilization of transport modeling to fluorescence deconvolution is widely recognized as being needed for accurate quantification of fluorescent agent concentrations [114]. Phantoms which have been shaped into regular geometries such as a slab, sphere or cylinder can most easily be fit to this type of modeling [115-117] for absolute extraction of tissue optical properties. Arbitrary shapes can also be used if the overall shape can be accurately fit to a numerical solution to Monte Carlo or diffusion theory[118] through numerical models, or if the shape is sufficiently large compared to the measurement area that a geometric simplification can be applied [116, 119].

3. Protocol

3.1 METHODS FOR PERFORMANCE TESTING

International consensus standards for established medical imaging modalities (e.g., CT, MRI, ultrasound) describe best practices for characterization of image quality based on objective, quantitative test methods [120-123]. These standards provide a core set of principles that can be applied across medical imaging, including image quality characteristics, test objects and their properties, experimental methods, and data analysis procedures for calculating figures of merit. These concepts have, to some extent and in various forms, been adopted for assessment of fluorescence imaging products, [92, 124, 125] however, consensus has not been established as with the aforementioned modalities, although some important research studies have recently come out[92, 126, 127]. This is the major goal of the current document.

In this section, we identify best practices for performance testing of fluorescence imaging devices used with exogenous fluorophore contrast agents. The intent is to provide a framework for objective assessment of image quality with quantitative metrics in a standardized manner applicable to a wide variety of devices. This framework includes test targets and tissue-simulating phantoms that are biologically relevant, consistent and “least burdensome” in terms of fabrication and implementation. However, given variations in clinical products (e.g., wide field vs. microscopy vs. endoscopic, wavelength, fluorophore), the specific embodiment of test methods and the significance of individual characteristics to clinical performance may vary from product to product. This section is divided into three parts: (1) fundamental system performance characteristics, (2) application or task-based characteristics, and (3) assessment of confounding factors/artifacts. A summary of these is provided in Table 4.

Table 4. System features and characteristics that require some level of performance testing

System Characteristics	Specific Feature to be tested
Fundamental Performance	Image Sharpness Depth of Field Signal Uniformity Distortion Field of View Spatial Co-Registration
Application-specific or task-oriented performance	Signal Sensitivity Limit of Detection Response Linearity & Dynamic Range Imaging Sensitivity Imaging Depth Sensitivity Tissue Absorption & Scattering Effects
Assessment of Confounding factors/artifacts	Crosstalk Off-target Fluorescence
Other Task Specific Tests	Geometric Accuracy Contrast-Detail Analysis Accuracy of Concentration Measurement Repeatability & Reproducibility

3.1.1 Fundamental System Performance Characteristics

Many of the most basic aspects of fluorescence image quality are identical to concepts used in white light imaging, and some of these can be measured with standard test targets or test fields commonly associated with white light imaging. Fluorescence test methods may be similar to white light tests, but they should be designed with the relevant fluorophores and testing performed in fluorescence imaging mode, as with the intended use of the system.

Specific test methods have been adapted to account for measurement of fluorescence rather than broadband reflected light, and include: 1) Image Sharpness, 2) Depth of Field, 3) Signal Uniformity, 4) Distortion, 5) Field of View and 6) Spatial Co-registration between imaging channels. Each of these are described briefly below.

Image Sharpness or High Contrast Spatial Resolution: Sharpness of features in an image is typically addressed in terms of spatial resolution, that is, the ability to resolve two distinct, high-contrast structures. This property is of primary importance for biological imaging due to the need to identify fine features such as tumor metastases. One of the most well established approaches for spatial resolution evaluation involves the Modulation Transfer Function (MTF), but it typically requires imaging target with sharp features and Fourier transforming the resultant signal intensities observed, which involves both measurement and computation. Instead, "bar chart" test targets with groups of black and white rectangular segments of decreasing size (e.g., USAF 1951) are commonly used to evaluate imaging device Contrast Transfer Function (CTF) by determining the contrast level for each spatial frequency (*f*, in line pairs/mm) based on the following equation:

$$C_I(f) = \frac{I_{\max}(f) - I_{\min}(f)}{I_{\max}(f) + I_{\min}(f)} \tag{1}$$

where C is Contrast, I_{max} represents values acquired for high intensity bars and I_{min} represents values for low intensity bars.

For fluorescence imaging, bar chart targets with alternating transparent and non-transparent regions (e.g., chrome on glass) can be used in front of a diffuse source of backlighting. This illumination can be produced using an integrating sphere, or a highly fluorescent object placed behind the target. While the former approach provides a uniform light distribution which isolates the effect of detection instrumentation, the latter approach includes the impact of illumination uniformity as well.

Once a CTF graph is generated (Figure 7), a spatial resolution metric can be obtained based on the *Rayleigh criterion*, in which the spatial frequency providing a contrast level of 26.4% is determined. Groups of bars in horizontal and vertical orientations should be used to evaluate resolution in each direction. CTF graphs should provide enough spatial frequencies to resolve all significant variations across a contrast range of 1.0 to 0.1 (100% to 10%).

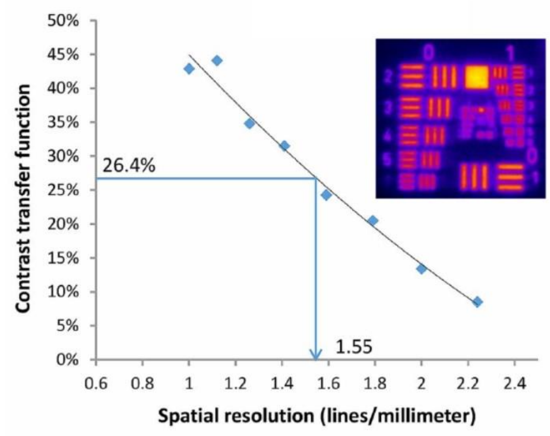


Figure 7. Image sharpness testing results including a back-illuminated bar chart (inset) and corresponding CTF curve for horizontal resolution, in terms of line pairs per mm, indicating the Rayleigh criterion [125].

Spatial resolution can vary with position in the image field due to optical system/component imperfections. Thus, as recommended in a prior endoscope image quality standard, [128] measuring "off-axis" resolution at four points located 70% of the distance from the center to the corner of a rectangular field of view – or the edge of a circular field of view – should be performed.

An alternate technique for CTF generation – the slanted edge method – can generate results more rapidly, but it has not been rigorously validated for near infrared fluorescence imaging in terms of its consistency with standard approaches [129]. This method involves imaging a light/dark edge at a slight angle from the vertical or horizontal, and taking the Fourier transform of the 1-D edge spread function. ISO standards based on this approach have been developed for camera systems. [128]

Depth of Field: Spatial resolution degrades rapidly as a function of distance from the imaging system focal plane. Practically, a short depth-of-field (DOF) reduces the performance of imaging systems where the device-to-tissue distance varies either temporally (e.g., handheld devices) or spatially (e.g., when tissue surface is irregular or not parallel with focal plane), thus causing parts of image to exhibit suboptimal sharpness. DOF can be measured using a bar-chart target placed at a range of working distances above and below the focal plane, as shown in Figure 8. Full CTF curves can be measured at each target depth, or, more simply, a specific spatial frequency that shows high contrast at the focal plane can be used to quantify variations in contrast with position. [130]

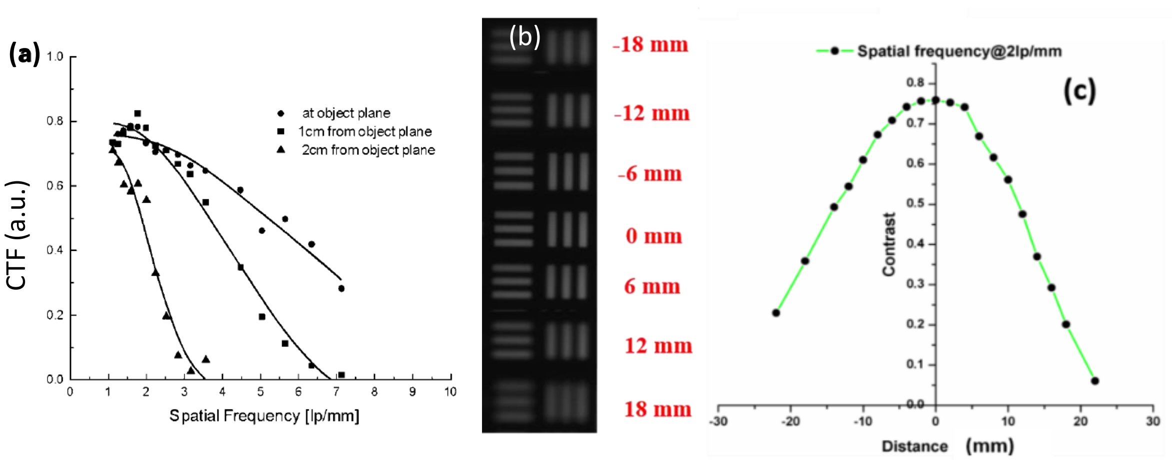


Figure 8. Depth of field measurements, including (a) CTF curves at different working distances [131]; and results using the 2 lp/mm resolution group, including (b) images of the group at seven positions and (c) a graph of contrast as a function of distance from best focus position. [132]

Signal Uniformity: Spatial variations in signal intensity across the image field unrelated to the interrogated tissue can reduce FGS device effectiveness. Non-Uniformity can arise from both illumination and detection path components. While it is possible to separate illumination and detection Uniformity, this is typically unnecessary for clinical device performance evaluation. Thus, FGS system Uniformity can be evaluated using a simple homogeneous, fluorophore-doped phantom [133] (Figure 9(a)). Signal intensity variation across the image field is then graphed along the horizontal and vertical midpoints of the image (Figure 9(b)), and a non-Uniformity metric can be obtained by determining the fractional decrease from maximum to minimum values. An alternate method for illumination uniformity has recently been reported evaluation based on reflective, yet non-fluorescent inclusions at the center and four edges of a square phantom (Figure 9(c-e)) [92]. In this approach, 5 localized regions of the same fluorescence intensity are placed in the 4 corners and center of the phantom, to provide individual measurement spots for the remitted fluorescence intensity across the imaging field. A more ideal approach for FGS systems would involve fluorescent inclusions. Most importantly, devices that perform non-uniformity correction, signal Uniformity should be evaluated both before and after correction and some repeated measures of system stability should be done on a regular basis or on a frequency commensurate with the expected change. Furthermore, the effect of non-Uniformity correction on local dynamic range and signal to noise ratio should also be identified, and all other performance data should be based on identically corrected results.

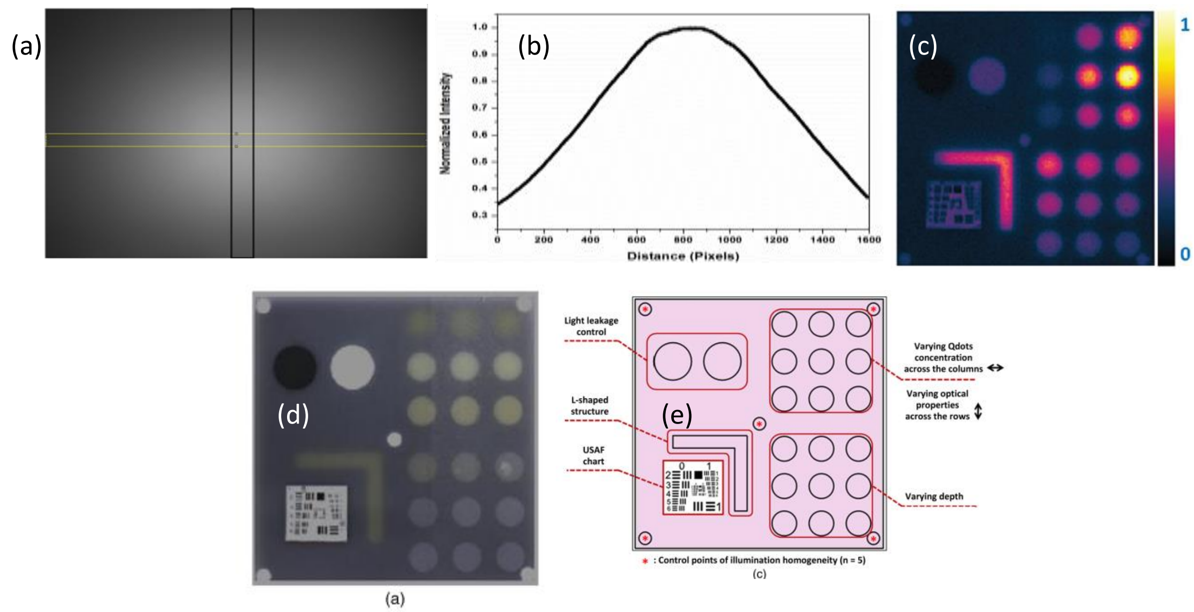


Figure 9. Illustration of Uniformity evaluation results, including (a) image of a homogeneous fluorescence target and (b) graph illustrating quantitative variations in signal intensity for a horizontal line through the center of the image and (c) image of a multi-parameter phantom for uniformity using five points (center and four corners). The photograph of the phantom in (c) is shown in (d) and the legend for the regions in (e). The uniformity acorss the imageing field was proposed to be tested by the signal from the dots in the 4 cornders of the phantom that match the central one, allowing for fluorescence intensity estimation across the field of view. [93, 126]

<u>Distortion:</u> When an image displays spatially-dependent variations in magnification – thus resulting in a deviation from rectilinear projection – it is considered to exhibit distortion. Typically presenting as a strong degree of radial symmetry from the center of the image, distortion is most evident in wide-angle lens assemblies used in endoscopes and other imagers designed to provide a large field of view (FOV). Given the potential for these variations in magnification to cause errors in estimation of tissue structure shape and size, device-to-device variations in this property may impact clinical device efficacy, of the type seen in Figure 10(a), and quantified in the graph (b).

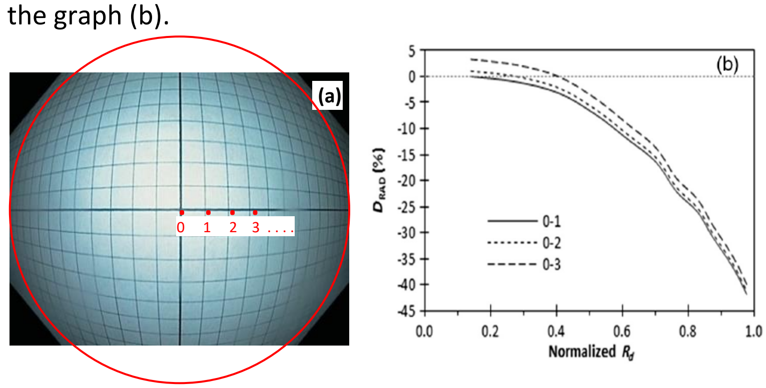


Figure 10. Distortion testing results, including (a) white light endoscopic image of a test target comprised of square grids and (b) distortion graph illustrating a typical curve for an image with barrel distortion, where Rd is radial distance [134]

Most commonly, a target comprised of square grids (based on lines or small individual points) is used. By determining change in magnification as a function of true position from the origin – based on an assumption of constant spacing between lines in a grid target – it is possible to generate a distortion curve. This curve should provide the maximum measured distortion (likely near the edge of the image). If a distortion correction algorithm is used for a device, results should be provided before and after its implementation. Furthermore, all other performance data generated for the device should include this correction.

Field of View (FOV): Imaging system FOV is a basic image quality characteristic that can be reported in terms of vertical and horizontal dimensions, or the angle subtended by the camera. In most cases, evaluating the former is a relatively simple exercise that can be performed by simply measuring distances with a fluorescent phantom. However, in the case of a device such as a surgical camera or endoscope, where a large field of view is achieved at the expense of strong image distortion, such an exercise becomes more difficult. Thus, angular field of view is sometimes preferred for high distortion imaging systems, but the simple distance measurement of FOV is at times easier. [134]

Spatial Co-Registration: Since fluorescence imaging systems are commonly implemented in conjunction with white light imaging – using composite overlay images for navigation and direction of treatment – accurate co-registration of features may impact safety and efficacy. Tests that contain features detectable with both modalities, white light and fluorescence, should be used to quantify spatial registration differences. Software processes to register them may be implemented in cases where there is significant mismatch or to fix changes in registration over time. Additionally, the use of testing approaches to ensure co-localization with other imaging modalities used for multi-modal surgical guidance – such as ultrasound, CT or MRI – may be appropriate as well. If a co-registration correction algorithm is used for a device, results should be provided before and after its implementation.

3.1.2 Application-Specific or Task-Oriented Performance Characteristics

Tests that are more specific to the nature of the purpose of the system will need some level of customization for the specific system, such as molecular probe measured, sensitivity range desired, depth of sensitivity needed, etc. The testing should be done in normal operation mode of the system, as would be used in surgery with appropriate focusing, frame rates, and all acquisition parameters as would be common in human usage. These tests are more likely to need a custom tissue-simulating phantom to perform analysis with excitation and emission in the band designed for the system.

The performance measures relevant are: 1) signal sensitivity and the related concepts, 2) concentration limit of detection, 3) response linearity, 4) dynamic range, 5) imaging detection sensitivity,

6) imaging depth sensitivity, and 7) effect of absorption and scattering changes. The first three are often defined on large regions of sample and the latter are tests of imaging detection where the size of the test region affects the outcome. Each are briefly described here.

Signal Sensitivity: Perhaps the most widely reported performance characteristic for NIRF imaging systems is sensitivity. However, this is a generalized term that addresses the relationship between contrast agent concentration and detected signal, and a range of approaches have been applied for evaluating this characteristic. Methods to distinguish sensitivity from characteristics like detection limit, linearity and dynamic range, are important as they can essentially be determined from the same set of measurements but provide different insights into NIRF product performance.

The clinical viability of a device depends on whether it is sufficiently sensitive to detect the levels of fluorophore concentration present in relevant tissue structures. While it is highly desirable for phantoms to have a form that is solid and stable over time – often achieved by using polymers such as silicone, polyurethane or epoxy – any solid phantom must be rigorously evaluated to ensure that its optical properties (fluorescence excitation/emission, scattering, absorption) are representative of the clinical scenario for which the product is intended. Typically, tests are performed using small, fluorophore-doped inclusions at a variety of fluorophore concentrations [92, 105, 124, 125], as shown in Figure 11(a). In the interest of consistency, generalized tissue/background values of $\mu_s' = 1\ mm^{-1}$, $\mu_a = 0.01\ mm^{-1}$ should be used, unless other distinct consensus values are warranted for a specific tissue type. Secondly, the boundaries of the phantom should reflect *in vivo* behavior, without unrealistic effects (e.g., highly reflective well boundaries). Additionally, the number of different fluorophore concentrations, the interval between each concentration, and the range to be covered should be designed such that the limit of detection can be accurately determined without excessive interpolation. In order to establish sensitivity and linearity, some studies have used over 20 concentrations [124]. No matter the range tested, ideally, two or more concentration levels that produce signal levels exceeding the mean background level by 1-5 standard deviations should be provided to establish detection limit. The measurement from each well should involve a region of interest that is within the interior of the region to encompasses about half the diameter, but avoiding any limb effect, of blurring in the edges that can be observed at the walls of the region. The relevant range of concentrations should span the concentration expected in tissue for the intended use, which can be high for blood vessels and considerably lower in tissue perfusion, for example. The spectrum of the dye used for testing might ideally match the emission of the dye intended for use in vivo, such as for indocyanine green being matched by IR125 having similar emission spectra to ICG. Although admittedly this criteria is a tradeoff with stability, and agents such as quantum dots, for example, have been shown to have similar emission spectra, but not similar excitation spectra, and yet have exceedingly high stability. So, this range of effects makes the choice of an ideal dye for a phantom to be an imperfect optimization process.

There are a variety of potential confounding factors that may impact sensitivity measurements. Spatial variations in sensitivity across the image due to non-uniformity or other effects may cause irregularity in measurements performed with a multi-phantom array. In these cases, it may be necessary to measure each element in the array near the center of the field of view. Nonlinear response in measurements of fluorophore-doped phantoms – due to quenching, inner filter effects or other concentration-dependent optical phenomena – can lead to misinterpretations of instrumentation behavior. In such cases, it may be useful to determine device sensitivity and linearity independently of the fluorophore (e.g., though the use of a single well and ND filters). One extreme example is illustrated in Figure 11(b) where ICG is known to aggregate or quench at higher concentrations, leading to a decreasing signal above concentrations of 10 μM.

A general graph of sensitivity should be generated which displays the measured fluorescence signal to noise ratio (SNR) as a function of known fluorophore concentration, where S is fluorescence signal intensity and C is fluorophore concentration, and signal S at C=0 is removed to prevent background from altering the interpretation. The σ is the standard deviation of the fluorescence signal at C=0:

$$SNR(C) = \frac{S_m(C) - S_m(0)}{\sigma(0)}$$

The first few data points should occur in a regime where background dominates the measured signal and is independent of any fluorescence, after which an increase in detected signal with fluorophore concentration is seen, as seen in Figure 11(c). Often, a linear region is followed by a decreasing slope, which may be due to nonlinear effects such as emission photon reabsorption by the dye itself. Such curves typically have a saturation regime at the top concentrations due to either probe or system saturation, and then a noise floor at the bottom where the system does not detect the probe anymore, and in between these saturation regions is the working range of detection.

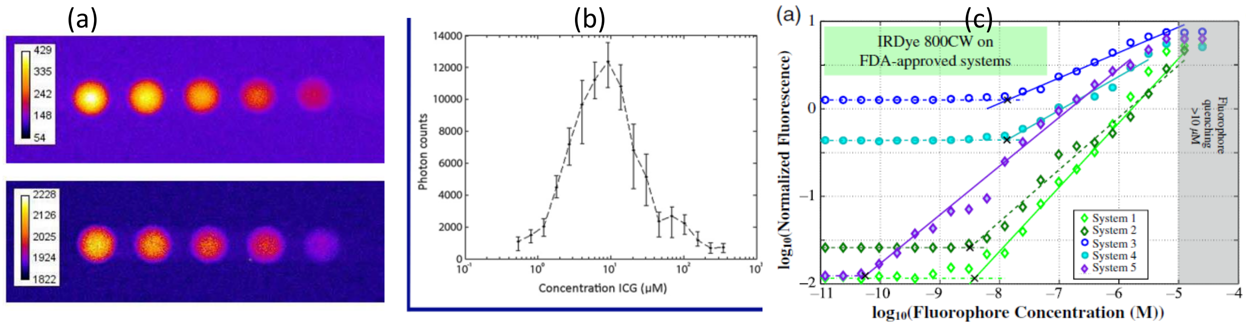


Figure 11. *Examples of sensitivity measurements, including: (a) fluorescence image of a multi-well phantom [125]; (2) graph of signal intensity as a function of fluorophore concentration [105]; and (3) graph of signal intensity as a function of fluorophore concentration for several imagers [124]. Limit of detection, linearity and dynamic range are also determined from these measurements.*

<u>Concentration Limit of Detection:</u> It is useful to define the ability of a system to accurately identify the presence of low concentrations of a fluorophore, as this can directly impact clinical effectiveness. The approach for determining detection limit commonly implemented in medical imaging standards has often involved subjective visualization by a reader (e.g., number of inclusions visible, where each has a different contrast level and/or size), rather than an objective measure. However, in clinical chemistry consensus documents, concentration *limit of detection* (LoD) is the metric commonly used to describe the detection capability of an instrument [135] [136]. While several analysis approaches may be suitable for such a test (e.g. Probit analysis or using the standard deviation and slope of the response), the simplest approach involves the determination of the point at which the detected CNR reaches 3.0. This threshold has been used previously in fluorescence imaging "as a surrogate measure for human detection of objects."[137] It should be noted that LoD defines the lowest amount of analyte in a sample that can be detected, but not necessarily quantified in an accurate manner. The aforementioned documents describe a second parameter, the limit of quantitation, which is the lowest concentration needed to determine analyte concentration with suitable precision and accuracy. Additionally, though the detection limit is coupled to the spatial resolution, and so the size of the targets should ideally be much larger than the limiting resolution to simplify the testing, and ideally near the size relevant to the use case of the system for detecting tissue regions. The LoD value itself should also be relevant to the use case of what concentrations are being detected with the standard medical need.

<u>Response Linearity and Dynamic Range:</u> The relationship between the concentration of an imaging biomarker and the detected signal is commonly called "linearity" in medical imaging literature, as these quantities are often proportional to one another under ideal conditions. Indeed, the relationship between fluorophore concentration and fluorescence signal detected is ideally linear and is complementary to sensitivity in that a similar approach based on a multi-concentration phantom can be used. Its significance lies in the ability to accurately estimate fluorophore concentrations as well as to accurately visualize tissue structures or spatial variations in fluorophore density; i.e., nonlinear response would decrease the contrast of a high intensity probe-labeled structure in a moderately fluorescent background.

Linearity can typically be derived from the same data used to determine sensitivity and LoD. The range of data used for linearity is defined at the lower end by the LoD and at the high end by the maximum intensity displayed by the device or a significant deviation from linear. Alternately, other points can be specified over which better linearity is achieved. Linearity can be defined in terms of a log-log- plot with equation [124]:

$$\log_{10} y = m \log_{10} x + C \quad \text{or} \quad y = 10^{C} x^{m}$$

Where m and C are the fitted slope and x-axis limit, respectively, and this approach assumes that the background signal has been removed. For a linear response m should be unity. The data used for sensitivity can also be used to determine dynamic range, a key inherent performance specification linked to the bit depth of a digital imaging device. However, dynamic range is typically defined as the ratio of the largest to smallest values of signal intensity that a system is capable of measuring.

Imaging Detection Sensitivity: During standard sensitivity measurements involving a set of targets with increasing fluorophore concentrations, nonlinearities may be introduced due to quenching from dye-molecule interactions, or from inner filter effects, where the dye self-absorbs its own emission. An alternate approach that minimizes these processes can be implemented to better characterize inherent device detection sensitivity. Well-controlled measurements either with or without a phantom can be implemented. A simple high-turbidity, fluorophore-doped phantom covered by a black plate with an aperture and neutral density filters that provide a wide range of attenuation levels. Thus, the limit of detection can be benchmarked in terms of a fraction of a moderate sample concentration. It would be necessary to standardize the phantom design so that the results are comparable between measurements. Graphing the detected signal intensity as a function of filter transmission squared (due to attenuation of excitation and emission light), it is possible to decouple nonlinear fluorophore effects from inherent device behavior. [132]

Imaging Depth Sensitivity: Differences in fluorophore optical properties (e.g., wavelength, quantum yield), optical instrumentation and processing methods can result in system-dependent variations in ability to image deeper structures, up to several millimeters below the tissue surface. These variations in penetration depth can impact clinical performance, particularly for applications such as lymph node localization and extraction, and subsurface tumor detection. A wide variety of phantom-based test methods have been used for penetration depth testing, typically involving fluorophore-doped inclusions located at different depths within a turbid, non-fluorescent matrix [105, 138]. While phantoms with solid fluorophore inclusions (e.g., Figure 12a) may provide longer stability for constancy testing, those more well suited to incorporation of liquid fluorophores may provide greater biological relevance and flexibility. When a single phantom with multiple inclusions at different depths is used, crosstalk between inclusions must be minimal. Results can be quantified by graphing signal vs. inclusion depth as in Fig. 12. However, a more standardized approach may involve graphing signal to noise ratio (where a blank sample is used to evaluate noise) as a function of depth. The point at which the contrast to noise ratio falls to 3.0 – based on the aforementioned detectability threshold, referred to as the 'Rose Criterion' – could be identified as the maximum imaging depth. Alternately, a metric based on changes in apparent inclusion size (e.g., full-width-half-maximum – FWHM – of the intensity across a channel) may be appropriate to characterize how products differ in their ability to image deep structures. While a Y-axis can be in units of pixels, a more optimal standardized approach would involve calibrated distance (e.g., mm).

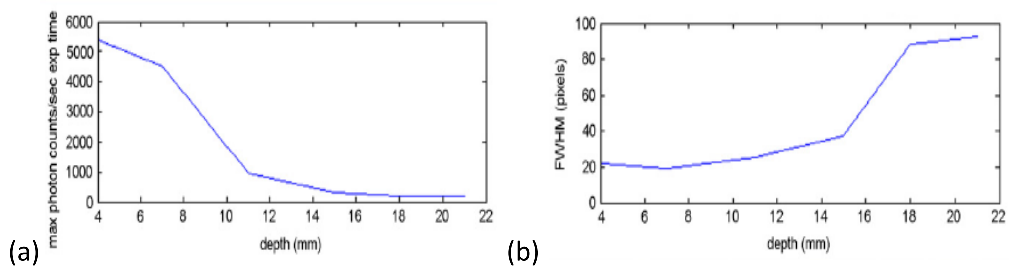


Figure 12. Imaging depth results based on a turbid agarose phantom with fluorophore-doped inclusions at different depths: (a) intensity vs. depth and (b) FWHM vs. depth. *[105]*

***<u>Tissue Absorption and Scattering Effects</u>*: A significant source of variability in biological tissue is heterogeneity of, and inter-patient variations in, *tissue optical properties* – particularly the impact of the reduced scattering coefficient and the absorption coefficient on the measured fluorescence intensity [92]. By measuring fluorophore-doped inclusions with constant concentrations but varying optical properties of the surrounding, it is possible to evaluate the robustness of a device to biologically realistic variations in these values. However, it is important to use biologically relevant values, because the extreme ranges of absorption or scattering can cause severe changes in signal that are highly non-linear, whereas there are also systems that are minimally affected across the typical human tissue range. So, it is important to have test phantoms that cover the range of typical human tissues.

Skin pigmentation, i.e., inter-patient variations in light absorption by epidermal melanin, is a specific subcase of absorption where the absorption is just in the very thin layer of the epidermis. Studies have indicated that high melanin concentrations can reduce detected signal intensity and affect clinical oximetry devices based on visible and near-infrared spectroscopy. Therefore, it would be appropriate for fluorescence imaging systems involving epidermal, dermal or trans-dermal measurements to be evaluated with phantoms that simulate a range of pigmentation levels, or at least for levels representing upper and lower bounds. This is not common in surgical systems though, and so while an important issue, it is more relevant for systems tasked for skin imaging and lymph node imaging. The dominant absorber throughout most of the surgical imaging world is clearly blood, and sometimes water in the mid to far NIR wavelengths above 800nm.

3.1.3 Assessment of Confounding Factors/Artifacts

This section addresses methods for quantifying the impact of specific well-known optical device limitations and tissue properties that can degrade image quality. Some of these issues could be considered as system specific tests, but in many cases the measurement tissue affects the presence or magnitude of the effect, and so they are not always strictly specific to just the imaging system itself, although the control over them is likely dictated by the system design and performance. The core issues to consider here are: 1) crosstalk, 2) off target fluorescence.

Crosstalk: is an undesired increase in measured fluorescence signals due to contributions from other sources, which can be a significant confounding factor under clinical conditions. One of the most common confounding factors in fluorescence imaging is excitation crosstalk, or light “leakage” from an excitation source that is detected by the camera, especially since low fluorescence yield often necessitates illumination intensity orders of magnitude greater than detected fluorescence intensity. Reflection of excitation light is particularly problematic at specular surfaces or locations of high scattering. Since this excitation light can be mistaken for fluorescence when viewing tissue, testing for excitation crosstalk under realistic scenarios is important to predicting clinical performance. A basic method for evaluating this effect is to image a highly scattering, yet non-fluorescent target (e.g., Spectralon®) and compare this value to an image of a non-fluorescent target with minimal scattering [92] and/or a dark image. This approach may also be useful to identify unwanted optical contributions from ambient light sources.

The second major category of light leakage or crosstalk is room light leakage into the fluorescence image. This is typically assessed by imaging fluorescence on a field without any fluorophore, and quantifying the background signal, with and without room lights present. The difference in the signal is then quantifiable as the contribution from the ambient room lighting.

The level of tolerable crosstalk is very challenging to quantify, and tends to be system-specific, and the reason that this is so challenging to diagnose is because of how it can appear as background or noise level or detector saturation, in different settings. However, some systems can deal with considerable crosstalk if the users expects to see this present in the image. So, this category of effect is perhaps one of the most challenging to deal with.

Off-Target Fluorescence: Other sources of fluorescence may also contaminate detected signals. For imaging products involving multiple exogenous fluorophores with overlapping spectral characteristics, the impact of one fluorophore on the measurement of another should be characterized. Furthermore, spurious fluorescence excited in filters or other optical components can contribute to the detected signal. Autofluorescence from the tissue can be a factor in some systems, where the background signal is low, and simulating this in a test target or tissue phantom is challenging. However, it is possible to create tissue phantoms that have low fluorescence background signals that mimic autofluorescence signals of tissue, if critical to assessing system performance.

3.1.4 Additional Task-Specific Tests

In addition to the test methods described above, there are several techniques that are typically of secondary importance for fluorescence imaging system evaluation but may be highly significant for specific devices and/or applications. A brief description of each approach is provided below.

Geometric Measurement Accuracy: This is the mean error in estimation of the diameter and/or area of a fluorescent structure of known dimensions – is important for devices used to quantify the size of biological structures (e.g., for evaluation of tumor treatment). Phantom-based approaches have been described in prior medical imaging standards [121], and the performance of this in fluorescence mode can likely be different than in white light imaging mode and can be affected by the concentration of the probe and the environment in which it is measured.

Contrast-Detail Analysis: This is very commonly used in medical imaging to evaluate the effect of inclusion size and target fluorophore concentration on detectability, and has been used previously in fluorescence imaging systems [137]. The assessment by a single target with varying properties or an array of targets to assess the detectable level of contrast that is required for each given size of a region. This type of assessment provides a comprehensive assessment of both resolution limits and contrast detection when done properly, as these two features are defined by the limits of the system performance testing. Examples of use of this technique are most common in systems such as x-ray CT or MRI where contrast detection is one of the major use cases [139].

Concentration Measurement Accuracy: This assesses the ability of a device to provide quantitative measurements of fluorophore content. This is only relevant for quantitative imaging or measurement systems. Those systems that have this as a task feature must employ stricter calibration methods to achieve this, ideally through well calibrated test phantoms with quantitative set of fluorescent regions with known concentrations.

<u>Repeatability and Reproducibility:</u> The reliability of measurements is significantly impacted by device repeatability and reproducibility. To evaluate repeatability, performance test methods should be performed at least three times on different days (within a short interval of time), under similar defined measurement conditions. Results provide error bars that illustrate measurement precision. Reproducibility involves consistency of measurements performed under different conditions. Ideally, performance testing should be executed under conditions that include different locations, operators, and devices, where relevant.

3.2. PERFORMANCE TESTING PARADIGMS

Some of the basic motivations and behavioral choices in performance testing require a bit more detail, as described here.

<u>Calibration and Initial Fluorescence Measurement Validation</u>**:** The amount of fluorescent light of biological significance that makes its way from inside the tissue to the sensor generating an output signal is affected by biological, chemical, and physical factors. Thus, it is most universal or fundamental to convert the measured sensor digital counts into the desired physical quantity – in this case, the amount or concentration of fluorophore of interest in the tissue, especially the amount that sets the LoD

For routine calibration, it is adequate to calibrate the system in optical terms. The components of a fluorescence-guided imaging system which specifically generate the signal output and need to be calibrated are the sensor, for light responsivity at the specified spectral band, the excitation light source and the amount of fluorophore. The excitation light source incident on the sample plane can be measured with commercial optical meters set at irradiance mode [W/m^2]. The fluorophore concentration [M] at the measurement plane is typically reported. The sensor's responsivity or digital counts corresponding to the amount of fluorescence is determined through calibration with well-known concentrations in test phantoms. Again, the detected signal can be distorted by a range of issues such as tissue optical properties and depth into tissue, so these factors need to be mitigated in this measurement. Alternatively, some systems may utilize more complex algorithms to compensate for tissue turbidity, however these would require phantom validation for accuracy.

<u>Direct Measurement of a Biological Fluorophore Versus Use of Surrogate Fluorophores:</u> The imaging device is calibrated using a working range of concentrations of the fluorophore it is intended to detect using material preparations that are of *in vivo relevance*. The sensor digital counts are proportional to the fluorophore concentration. This system-level calibration, with the imager set at operational parameters, is a direct calibration route but may not be straight forward [124]. Material preparations with the fluorophore embedded in a matrix closely resembling the optical characteristics of *in vivo* measurements of fluorescence in tissue is likely important [84, 92, 98, 140-143]. The key problem with this approach is that most biological fluorophores are unstable in time and with respect to their environment, and so while measurement with say ICG would be desirable for a true test of a system, it would require preparation of the agent fresh for each test. While this is feasible, the likelihood of mistakes in such a labor intensive process is likely high, and so this is more common in research studies rather than in routine system performance evaluation.

Calibration of the imager using the fluorophore of interest directly may not be possible due to various constraints such as stability, cost or practical difficulty. Surrogate fluorophores such as quantum dots in phantom preparations have been successfully used as a convenient material for instrument characterization [84, 92], although laser dyes can also be used and have similar resistance to photobleaching. Sensor digital counts are proportional to the surrogate fluorophore concentration as long as the concentration is within a dilute linear range. Since the surrogate fluorophore concentration is of no measurement interest, the equivalency between the fluorescence emission level from the surrogate and the fluorophore of system interest within its in vivo environment, need to be established. The surrogate material preparations can then be used as quality control and quality assurance material working standard. This is analogous to fluorescent microspheres used by the flow cytometry community to standardize measurements of fluorescence in cellular samples [144]. In recent studies IR125 was found to be a reasonable surrogate for ICG, with similar absorption and emission spectra [95].

Reference Light Sources: The imager is calibrated against an electrical light source, that is emitting within the spectral band expected for the fluorophore of interest, at light levels matching the fluorescence output, from the working range of concentrations of the fluorophore [98, 140-143]. Sensor digital counts are proportional to light source output. As with the surrogate fluorophore, equivalency between the quasi-fluorescence light levels from the light source to the fluorescence levels from the fluorophore of interest in its biological environment should be established. This has the advantage that an electrical source is easy to operate, quantifiable and can be made SI-traceable using commercial optical power meters. It does not however, approximate properties of fluorescence emission from inside a tissue; the light source spectral band may be broader (eg. white light) or narrower (eg. laser line) which will affect the calibration of the imaging system, attention to this issue of the spectral band and center wavelength is critically important to establishing a good reference. Electrical light sources are typical checked by standards for many optical benchtop devices.

Routine or Initial Quality Assurance & System Parameters that can be Automatically Set/Saved: In order to summarize some of the above discussion, the key components and their most important measures are:

- *Excitation source:* intensity & uniformity measured at a specified distance from the imager
- *Fluorophore quantity:* via specified mass or volume calibration
- *Surrogate light source:* light level equivalency to fluorophore of interest
- *Power meters:* as specified by manufacturer (referencing a standard calibration, i.e. NIST)

Many system level parameters can or may be automatically stored in the image file as metadata or associated text file, automatically set or read from the system. Some of these include:

- Excitation irradiance/exposure
- Exposure time & camera timing settings
- f-stop or aperture size
- sensor output from a calibrant/reference object

3.3 CONSIDERATIONS FOR CLINICAL IMPLEMENTATION OF QUALITY PROCEDURES

3.3.1 The Role of Manufacturing Quality Control (QC)

In the medical device industry, the US Food and Drug Administration (FDA) sets requirements for quality systems[145]. While quality is often considered a subjective attribute that is perceived differently by different people, generally speaking, there are several types of quality that can be discussed. For the final customer, or user of the product, there is the idea of comparative quality of one product over another, which might include issues of form as well as function, convenience over performance, or cost over speed. Also included are quality features such as reliability, maintainability, and sustainability for customer satisfaction. It is feasible that measures of system performance realized by tissue phantoms would be a part of many stages of the quality procedures.

For the manufacturer, market analysis of customer quality perceptions and requirements is a vital part of determining exactly what product features are needed and what standard of performance quality is best incorporated into a final product. Conformance quality, describing the degree to which a device is correctly produced from the specifications, is the first quality consideration to be considered by the manufacturer. In the previous section, several fundamental performance characteristics such as image sharpness, depth of field, signal uniformity, field of view, distortion, and imaging depth for fluorescence guided surgical instruments were presented. There may be others, depending on the extended mission of the device. The degree to which these measurable attributes are required in a certain product must be obtained through careful market analysis at the outset. Once specified, the correct attainment of each product attribute is the goal of a conformance quality plan. Of course, this attainment must be reproduced in each model constructed, so the specifications must allow for a certain amount of acceptable variation about the nominal accepted value for each attribute while still providing acceptable performance for the customer. The ability to create repetitive instruments within the established error limits for all conformance attributes is the role of the conformance quality plan for the entire manufacturing process.

Satisfying conformance quality levels is necessary but not sufficient for achieving total quality in device function. A second quality, performance quality, is also needed. Here, specific performance characteristics such as those presented in the previous section are important. Sensitivity, minimum detectable concentration, linearity and dynamic rage, and detection sensitivity were discussed at length. Together, these device attributes determine whether the device will perform for the customer as required.

Consider the two types of quality as a hierarchy. The conformance attributes are needed to determine that the device performs according to the design specifications. Confirmation of this is called device verification. Validation, on the other hand, comes when the device is shown to perform the function for which it was constructed. This is the role of performance quality. It is entirely possible that a device can be verified through a number of quality steps, yet fail to be validated through performance quality testing. If this occurs, the original specifications are more than likely at fault, and a redesign of the device from the specifications and up is needed.

The quality activities for medical device manufacturing in the United States are regulated by 21 CFR 820, QUALITY SYSTEM REGULATION, by the FDA[145]. There are many quality system packages on the market today, covering a variety of industry requirements, but for medical device manufacturing, the International Organization for Standardization (ISO) Standard 13485 is closely harmonized with the requirements regulated by the FDA. Recently revised in 2016, this standard entered into a three-year transition period ending Feb 28, 2019. In addition, the European Parliament published new regulations for medical devices (MDR) and in-vitro diagnostics (IVDR) in May 2017 [146]. The MDR will take effect in 2020, and the new IVDR will begin in 2022.

The 21CFR 820 document discusses many aspects of a quality system, covering activities that would be a part of the conformance quality and performance quality characteristics. The extent and detail of each of these sections is beyond the scope of this article. While this section of the Code of Federal Regulations discusses the components that must be included in a quality system for medical device manufacturing, it does not specify exactly which quality system must be used. The manufacturer is free to use the ISO 13485 or any other method so long as it is commensurate with the items above and is in line with: (i) risks presented by the device; (ii) complexity of the device and the manufacturing process; (iii) extent of the activities to be carried out; and (iv) size and complexity of the manufacturer.

However, it is implemented in a manufacturing environment, a quality management plan must contain both quality control (QC) and quality assurance (QA). QC is that part of overall quality management that focuses on the activities that fulfill the requirements, while QA consists of those activities that provide confidence that the requirements have or will be fulfilled. Information regarding all activities associated with the design, construction, testing, analysis, and corrective actions involving a medical device can be requested by the FDA when market approval is sought by the manufacturer. Therefore, it is important to implement the quality management plan early in the device planning and design, and to carry it through to the end. This is a management burden that most academic institutions and research facilities are unwilling to bear, and manufacturing firms find acceptable only if the expected financial return is sufficiently high to warrant it.

3.3.2 Guidelines for Failure Mode and Effects Analysis (FMEA)

Failure Modes and Effects Analysis (FMEA) is a method to examine design and manufacturing processes to identify causes of potential device defects and suggest methods for corrective action, as well as provide logical methods for continuous quality improvement and use throughout the lifespan of the system. It should start as an early step in an overall product reliability study, and provide a flow chart for all use. FMEA is designed to identify potential failure modes of a device based on experience with other similar products or commonly understood engineering principles. There are two aspects to this analysis: the first is a projection of possible failure modes of the device, and the second is a probability analysis of the effects that the projected failures might have on device performance or customer acceptance. Good manufacturing practice (GMP) suggests that FMEA be performed at the system through to the subassembly or part level whenever possible. For surgical fluorescence devices, the system level would consist of the entire optical excitation and detection functions along with the display hardware and any software used to provide information to the surgeon. Subsystem components would include the excitation source, the detection equipment and display mode hardware, among others. At the assembly level, optical components such as completed lens configurations and beamsplitter assemblies are all relevant. Electronic assemblies that automate system performance, collect and display images, and record data are also assembly-level components. The individual lenses, filters, shutters, etc. are subassemblies or parts that would require failure mode study.

In the context of FMEA, the term "failure modes" represents loss of function of the system, subsystem, assembly, subassembly, or part under operating conditions. It does not mean the inability of the manufacturer to conform to the performance goals specified in the design. In fact, FMEA is intended to impact hardware design considerations. Therefore, a timely failure mode study should be performed before fabrication of the system is started. This process can help to specify certain components and subassemblies before construction begins. Functional analysis performed through careful experiments prior to construction can help to identify potential failure modes that might arise through choice of individual parts or through component integration. The process of FMEA ideally requires the analysis of all possible failure modes for each component and assembly of the final system, but because the analysis is best performed before final system construction, it is difficult to capture all possible pathways of failure. The creation of a product FMEA spreadsheet is an exercise that involves design, construction, field, and software engineers. The use of test targets or tissue phantoms used in tests to avoid failure mode is a very realizable possibility and this consideration is something that manufacturers should take into their design process.

3.3.3 Clinical Translation and Standardization

Clinical translation has come to mean the harnessing of knowledge from basic science to produce new devices, drugs and treatment options for patients. Former NIH Director Elias Zerhouni wrote [147, 148]: "*It is the responsibility of those involved in today's biomedical research enterprise to translate the remarkable scientific innovations we are witnessing into health gains for the nation."* There is clear motivation for translation from the laboratories of basic research to the domain of clinical care, and quality management is the vehicle by which the translation is made. Quality Management Systems (QMS) approaches include good laboratory practice (GLP) with its use of established standards and procedures for the design, performance, monitoring, and auditing of clinical trials or studies. At the design and development stage, GLP involves control of the manufacturing and verification processes to ensure the device, drug, or software meets all specifications. In the clinical environment, the protection of patients is critically important. Good clinical practice (GCP) regulations and standards are used to ensure excellence in clinical research, providing a standard for clinical conduct and analysis.

The International Council for Harmonization (ICH) is an organization created to achieve worldwide harmonization of the development and clinical validation of safe and effective clinical trials [149]. Good clinical practice (GCP) is founded on a program of good laboratory practice (GLP) for the creation and verification of devices and imaging agents (fluorophores) and includes

- An Internal Review Board (IRB)-approved protocol

- A valid informed consent form
- A data and safety monitoring plan
- Adverse Event (AE) reporting (device and drug)
- Proper device documentation
- Valid data collection, data storage and reporting procedures

The FDA insists that GCP be enforced in products, and a number of 21 CFR sections are relevant to appropriate GCP [150]. One important harmonization and standardization tool used in both the development phase and the clinical validation phase is an appropriate phantom target. The phantom takes the place of the targeted tissue to test the performance characteristic of the device. Risks in ignoring the use of phantoms in development and testing of fluorescence-guided surgical devices can be serious. Failures in the device operation, as suggested in the section above, can mislead the surgeon to thinking the tumor has been completely resected when, in fact, tumor remains at the margins. Other possible dangers might include improper light intensity on the tissue that could be dangerous to the patient. A full range of possible events could potentially be mitigated by the appropriate test procedures.

The use of phantoms before or during the surgical process serves as calibration to ensure proper performance of the device. A caution should be expressed, however. The use of phantoms implies that the phantoms themselves have been standardized. All aspects of usage and environmental conditions that can alter the optical characteristic of phantoms need to be accounted for, given that this could cause the operator to adjust the operating conditions of the fluorescence device, leading to a possible failure mode. Thus, the phantom and its use must be part of the FMEA design.

4. Recommendations for technical evaluation and guidance of new systems

4.1 RECOMMENDATIONS FOR TECHNICAL EVALUATION

Tissue-simulating phantoms should be used to test the pertinent task-specific performance characteristics of an FGS system. These should be designed to allow for ease of use and longitudinal comparisons of a single system and for comparisons of performance between different systems. Phantom longevity and robust performance are critical to make them useful rather than burdensome, which points to solid phantoms with a long stable life of use. This approach to testing should be considered as part of ongoing system QA needs, where the measurements are able to test features of the intended use.

A **minimum set of requirements** is as follows.

- Confirm system imaging performance or allow system self-calibration in terms of: image sharpness, depth of field, signal uniformity, distortion and field of view. These tests simply require stable test objects to image, not necessarily a phantom.
- Confirm task-specific performance, including quantitative assessment of signals in the intended wavelength range, and intended frame rate, for: signal sensitivity, linearity, dynamic range, depth sensitivity in a tissue-like medium, and effects of tissue scatter and absorption on the signal. These tests require phantoms that mimic the tissue optical properties, conditions and fluorescence.
- Assess confounding issues of light leakage through the optical filters, as related to limits of detection and performance under ambient lighting. These tests should ideally use tissue phantoms that mimic the fluorescence, reflectance and autofluorescence of the human tissue that will be imaged in the indicated use.
- Anthropomorphic phantoms should be used if the geometry of the biological tissue affects the observed signal interpretation or if user training in this geometry is critical. The type and composition of these phantoms would be ideally designed with optimal training and testing in mind.

Each type of measurements could be simple verifications and ideally they could be common across each class of imaging indications, to allow for inter-system comparison by the users and even sharing data across clinical centers. The most ideal situation is to have them integrated with software for automated calibration, electronic documentation in metadata. The frequency of testing is not specified here, and each manufacturer and user should consider the needs of this based upon the expected and tested variation in the values. The technology of FGS is rapidly evolving and the stability and repeatability has improved. Future consideration to specify the frequency of each test should be done, with specific requirements in regulated use perhaps.

4.2 RECOMMENDATION ON TECHINCAL GUIDANCE FOR NEW SYSTEMS

New systems qualified and supplied from the vendor should ideally include the test targets and phantoms needed for internal quality processes and as needed for routine audit by the user. FMEA processes can be utilized to establish guidelines that incorporate tissue phantoms of appropriate complexity to test for relevant failure modes and it would be ideal to include automated processes in the software to perform verification checks. Intersystem performance should be verified to some level of defined tolerance based on sensitivity, contrast and background suppression, thereby allowing use across vendor platforms, similar to the way CT, MRI and ultrasound are used now with interchangeability between vendors by the user.

4.3 RECOMMENDATION ON QUALIFIED PERSONNEL USING SYSTEMS

The most appropriate qualified personnel to I) use and to ii) measure performance are likely to be two separate individuals, although they could be the same person in certain systems where performance assessment does not overly impact the user's job function. However, in most cases qualified personnel to measure performance will be those with the technical expertise to recognize when a performance test is appropriate and if the data provided indicate acceptable function. The results of FMEA analysis can point to the needs and frequency for performance assessments and to the depth of technical knowledge needed for each system. Generally, the more serious the repercussion of mis-performance, the greater the need for testing to be performed by a trained technical expert. In most systems, fluorescence imaging tools require calibration and regular maintenance checking by the manufacturer or supplier. When used in the conjunction with a surgical procedure that relies upon the imaging performance, regular checks would be more frequent. If there is need for substantial physical insight or for frequent calibration at the user institution, then technically trained personnel onsite are likely required. In many cases this could be a bioengineering technician with specific training on the device. If the system requires standardization between centers, or interpretation of the imaging quality, then onsite trained personnel would also be required.

Certification requirements for qualified staff who are appropriately trained to use a particular clinical FGS system should be developed (e.g., observed 5 cases and performed 5 FGS services under supervision). Because of the diversity of systems and performance measures necessary, this is expected to be an evolving issue requiring interactions between the manufacturer, regulatory agencies and the user community.

5. Limitations of this report

This report is not intended to be a guidance document, but rather to provide scientific advice to developers, users and regulatory bodies who manage FGS systems. The implementation of these procedures is not intended to increase the financial or logistical burden of getting a product to market but rather, when implemented properly, should help optimize and simplify the quality-system approach and the qualification processes. Tissue phantoms are only one aspect of a whole quality system process and can directly address the intended use-testing of these systems. Current quality systems tend to focus much more on standard device issues such as electrical and optical performance checks and component function, whereas a well-designed phantom and set of tests can actually simplify the performance evaluation of these system issues as well.

Access to viable well-controlled manufactured phantoms remains an issue to be solved, both commercially and in terms of regulatory value to this advice.

Procedures, users, training and requirements are all things that need to be worked out, but at this preliminary stage of professional society guidance, it would not be appropriate to be too specific about these. Rather it should be expected that this will evolve as the field evolves and more clinical indications are developed or more multi-center trials are developed.

6. Summary

Fluorescence-guided surgery systems are being developed and used in a manner that is largely uncoordinated by any professional group, being driven rather by industrial and scientific opportunities in perfusion imaging and molecular medicine that influence surgical practice. The goals of this document are to outline key performance factors relevant to the intended clinical uses and to provide advice on calibrations and standards for optimal quality assurance processes. Ideally, tissue-simulating phantoms will be used to validate and calibrate the systems for pertinent task-specific goals and will be specified with a suitable longevity. They might ideally allow for use within a QMS system, possibly initial release testing, user training, and most importantly for ongoing QA for long term system performance. The minimum set of measurements considered important for basic device performance are: Image Sharpness, Depth of Field, Spatial Resolution, Signal Uniformity, Distortion and Field of View. The recommended task-specific performance measures for the real-time or video use of fluorescence signal are signal sensitivity, linearity, dynamic range, depth sensitivity, and scatter & absorption effects, each measured in the standard use case of the system. Confounding issues of ambient light leakage and filtering efficiency should be assessed as they relate the task-specific performance. Anthropomorphic phantoms should be considered if the geometry of the biological tissue affects the observed signal interpretation or if physician training in the tissue geometry is critical to proper use. The need for each of these measures should appear in a complete design with appropriate FMEA. Following this, new systems would ideally be qualified and supplied by the vendor with test targets with phantoms developed as part of their internal quality process or obtained from a validated vendor. Intersystem performance should be verified to some level of tolerance based upon sensitivity, contrast and background suppression. As the field progresses, some consideration should be put into identifying and training the appropriate qualified personnel to carry out on-site performance testing.

Disclosure Statement

The members of TG 311 listed attest that they have no potential Conflicts of Interest related to the subject matter or materials presented in this document.

7. Acknowledgements

None.

8. Conflicts of Interest

At the time of writing of the report, none of the authors had any conflict of interests to declare as related to this work. Subsequently to the completion of the manuscript, Prof Sylvain Gioux transitioned to Intuitive Surgical, which is a company that markets fluorescence guided robotic surgery.

9. References

1. Polom, K., D. Murawa, Y.S. Rho, P. Nowaczyk, M. Hunerbein, and P. Murawa, *Current trends and emerging future of indocyanine green usage in surgery and oncology: a literature review.* Cancer, 2011. **117**(21): p. 4812-22.
2. Kogon, B., J. Fernandez, K. Kanter, P. Kirshbom, B. Vincent, K. Maher, and N. Guzetta, *The role of intraoperative indocyanine green fluorescence angiography in pediatric cardiac surgery.* Ann Thorac Surg, 2009. **88**(2): p. 632-6.
3. Miyashiro, I., K. Kishi, M. Yano, K. Tanaka, M. Motoori, M. Ohue, H. Ohigashi, A. Takenaka, Y. Tomita, and O. Ishikawa, *Laparoscopic detection of sentinel node in gastric cancer surgery by indocyanine green fluorescence imaging.* Surg Endosc, 2011. **25**(5): p. 1672-6.
4. Tsuzuki, S., Y. Aihara, S. Eguchi, K. Amano, T. Kawamata, and Y. Okada, *Application of indocyanine green (ICG) fluorescence for endoscopic biopsy of intraventricular tumors.* Childs Nerv Syst, 2014. **30**(4): p. 723-6.
5. Daskalaki, D., E. Fernandes, X. Wang, F.M. Bianco, E.F. Elli, S. Ayloo, M. Masrur, L. Milone, and P.C. Giulianotti, *Indocyanine Green (ICG) Fluorescent Cholangiography During Robotic Cholecystectomy: Results of 184 Consecutive Cases in a Single Institution.* Surg Innov, 2014.
6. Imai, K., Y. Minamiya, H. Saito, T. Nakagawa, M. Ito, T. Ono, S. Motoyama, Y. Sato, H. Konno, and J. Ogawa, *Detection of pleural lymph flow using indocyanine green fluorescence imaging in non-small cell lung cancer surgery: a preliminary study.* Surg Today, 2013. **43**(3): p. 249-54.
7. Yoshida, M., K. Kubota, J. Kuroda, K. Ohta, T. Nakamura, J. Saito, M. Kobayashi, T. Sato, Y. Beck, Y. Kitagawa, and M. Kitajima, *Indocyanine green injection for detecting sentinel nodes using color fluorescence camera in the laparoscopy-assisted gastrectomy.* J Gastroenterol Hepatol, 2012. **27 Suppl 3**: p. 29-33.
8. Manny, T.B. and A.K. Hemal, *Fluorescence-enhanced robotic radical cystectomy using unconjugated indocyanine green for pelvic lymphangiography, tumor marking, and mesenteric angiography: the initial clinical experience.* Urology, 2014. **83**(4): p. 824-9.

9. Pessaux, P., M. Diana, L. Soler, T. Piardi, D. Mutter, and J. Marescaux, *Robotic duodenopancreatectomy assisted with augmented reality and real-time fluorescence guidance.* Surg Endosc, 2014. **28**(8): p. 2493-8.
10. Bjurlin, M.A., M. Gan, T.R. McClintock, A. Volpe, M.S. Borofsky, A. Mottrie, and M.D. Stifelman, *Near-infrared fluorescence imaging: emerging applications in robotic upper urinary tract surgery.* Eur Urol, 2014. **65**(4): p. 793-801.
11. Pogue, B.W. and E.L. Rosenthal, *Review of successful pathways for regulatory approvals in open-field fluorescence-guided surgery.* J Biomed Opt, 2021. **26**(3).
12. Lauwerends, L.J., P. van Driel, R.J. Baatenburg de Jong, J.A.U. Hardillo, S. Koljenovic, G. Puppels, L. Mezzanotte, C. Lowik, E.L. Rosenthal, A.L. Vahrmeijer, and S. Keereweer, *Real-time fluorescence imaging in intraoperative decision making for cancer surgery.* Lancet Oncol, 2021. **22**(5): p. e186-e195.
13. Alander, J.T., I. Kaartinen, A. Laakso, T. Patila, T. Spillmann, V.V. Tuchin, M. Venermo, and P. Valisuo, *A review of indocyanine green fluorescent imaging in surgery.* Int J Biomed Imaging, 2012. **2012**: p. 940585.
14. van Dam, G.M., G. Themelis, L.M. Crane, N.J. Harlaar, R.G. Pleijhuis, W. Kelder, A. Sarantopoulos, J.S. de Jong, H.J. Arts, A.G. van der Zee, J. Bart, P.S. Low, and V. Ntziachristos, *Intraoperative tumor-specific fluorescence imaging in ovarian cancer by folate receptor-alpha targeting: first in-human results.* Nat Med, 2011. **17**(10): p. 1315-9.
15. Ito, A., Y. Ito, S. Matsushima, D. Tsuchida, M. Ogasawara, J. Hasegawa, K. Misawa, E. Kondo, N. Kaneda, and H. Nakanishi, *New whole-body multimodality imaging of gastric cancer peritoneal metastasis combining fluorescence imaging with ICG-labeled antibody and MRI in mice.* Gastric Cancer, 2014. **17**(3): p. 497-507.
16. Snoeks, T.J., P.B. van Driel, S. Keereweer, S. Aime, K.M. Brindle, G.M. van Dam, C.W. Lowik, V. Ntziachristos, and A.L. Vahrmeijer, *Towards a successful clinical implementation of fluorescence-guided surgery.* Mol Imaging Biol, 2014. **16**(2): p. 147-51.
17. Sevick-Muraca, E.M., *Translation of near-infrared fluorescence imaging technologies: emerging clinical applications.* Annu Rev Med, 2012. **63**: p. 217-31.
18. Chen, K., L.P. Yap, R. Park, X. Hui, K. Wu, D. Fan, X. Chen, and P.S. Conti, *A Cy5.5-labeled phage-displayed peptide probe for near-infrared fluorescence imaging of tumor vasculature in living mice.* Amino Acids, 2012. **42**(4): p. 1329-37.
19. Nakajima, T., M. Mitsunaga, N.H. Bander, W.D. Heston, P.L. Choyke, and H. Kobayashi, *Targeted, activatable, in vivo fluorescence imaging of prostate-specific membrane antigen (PSMA) positive tumors using the quenched humanized J591 antibody-indocyanine green (ICG) conjugate.* Bioconjug Chem, 2011. **22**(8): p. 1700-5.
20. Pogue, B.W., K.S. Samkoe, S. Hextrum, J.A. O'Hara, M. Jermyn, S. Srinivasan, and T. Hasan, *Imaging targeted-agent binding in vivo with two probes.* J Biomed Opt, 2010. **15**(3): p. 030513.
21. Tichauer, K.M., K.S. Samkoe, J.R. Gunn, S.C. Kanick, P.J. Hoopes, R.J. Barth, P.A. Kaufman, T. Hasan, and B.W. Pogue, *Microscopic lymph node tumor burden quantified by macroscopic dual-tracer molecular imaging.* Nat Med, 2014.
22. Pogue, B.W., T.C. Zhu, V. Ntziachristos, K.D. Paulsen, B.C. Wilson, J. Pfefer, R.J. Nordstrom, M. Litorja, H. Wabnitz, Y. Chen, S. Gioux, B.J. Tromberg, and A.G. Yodh, *Fluorescence-guided surgery and intervention - An AAPM emerging technology blue paper.* Med Phys, 2018. **45**(6): p. 2681-2688.
23. Jacques, S.L., *Optical properties of biological tissues: a review.* Phys Med Biol, 2013. **58**(11): p. R37-61.

24. Alexandrakis, G., T.J. Farrell, and M.S. Patterson, *Accuracy of the diffusion approximation in determining the optical properties of a two-layer turbid medium.* Appl Opt, 1998. **37**(31): p. 7401-9.
25. Star, W.M., *Light dosimetry in vivo.* Phys Med Biol, 1997. **42**(5): p. 763-87.
26. Karagiannes, J.L., Z. Zhang, B. Grossweiner, and L.I. Grossweiner, *Applications of the 1-D diffusion approximation to the optics of tissues and tissue phantoms.* Appl Opt, 1989. **28**(12): p. 2311-7.
27. Cuccia, D.J., F. Bevilacqua, A.J. Durkin, F.R. Ayers, and B.J. Tromberg, *Quantitation and mapping of tissue optical properties using modulated imaging.* J Biomed Opt, 2009. **14**(2): p. 024012.
28. Doornbos, R.M., R. Lang, M.C. Aalders, F.W. Cross, and H.J. Sterenborg, *The determination of in vivo human tissue optical properties and absolute chromophore concentrations using spatially resolved steady-state diffuse reflectance spectroscopy.* Phys Med Biol, 1999. **44**(4): p. 967-81.
29. Farrell, T.J., M.S. Patterson, and M. Essenpreis, *Influence of layered tissue architecture on estimates of tissue optical properties obtained from spatially resolved diffuse reflectometry.* Appl Opt, 1998. **37**(10): p. 1958-72.
30. Tromberg, B.J., O. Coquoz, J.B. Fishkin, T. Pham, E.R. Anderson, J. Butler, M. Cahn, J.D. Gross, V. Venugopalan, and D. Pham, *Non-invasive measurements of breast tissue optical properties using frequency-domain photon migration.* Philos Trans R Soc Lond B Biol Sci, 1997. **352**(1354): p. 661-8.
31. Tran, A.P. and S. Jacques, *Modeling voxel-based Monte Carlo light transport with curved and oblique boundary surfaces.* J Biomed Opt, 2020. **25**(2): p. 1-13.
32. Ntziachristos, V., G. Turner, J. Dunham, S. Windsor, A. Soubret, J. Ripoll, and H.A. Shih, *Planar fluorescence imaging using normalized data.* J Biomed Opt, 2005. **10**(6): p. 064007.
33. Mohajerani, P. and V. Ntziachristos, *Compression of Born ratio for fluorescence molecular tomography/x-ray computed tomography hybrid imaging: methodology and in vivo validation.* Opt Lett, 2013. **38**(13): p. 2324-6.
34. Vinegoni, C., D. Razansky, J.L. Figueiredo, M. Nahrendorf, V. Ntziachristos, and R. Weissleder, *Normalized Born ratio for fluorescence optical projection tomography.* Opt Lett, 2009. **34**(3): p. 319-21.
35. Soubret, A., J. Ripoll, and V. Ntziachristos, *Accuracy of fluorescent tomography in the presence of heterogeneities: study of the normalized Born ratio.* IEEE Trans Med Imaging, 2005. **24**(10): p. 1377-86.
36. Graves, E.E., D. Yessayan, G. Turner, R. Weissleder, and V. Ntziachristos, *Validation of in vivo fluorochrome concentrations measured using fluorescence molecular tomography.* J Biomed Opt, 2005. **10**(4): p. 44019.
37. Ntziachristos, V., *Fluorescence molecular imaging.* Annu Rev Biomed Eng, 2006. **8**: p. 1-33.
38. Richards-Kortum, R. and E. Sevick-Muraca, *Quantitative optical spectroscopy for tissue diagnosis.* Annu Rev Phys Chem, 1996. **47**: p. 555-606.
39. Pogue, B.W., E.L. Rosenthal, S. Achilefu, and G.M. van Dam, *Perspective review of what is needed for molecular-specific fluorescence-guided surgery.* J Biomed Opt, 2018. **23**(10): p. 1-9.
40. Hadjipanayis, C.G. and W. Stummer, *5-ALA and FDA approval for glioma surgery.* J Neurooncol, 2019. **141**(3): p. 479-486.
41. Francois, A., S. Battah, A.J. MacRobert, L. Bezdetnaya, F. Guillemin, and M.A. D'Hallewin, *Fluorescence diagnosis of bladder cancer: a novel in vivo approach using 5-aminolevulinic acid (ALA) dendrimers.* BJU Int, 2012. **110**(11 Pt C): p. E1155-62.
42. Krammer, B. and K. Plaetzer, *ALA and its clinical impact, from bench to bedside.* Photochem Photobiol Sci, 2008. **7**(3): p. 283-9.

43. Patel, G., A.W. Armstrong, and D.B. Eisen, *Efficacy of photodynamic therapy vs other interventions in randomized clinical trials for the treatment of actinic keratoses: a systematic review and meta-analysis.* JAMA Dermatol, 2014. **150**(12): p. 1281-8.
44. McCaughan, J.S., Jr., *Photodynamic therapy: a review.* Drugs Aging, 1999. **15**(1): p. 49-68.
45. Celli, J.P., B.Q. Spring, I. Rizvi, C.L. Evans, K.S. Samkoe, S. Verma, B.W. Pogue, and T. Hasan, *Imaging and photodynamic therapy: mechanisms, monitoring, and optimization.* Chem Rev, 2010. **110**(5): p. 2795-838.
46. Zhu, B., J.C. Rasmussen, Y. Lu, and E.M. Sevick-Muraca, *Reduction of excitation light leakage to improve near-infrared fluorescence imaging for tissue surface and deep tissue imaging.* Med Phys, 2010. **37**(11): p. 5961-70.
47. Hwang, K., J.P. Houston, J.C. Rasmussen, A. Joshi, S. Ke, C. Li, and E.M. Sevick-Muraca, *Improved excitation light rejection enhances small-animal fluorescent optical imaging.* Mol Imaging, 2005. **4**(3): p. 194-204.
48. Angelo, J., V. Venugopal, F. Fantoni, V. Poher, I.J. Bigio, L. Herve, J.M. Dinten, and S. Gioux, *Depth-enhanced fluorescence imaging using masked detection of structured illumination.* J Biomed Opt, 2014. **19**(11): p. 116008.
49. Mazhar, A., D.J. Cuccia, S. Gioux, A.J. Durkin, J.V. Frangioni, and B.J. Tromberg, *Structured illumination enhances resolution and contrast in thick tissue fluorescence imaging.* J Biomed Opt, 2010. **15**(1): p. 010506.
50. Gioux, S., V. Kianzad, R. Ciocan, S. Gupta, R. Oketokoun, and J.V. Frangioni, *High-power, computer-controlled, light-emitting diode-based light sources for fluorescence imaging and image-guided surgery.* Mol Imaging, 2009. **8**(3): p. 156-65.
51. Sexton, K.J., Y. Zhao, S.C. Davis, S. Jiang, and B.W. Pogue, *Optimization of fluorescent imaging in the operating room through pulsed acquisition and gating to ambient background cycling.* Biomed Opt Express, 2017. **8**(5): p. 2635-2648.
52. Tichauer, K.M., M. Migueis, F. Leblond, J.T. Elliott, M. Diop, K. St Lawrence, and T.Y. Lee, *Depth resolution and multiexponential lifetime analyses of reflectance-based time-domain fluorescence data.* Appl Opt, 2011. **50**(21): p. 3962-72.
53. Yamamoto, S., P. Kim, R. Kurokawa, K. Itoki, and S. Kawamoto, *Selective intraarterial injection of ICG for fluorescence angiography as a guide to extirpate perimedullary arteriovenous fistulas.* Acta Neurochir (Wien), 2012. **154**(3): p. 457-63.
54. Boni, L., G. David, A. Mangano, G. Dionigi, S. Rausei, S. Spampatti, E. Cassinotti, and A. Fingerhut, *Clinical applications of indocyanine green (ICG) enhanced fluorescence in laparoscopic surgery.* Surg Endosc, 2014.
55. Tobis, S., J.K. Knopf, C.R. Silvers, J. Marshall, A. Cardin, R.W. Wood, J.E. Reeder, E. Erturk, R. Madeb, J. Yao, E.A. Singer, H. Rashid, G. Wu, E. Messing, and D. Golijanin, *Near infrared fluorescence imaging after intravenous indocyanine green: initial clinical experience with open partial nephrectomy for renal cortical tumors.* Urology, 2012. **79**(4): p. 958-64.
56. Fan, W., B. Shen, W. Bu, X. Zheng, Q. He, Z. Cui, D. Ni, K. Zhao, S. Zhang, and J. Shi, *Intranuclear biophotonics by smart design of nuclear-targeting photo-/radio-sensitizers co-loaded upconversion nanoparticles.* Biomaterials, 2015. **69**: p. 89-98.
57. Venugopal, V., M. Park, Y. Ashitate, F. Neacsu, F. Kettenring, J.V. Frangioni, S.P. Gangadharan, and S. Gioux, *Design and characterization of an optimized simultaneous color and near-infrared fluorescence rigid endoscopic imaging system.* J Biomed Opt, 2013. **18**(12): p. 126018.
58. Lu, Y., B. Zhu, C. Darne, I.C. Tan, J.C. Rasmussen, and E.M. Sevick-Muraca, *Improvement of fluorescence-enhanced optical tomography with improved optical filtering and accurate model-based reconstruction algorithms.* J Biomed Opt, 2011. **16**(12): p. 126002.

59. Soubret, A. and V. Ntziachristos, *Fluorescence molecular tomography in the presence of background fluorescence.* Phys Med Biol, 2006. **51**(16): p. 3983-4001.
60. Gao, M., G. Lewis, G.M. Turner, A. Soubret, and V. Ntziachristos, *Effects of background fluorescence in fluorescence molecular tomography.* Appl Opt, 2005. **44**(26): p. 5468-74.
61. Sandison, D.R. and W.W. Webb, *Background Rejection and Signal-to-Noise Optimization in Confocal and Alternative Fluorescence Microscopes.* Applied Optics, 1994. **33**(4): p. 603-615.
62. Frangioni, J.V., *The problem is background, not signal.* Mol Imaging, 2009. **8**(6): p. 303-4.
63. Frangioni, J.V., *In vivo near-infrared fluorescence imaging.* Curr Opin Chem Biol, 2003. **7**(5): p. 626-34.
64. Gurfinkel, M., T. Pan, and E.M. Sevick-Muraca, *Determination of optical properties in semi-infinite turbid media using imaging measurements of frequency-domain photon migration obtained with an intensified charge-coupled device.* J Biomed Opt, 2004. **9**(6): p. 1336-46.
65. Alexander, D., P. Bruza, C. Farwell, V. Krishnaswamy, D.J. Gladstone, and B.W. Pogue, *Detective quantum efficiency of intensified CMOS cameras for use in Cherenkov and scintillation-based radiotherapy dosimetry.* Med Phys, 2020: p. (in review).
66. Becker, W., A. Bergmann, M.A. Hink, K. Konig, K. Benndorf, and C. Biskup, *Fluorescence lifetime imaging by time-correlated single-photon counting.* Microsc Res Tech, 2004. **63**(1): p. 58-66.
67. Erkkila, M.T., B. Bauer, N. Hecker-Denschlag, M.J. Madera Medina, R.A. Leitgeb, A. Unterhuber, J. Gesperger, T. Roetzer, C. Hauger, W. Drexler, G. Widhalm, and M. Andreana, *Widefield fluorescence lifetime imaging of protoporphyrin IX for fluorescence-guided neurosurgery: An ex vivo feasibility study.* J Biophotonics, 2019. **12**(6): p. e201800378.
68. Nakajima, T., K. Sano, K. Sato, R. Watanabe, T. Harada, H. Hanaoka, P.L. Choyke, and H. Kobayashi, *Fluorescence-lifetime molecular imaging can detect invisible peritoneal ovarian tumors in bloody ascites.* Cancer Sci, 2014. **105**(3): p. 308-14.
69. DSouza, A., H. Lin, J. Gunn, and B.W. Pogue, *Logarithmic intensity compression in fluorescence guided surgery applications.* J Biomed Opt, 2015. **20**(8): p. 80504.
70. Pogue, B.W. and M.S. Patterson, *Review of tissue simulating phantoms for optical spectroscopy, imaging and dosimetry.* J Biomed Opt, 2006. **11**(4): p. 041102.
71. Pogue, B.W., S.C. Davis, X. Song, B.A. Brooksby, H. Dehghani, and K.D. Paulsen, *Image analysis methods for diffuse optical tomography.* J Biomed Opt, 2006. **11**(3): p. 33001.
72. Ntziachristos, V., B. Chance, and A. Yodh, *Differential diffuse optical tomography.* Opt Express, 1999. **5**(10): p. 230-42.
73. Di Ninni, P., F. Martelli, and G. Zaccanti, *Intralipid: towards a diffusive reference standard for optical tissue phantoms.* Physics in Medicine and Biology, 2011. **56**(2): p. N21-N28.
74. van Staveren, H.J., C.J.M. Moes, J. van Marie, S.A. Prahl, and M.J.C. van Gemert, *Light scattering in Intralipid-10\% in the wavelength range of 400--1100 nm.* Appl. Opt, 1991. **30**(31): p. 4507--4514.
75. Flock, S.T., S.L. Jacques, B.C. Wilson, W.M. Star, and M.J.C. Vangemert, *Optical-Properties of Intralipid - a Phantom Medium for Light-Propagation Studies.* Lasers in Surgery and Medicine, 1992. **12**(5): p. 510-519.
76. Kienle, A., L. Lilge, M.S. Patterson, R. Hibst, R. Steiner, and B.C. Wilson, *Spatially resolved absolute diffuse reflectance measurements for noninvasive determination of the optical scattering and absorption coefficients of biological tissue.* Appl Opt, 1996. **35**(13): p. 2304-14.
77. Spinelli, L., M. Botwicz, N. Zolek, M. Kacprzak, D. Milej, P. Sawosz, A. Liebert, U. Weigel, T. Durduran, F. Foschum, A. Kienle, F. Baribeau, S. Leclair, J.P. Bouchard, I. Noiseux, P. Gallant, O. Mermut, A. Farina, A. Pifferi, A. Torricelli, R. Cubeddu, H.C. Ho, M. Mazurenka, H. Wabnitz, K. Klauenberg, O. Bodnar, C. Elster, M. Benazech-Lavoue, Y. Berube-Lauziere, F. Lesage, D. Khoptyar, A.A. Subash, S. Andersson-Engels, P. Di Ninni, F. Martelli, and G. Zaccanti,

Determination of reference values for optical properties of liquid phantoms based on Intralipid and India ink. Biomed Opt Express, 2014. **5**(7): p. 2037-53.

78. Jiang, S., B.W. Pogue, A.M. Laughney, C.A. Kogel, and K.D. Paulsen, *Measurement of pressure-displacement kinetics of hemoglobin in normal breast tissue with near-infrared spectral imaging.* Appl Opt, 2009. **48**(10): p. D130-6.
79. Ren, W., Q. Gan, Q. Wu, S. Zhang, and R. Xu, *Quasi-simultaneous multimodal imaging of cutaneous tissue oxygenation and perfusion.* J Biomed Opt, 2015. **20**(12): p. 121307.
80. Liu, K., Y. Wang, X. Kong, X. Liu, Y. Zhang, L. Tu, Y. Ding, M.C. Aalders, W.J. Buma, and H. Zhang, *Multispectral upconversion luminescence intensity ratios for ascertaining the tissue imaging depth.* Nanoscale, 2014. **6**(15): p. 9257-63.
81. Sukowski, U., F. Schubert, D. Grosenick, and H. Rinneberg, *Preparation of solid phantoms with defined scattering and absorption properties for optical tomography.* Phys Med Biol, 1996. **41**(9): p. 1823-44.
82. Firbank, M. and D.T. Delpy, *A phantom for the testing and calibration of near infra-red spectrometers.* Phys Med Biol, 1994. **39**(9): p. 1509-13.
83. Firbank, M., M. Oda, and D.T. Delpy, *An improved design for a stable and reproducible phantom material for use in near-infrared spectroscopy and imaging.* Phys Med Biol, 1995. **40**(5): p. 955-61.
84. Zhu, B., I.C. Tan, J.C. Rasmussen, and E.M. Sevick-Muraca, *Validating the sensitivity and performance of near-infrared fluorescence imaging and tomography devices using a novel solid phantom and measurement approach.* Technol Cancer Res Treat, 2012. **11**(1): p. 95-104.
85. Hebden, J.C., B.D. Price, A.P. Gibson, and G. Royle, *A soft deformable tissue-equivalent phantom for diffuse optical tomography.* Phys Med Biol, 2006. **51**(21): p. 5581-90.
86. Saager, R.B., C. Kondru, K. Au, K. Sry, F. Ayers, and A.J. Durkin, *Multi-layer silicone phantoms for the evaluation of quantitative optical techniques in skin imaging.* Design and Performance Validation of Phantoms Used in Conjunction with Optical Measurement of Tissue Ii, 2010. **7567**.
87. Ayers, F., A. Grant, D. Kuo, D.J. Cuccia, and A.J. Durkin, *Fabrication and characterization of silicone-based tissue phantoms with tunable optical properties in the visible and near infrared domain - art. no. 687007.* Design and Performance Validation of Phantoms Used in Conjunction with Optical Measurements of Tissue, 2008. **6870**: p. 87007-87007.
88. Krauter, P., S. Nothelfer, N. Bodenschatz, E. Simon, S. Stocker, F. Foschum, and A. Kienle, *Optical phantoms with adjustable subdiffusive scattering parameters.* Journal of Biomedical Optics, 2015. **20**(10).
89. Culver, J.P., R. Choe, M.J. Holboke, L. Zubkov, T. Durduran, A. Slemp, V. Ntziachristos, B. Chance, and A.G. Yodh, *Three-dimensional diffuse optical tomography in the parallel plane transmission geometry: evaluation of a hybrid frequency domain/continuous wave clinical system for breast imaging.* Med Phys, 2003. **30**(2): p. 235-47.
90. Schulz, R.B., J. Ripoll, and V. Ntziachristos, *Noncontact optical tomography of turbid media.* Opt Lett, 2003. **28**(18): p. 1701-3.
91. Zacharakis, G., H. Shih, J. Ripoll, R. Weissleder, and V. Ntziachristos, *Normalized transillumination of fluorescent proteins in small animals.* Mol Imaging, 2006. **5**(3): p. 153-9.
92. Anastasopoulou, M., M. Koch, D. Gorpas, A. Karlas, U. Klemm, P.B. Garcia-Allende, and V. Ntziachristos, *Comprehensive phantom for interventional fluorescence molecular imaging.* J Biomed Opt, 2016. **21**(9): p. 091309.
93. Gorpas, D., M. Koch, M. Anastasopoulou, U. Klemm, and V. Ntziachristos, *Benchmarking of fluorescence cameras through the use of a composite phantom.* J Biomed Opt, 2017. **22**(1): p. 16009.

94. Pleijhuis, R., A. Timmermans, J. De Jong, E. De Boer, V. Ntziachristos, and G. Van Dam, *Tissue-simulating phantoms for assessing potential near-infrared fluorescence imaging applications in breast cancer surgery.* J Vis Exp, 2014(91): p. 51776.
95. Ruiz, A.J., M. Wu, E.P.M. LaRochelle, D. Gorpas, V. Ntziachristos, T.J. Pfefer, and B.W. Pogue, *Indocyanine-green matching phantom for fluorescence guided surgery imaging system characterization and performance assessment.* J. Biomed. Opt., 2020: p. (in press).
96. Firbank, M., M. Hiraoka, M. Essenpreis, and D.T. Delpy, *Measurement of the optical properties of the skull in the wavelength range 650-950 nm.* Phys Med Biol, 1993. **38**(4): p. 503-10.
97. Diamond, K.R., M.S. Patterson, and T.J. Farrell, *Quantification of fluorophore concentration in tissue-simulating media by fluorescence measurements with a single optical fiber.* Appl Opt, 2003. **42**(13): p. 2436-42.
98. Diamond, K.R., T.J. Farrell, and M.S. Patterson, *Measurement of fluorophore concentrations and fluorescence quantum yield in tissue-simulating phantoms using three diffusion models of steady-state spatially resolved fluorescence.* Phys Med Biol, 2003. **48**(24): p. 4135-49.
99. Pfefer, T.J., L.S. Matchette, C.L. Bennett, J.A. Gall, J.N. Wilke, A.J. Durkin, and M.N. Ediger, *Reflectance-based determination of optical properties in highly attenuating tissue.* J Biomed Opt, 2003. **8**(2): p. 206-15.
100. Stasic, D., T.J. Farrell, and M.S. Patterson, *The use of spatially resolved fluorescence and reflectance to determine interface depth in layered fluorophore distributions.* Phys Med Biol, 2003. **48**(21): p. 3459-74.
101. Roy, H.K., M.J. Goldberg, S. Bajaj, and V. Backman, *Colonoscopy and optical biopsy: bridging technological advances to clinical practice.* Gastroenterology, 2011. **140**(7): p. 1863-7.
102. Leh, B., R. Siebert, H. Hamzeh, L. Menard, M.A. Duval, Y. Charon, and D.A. Haidar, *Optical phantoms with variable properties and geometries for diffuse and fluorescence optical spectroscopy.* Journal of Biomedical Optics, 2012. **17**(10).
103. Zhu, H., S.O. Isikman, O. Mudanyali, A. Greenbaum, and A. Ozcan, *Optical imaging techniques for point-of-care diagnostics.* Lab Chip, 2013. **13**(1): p. 51-67.
104. Szyc, L., S. Bonifer, A. Walter, U. Jagemann, D. Grosenick, and R. Macdonald, *Development of a handheld fluorescence imaging camera for intraoperative sentinel lymph node mapping.* J Biomed Opt, 2015. **20**(5): p. 051025.
105. Pleijhuis, R.G., G.C. Langhout, W. Helfrich, G. Themelis, A. Sarantopoulos, L.M. Crane, N.J. Harlaar, J.S. de Jong, V. Ntziachristos, and G.M. van Dam, *Near-infrared fluorescence (NIRF) imaging in breast-conserving surgery: assessing intraoperative techniques in tissue-simulating breast phantoms.* Eur J Surg Oncol, 2011. **37**(1): p. 32-9.
106. Wang, J., J. Coburn, C.P. Liang, N. Woolsey, J.C. Ramella-Roman, Y. Chen, and T.J. Pfefer, *Three-dimensional printing of tissue phantoms for biophotonic imaging.* Opt Lett, 2014. **39**(10): p. 3010-3.
107. Bentz, B.Z., A.G. Bowen, D. Lin, D. Ysselstein, D.H. Huston, J.C. Rochet, and K.J. Webb, *Printed optics: phantoms for quantitative deep tissue fluorescence imaging.* Opt Lett, 2016. **41**(5230-33).
108. Baeten, J., M. Niedre, J. Dunham, and V. Ntziachristos, *Development of fluorescent materials for Diffuse Fluorescence Tomography standards and phantoms.* Opt Express, 2007. **15**(14): p. 8681-94.
109. Star, W.M., J.P.A. Marijnissen, and M.J.C. Vangemert, *Light Dosimetry in Optical Phantoms and in Tissues .1. Multiple Flux and Transport-Theory.* Physics in Medicine and Biology, 1988. **33**(4): p. 437-454.
110. Patterson, M.S., B. Chance, and B.C. Wilson, *Time resolved reflectance and transmittance for the non-invasive measurement of tissue optical properties.* Appl Opt, 1989. **28**(12): p. 2331-6.

111. Grosenick, D., H. Wabnitz, and H. Rinneberg, *Time-resolved imaging of solid phantoms for optical mammography.* Applied Optics, 1997. **36**(1): p. 221-231.
112. Madsen, S.J., E.R. Anderson, R.C. Haskell, and B.J. Tromberg, *Portable, high-bandwidth frequency-domain photon migration instrument for tissue spectroscopy.* Opt Lett, 1994. **19**(23): p. 1934-6.
113. Corlu, A., T. Durduran, R. Choe, M. Schweiger, E.M. Hillman, S.R. Arridge, and A.G. Yodh, *Uniqueness and wavelength optimization in continuous-wave multispectral diffuse optical tomography.* Opt Lett, 2003. **28**(23): p. 2339-41.
114. Sevick-Muraca, E.M., G. Lopez, J.S. Reynolds, T.L. Troy, and C.L. Hutchinson, *Fluorescence and absorption contrast mechanisms for biomedical optical imaging using frequency-domain techniques.* Photochem Photobiol, 1997. **66**(1): p. 55-64.
115. Arridge, S.R., M. Cope, and D.T. Delpy, *The theoretical basis for the determination of optical pathlengths in tissue: temporal and frequency analysis.* Phys Med Biol, 1992. **37**(7): p. 1531-60.
116. Pogue, B.W. and M.S. Patterson, *Frequency-domain optical absorption spectroscopy of finite tissue volumes using diffusion theory.* Phys Med Biol, 1994. **39**(7): p. 1157-80.
117. Sassaroli, A., F. Martelli, D. Imai, and Y. Yamada, *Study on the propagation of ultra-short pulse light in cylindrical optical phantoms.* Physics in Medicine and Biology, 1999. **44**(11): p. 2747-2763.
118. Gibson, A., R.M. Yusof, H. Dehghani, J. Riley, N. Everdell, R. Richards, J.C. Hebden, M. Schweiger, S.R. Arridge, and D.T. Delpy, *Optical tomography of a realistic neonatal head phantom.* Appl Opt, 2003. **42**(16): p. 3109-16.
119. Pogue, B.W. and M.S. Patterson, *Error assessment of a wavelength tunable frequency domain system for noninvasive tissue spectroscopy.* J Biomed Opt, 1996. **1**(3): p. 311-323.
120. Samei, E., A. Rowberg, E. Avraham, and C. Cornelius, *Toward clinically relevant standardization of image quality.* J Digit Imaging, 2004. **17**(4): p. 271-8.
121. Goodsitt, M.M., P.L. Carson, S. Witt, D.L. Hykes, and J.M. Kofler, Jr., *Real-time B-mode ultrasound quality control test procedures. Report of AAPM Ultrasound Task Group No. 1.* Med Phys, 1998. **25**(8): p. 1385-406.
122. McCollough, C.H., M.R. Bruesewitz, M.F. McNitt-Gray, K. Bush, T. Ruckdeschel, J.T. Payne, J.A. Brink, R.K. Zeman, and R. American College of, *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, 2004. **31**(9): p. 2423-42.
123. Pfefer, J. and A. Agrawal, *A review of consensus test methods for established medical imaging modalities and their implications for optical coherence tomography.* Proc. SPIE BiOS, 2012. **8215**((28 February 2012);): p. 82150D.
124. DSouza, A.V., H. Lin, E.R. Henderson, K.S. Samkoe, and B.W. Pogue, *Review of fluorescence guided surgery systems: identification of key performance capabilities beyond indocyanine green imaging.* J Biomed Opt, 2016. **21**(8): p. 80901.
125. Zhu, B., J.C. Rasmussen, M. Litorja, and E.M. Sevick-Muraca, *Determining the Performance of Fluorescence Molecular Imaging Devices Using Traceable Working Standards With SI Units of Radiance.* IEEE Trans Med Imaging, 2016. **35**(3): p. 802-11.
126. Gorpas, D., M. Koch, M. Anastasopoulou, D. Bozhko, U. Klemm, M. Nieberler, and V. Ntziachristos, *Multi-Parametric Standardization of Fluorescence Imaging Systems Based on a Composite Phantom.* IEEE Trans Biomed Eng, 2020. **67**(1): p. 185-192.
127. Kanniyappan, U., B. Wang, C. Yang, P. Ghassemi, M. Litorja, N. Suresh, Q. Wang, Y. Chen, and T.J. Pfefer, *Performance test methods for near-infrared fluorescence imaging.* Medical physics, 2020: p. (in press).

128. *Standard: ISO 12233 — Photography— Electronic still picture imaging — Resolution and spatial frequency responses*. 2017, International Standards Organization / Technical 42 Photography.
129. Kittle D S, M., md A, Parrish-novak J E, et al., *Fluorescence-Guided Tumor Visualization Using the Tumor Paint BLZ-100.* Cureus, 2014. **6**(9): p. e210.
130. Sellors, J.W., J.L. Winkler, and D.F. Kreysar, *Illumination, optics, and clinical performance of a hand-held magnified visual inspection device (AviScope): a comparison with colposcopy.* J Acquir Immune Defic Syndr, 2004. **37 Suppl 3**: p. S160-6.
131. Yang, V.X., P.J. Muller, P. Herman, and B.C. Wilson, *A multispectral fluorescence imaging system: design and initial clinical tests in intra-operative Photofrin-photodynamic therapy of brain tumors.* Lasers Surg Med, 2003. **32**(3): p. 224-32.
132. Kanniyappan, U., B. Wang, C. Yang, P. Ghassemi, Q. Wang, Y. Chen, and J. Pfefer. *Near-infrared fluorescence image quality test methods for standardized performance evaluation*. in *Proc. of SPIE Vol*. 2017.
133. Tang, J., N. Huang, X. Zhang, T. Zhou, Y. Tan, J. Pi, L. Pi, S. Cheng, H. Zheng, and Y. Cheng, *Aptamer-conjugated PEGylated quantum dots targeting epidermal growth factor receptor variant III for fluorescence imaging of glioma.* Int J Nanomedicine, 2017. **12**: p. 3899-3911.
134. ISO, *ISO 8600-3: Optics and optical instruments - Medical endoscopes and endoscopic accessories - Part 3: Determination of field of view and direction of view of endoscopes with optics*. 2003, International Organization for Standardization.
135. CLSI, *Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures; Approved. Guideline—Second Edition*, in *CLSI document EP17-A2*. 2012, Clinical and Laboratory Standards Institute.
136. Harmonisation, I.C.o., *VALIDATION OF ANALYTICAL PROCEDURES: TEXT AND METHODOLOGY Q2(R1).* 1996.
137. Davis, S.C., B.W. Pogue, H. Dehghani, and K.D. Paulsen, *Contrast-detail analysis characterizing diffuse optical fluorescence tomography image reconstruction.* J Biomed Opt, 2005. **10**(5): p. 050501.
138. Roy, M., A. Kim, F. Dadani, and B.C. Wilson, *Homogenized tissue phantoms for quantitative evaluation of subsurface fluorescence contrast.* J Biomed Opt, 2011. **16**(1): p. 016013.
139. AAPM, *Acceptance Testing and Quality Assurance Procedures for Magnetic Resonance Imaging Facilities*, in *AAPM REPORT NO. 100. Report of MR SubcommitteeTask Group I.* . 2010, American Association of Physicists in Medicine.
140. Bouchard, J.P., F. Baribeau, and O. Mermut, *Calibration of fluorescence reflectance reference phantoms.* Design and Performance Validation of Phantoms Used in Conjunction with Optical Measurement of Tissue V, 2013. **8583**.
141. Noiseux, I., M. Fortin, S. Leclair, J. Osouf, and O. Mermut, *Development of optical phantoms for use in fluorescence-based imaging.* Design and Performance Validation of Phantoms Used in Conjunction with Optical Measurement of Tissue Ii, 2010. **7567**.
142. Pogue, B.W. and M.S. Patterson, *Review of tissue simulating phantoms for optical spectroscopy, imaging and dosimetry.* Journal of Biomedical Optics, 2006. **11**(4).
143. De Grand, A.M., S.J. Lomnes, D.S. Lee, M. Pietrzykowski, S. Ohnishi, T.G. Morgan, A. Gogbashian, R.G. Laurence, and J.V. Frangioni, *Tissue-like phantoms for near-infrared fluorescence imaging system assessment and the training of surgeons.* J Biomed Opt, 2006. **11**(1): p. 014007.
144. Wang, F., J. Kong, L. Cooper, T. Pan, T. Kurc, W. Chen, A. Sharma, C. Niedermayr, T.W. Oh, D. Brat, A.B. Farris, D.J. Foran, and J. Saltz, *A data model and database for high-resolution pathology analytical image informatics.* J Pathol Inform, 2011. **2**: p. 32.

145. *Medical Devices; Current Good Manufacturing Practice (CGMP) Final Rule; Quality System Regulation*, C.f.D.a.R. Health, Editor. 1996, US Food and Drug Administration, HHS: Rockville MD. p. 21 CFR part 820.
146. *REGULATION (EU) 2017/745 amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009, Union regulatory framework for medical devices, other than in vitro diagnostic medical devices* T.C.o.m. devices, Editor. 2017, Official Journal of the European Union.
147. Zerhouni, E.A., *US biomedical research: basic, translational, and clinical sciences.* JAMA, 2005. **294**(11): p. 1352-8.
148. Zerhouni, E.A., *Translational and clinical science--time for a new vision.* N Engl J Med, 2005. **353**(15): p. 1621-3.
149. *Good Clinical Practice (GCP) E6*, E. Guidelines, Editor., International Council for Harmonization.
150. *Good Clinical Practice*. Available from: https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/good-clinical-practice.