Repeatability of tissue fluorescence measurements for the detection of cervical intraepithelial neoplasia

José-Miguel Yamal,[1,*] Dennis D. Cox,[2] E. Neely Atkinson,[3] Calum MacAulay,[4] Roderick Price,[5] and Michele Follen[5]

[1]*Division of Biostatistics, The University of Texas School of Public Health, 1200 Herman Pressler, Houston, TX 77030, USA*
[2]*Department of Statistics, Rice University, 6100 Main St, Houston, Texas 77030, USA*
[3] *Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd, Houston, Texas 77030, USA*
[4]*Department of Cancer Imaging, British Columbia Cancer Research Centre, 600 West 10 Ave.,Vancouver, British Columbia V5Z 4E6, Canada*
[5]*Department of Obstetrics and Gynecology, The Institute for Women's Health, 245 N. 15th St., Philadelphia, Pennsylvania 19102, USA*
**Jose-Miguel.Yamal@uth.tmc.edu*

Abstract: We examined intensity and shape differences in 378 repeated spectroscopic measures of the cervix. We examined causes of variability such as presence of precancer or cancer, pathologic tissue type, menopausal status, hormone or oral contraceptive use, and age; as well as technology related variables like generation of device and provider making exam. Age, device generation, and provider were statistically significantly related to intensity differences. Provider and device generation were related to shape differences. We examined the order of measurements and found a decreased intensity in the second measurement due to hemoglobin absorption. 96% of repeat measurements had classification concordance of cervical intraepithelial neoplasia.

©2010 Optical Society of America

OCIS codes: (120.0120) Instrumentation, measurement, and metrology; (170.0170) Medical optics and biotechnology; (300.0300) Spectroscopy.

References and links

1. J. Freeberg, D. Serachitopol, N. McKinnon, R. Price, E. N. Atkinson, D. D. Cox, C. MacAulay, R. Richards-Kortum, M. Follen, and B. Pikkula, "Fluorescence and reflectance device variability throughout the progression of a phase II clinical trial to detect and screen for cervical neoplasia using a fiber optic probe," J. Biomed. Opt. **12**(3), 034015 (2007).
2. N. Marín, N. MacKinnon, C. MacAulay, S. K. Chang, E. N. Atkinson, D. D. Cox, D. Serachitopol, B. Pikkula, M. Follen, and R. Richards-Kortum, "Calibration standards for multicenter clinical trials of fluorescence spectroscopy for in vivo diagnosis," J. Biomed. Opt. **11**(1), 014010 (2006).
3. S.B. Cantor, J.M. Yamal, M. Guillaud, D.D. Cox, E.N. Atkinson, J.L. Benedet, D. Miller, T. Ehlen, J. Matisic, D. van Niekerk, M. Bertrand, A. Milbourne, H. Rhodes, A. Malpica, G. Staerkel, S. Nader-Eftekhari, K. Adler-Storthz, M.E. Scheurer, K. Basen-Engquist, E. Shinn, L.A. West, A.T. Vlastos, X. Tao, J.R. Beck, C. MacAulay, and M. Follen, "Accuracy of optical spectroscopy for the detection of cervical intraepithelial neoplasia: testing a device as an adjunct to colposcopy," Int. J. Cancer (accepted for resubmission).
4. S. K. Chang, M. Dawood, G. Staerkel, U. Utzinger, E. N. Atkinson, R. Richards-Kortum, and M. Follen, "Fluorescence spectroscopy for cervical precancer detection: Is there variance across the menstrual cycle?" J. Biomed. Opt. **7**(4), 595–602 (2002).
5. D. D. Cox, S. K. Chang, M. Dawood, G. Staerkel, U. Utzinger, R. Richards-Kortum, and M. Follen, "Detecting the signal of the menstrual cycle in fluorescence spectroscopy of the cervix," Appl. Spectrosc. **57**(1), 67–72 (2003).
6. A. Nath, K. Rivoire, S. Chang, D. Cox, E. N. Atkinson, M. Follen, and R. Richards-Kortum, "Effect of probe pressure on cervical fluorescence spectroscopy measurements," J. Biomed. Opt. **9**(3), 523–533 (2004).

#129414 - $15.00 USD Received 2 Jun 2010; revised 5 Aug 2010; accepted 17 Aug 2010; published 19 Aug 2010
(C) 2010 OSA

7. K. Rivoire, A. Nath, D. D. Cox, E. N. Atkinson, R. Richards-Kortum, and M. Follen, "The effects of repeated spectroscopic pressure measurements on fluorescence intensity in the cervix," Am. J. Obstet. Gynecol. **191**(5), 1606–1617 (2004).
8. C. Brookner, U. Utzinger, M. Follen, R. Richards-Kortum, D. D. Cox, and E. N. Atkinson, "Effects of biographical variables on cervical fluorescence emission spectra," J. Biomed. Opt. **8**(3), 479–483 (2003).
9. C. Brookner, M. Follen, I. Boiko, J. Galvan, S. Thomsen, A. Malpica, S. Suzuki, R. Lotan, and R. Richards-Kortum, "Autofluorescence patterns in short-term cultures of normal cervical tissue," Photochem. Photobiol. **71**(6), 730–736 (2000).
10. N. Ramanujam, R. Richards-Kortum, S. Thomsen, A. Mahadevan-Jansen, M. Follen, and B. Chance, "Low Temperature Fluorescence Imaging of Freeze-trapped Human Cervical Tissues," Opt. Express **8**(6), 335–343 (2001).
11. R. Drezek, C. Brookner, I. Pavlova, I. Boiko, A. Malpica, R. Lotan, M. Follen, and R. Richards-Kortum, "Autofluorescence microscopy of fresh cervical-tissue sections reveals alterations in tissue biochmistry with dysplasia," Photochem. Photobiol. **73**(6), 636–641 (2001).
12. R. Drezek, K. Sokolov, U. Utzinger, I. Boiko, A. Malpica, M. Follen, and R. Richards-Kortum, "Understanding the contributions of NADH and collagen to cervical tissue fluorescence spectra: modeling, measurements, and implication," J. Biomed. Opt. **6**(4), 385–396 (2001).
13. R. Drezek, M. Guillaud, T. Collier, I. Boiko, A. Malpica, C. Macaulay, M. Follen, and R. Richards-Kortum, "Light scattering from cervical cells throughout neoplastic progression: influence of nuclear morphology, DNA content, and chromatin texture," J. Biomed. Opt. **8**(1), 7–16 (2003).
14. D. Arifler, I. Pavlova, A. Gillenwater, and R. Richards-Kortum, "Light scattering from collagen fiber networks: micro-optical properties of normal and neoplastic stroma," Biophys. J. **92**(9), 3260–3274 (2007).
15. I. Pavlova, K. Sokolov, R. Drezek, A. Malpica, M. Follen, and R. Richards-Kortum, "Microanatomical and biochemical origins of normal and precancerous cervical autofluorescence using laser-scanning fluorescence confocal microscopy," Photochem. Photobiol. **77**(5), 550–555 (2003).
16. J. S. Lee, O. Shuhatovich, R. Price, B. Pikkula, M. Follen, N. McKinnon, C. Macaulay, B. Knight, R. Richards-Kortum, and D. D. Cox, "Design and preliminary analysis of a study to assess intra-device and inter-device variability of fluorescence spectroscopy instruments for detecting cervical neoplasia," Gynecol. Oncol. **99**(3), S98–S111 (2005).
17. B. M. Pikkula, O. Shuhatovich, R. L. Price, D. M. Serachitopol, M. Follen, N. McKinnon, C. MacAulay, R. Richards-Kortum, J. S. Lee, E. N. Atkinson, and D. D. Cox, "Instrumentation as a source of variability in the application of fluorescence spectroscopic devices for detecting cervical neoplasia," J. Biomed. Opt. **12**(3), 034014 (2007).
18. B. Pikkula, D. Serachitopol, C. MacAulay, N. Mackinnon, J. S. Lee, D. D. Cox, E. N. Atkinson, M. Follen, and R. Richards-Kortum, "Multicenter clinical trials of in vivo fluorescence: are the measurements equivalent?" Proc SPIE 6430–64301Q (2007).
19. D. M. Gershensen, A. H. DeCherney, S. L. Curry, and L. Brubaker, *Operative Gynecology*, 2nd edition, (Saunders, 2001).
20. H. Zhu, and D. D. Cox, "A Functional Generalized Linear Model with Curve Selection in Cervical Pre-cancer Diagnosis Using Fluorescence Spectroscopy," Optimality: The Third Erich L. Lehmann Symposium **57**, 173–189 (2009).
21. H. Zhu, M. Vannucci, and D. D. Cox, "A Bayesian Hierarchical Model for Classification with Selection of Functional Predictors," Biometrics **66**(2), 463–473 (2010).

1. Introduction

The spectroscopic device that we analyze here is a candidate technology for automated detection of cervical cancer and could be used in a clinic to replace biopsies and permit diagnosis and treatment in a single visit.

Studies confirm that optical technologies can potentially provide a real-time diagnosis of tissue condition based on the molecular and morphologic changes associated with precancer. Inexpensive, small and portable optical sensors coupled with software for automated signal analysis could potentially yield an objective and reproducible diagnosis in the hands of the non-expert. Thus, the potential of optical technologies is enormous.

Many factors exist that possibly increase the measurement error including the coupling of the probe and tissue, environmental factors, and movement of the probe or of the tissue during the measurement process. The variability in the measurements can directly impact the classification accuracy of the device. It is therefore desirable to quantify the amount of variability and seek to identify ways to minimize this.

Although there have been many studies using optical spectroscopy, there is relatively little investigation into the variability within a patient and the biological and environmental factors

#129414 - $15.00 USD Received 2 Jun 2010; revised 5 Aug 2010; accepted 17 Aug 2010; published 19 Aug 2010
(C) 2010 OSA

that affect this. Our goal is to assess the variability between spectroscopic measurements taken at the same location of the cervix in the same patient and to identify factors that contribute to this variability. We then propose ways to minimize this variability in future studies.

2. Methods

Overview of Study Procedures

Details of the research grade, fiber-optic spectrometers used during the trial can be found in Freeberg et al. [1]. The devices measured fluorescence emission spectra at 16 different excitation wavelengths ranging from 330 nm to 480 nm and collected at a range of emission wavelengths between 360 nm to 800 nm. These data are referred to as an excitation-emission matrix (EEM). There were two generations of the device used during the seven years of the trial. The second-generation device improved over the first generation in that it was cheaper to construct and took less time to make measurements. The details of the processing of the data from these devices can be found in Marin, et al. [2].

Details of the study procedures can be found in Cantor, et al. [3]. Following colposcopic examination, a fiber optic probe 5.1 mm in diameter (2mm optically active center window) was advanced through the speculum and placed in gentle contact with the cervix. Spectroscopic measurements were obtained from one or two colposcopically normal cervical sites covered with squamous epithelium and, when visible, one colposcopically normal cervical site with columnar epithelium. If abnormalities were present and visible, measurements were taken from one or more colposcopically-abnormal sites. Thus, all patients had sampling of both abnormal and normal areas, if colposcopic abnormalities were present. Following spectroscopic measurements, all sites interrogated with the fiber optic probe were biopsied with a biopsy forceps yielding specimens that were 2 mm long by 1 mm wide by 1 mm deep, approximately the same volume interrogated by the probe. The histopathologic consensus diagnosis among pathologists was used as the gold standard for the trial.

Ten percent of all spectra, throughout the duration of the study, were explicitly repeated for purposes of investigating the variability from measurement to measurement. The probe left a 2mm circular impression on the cervix. The center of the probe has both the light-emitting and light-detection systems. The repeat measurement was registered to the first measurement by using the impression to direct the placement of the probe, approximately 30-60 seconds after the first measurement. The biopsy device was then used in the center of the circular impression so that the biopsy site was as close as possible to the spectroscopic site. After removing any pairs where at least one measurement did not pass quality assurance, we were left with 378 sites with repeat measurements (267 unique patients). 158 patients had repeat measurements at one site, 107 had repeat measurements at two sites, and two patients had repeat measurements at three sites.

Previous experiments examining the importance of the day of the menstrual cycle showed only that blood from menstruation affected measurement; thus, similarly to clinical practice, patients were rescheduled if menstruating [4,5]. Similarly, experiments examining probe pressure were conducted and showed no statistically significant effect from different degrees of pressure over a range of values that approximated those used by providers [6,7].

We were interested in seeing the possible effect of several factors on the variability between pairs of measurements at the same site. Factors that we believe might cause changes to the vasculature in the cervix and therefore cause variability in the spectroscopic measurements include 1) the severity of disease measured by the histologic grade of the biopsy into the three categories normal, low-grade squamous intraepithelial lesion (SIL), and high-grade SIL or cancer, 2) age, 3) colposcopic tissue type, using five categories ranging from squamous to columnar tissue, 4) menopausal status (pre-, peri-, or post-menopausal), and 5) oral contraceptive use [8–15]. Two additional factors that are technology related

#129414 - $15.00 USD Received 2 Jun 2010; revised 5 Aug 2010; accepted 17 Aug 2010; published 19 Aug 2010
(C) 2010 OSA

include the device/generation used (first or second generation), and the identification of which provider made the measurements (eight providers) [16–18].

Statistical Methods

Our aim is to characterize the similarity between the repeat measurements, and if they are not similar, then to determine if the differences are more in shape or intensity (or both). Let y_{ijk} = measurement for site i, with j denoting the order of the measurement (first or second), and k labeling the excitation-emission wavelength pair. We will denote the vector of all intensities for patient i and measurement j by $y_{ij\bullet}$. We define a relative measure of the squared distance between the two spectra:

$$d_i^2 = \frac{\left\| y_{i1\bullet} - y_{i2\bullet} \right\|^2}{\left\| y_{i1\bullet} \right\| \left\| y_{i2\bullet} \right\|}.$$

Note that this can be written

$$d_i^2 = \frac{\left\| y_{i1\bullet} \right\|^2 - 2y^T{}_{i1\bullet} y_{i2\bullet} + \left\| y_{i2\bullet} \right\|^2}{\left\| y_{i1\bullet} \right\| \left\| y_{i2\bullet} \right\|} = \frac{\left\| y_{i1\bullet} \right\|^2 + \left\| y_{i2\bullet} \right\|^2}{\left\| y_{i1\bullet} \right\| \left\| y_{i2\bullet} \right\|} - 2r_i = \left(\frac{\left\| y_{i1\bullet} \right\|^2 + \left\| y_{i2\bullet} \right\|^2}{\left\| y_{i1\bullet} \right\| \left\| y_{i2\bullet} \right\|} - 2 \right) + 2(1 - r_i).$$

Here $r_i = \frac{y_{i1\bullet}{}^T y_{i2\bullet}}{\left\| y_{i1\bullet} \right\| \left\| y_{i2\bullet} \right\|}$. The first term $I_i = \left(\frac{\left\| y_{i1\bullet} \right\|^2 + \left\| y_{i2\bullet} \right\|^2}{\left\| y_{i1\bullet} \right\| \left\| y_{i2\bullet} \right\|} - 2 \right)$ is a measure of the intensity difference. Note that $I_i \geq 0$ and $I_i = 0$ if and only if $\left\| y_{i1\bullet} \right\|^2 = \left\| y_{i2\bullet} \right\|^2$. Also, if $\left\| y_{i1\bullet} \right\| / \left\| y_{i2\bullet} \right\| \to \infty$ or $\to 0$ then $I_i \to \infty$ so the intensity difference depends on the ratio of the intensities and not strictly speaking on the differences. The second term in the expansion for d_i^2, namely $s_i = 2(1 - r_i)$ can be thought of as a measure of the difference in shapes. Note that r_i is similar to a correlation between the two EEMs. If the two EEMs have the same shape (i.e., one is a positive multiple of the other), then $s_i = 0$. It takes its largest value $s_i = 2$ if all excitation-emission wavelength pairs where one EEM is positive occur only where the other is 0. A higher value for either difference measure indicates a larger difference.

A linear mixed-effects model was used to compare the effect of covariates on the log of the difference measures. The $\log_{10}$ transformation was applied to the intensity and shape differences in order to satisfy the normality assumption in the model. The variables histologic grade, device generation (first or second generation), age, menopause status, provider, oral contraceptive use, and histologic tissue type (with five levels) were selected as predictors (although the device generation and provider identification were confounded). We included the variables that were previously found to be important for classification and included others, based on clinical judgement, that had the potential of affecting how the tissue responded to the probe pressure (or some other perturbation of the tissue due to the first measurement) and therefore affect the second measurement. For example, the histologic grade was chosen since there is increased neovasculature (and increased blood flow) with neoplasia. The increased blood flow in the area of measurement can increase variability of repeated measurements. Similarly, as tissue ages, there are changes in the elastin and collagen, leading to firmer tissue and less vessels present. Generally, with increased estrogenization, there is more blood flow (menopause status and oral contraceptive use). The histologic tissue type was chosen since there is entirely different vasculature in each tissue type.

The patient identifier was modeled as a random effect and restricted maximum likelihood was used to fit the model. The linear mixed-effects model can be represented as

#129414 - $15.00 USD Received 2 Jun 2010; revised 5 Aug 2010; accepted 17 Aug 2010; published 19 Aug 2010
(C) 2010 OSA

$$\log_{10} I_{pi} = \beta_0 + \sum_{m=1}^{7} \beta_m x_{mpi} + b z_p + \varepsilon_{pi},$$

where β_m and x_{mpi}, m = 1,…,7, represent the coefficients and variables of the predictors histologic grade, device generation, age, menopause status, provider, oral contraceptive use, and histologic tissue type for patient p and site i. The coefficient b and variable z_p correspond to the random effect of the patient identification, with $b \sim N(0, \sigma_p^2)$. This model produces a covariance matrix with compound symmetry structure. The same model was used for the shape difference measure.

Our second objective is to detect if there is an increase or decrease in intensities between the first and second measurements taken. Let $\delta_{ik} = y_{i2k} - y_{i1k}$ denote the difference between the first and second repeat measurements for patient i and excitation-emission pair k. We will denote $\delta_{\bullet k}$ as the vector of differences $\delta_{ik} \forall i$ and $\bar{\delta}_{\bullet k}$ as the mean of the vector $\delta_{\bullet k}$. We calculated a z score defined by

$$z_k = \frac{(\bar{y}_{\bullet 2k} - \bar{y}_{\bullet 1k})}{\frac{1}{\sqrt{N}}[\frac{1}{N-1}\sum_{i=1}^{N}(\delta_{ik} - \bar{\delta}_{\bullet k})^2]^{1/2}},$$

taking the difference of the means of the excitation-emission pair across all N pairs of EEMs and dividing by the standard error of the difference of each pair for that excitation-emission pair. The z-score is a dimensionless standardized score that gives information about how many standard deviations an observation is above or below the mean.

To determine which excitation-emission pairs had the most variability between repeat measurements, we computed a measure of the variance for every excitation-emission pair. For every excitation-emission pair k, we computed the standard deviation of the difference between the two repeated measures among all EEMs, $[\frac{1}{N-1}\sum_{i=1}^{N}(\delta_{ik} - \bar{\delta}_{\bullet k})^2]^{1/2}$. This was then standardized by dividing by the mean intensity for that excitation-emission pair.

Statistical analysis was performed using the statistical packages R version 2.6.2 (R Foundation for Statistical Computing, Vienna, Austria) and JMP version 7.0.1 (SAS Institute, Cary, North Carolina, USA). Confidence intervals for proportions were calculated using the exact binomial test.

3. Results

We had repeat measurement data on 378 sites. The distributions of the covariates are shown in Tables 1–7.

Table 1. Table of the distribution of the histologic grade of the biopsies

Histologic grade of biopsy		
normal	Low-grade SIL	High-grade SIL or cancer
201	116	61

Table 2. Table of the age distribution

Age					
minimum	1st quartile	median	mean	3rd quartile	Max
18	28	36	38.05	47	70

#129414 - $15.00 USD Received 2 Jun 2010; revised 5 Aug 2010; accepted 17 Aug 2010; published 19 Aug 2010
(C) 2010 OSA

Table 3. Table of the distribution of the pathology tissue type. The pathology tissue types range from ecto-cervical to endo-cervical, coded numerically as one to five

Pathology tissue type				
Ecto-cervical	Primarily ecto-cervical	Both endo- and ecto-cervical	Primarily endo-cervical	Endo-cervical
177	36	109	16	40

Table 4. Table of the distribution of the menopausal status

Menopausal status		
Pre-menopausal	Peri-menopausal	Post-menopausal
295	19	64

Table 5. Table of the distribution of the generation of the device used to obtain the spectroscopic measurements

Device generation	
1st generation	2nd generation
204	174

Table 6. Table of the distribution of the provider obtaining the measurements

Provider							
1st	2nd	3rd	4th	5th	6th	7th	8th
39	64	113	11	36	14	37	64

Table 7. Table of the distribution of the oral contraceptive use

Oral contraceptive use or HRT use	
No	Yes
283	95

We computed the overall difference for the 378 sites that had a repeated measurement. The overall difference measure is right-skewed, with most of the points close to 0. The $\log_{10}$ transformation is approximately normally distributed (Fig 1). The median overall difference of d_i^2 is 0.12, a median of $\sqrt{0.12}$ = 35% difference between the repeat measurements. The range of the d_i^2 values was 0.0005 to 16.8 (1st quartile 0.003, 3rd quartile 0.32).

We present separate analyses for the intensity and shape differences.

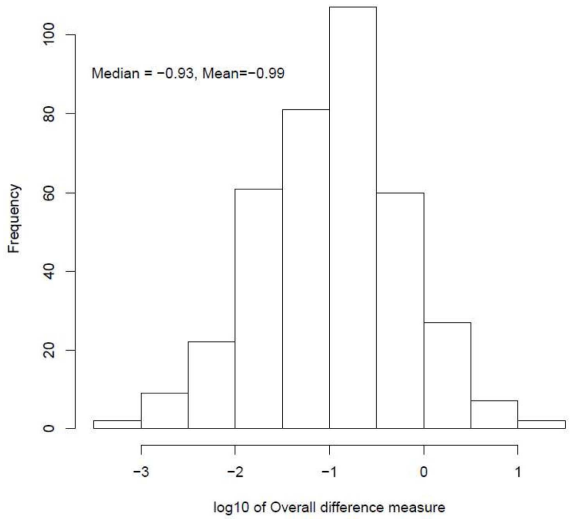

Fig. 1. Histogram of the overall $\log_{10}$ difference between repeat EEMs showing a median of 35% difference.

#129414 - $15.00 USD Received 2 Jun 2010; revised 5 Aug 2010; accepted 17 Aug 2010; published 19 Aug 2010
(C) 2010 OSA

Intensity difference

The intensity difference component of the overall difference measure had a median of 0.08. Some examples of the lowest, median, and highest intensity differences are shown in Fig. 2. Each plot is a pair of measured EEMs taken at the same site of a patient. Each EEM matrix has been concatenated into a single vector, by excitation wavelength, for all excitation-emission pairs – the resulting plotted line has 16 modes for the 16 excitation wavelengths. The measurement pairs with the highest intensity difference identified pairs where one measurement possibly should not have passed quality control. The median values identified pairs that had a moderate intensity difference at some excitation wavelengths.

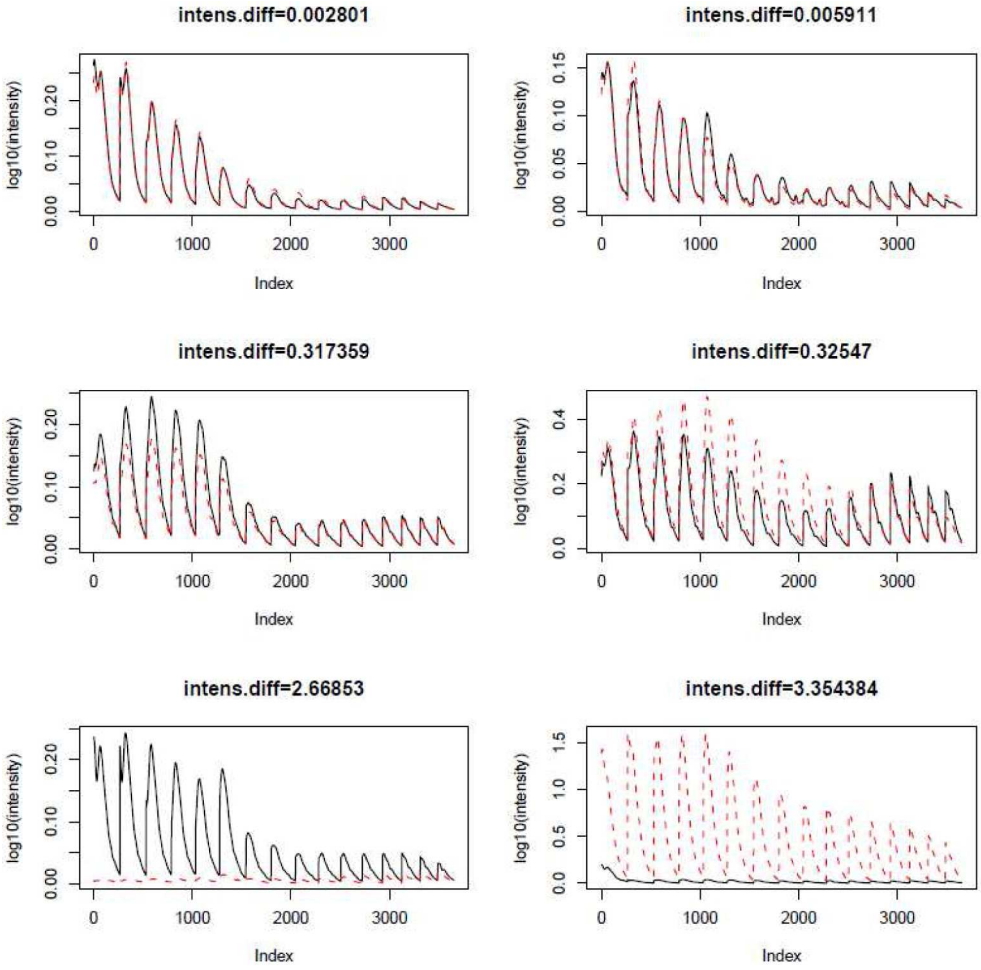

Fig. 2. Examples of repeat EEMs that have the lowest intensity difference (first row), median intensity differences (second row), and highest intensity difference between them (third row). The black (solid) and red (dotted) lines denote the first and second EEMs at the same barcode and clock position, respectively. Each EEM matrix was concatenated (by the 16 excitation wavelengths) to form a vector of excitation-emission pairs.

Linear mixed-effects model results for intensity difference

The variables device generation and provider identification were confounded; hence, the linear mixed-effects model failed to converge. When the device generation variable was removed from the model, both patient age and provider identification were statistically significant in the linear mixed-effects model ($p = 0.047$ and $p = 0.023$, respectively). When

#129414 - $15.00 USD Received 2 Jun 2010; revised 5 Aug 2010; accepted 17 Aug 2010; published 19 Aug 2010
(C) 2010 OSA

the provider identification variable was removed from the model, age was marginally significant (p = 0.057) and device generation was highly statistically significant (p = 0.001). Figures 3–5 show plots of intensity difference versus these statistically significant factors.

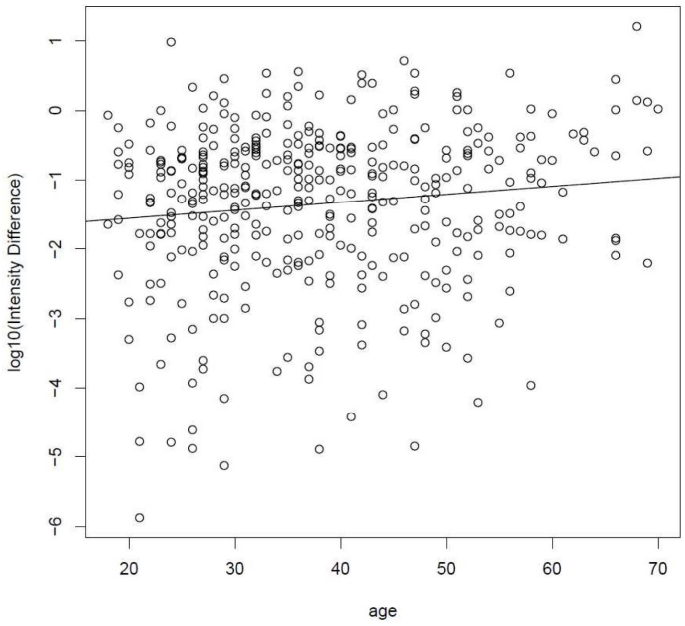

Fig. 3. Scatter plot of $\log_{10}$ of intensity difference between repeats of EEMs versus age. This shows an increase in the intensity difference for older women.

A regression line is drawn on the scatterplot, showing a slight upward trend in the intensity difference as age increases.

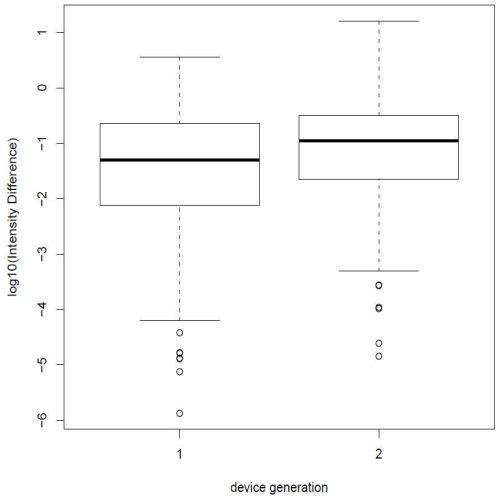

Fig. 4. Device generation versus $\log_{10}$ intensity difference between repeats of EEMs.

#129414 - $15.00 USD Received 2 Jun 2010; revised 5 Aug 2010; accepted 17 Aug 2010; published 19 Aug 2010
(C) 2010 OSA

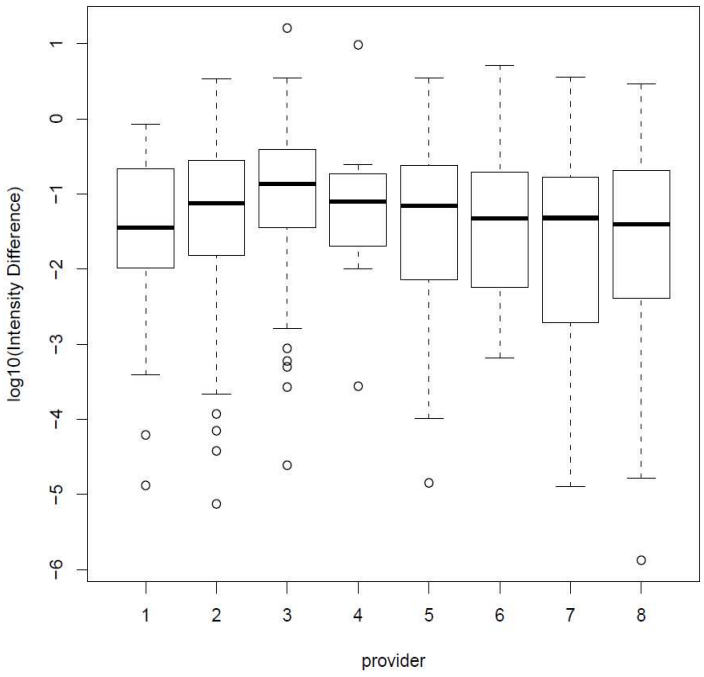

Fig. 5. Boxplots of $\log_{10}$ (intensity difference) by provider.

Shape difference

The shape measure is plotted for the two EEM pairs with the lowest values, median values, and with the highest values (Fig. 6). To aid the visual comparison in these graphs, we normalized shape log intensities by dividing by the square root of the sum of squares for that observation. The black (solid) and red (dotted) lines are the concatenated vector of $\frac{y_{i1\bullet}}{\|y_{i1\bullet}\|}$ and $\frac{y_{i2\bullet}}{\|y_{i2\bullet}\|}$, respectively.

Linear mixed-effects model results for shape difference

Again, variables device generation and provider identification were confounded. We fit the model omitting each variable one at a time. When the provider variable was omitted, the device generation variable was statistically significant ($p = 0.0002$, Fig. 7). When the device generation variable was omitted, the provider identification variable was statistically significant ($p<0.0001$, Fig. 8). There is more of a shape difference for the second-generation device than the first generation. There is more of a shape difference for providers 3, 4, and 8. There does not appear to be any association with the number of measurements each provider obtained (Table 6).

#129414 - $15.00 USD Received 2 Jun 2010; revised 5 Aug 2010; accepted 17 Aug 2010; published 19 Aug 2010
(C) 2010 OSA

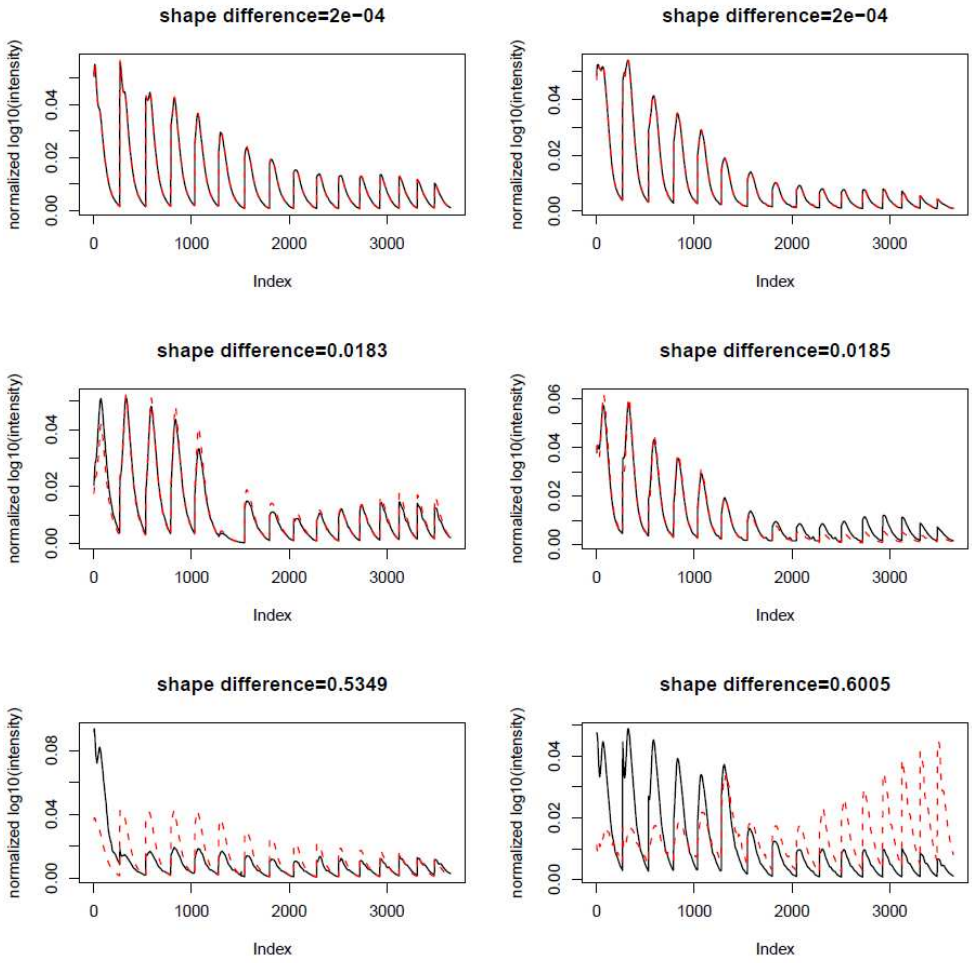

Fig. 6. Examples of repeat EEMs which have the lowest (top row), median (middle row), and highest (bottom row) shape difference between them. The black (solid) and red (dotted) lines denote the first and second EEMs at the same barcode and clock position, respectively. Each EEM matrix was concatenated (by the 16 excitation wavelengths) to form a vector of excitation-emission pairs. A low shape difference value denotes two EEMs that have similar shape.

#129414 - $15.00 USD Received 2 Jun 2010; revised 5 Aug 2010; accepted 17 Aug 2010; published 19 Aug 2010
(C) 2010 OSA

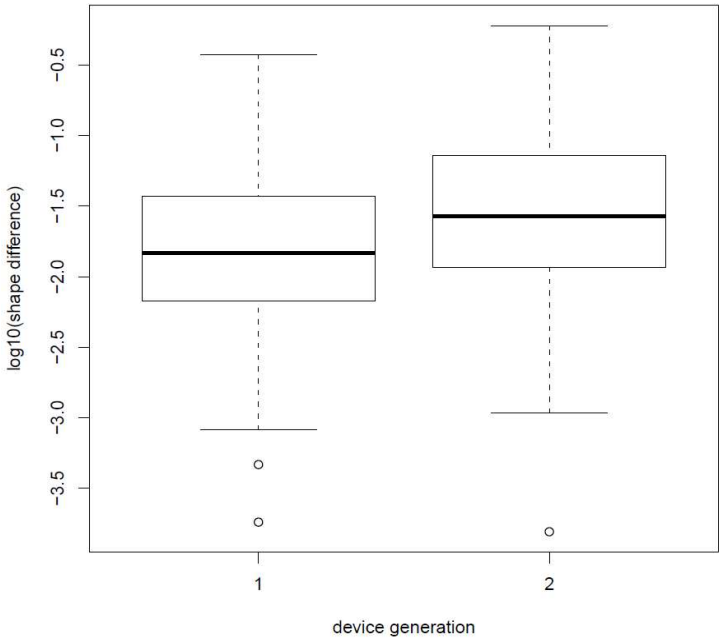

Fig. 7. Boxplots of $\log_{10}$ shape difference by device generation. There is more of a shape difference for pairs of EEMs in the 2nd generation device.

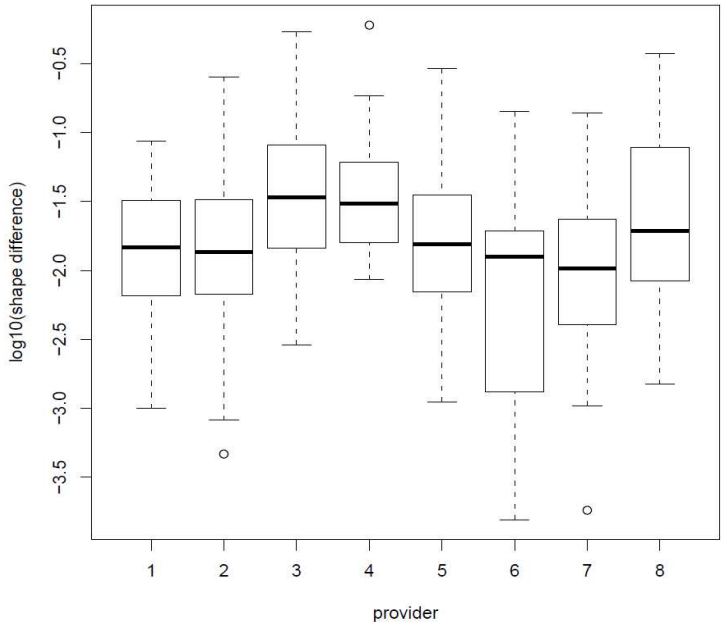

Fig. 8. Boxplots of $\log_{10}$ (shape difference) by provider.

Effect of the order the measurements were taken

The process of obtaining a measurement can cause temporary changes in the tissue. Our conjecture was that a second reading done right after the first might be influenced by possible after-effects on the tissue from having just had the probe pressed against the tissue or the

#129414 - $15.00 USD Received 2 Jun 2010; revised 5 Aug 2010; accepted 17 Aug 2010; published 19 Aug 2010
(C) 2010 OSA

recent excitement of the fluorescence tissue. When applied to all pairs of EEMs, the z values of the difference between the first and second measurements contained no outliers, yet were overwhelmingly negative (range from −1.92 to 0.08). Figure 9 shows a heat map and contours of the EEM z values.

Since we found statistically significant differences between the two generations of device, we stratified the analysis separately by each generation of device.

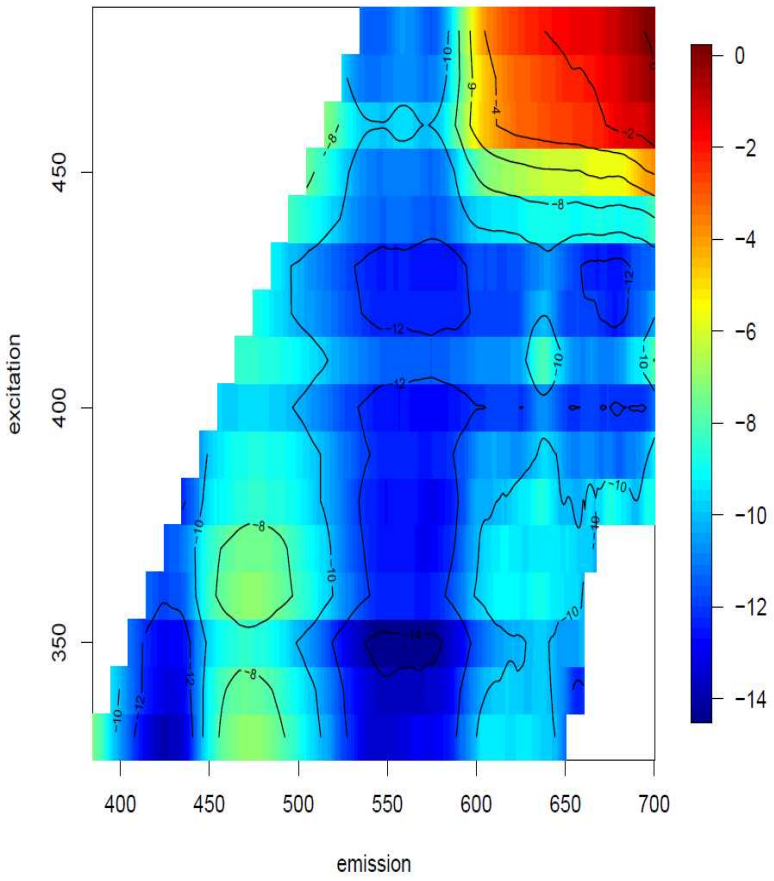

Fig. 9. Plot of z values for 2nd measurement - 1st measurement.

The mean of the z values decreased from the first to the second measurement.

#129414 - $15.00 USD Received 2 Jun 2010; revised 5 Aug 2010; accepted 17 Aug 2010; published 19 Aug 2010
(C) 2010 OSA

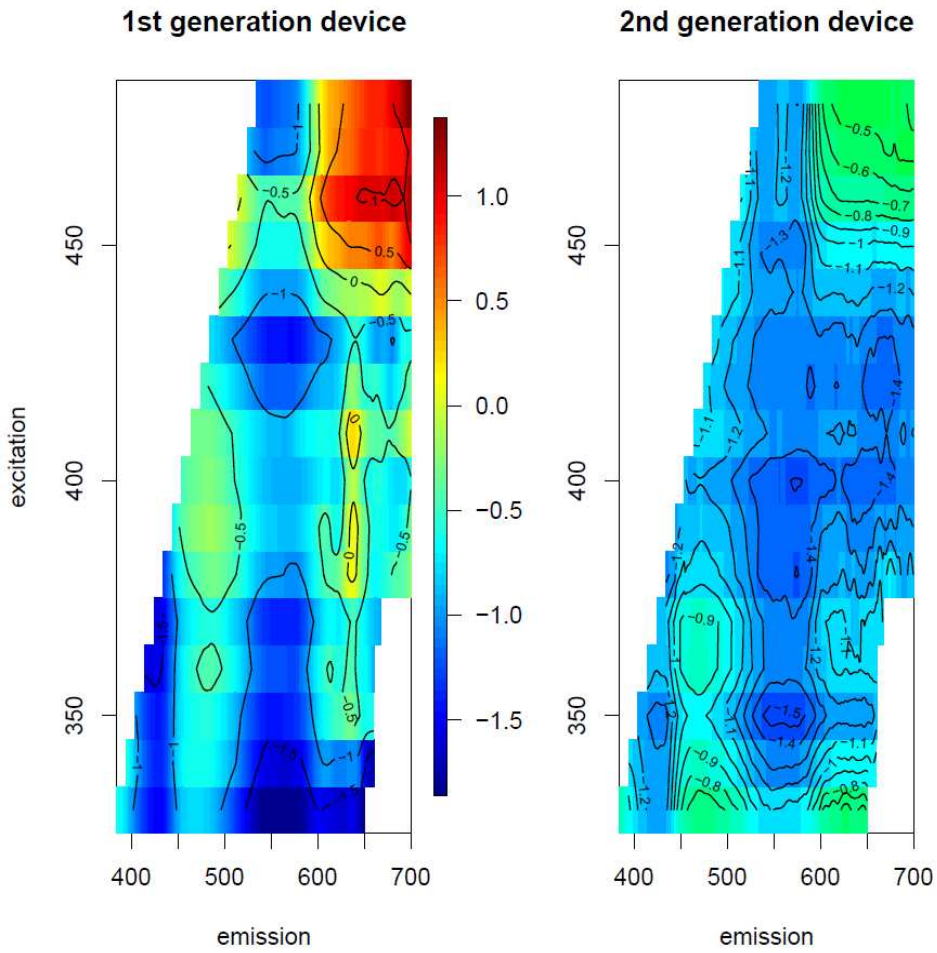

Fig. 10. Plot of z values for (2nd measurement - 1st measurement) by device generation.

We observed a difference between the first and second generation devices (Fig. 12). The second generation device tended to have a smaller difference between the first and second measurements, yet the differences were mostly negative indicating a decrease in intensity for the second measurement. The majority of the first generation device z values were negative – most EEM pairs z values decreased in the second measurements except for regions with high excitation and emission values.

Excitation-Emission Pairs having the most amount of variation

Figure 11, an EEM plot of the standard deviation between the first and second measurements, shows that the greatest standard deviation is around excitation 410, and emission 460. The standard deviation is not constant throughout the EEM but increases around the area where hemoglobin absorption plays a role in the intensities.

#129414 - $15.00 USD Received 2 Jun 2010; revised 5 Aug 2010; accepted 17 Aug 2010; published 19 Aug 2010
(C) 2010 OSA

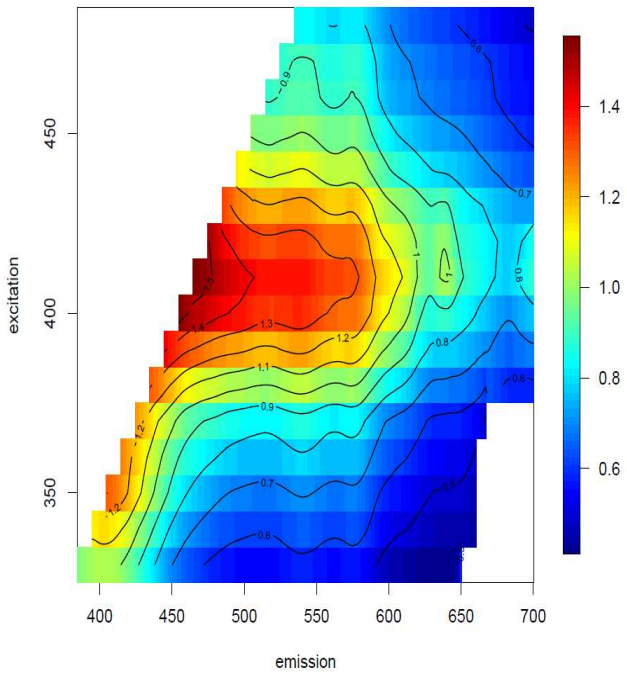

Fig. 11. EEM plot of the normalized standard deviation between the 1st and 2nd measurements.

Repeatability and concordance

To further assess how repeatability affected classification, we examined the classification concordance of the repeat measurements given the classification algorithm developed in [3]. We calculated the percentage of repeat measures that were classified into different classes (low-grade SIL or better versus high-grade SIL or cancer) than its repeat measurement. Overall, 4% of the repeat measurements were not concordant. The same was done for repeat pairs in each quartile of the intensity difference measure and the shape difference measure. The barplots for intensity difference and shape difference are shown in Figs. 12 and 13, respectively. There is no obvious trend between the percentage of concordance and the quartiles of the difference measures.

#129414 - $15.00 USD Received 2 Jun 2010; revised 5 Aug 2010; accepted 17 Aug 2010; published 19 Aug 2010
(C) 2010 OSA

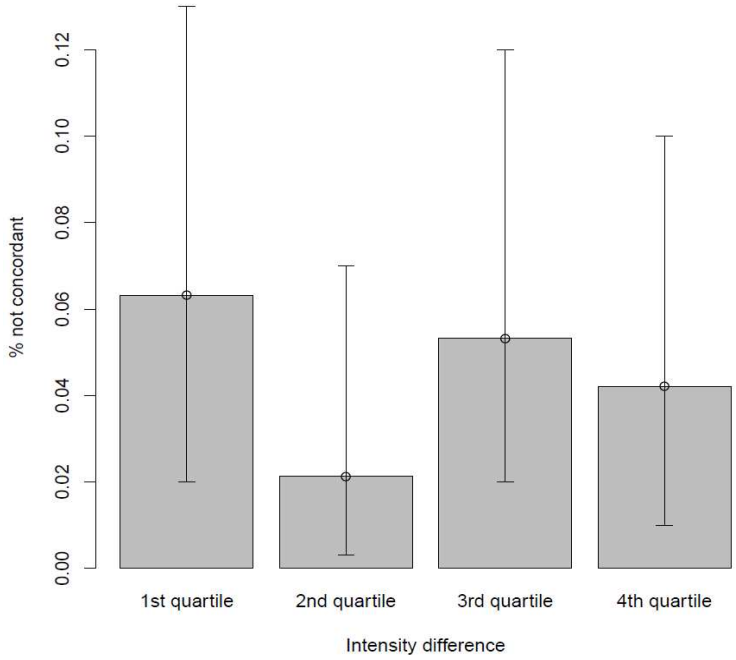

Fig. 12. Barplot of percentage of repeat measurements whose classification class was not concordant by intensity difference. The error bar gives the 95% confidence interval.

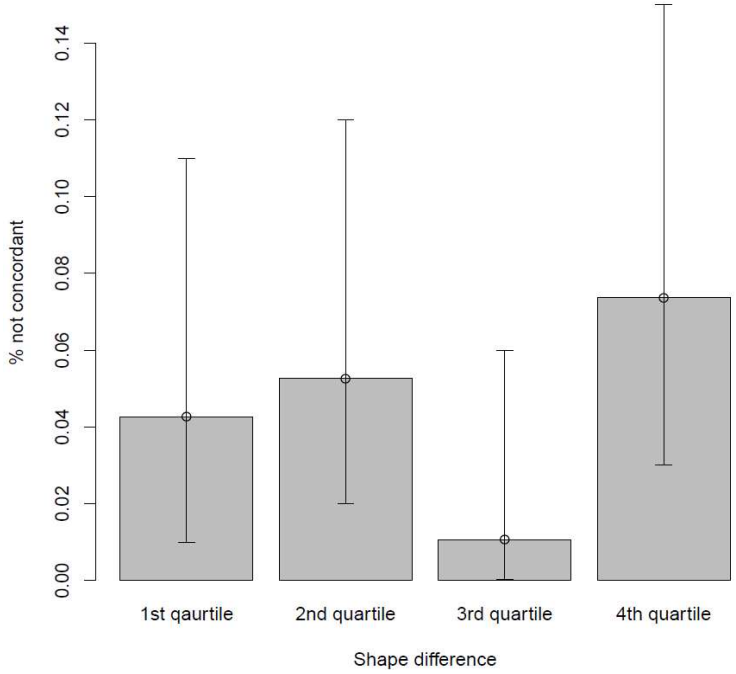

Fig. 13. Barplot of percentage of repeat measurements whose classification class was not concordant by shape difference. The error bar gives the 95% confidence interval.

#129414 - $15.00 USD Received 2 Jun 2010; revised 5 Aug 2010; accepted 17 Aug 2010; published 19 Aug 2010
(C) 2010 OSA

4. Conclusions

Several factors influenced the repeatability of the EEMs. Measurements taken at the same location of the same patient should look identical yet many have substantial intensity and shape differences between them. The generation of device, age, and provider identification are associated with the degree of intensity difference. The generation of device and provider are associated with the shape differences.

We observed more differences for the second generation device than in the first generation device. All first generation device measurements were taken in Houston and all second generation device measurements were taken in Vancouver, causing confounding of what the source of the observed differences is. The difference could include engineering differences between the two generations of devices and differences between providers or test conditions. The second generation device had the advantage of being able to obtain measurements more quickly and the device itself was less expensive to construct. However, each generation of device has a different way of resetting the filter wheels before acquiring another measurement. We will examine the possible filter wheel effect in a forthcoming manuscript. Providers 3-6 were in Vancouver and the rest were in Houston. The Vancouver measurements were dominated by Provider 3, having the lion's share of repeat measurements ($n = 113$) and the highest shape difference among all providers. This is a possible explanation for the observed generation difference.

Many additional factors could potentially explain the provider differences. Each provider used varying amounts of pressure when pressing the probe on the tissue and the imprint on the tissue could help direct the placement of the probe in the same location for the second measurement. We are surprised by the provider effect because previous studies of pressure didn't find any statistically significant differences between different pressure amounts. We are going to re-explore this issue in future studies. Steadiness of the hand might also influence the repeatability, although the second generation device, which obtained measurements faster than the first, exhibited more variation. We thought that having to hold the probe to the tissue for less time would result in more repeatable measurements. Interestingly, we don't think that experience influenced the provider differences. The providers that had the largest difference between their first and second measurements had widely varying levels of experience obtaining spectroscopic measurements in our study.

As women age, there is less estrogen in cervical tissue and this leads to atrophy. The atrophic cervix is often firmer than the well-estrogenized cervix of younger women [19]. We have not yet sorted out what changes are due to age and what are due to decreased estrogen. We hypothesize that the increase in intensity difference is due to gradual changes in the epithelium as the woman becomes postmenopausal. There was a slightly larger intensity difference for postmenopausal women on hormone replacement therapy than postmenopausal women not on hormone replacement therapy ($p = 0.08$).

We observed a decrease in EEM intensities from the first to the second measurements. We hypothesize that this is due to increased hemoglobin absorption (Figs. 9 and 10). We note more of a difference in the areas of the EEM matrix where blood is absorbed, mainly around excitation wavelengths 420-430 and separately at the band at emission wavelengths 400-450 nm. We believe that the compression of the tissue from the first measurement leads to revascularization of the tissue, causing more hemoglobin absorption during the second measurement. In our study, the second measurement was taken 30-60 seconds after the first, which may have caused increased hemoglobin absorption in the second measurement. Systematic differences between the first and second measurement are hypothesized to be an artifact of making an additional measurement rather than of the repeatability of the device. Classification algorithms that either use only one measurement at the same site or use excitation-emission pairs that are invariant to hemoglobin absorption would not be affected by this artifact.

#129414 - $15.00 USD Received 2 Jun 2010; revised 5 Aug 2010; accepted 17 Aug 2010; published 19 Aug 2010
(C) 2010 OSA

The most amount of variation in the EEM occurred around excitation 410 and emission 460. Interestingly, Zhu demonstrated that the most important excitation wavelengths for classification using these devices are, in order, 340 nm, 460 nm, 420 nm, and 410 nm [20]. The two most important wavelengths have a low amount of variability. In another anlaysis, the two most important excitation wavelengths are 360 nm and 400 nm [21]. The second most important wavelength has a high degree of variability although it still retained classification concordance.

Classification algorithms depend on clean, reproducible data to be able to accurately predict independent data. The degree of repeatability of new devices can be a major source of variation. If the device variation exists in areas of the variable space that is being used by a classifier, this decrease in repeatability can adversely affect the accuracy to detect disease. We found that the top wavelengths that have been used in classifiers have a relatively low variance between repeat measurements – however, many excitation wavelengths with high variance have also been found to be predictive of disease. Minimizing variance by controlling factors that might introduce additional variability might help increase the predictive accuracy of classification algorithms.

Although we observed a 35% overall median difference between the repeat measurements, only 4% were classified into different classes than their repeat pair by the classification algorithm. Several factors influenced the repeatability. However, the high concordance suggests high utility/repeatability of spectroscopy for the classification of cervical neoplasia.

The contribution of this manuscript is twofold. First, it proposes methodology to assess the repeatability of spectroscopic data, with respect to covariates. We are not aware of other work that uses a similar approach to characterize shape and intensity differences of functional data. Second, repeatability studies of emerging technologies, incorporating patient covariate information, are extremely important in the transition from a research device to a usable device. Emerging devices should be repeatable, especially for classification. If measuring at roughly the same location results in wildly varying measurements, it would have important implications about the utility of the device. Our clinical trial was designed with a repeatability study built-in, and we recommend this type of study design for emerging technologies.

Acknowledgements

This work was supported by the National Cancer Institute, grant P01-CA-82710-09.

#129414 - $15.00 USD Received 2 Jun 2010; revised 5 Aug 2010; accepted 17 Aug 2010; published 19 Aug 2010
(C) 2010 OSA