Performance Assessment and Quality Control of Fluorescence Molecular Endoscopy with a Multi-Parametric Rigid Standard

Anna Tenditnaya, Ruben Y. Gabriëls, Wouter T.R. Hooghiemstra, Uwe Klemm, Wouter B. Nagengast, Vasilis Ntziachristos, *Fellow, IEEE*, and Dimitris Gorpas, *Member, IEEE*

***Abstract*— Fluorescence molecular endoscopy (FME) is emerging as a "red-flag" technique with potential to deliver earlier, faster, and more personalized detection of disease in the gastrointestinal tract, including cancer, and to gain insights into novel drug distribution, dose finding, and response prediction. However, to date, the performance of FME systems is assessed mainly by endoscopists during a procedure, leading to arbitrary, potentially biased, and heavily subjective assessment. This approach significantly affects the repeatability of the procedures and the interpretation or comparison of the acquired data, representing a major bottleneck towards the clinical translation of the technology. Herein, we propose a robust methodology for FME performance assessment and quality control that is based on a novel multi-parametric rigid standard. This standard enables the characterization of an FME system's sensitivity through a single acquisition, performance comparison of multiple systems, and, for the first time, quality control of a system as a function of time and number of usages. We show the photostability of the standard experimentally and demonstrate how it can be used to characterize the performance of an FME system. Moreover, we showcase how the standard can be employed for quality control of a system. In this study, we find that the use of composite fluorescence standards before endoscopic procedures can ensure that an FME system meets the performance criteria and that components prone to performance degradation are replaced in time, avoiding disruption of clinical endoscopy logistics. This will help overcome a major barrier for the translation of FME into the clinics.**

***Index Terms*— Fluorescence molecular endoscopy, quality control, rigid fluorescence standards, standardization, system benchmarking.**

I. INTRODUCTION

NEAR infrared fluorescence molecular endoscopy (FME) with its "red-flag" capability (i.e., real-time highlighting of lesions) is a promising tool for earlier, faster, and more personalized detection of esophageal and colorectal cancer and for gaining insights into novel drug distribution, optimal dose, and response [1]. In a preliminary Phase I study (NCT02129933), we have shown that by targeting and thus, "highlighting", dysplastic and malignant cells with the vascular endothelial growth factor A (VEGF-A) specific tracer Bevacizumab-IRDye800, FME in tandem with high-definition white-light endoscopy (HD-WLE) can identify an impressive 33% more cancerous lesions in Barrett's esophagus than expert endoscopists [2]. Similarly, Hartmans et al. presented the first feasibility and dose-finding study (NCT02113202) that showcased the identification of VEGF-A-positive colorectal adenomas using FME [3]. Besides its clinical potential, FME has also been proven to be a valuable tool for the discovery of novel fluorescent tracers. We have recently reported the application of this technology in animal models for the validation of tracers specific for esophageal [4, 5] and colorectal cancer [6, 7]. Despite the anticipated clinical benefits, a more systematic use of FME in clinical and pre-clinical applications is impeded without a way to assess performance and carry out quality control of FME systems [8].

Quality control, for example, is particularly critical for FME, as cycles of usage and cleaning gradually affect the optics of the employed endoscopes, frequently leading to image fidelity degradation [8]. Although most clinical HD-WLE systems have counters to determine the endpoint of the endoscope's lifetime, this is not the case for FME systems. To date, the endoscopist decides when to replace an FME endoscope. As such, the decision is arbitrary and highly dependent on the experience of the endoscopist. Moreover, in 2018, we introduced the concept of high-fidelity fluorescence imaging (HiFFi) and its dependence on invariable (system-related) and variable (application-related) parameters [9]. In that study, we proposed

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101046923 (msGUIDE) and by the Dutch Cancer Society, Amsterdam, the Netherlands (Project 8306) and the Bundesministerium für Bildung und Forschung (BMBF), Bonn, Germany (Project ESCEND, 01KT1809), under the framework of the ERA-NET TRANSCAN-2 initiative under Grant Agreement No 643638. (Anna Tenditnaya and Ruben Y. Gabriëls contributed equally to this work). (Corresponding author: Dimitris Gorpas).

Anna Tenditnaya (anna.tenditnaya@helmholtz-munich.de), Uwe Klemm (uwe.klemm@helmholtz-munich.de), Vasilis Ntziachristos (bioimaging.translatum@tum.de), and Dimitris Gorpas (dimitrios.gkorpas@helmholtz-munich.de) are with the Institute of Biological and Medical Imaging, Helmholtz Zentrum München, Neuherberg, Germany, and Chair of Biological Imaging at the Central Institute for Translational Cancer Research (TranslaTUM), School of Medicine and Health, Technical University of Munich, Munich, Germany.

Ruben Y. Gabriëls (r.y.gabriels@umcg.nl), Wouter T.R. Hooghiemstra (w.t.r.hooghiemstra@umcg.nl), and Wouter B. Nagengast (w.b.nagengast@umcg.nl) are with the Department of Gastroenterology and Hepatology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.

that there should be end-user- and application-independent readouts even when using markedly different fluorescence molecular imaging (FMI) implementations for consistent data interpretation and easy system performance assessment [9]. Expanding on this work, in 2022 we highlighted the need for performance assessment and quality control in FME applications as an important prerequisite for high fidelity FME that could potentially lead to unbiased interpretation of the acquired images [8].

The development of a reliable and repeatable means for FME system performance assessment and quality control would allow for readout comparisons, system benchmarking, and performance monitoring, both before each usage and over time. All these capabilities are important prerequisites for successful clinical translation of the technology and its establishment as an important tool for clinical and pre-clinical studies.

To date, the standardization approaches proposed in the literature are mostly focused on the characterization [10, 11] or the comparison and benchmarking [12-14] of markedly different systems. Several groups have designed multi-parametric rigid phantoms for the establishment of sensitivity limits, performance assessment of cameras, and quantification of fluorescence data [12, 15-17]. Usually these phantoms consist of polyurethane as a hardener, with the addition of titanium dioxide (TiO_2) nanoparticles to simulate scattering, hemin or nigrosin for absorption, and quantum dots (QDot) for fluorescence [10, 11, 14]. Recently, 3D printed phantoms have been introduced as an alternative to polyurethane-based ones, with the advantage of enhanced repeatability in manufacturing procedures [18, 19]. A common characteristic of all these approaches is that they present multiple structures within the field of view (FoV), so various system parameters can be quantified using just one or a few acquired frames [14]. Nonetheless, most of these phantoms are tailored towards standardization of FMI systems and are not fully compatible with the requirements of FME [8]. On the other hand, most FME phantoms proposed to date have been designed for training purposes, limiting their application as fluorescence standards [20, 21]. For example, the anthropomorphic elastomer phantom proposed by Yang et al. mimics healthy and Barrett's esophagus mucosal layers as a training aid for using an FME system rather than for the assessment of a system's performance [20]. Thus, the need for comprehensive FME characterization is still unmet, hindering and delaying the clinical translation of the technology [8].

Based on the very encouraging findings of our previous work for FMI performance assessment [10, 13, 14, 18], we believe that a multi-parametric standard, tailored to the requirements of FME, can fulfil the need mentioned above and potentially define the path towards achieving endoscopic HiFFi.

In this work, we propose a multi-parametric fluorescence standard and demonstrate the first systematic performance assessment and quality control of a fiberscope-based flexible FME system. In contrast to the square-shaped phantoms for FMI [10, 18], we opted for a cylindrical design that best matches the FoV of FME systems. The standard allows system characterization using only a single image and can generate reports on i) sensitivity; ii) spatial distribution of illumination; iii) fluorescence and reflectance resolution; and iv) cross-talk between excitation and detection optical paths. We further introduce an analysis platform to calculate descriptive benchmarking score metrics that can be employed for rapid and comprehensive quality control of an FME system as a function of time and number of usages. This methodology is based on the quantification of signal-to-noise ratio (SNR) and contrast for the assessment of the system's performance and the computation of the benchmarking metrics [14]. By incorporating the proposed methodology into a clinical study by our group, we assessed the performance of the FME system before each procedure and the findings indicate inverse correlation between the performance of the fiberscopes and the number of usages. To the best of our knowledge, this is the first FME quality control study that showcases the necessity and significance of a standardization approach for repeated procedures. This study can potentially have a great impact on how composite standards can be used for routine system performance assessment and quality control, and, thus, further accelerate the clinical translation of FME.

II. Materials and Methods

A. Components and design of the FME standard

The proposed FME standard was built with transparent polyurethane (WC-783 A/B, BJB Enterprises, Tustin, United States) as the main material of its matrix and hardener. To simulate scattering, TiO_2 nanoparticles (Titanium IV Oxide; Sigma Aldrich, St. Louis, MO, USA) were used, while absorption was enabled by alcohol-soluble Nigrosine (Sigma Aldrich) in the matrix and bovine hemin (≥90% pure; Sigma Aldrich) in the different wells. Finally, organic quantum dots (QDot 800 ITK, Thermofisher Scientific, Waltham, MA, USA) were used for fluorescence due to their excellent stability [10, 22]. The components of the standard are summarized in Fig. 1a, and the standard was made following the guidelines proposed by our group in a previous study [10]. To ensure homogeneity of each element of the standard the QDots 800, TiO_2 nanoparticles, and polyurethane hardener were sonicated before adding the polyurethane base, while the bubbles were removed from the mixture using degassing, as proposed by the previously published protocols [11, 14]. It has been shown by Zhu et al. [11], and also confirmed in our previous work [14], that the adopted protocol for the development of the fluorescence standard warrants minimum uncertainty in the uniformity of the fluorescence emitted by the QDots [14]. As this standard is not an anthropomorphic phantom, its optical properties were selected for assessing the optimal performance of the system and not to simulate a specific tissue type. Nevertheless, the quantification of the optical properties was implemented as previously described [10].

For the quantification of the different invariable parameters that characterize the performance of an FME system [8], the following components (see Fig. 1a) were incorporated into the design of the standard:

Matrix. In order to limit diffusion and cross-talk between neighboring wells, the matrix of the standard is highly absorbing (μ_a = 55 cm^{-1}). Its diameter is 20 mm and its height is 24 mm (Fig. 1)

Background. To maintain consistency between different measurements, background was taken from a location where the matrix was not drilled to create a well. This means that this so-called "well" used for ascertaining background is actually the matrix of the standard.

Flat fielding / Resolution. A cross-shaped well with 20 nM QDot (μ_a = 0.25 cm^{-1} and m_s' = 7 cm^{-1}) serves as reference for the estimation of the spatial distribution of illumination and to address any uncertainties of both illumination sources. The high TiO_2 and QDot concentrations make the well clearly visible by both color and fluorescence imaging sensors.

Sensitivity. The rest of the wells were used for the quantification of the system's sensitivity. These wells had the same optical properties as the flat fielding well, but QDot concentrations ranging from 0 nM to 18 nM. These wells were used for the quantification of the SNR, contrast, linearity, and dynamic range of the FME system.

Working distance (WD). Finally, a cylindrical depression, 10 mm in diameter and 5 mm depth, enables the quantification of all sensitivity parameters simultaneously at two working distances (i.e., WD1 and WD2 = WD1 + 5 mm). This accounted for the narrow Gaussian-like illumination profile of most FME systems. As the WD is a dynamically changing parameter during an FME imaging session, the introduction of this feature enables the performance assessment of a system at two WDs within a single frame. The 5 mm difference corresponds to an approximate average translation of the fiberscope due to the manual navigation of the clinical endoscope by the endoscopist.

All wells were milled out of the polyurethane matrix at a depth of 7 mm from the top surface of the standard. Moreover, each well has 6 mm length and 1 mm width. The cross-shaped well for flat fielding and resolution estimation has a length of 16 mm and a width of 1 mm.

B. Photostability measurements

A light-tight enclosure was developed and the proposed standard was positioned at a fixed location and imaged periodically over two months twice or thrice a week. All acquisition parameters, including camera gain, exposure time (i.e., 300 a.u. and 500 ms respectively), and excitation light power (52 mW), remained constant between measurements. The standard remained in darkness while not measuring, with the excitation source turned off. Moreover, the fiberscope was kept at a fixed position 10 mm above the top surface of the standard. These measurements were implemented with an FME system previously described by our group [4, 7] and the photostability was assessed as the intensity normalized to the exposure power at the center of the flat fielding element as shown in Fig. 1.

C. FME system and imaging protocol

All measurements of the fluorescence standard were performed at the University Medical Center Groningen (UMCG) using the SurgVision Explorer Endoscope (SVEE, SurgVision BV., Groningen, The Netherlands), that operates with exchangeable fiberscopes (Schölly Fiberoptic GmbH, Denzlingen, Germany) that have a depth of field up to 60 mm, making them ideal for endoscopic applications. The system is based on a continuous wave 750 nm laser diode with an average power of 85 mW at the fiberscope output, while the average power of the white-light source was measured at 120 mW. The system control and data acquisition was implemented with the built-in software, which does not allow for independent operation of the color and fluorescence light soursces. Thus, due to the wide absorption spectrum of the used QDots, all measurements were augmented with the fluorescence induced by the white-light source. This system has been extensively used by our group in numerous clinical studies in the gastrointestinal (GI) tract and is capable of acquiring concurrent fluorescence and color images [2, 3, 23].

For the needs of this study, the FME standard was imaged in the endoscopy suite during the preparation of the FME system for use within our Phase II clinical trial ESCEND (NCT03877601). The fiberscope was inserted into a custom-made post-based holder encapsulated into a black cage to isolate all measurements from the ambient illumination of the endoscopy suite. The holder held the fiberscope at a minimum of 11 mm away from the standard's top surface in a perpendicular orientation as shown in Fig. 1b. In total, we imaged the standard using 8 different fiberscopes before endoscopic procedures done on 60 patients. However, some of these 8 fiberscopes were also used for other clinical trials by our group that were outside the scope of this study, but increased the overall usage number of the 8 fiberscopes.

D. Data processing

Following data acquisition, all images were normalized to the corresponding camera gain values and exposure times (i.e., both linear to the acquired signal) before being registered to the standard template shown in Fig. 2a. All procedures described below were implemented in MATLAB (Mathworks, Natick, MA, USA).

The transformation matrices were approximated by manually selecting at least four fiduciary points present in all three images (i.e., fluorescence, color, and template) and applying the maximum likelihood estimator sample consensus (MLESAC) algorithm, as implemented in MATLAB. Two transformation matrices were derived; one for translating the fluorescence images and one for translating the color images to the template's coordinate system. Since this was a template-based approach, the segmentation of the fluorescence and color images was straightforward due to the use of the template as a mask. In addition, the template was labeled to describe the different elements of the standard. These labels were then used to identify the different standard elements on the acquired fluorescence and color images, as shown in Fig. 2a. After provision of the two transformation matrices and the standard labels, the following metrics were computed by adopting the methodology previously proposed by our group [13, 14]:

Magnification/Working distance. The physical length of

the flat-fielding element (17.5 mm) shown in Fig. 1a was divided by the element's length in the color camera coordinates [13]. The WD was approximated by the equation:

$$WD = (Dia \cdot PixelSize \cdot Magn)/[2 \cdot \tan(\theta/2)] \quad (1)$$

where *Dia* is the diameter of the fiberscope's FoV in pixels as measured in the color images, *PixelSize* is the dimension of the camera's pixels (0.0055 mm for the color camera), *Magn* is the estimated magnification, and θ is the viewing angle of the fiberscope (i.e., 85°). The WD corresponds to the distance between the standard's top surface and the fiberscope (i.e., WD1), while the lower surface of the standard is an additional 5 mm further from the fiberscope (WD2 = WD1 + 5 mm).

Flat-fielding. The illumination profile was estimated by application of bi-harmonic interpolation to the normalized average intensity from the flat-fielding element for both fluorescence and color images. Before the interpolation, the intensity of the flat-fielding element at WD2 (I_{WD2}) was projected to WD1 ($I_{WD2\rightarrow WD1}$) assuming the inverse square law:

$$I_{WD2\rightarrow WD1} = [(WD1 + 5)/WD1]^2 \cdot I_{WD2} \quad (2)$$

Flat-fielding was then achieved by normalizing the acquired images with the approximated illumination profile [14]. Although flat-fielding cannot be employed during the endoscopy, where the viewing angle and working distance dynamically change, it can provide valuable information about the illumination profile of the used fiberscope.

Resolution. By adopting the diffused resolution quantification approach proposed by our group for FMI standardization [14], a line segment was scanned in the flat-fielding element at WD2 in the direction shown in Fig. 3. The resolution of the camera was inferred as the line distance corresponding to the average contrast transfer function (CTF) of 0.264 a.u. This process was applied to (i) all four quadrants of the flat-fielding element with the average resolution reported as the resolution of the camera, and (ii) both fluorescence and color cameras.

Cross-talk. The ratio between the average intensities of two wells without any QDots (i.e., sensitivity well with 0 nM QDot and background well, as shown in Fig. 1a) was used to quantify the cross-talk of the fluorescence camera. This corresponds to the excitation light propagating through the detection optical path of the system and was calculated for both working distances of the standard. In principle, since this is a ratio of the signal to the background, a cross-talk equal to or smaller than one implies reduced leakage of the excitation light towards the detection path.

Sensitivity. The sensitivity of the system was quantified through the SNR, contrast, and linearity. These parameters were estimated for all the sensitivity wells of the standard at both working distances (see Fig. 1a) as detailed in a previous study from our group [14]. The SNR was calculated in dB through the ratio [24]:

$$\mathrm{SNR} = 20 \cdot \log((\bar{S} - \bar{N})/\mathrm{std}_N) \quad (3)$$

where $\bar{S}$ is the average intensity within each sensitivity well, whereas $\bar{N}$ and std_N are the average and standard deviation, respectively, of the intensity within the 0 nM QDot concentration sensitivity well (see Fig. 1a). Adopting the Rose criterion definition for computed tomography [25], the background corrected signal for the detection of a well should be greater than 5 times the standard deviation of the

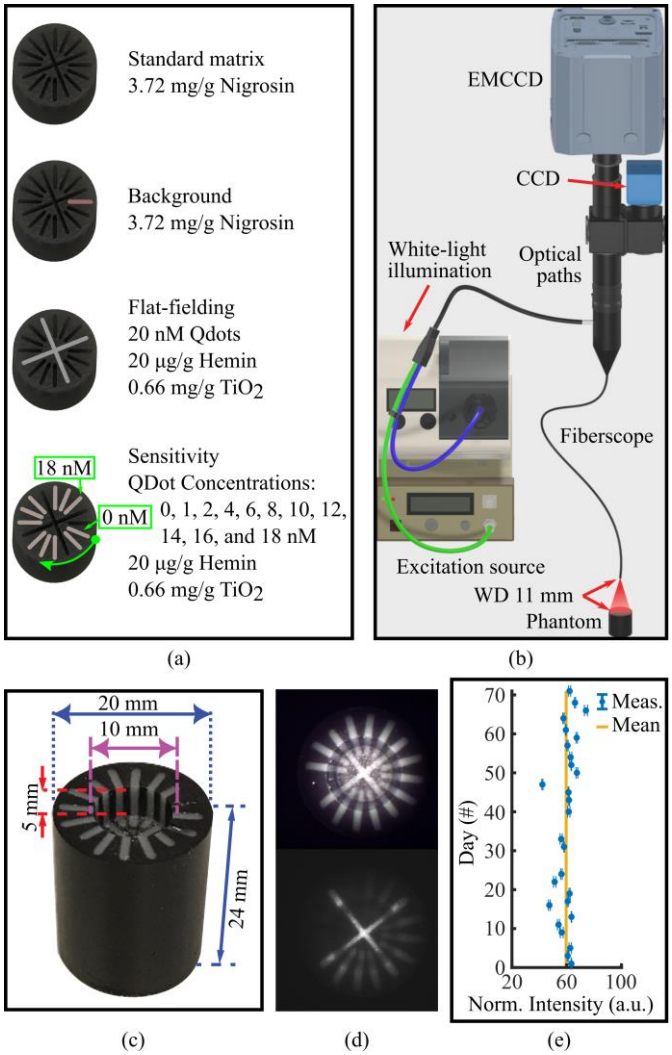

Fig. 1. Multi-parametric rigid standard for FME system performance assessment and quality control. (a) The standard design and its components. The different wells of the standard can be used for the quantification of various system parameters after a single acquisition. In the dynamic range wells, the red arrow indicates the clockwise order of the different QDot concentrations. (b) Imaging set up for FME system performance assessment and quality control. Fluorescence images are acquired by the EMCCD, while true color images by the CCD. The main prerequisites are the perpendicular viewing angle and the maximum coverage of the system's field of view, as defined by the working distance. (c) The physical dimensions of the standard and the two working distances that are 5 mm apart. (d) Exemplary color (top) and fluorescence (bottom) images of the standard. (e) The mean intensity (normalized to the excitation source power) of the resolution target shown in (a), as a function of time. The error bars correspond to the standard deviation of 20 measurements per time point. EMCCD: electron-multiplying charge-coupled device; CCD: charge-coupled device; WD: working distance.

background. In decibel units this value becomes equal to SNR = 14 dB and corresponds to the threshold adopted herein. On the other hand, the Weber fraction was adopted for the estimation of the contrast [13]:

$$C = (\bar{S} - \bar{N})/\bar{N} \quad (4)$$

with a threshold equal to 1 that corresponds to a sensitivity well having twice the average signal of the background [13, 26].

Moreover, the dynamic range was also quantified as the ratio between the detected (i.e., SNR > 14 dB and C > 1) and the total number (N = 10) of the sensitivity wells (see Fig. 1a).

Benchmarking scores. The benchmarking scores were estimated for the system's fluorescence and optical resolution, sensitivity, and cross-talk according to the definition proposed in our previous work [14]. Briefly, for each performance metric the symmetric mean absolute percentage error (sMAPE) was quantified. For the SNR the sMAPE was quantified in respect to the 14 dB threshold and for contrast in respect to the threshold value of 1. Similarly, the resolution sMAPE was quantified in respect to the worst resolution of a system which corresponds to the maximum line distance as shown in Fig. 3. The overall benchmarking score of a system was the sum of all the individual benchmarking metrics [14].

III. Results

A. Design, imaging protocol, and photostability

The components and design of the cylindrical multi-parametric standard for FME performance assessment and quality control are shown in Fig. 1a. The standard's top surface encompasses numerous elements for the quantification of invariable system parameters [9], including i) excitation light cross-talk towards the camera, ii) dynamic range, iii) fluorescence and reflectance resolution, and iv) spatial distribution of the excitation illumination.

The image acquisition protocol is shown in Fig. 1b. The working distance (WD) of approximately 11 mm was chosen so that the standard fully covered the FoV of an 85° forward viewing fiberscope when perpendicular to the standard's top surface. Finally, due to the narrow FoV of the fiberscopes used, the standard has the same elements arranged concentrically at two WDs as shown in Fig. 1a and Fig. 1c.

Fig. 1c depicts the physical dimensions of the standard. For the development of the 20 mm diameter cylindrical standard, we adapted the manufacturing process previously proposed for wide-field intraoperative FMI standardization phantoms [10, 13]. This diameter is suitable for the FoV of the fiberscopes usually employed in FME applications. Exemplary color (top) and fluorescence (bottom) images of the standard are shown in Fig. 1d.

Although the photostability of the various materials used for the development of the standard has been previously reported [10, 17, 27], we also confirmed it for our specific design. Repeated imaging of our standard showed that it was photostable over the course of more than two months. The average intensity (normalized to the power of the excitation light) of the resolution target (Fig. 1a) across the multiple measurements is shown in Fig. 1e. The coefficient of variance (CV) is estimated to be 10.6%, which indicates that the fluorescence signal remains stable to within approximately 10% of its mean value and, in full agreement with previous studies [22], demonstrates the excellent performance of the standard. The small disparities recorded could possibly be due to system temporal variability or the positioning of the phantom within the system's FoV.

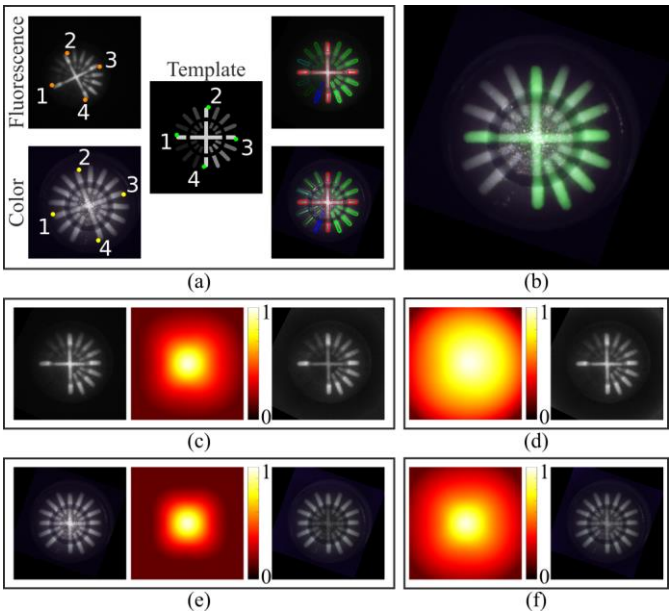

Fig. 2. Segmentation of the fluorescence and color images and illumination pattern correction. (a) The fluorescence (top left) and color (bottom left) images are registered to the template (center) through the use of at least 4 manually-selected fiduciary points per image. This aligns the coordinates of the acquired data to the template's coordinate system, and allows the automatic segmentation of the different elements of the standard (top right for fluorescence and bottom right for color image). (b) The alignment of the acquired fluorescence and color data to the template's coordinate system enable the rapid overlay and visualization of the composite fluorescence and color images. (c) The acquired fluorescence image (left) is normalized using the intensity profile derived from the flat-fielding element (center) for the two working distances (see Fig. 1a and Methods). Through this normalization, the spatial inhomogeneity of the excitation source is corrected (right) and the different wells can be used for performance assessment of the FME system. (d) When the intensity profile is not normalized to the working distance (left), the resultant fluorescence image does not account for the two working distances and the intensity levels of the different QDot concentrations appear equal (right). (e) Similar to (c), but for the color image. (f) Similar to (d), but for the color image.

B. Co-registration and segmentation

Following the manual selection of at least 4 common fiduciary points, the fluorescence and color images were geometrically referenced to a template of the standard's design (Fig. 2a). By leveraging on the known coordinates of the template, the segmentation and labeling of the standard's wells is straightforward, as shown in Fig. 2a [13]. Moreover, this template serves as the reference coordinate system for the visualization of the composite fluorescence and color image (Fig. 2b).

C. Quantification of invariable parameters and FME system benchmarking

In a recent work from our group, we defined the invariable parameters of an FME system as (i) the magnification and working distance of the system, (ii) the homogeneity of the illumination (i.e., flat-fielding), (iii) the resolution of the imaging sensor, (iv) the cross-talk between the optical paths, and (v) the sensitivity of the system [8]. Following the segmentation process shown in Fig. 2a and the quantification processes described in Methods, the different wells of the standard imaged in Fig. 2 were used to quantify these five parameters:

Magnification/Working distance. The magnification was approximated to 1:4.6 and the working distance to the standard's upper surface was estimated from (1) equal to 12.3 mm (WD1). This means that WD2 was 17.3 mm as it is 5 mm below the top surface.

Flat-fielding. Fig. 2c and Fig. 2e show the results of the flat-fielding process on the fluorescence and color images, respectively, after fitting the normalized intensity profiles of the flat-fielding element (Fig. 1a) to a surface with application of bi-harmonic interpolation. The normalization of the intensity

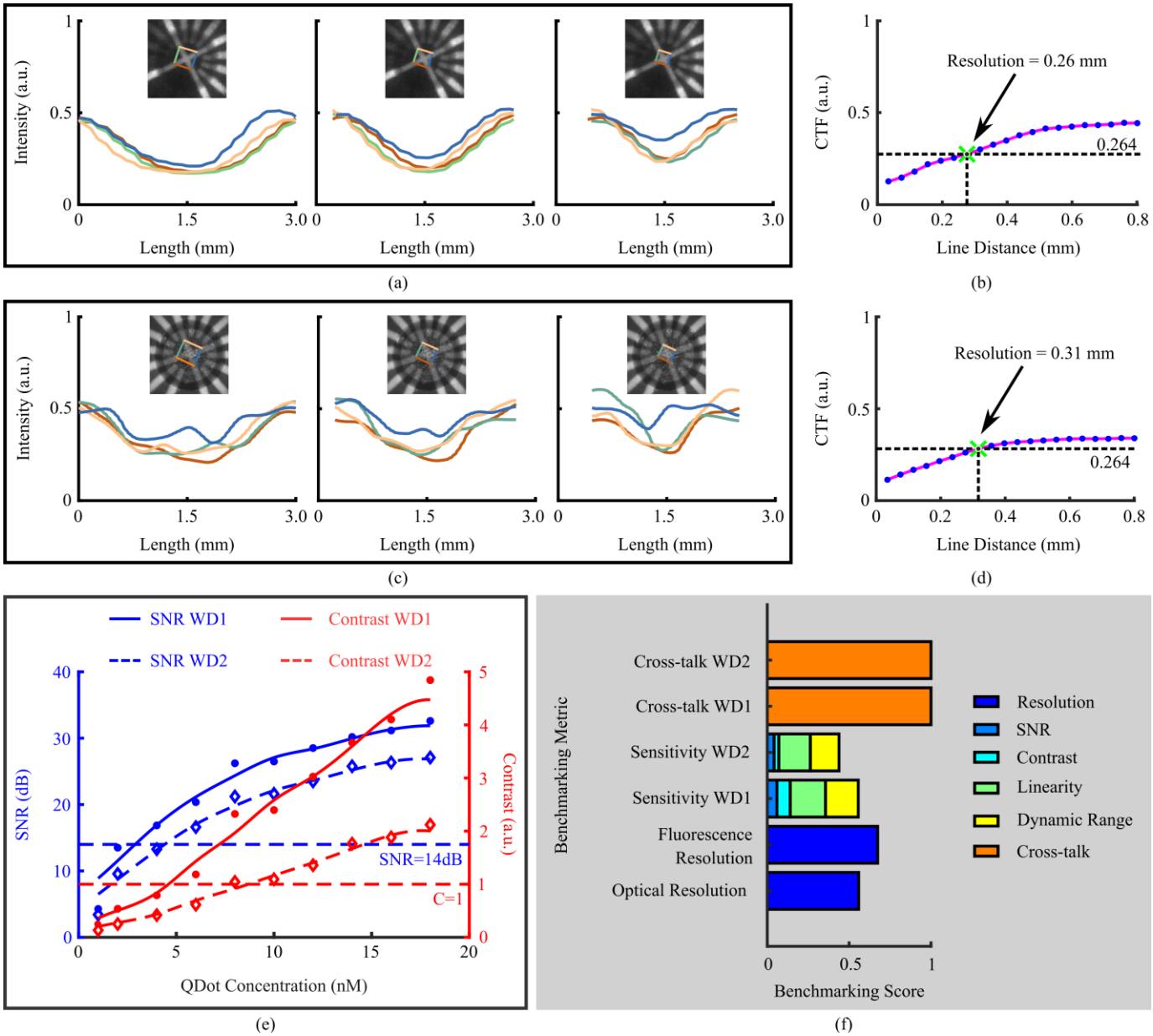

Fig. 3. Sensitivity assessment and FME system benchmarking. (a) The diffused fluorescence resolution is defined by the intensity profile scanned across four lines positioned over the edges of the four concave vertices. The insets from left to right depict three exemplary positions of the four lines scanned. (b) Contrast transfer function (CTF) plot derived from the intensity profiles shown in (a) for the fluorescence camera. The resolution of the camera (0.28 mm) is inferred as the line distance from the concave angle of the flat-fielding cross that corresponds to the average CTF of 0.264 a.u. (c) The scanning process as described in (a), but for the color camera. (d) The resolution of the color camera is approximated in a manner similar to (b). (e) The signal-to-noise ratio (SNR) and contrast for the wells with different QDot concentrations (shown in Fig. 1a) for the two working distances of the standard (see Fig. 1c). (f) The benchmarking scores for all the invariable parameters that can be quantified by the standard. WD: working distance; WD1 = 12.3 mm; WD2 = 17.3 mm.

profile for the fluorescence and color images was implemented with regards to the two working distances and applying (2). If the working distances were not considered, the flat-fielding process would result in the erroneous equalization of the well's intensities at the two surfaces, as shown in Fig. 2d and Fig. 2f for the fluorescence and color images, respectively.

Resolution. Adopting our previously reported approach [13], the resolution of the system was quantified for the four quadrants of the flat-fielding element, as shown in Fig. 3a for the fluorescence camera and in Fig. 3c for the color camera. The corresponding average contrast transfer function is shown in Fig. 3b and Fig. 3d, indicating a resolution of 0.26 mm for the fluorescence camera and 0.31 mm for the color camera.

Cross-talk. The cross-talk of the system was estimated at 0.82 for WD1 and at 0.88 for WD2, meaning that there is no leakage of the excitation light into the detection optical path.

Sensitivity. The sensitivity of the system was assessed by means of SNR, contrast, linearity, and dynamic range. The SNR and contrast as a function of the various QDot concentrations (Fig. 1a) are depicted in Fig. 3e, while the dynamic range represents the number of concentrations that the system can detect (i.e., SNR > 14 dB according to the Rose criterion [25] and contrast > 1 according to the Weber fraction [13], see Methods). Based on Fig. 3e, the dynamic range of the system is equal to 7 for WD1 (i.e., 8 wells above the SNR threshold and 7 wells above the contrast threshold) and 6 for WD2 (i.e., 7 wells above the SNR and 6 wells above the contrast threshold).

The various invariable parameters can then be combined into the different benchmarking scores shown in Fig. 3f. These scores provide a "single-value" description of the FME system's performance for the optical and fluorescence resolutions, the sensitivity, and the cross-talk at the two working distances. Moreover, all the scores can be merged to a single benchmarking score to describe the performance of the system with only one number [14]. This approach enables rapid (i) comparison of markedly different FME systems and (ii) monitoring of a system's performance as a function of time (i.e., quality control). For the system examined in this study, the unified benchmarking score was 0.73, which meant the system could achieve 73% of the ideal system's performance.

D. Quality control

Besides assessing the performance of an FME system, the

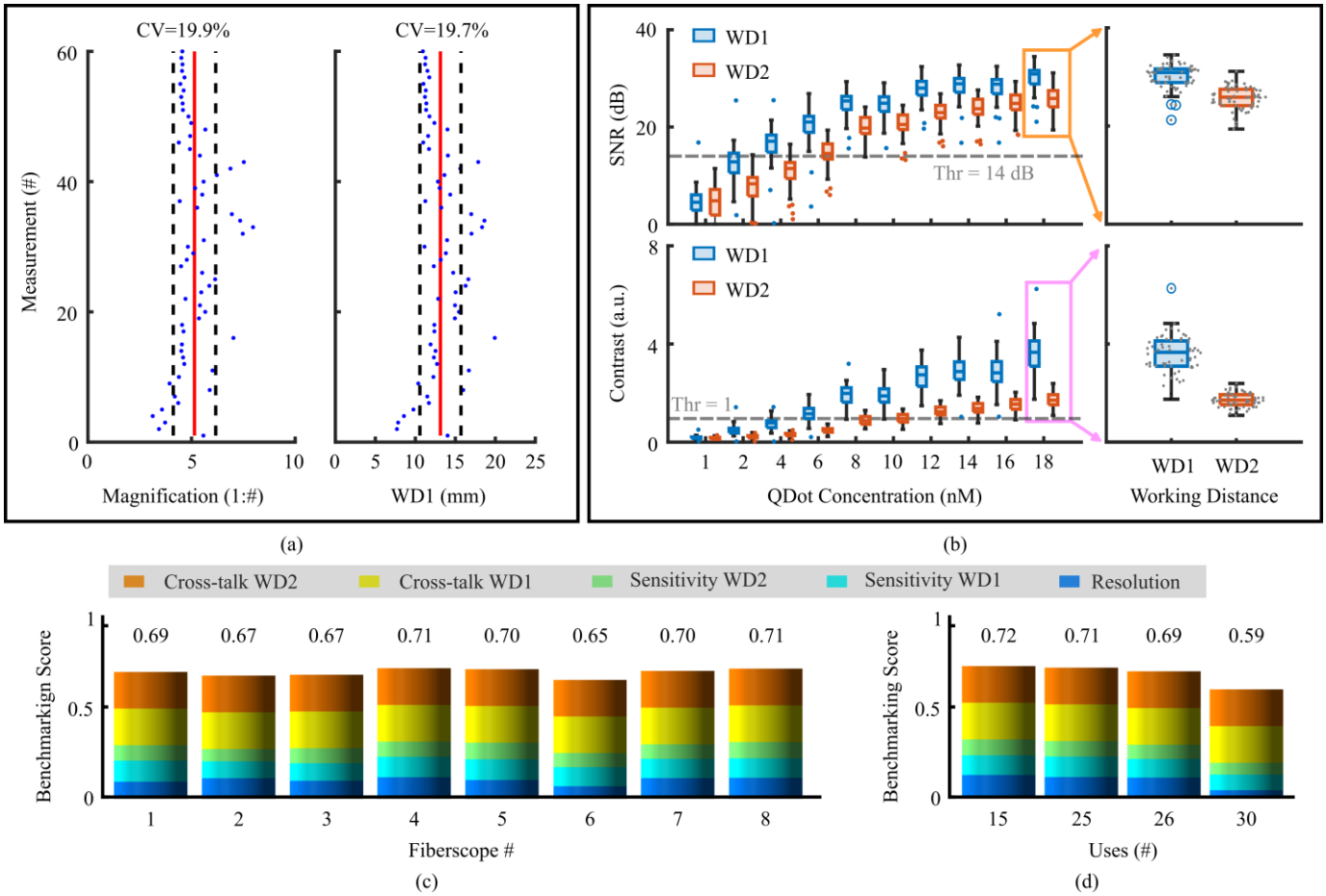

Fig. 4. FME quality control. (a) The magnification and working distance quantified before endoscopy sessions for 60 patients. (b) The signal-to-noise ratio (SNR - top charts) and contrast (bottom charts) for wells with varying QDot concentrations are shown on the left. Enlarged charts for SNR and contrast for 18 nM QDot are shown on the right. At 18 nM QDot, a SNR of 30.16±2.49 dB and contrast of 3.57±0.77 for WD1 and the corresponding 25.70±2.33 dB and 1.72±0.29 for WD2 indicate consistent measurements. (c) The benchmarking scores quantified for the 8 fiberscopes before endoscopy of randomly selected patients indicate consistent system performance regardless of the fiberscope used. (d) The degradation of a fiberscope's performance due to the usage and cleaning cycles are indicated by the negative trend of the benchmarking scores. CV: coefficient of variation; WD: working distance; WD1 = 13.14±2.57 mm; WD2 = 18.14±2.57 mm.

proposed standard enables the quality control of a system across numerous measurements. For a period of almost two years (July 2019 to April 2022), we measured the standard before endoscopy procedures on 60 Barrett's esophagus patients who were recruited for the ESCEND clinical trial (NCT03877601). These measurements gave us valuable insights on the application of the standard for assessing the performance stability of the system, as well as the quality of the used fiberscopes.

During all measurements of the standard, the magnification and working distance were kept constant at 1:5.1 (standard deviation = 1.02) and 13.14±2.57 mm, respectively, as shown in Fig. 4a. The small CV (<20%) of these first measurements of an FME standard in a clinical environment highlights the feasibility of the proposed methodology under consistent acquisition settings. This is further demonstrated by the quantification of the SNR and contrast in Fig. 4b. In specifics, the SNR from the 18 nM well (see Fig. 1a) was estimated to be 30.16±2.49 dB (CV=8.26%) for WD1 and 25.70±2.33 dB (CV=9.07%) for WD2, while the corresponding contrast was 3.57±0.77 (CV=21.42%) and 1.72±0.29 (CV=16.57%). We also observed equally small variability in all the other wells with varying concentrations of QDots, as shown in Fig. 3e.

We could achieve the performance assessment and the quality control of the FME system across all 60 measurements by observing and comparing all the metrics shown in Fig. 3. However, such a strategy will not be possible in a clinical environment, where the endoscopy systems are used multiple times a day with quick turnaround times between patients, and the medical staff is not trained to assess such metrics. For successful translation, sensitivity and quality control metrics have to be simplified. We have achieved this by provision of a single score for easy assessment and comparison. Fig. 4 shows the benchmarking scores of 8 fiberscopes used on random patients during the study. It can be seen that the scores remain relatively constant for these measurements despite the use of different fiberscopes. This observation is further supported by the consistency observed in the SNR and contrast metrics shown in Fig. 4b.

Besides comparing different fiberscopes, it is also important to monitor a single fiberscope throughout its lifespan as it is well known that repeated measurement and cleaning cycles will impact its performance. However, despite knowing this drawback, fiberscope suitability is still mainly assessed by endoscopists during endoscopy procedures. This means that rejection of a poorly performing system is wholly based on the experience of the endoscopist. Furthermore, given that the assessment happens during the procedure, the replacement of the fiberscope leads to unnecessary delays that can badly disrupt the tightly-packed operation schedule. In contrast, a nurse could measure the proposed standard before the endoscopy procedure and quickly identify the fiberscopes with insufficient performance, allowing appropriate measures to be taken without affecting the standard clinical practice. Fig. 4d shows an example of such quality control measurements, where a fiberscope shows a declining trend in the benchmarking scores up to 26 usages and a markedly decreased score after 30 consecutive usages. An equivalent performance degradation was further observed in one additional fiberscope used in this study. As the fiberscopes can unexpectedly break or present reduced performance at any point, a sharp reduction in the benchmarking score can indicate the need to replace the fiberscope and, thus, ensure optimal clinical measurements. Moreover, the fact that no other of the remaining 6 fiberscopes presented a similar trend as the one shown in Fig. 4d indicates that this trend is not due to changes in the fluorescence standard's performance, but due to the actual degradation of the fiberscopes.

IV. Discussion

In this study, we introduce a novel standard for FME performance assessment and quality control, and we showcase its great potential for integration into FME procedures, where the system performance depends strongly on the specifications of the used fiberscopes. The design of the standard is tailored to the imaging requirements of FME systems and allows for rapid quantification of a system's invariable parameters [8, 9]. Importantly, and to the best of our knowledge, this is the first time that an FME standard has been measured in the endoscopy suite during a Phase II clinical trial and the potential for FME system benchmarking and quality control over time has been underscored by the results of this study.

While numerous standards have been proposed for FMI performance assessment and there are many recommendations for their integration into clinical FMI procedures [10, 18, 28, 29], there is still a lack of such efforts for FME [1, 8]. The standard and methodology proposed herein show a clear potential to address this lack through the use of photostable materials for manufacturing the standard, the multi-parametric design for thorough performance assessment, and the straightforward integration of the standard into FME procedures.

In short, a single image of the standard allows quantification of the optical and fluorescence resolutions, the sensitivity by means of SNR, contrast, linearity and dynamic range, and cross-talk between fluorescence and color optical paths at two working distances (Fig. 3). To ensure consistency between the performance assessment measurements, all images acquired were corrected for the spatial distribution of both the fluorescence excitation and the white-light source, with the working distance between the tip of the endoscope and the surface of the standard accounted for as shown in Fig. 2. Although these approaches are proposed for the assessment of FMI system performance [14, 18], they have never been employed for the narrow FoV of the endoscopes. In fact, miniaturization of existing FMI standards will not yield a standard suitable for FME and a bespoke standard has to be made instead. As a result, we have established and characterized a standard, where all elements of the matrix have been designed specifically for endoscopy. Moreover, this standard can be used either for fluorescence only systems or for hybrid fluorescence and color systems. The adopted registration approach for registering the acquired images to the template of the standard accounts for rotation, shear, scale, and translation

distortions. For fisheye fiberscopes a new template would be needed or the additional distortion should first be corrected algorithmically before using the proposed fluorescence standard.

Currently, it is well known that clinically approved endoscopes for the gastrointestinal (GI) tract have an internal counter to indicate the lifespan of the endoscopes. All endoscopes reaching the counter's limit need to be replaced in order to ensure optimal imaging quality during the clinical procedures. However, this safeguard does not exist for investigational FME endoscopes. Instead, the end of an endoscope's lifespan is dictated by an endoscopist, making decisions for replacement arbitrary and dependent on acquired experience. This study, consisting of 60 measurements before clinical endoscopies, shows for the first time that unbiased quality control is possible with a single image of the proposed standard. This image can be used to define the performance of a system by means of the individual invariable parameters (Fig. 3) or by the benchmarking scores [14] (Fig. 4), allowing quick quality assessments before each usage without the need for trained personnel. This is of upmost importance for fiberscope-based FME systems, where the usage and cleaning cycles have a negative and unpredictable impact on the optics of the fiberscope [8], as seen in Fig. 4d.

With this study we set the paradigm for performing quality control of FME systems, ensuring consistence of the invariant acquisition parameters across multiple studies. In addition, with the proposed fluorescence standard and methodology, the acquisition parameters of markedly different FME systems can be set up to provide equivalent readouts, which is an important milestone towards the repeatability of FME procedures. However, for comparisons between fluorescence only and hybrid systems, fluorescence and color imaging should be performed separately or the acquired fluorescence signal should be corrected for the one induced by the white-light source, due to the broad excitation spectrum of the QDots [14]. Although the system employed herein does not allow for such corrections, by maintaining both imaging and illumination modules consistent throughout the study, the performance assessment and quality control were possible as all metrics were always referred to the same system and fluorescence standard.

Despite the clear evidence in Fig. 4a that the imaging of the proposed standard can be achieved under similar working distances and magnifications through a simple post-based mount, the next steps following this study would be to design and develop a specialized holder. This holder will ensure a constant working distance and viewing angle and, thus, will enable faster and more consistent measurements, strengthening the potential of the proposed standard and methodology for clinical acceptance. Moreover, a larger number of sensitivity wells would certainly provide a better understanding of a system's sensitivity, similar to the paradigm of FMI standards [14, 17, 18]. Finally, a study focused on the performance degradation of multiple fiberscopes as a function of usage would be pivotal for the identification of the benchmarking threshold that would define the time point for replacing a fiberscope. Such a study will also allow us to identify the order of importance that the different performance metrics bear for the endoscopists and, thus, weight accordingly the individual benchmarking metrics (Fig. 3).

In summary, as the number of clinical studies using FME technology continuously rises, the requirement for performance assessment and quality control becomes increasingly pressing. Building on the experiences gained from FMI, we designed the first FME-tailored standard and methodology for characterizing and monitoring the performance of FME systems. By showcasing the process for quantification and interpretation of the different invariable system parameters, we set out the first step towards high fidelity fluorescence endoscopy.

ACKNOWLEDGMENT

The authors would like to thank Dr. Serene Lee and Dr. Elisa Bonnin for their attentive reading and improvements of the manuscript.

REFERENCES

[1] J. A. Stibbe *et al.*, "Highlighting the undetectable - Fluorescence molecular imaging in gastrointestinal endoscopy," *Mol. Imaging. Biol.,* vol. 25, no. 1, pp. 18-35, Feb 2023.

[2] W. B. Nagengast *et al.*, "Near-infrared fluorescence molecular endoscopy detects dysplastic oesophageal lesions using topical and systemic tracer of vascular endothelial growth factor A," *Gut,* vol. 68, no. 1, pp. 7-10, Jan. 2019.

[3] E. Hartmans *et al.*, "Potential red-flag identification of colorectal adenomas with wide-field fluorescence molecular endoscopy," *Theranostics,* vol. 8, no. 6, pp. 1458-1467, Feb. 2018.

[4] H.-Y. Fang *et al.*, "Targeted Hsp70 fluorescence molecular endoscopy detects dysplasia in Barrett's esophagus," *Eur. J. Nucl. Med. Mol. Imaging,* vol. 49, no. 6, pp. 2049-2063, May 2022.

[5] S. Marcazzan *et al.*, "CXCR4 peptide-based fluorescence endoscopy in a mouse model of Barrett's esophagus," *EJNMMI Res.,* vol. 12, no. 1, pp. 2, Jan. 2022.

[6] J. J. Yim *et al.*, "A protease-activated, near-infrared fluorescent probe for early endoscopic detection of premalignant gastrointestinal lesions," *Proc. Natl. Acad. Sci. USA,* vol. 118, no. 1, pp. e2008072118, Jan. 2021.

[7] S. Rogalla *et al.*, "Biodegradable fluorescent nanoparticles for endoscopic detection of colorectal carcinogenesis," *Adv. Funct. Mater.,* vol. 29, no. 51, pp. 1904992, Dec. 2019.

[8] A. J. Sterkenburg, W. T. R. Hooghiemstra, I. Schmidt, V. Ntziachristos, W. B. Nagengast, and D. Gorpas, "Standardization and implementation of fluorescence molecular endoscopy in the clinic," *J. Biomed. Opt.,* vol. 27, no. 7, pp. 074704, Feb. 2022.

[9] M. Koch, P. Symvoulidis, and V. Ntziachristos, "Tackling standardization in fluorescence molecular imaging," *Nat Photonics,* vol. 12, no. 9, pp. 505-515, Sep. 2018.

[10] M. Anastasopoulou *et al.*, "Comprehensive phantom for interventional fluorescence molecular imaging," *J. Biomed. Opt.,* vol. 21, no. 9, pp. 091309, Sep. 2016.

[11] B. Zhu, 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,* vol. 35, no. 3, pp. 802-811, Mar. 2016.

[12] B. Zhu, and E. M. Sevick-Muraca, "A review of performance of near-infrared fluorescence imaging devices used in clinical studies," *Brit. J. Radiol.,* vol. 88, no. 1045, pp. 20140547, Jan. 2015.

[13] D. Gorpas, M. Koch, M. Anastasopoulou, U. Klemm, and V. Ntziachristos, "Benchmarking of fluorescence cameras through the use of a composite phantom," *J. Biomed. Opt.,* vol. 22, no. 1, pp. 016009, Jan. 2017.

[14] D. Gorpas *et al.*, "Multi-parametric standardization of fluorescence imaging systems based on a composite phantom," *IEEE Trans. Biomed. Eng.,* vol. 67, no. 1, pp. 185-192, Jan. 2020.

[15] J. Baeten, M. Niedre, J. Dunham, and V. Ntziachristos, "Development of fluorescent materials for diffuse fluorescence tomography standards and phantoms," *Opt. Express,* vol. 15, no. 14, pp. 8681-94, Jul. 2007.

[16] A. V. Dsouza, 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.,* vol. 21, no. 8, pp. 080901, Aug. 2016.

[17] B. Zhu, J. C. Rasmussen, and E. M. Sevick-Muraca, "A matter of collection and detection for intraoperative and noninvasive near-infrared fluorescence molecular imaging: To see or not to see?," *Med. Phys.,* vol. 41, no. 2, pp. 022105, Feb. 2014.

[18] A. J. Ruiz *et al.*, "Indocyanine green matching phantom for fluorescence-guided surgery imaging system characterization and performance assessment," *J. Biomed. Opt.,* vol. 25, no. 5, pp. 1-15, May 2020.

[19] U. Kanniyappan *et al.*, "Performance test methods for near-infrared fluorescence imaging," *Med. Phys.,* vol. 47, no. 8, pp. 3389-3401, Aug. 2020.

[20] C. Yang, V. Hou, L. Y. Nelson, and E. J. Seibel, "Color-matched and fluorescence-labeled esophagus phantom and its applications," *J. Biomed. Opt.,* vol. 18, no. 2, pp. 1-12, Feb. 2013.

[21] Y. Jiang, D. Cooper, M. D. Carson, and E. J. Seibel, "Custom bile duct phantom for first-in-human multiplexed NIR fluorescence peptide imaging," in Proc. SPIE, Design and Quality for Biomedical Technologies XII, San Francisco, California, United States, 2019, pp. 108700M.

[22] B. Zhu, 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.,* vol. 11, no. 1, pp. 95-104, Feb. 2012.

[23] R. Y. Gabriels *et al.*, "Detection of early esophageal neoplastic Barrett lesions with quantified fluorescence molecular endoscopy using cetuximab-800CW," *J. Nucl. Med.,* vol. 64, no. 5, pp. 803-808, May 2023.

[24] B. W. Pogue *et al.*, "AAPM Task Group Report 311: Guidance for performance evaluation of fluorescence-guided surgery systems," *Med. Phys.*, Epub ahead of print, Dec. 2023.

[25] S. S. Hsieh, S. Leng, L. Yu, N. R. Huber, and C. H. McCollough, "A minimum SNR criterion for computed tomography object detection in the projection domain," *Med. Phys.,* vol. 49, no. 8, pp. 4988-4998, Aug. 2022.

[26] E. Peli, "Contrast in complex images," *J. Opt. Soc. Am. A,* vol. 7, no. 10, pp. 2032-2040, Oct. 1990.

[27] P. Krauter *et al.*, "Optical phantoms with adjustable subdiffusive scattering parameters," *J. Biomed. Opt.,* vol. 20, no. 10, pp. 105008, Oct. 2015.

[28] B. W. Pogue *et al.*, "Fluorescence-guided surgery and intervention — An AAPM emerging technology blue paper," *Med. Phys.,* vol. 45, no. 6, pp. 2681-2688, Jun. 2018.

[29] L. Hacker, H. Wabnitz, A. Pifferi, T. J. Pfefer, B. W. Pogue, and S. E. Bohndiek, "Criteria for the design of tissue-mimicking phantoms for the standardization of biophotonic instrumentation," *Nat. Biomed. Eng.,* vol. 6, no. 5, pp. 541-558, May 2022.