Detecting the Signal of the Menstrual Cycle in Fluorescence Spectroscopy of the Cervix

DENNIS D. COX,* SUNG K. CHANG, M. YUSOFF DAWOOD, GREGG STAERKEL, URS UTZINGER, REBECCA R. RICHARDS-KORTUM, and MICHELE FOLLEN

Department of Statistics, Rice University, Houston, Texas 77005 (D.D.C.); Biomedical Engineering Program, University of Texas, Austin, Texas 78712 (S.K.C., U.U., R.R.R.-K., M.F.); Department of Obstetrics, Gynecology and Reproductive Sciences, The University of Texas Health Sciences Center, Houston, Texas 77030 (M.Y.D., M.F.); Department of Anatomic Pathology, University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030 (G.S.); and Department of Gynecologic Oncology, University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030 (M.F.)

Fluorescence spectroscopy of the cervix has been shown to be an effective noninvasive diagnostic tool for cervical intraepithelial neoplasia (precancer). To assess the effect of the menstrual cycle on fluorescence spectroscopy, daily measurements were made on ten subjects for the length of their cycle. These measurements were analyzed to determine if there was a statistically significant signal associated with the menstrual cycle. A signal was found for emission wavelengths between 425 and 445 nm inclusive—near the main hemoglobin absorption band, the Soret band, at 420 nm. We suspect that the slight displacement of the Soret band is due to the nearby dominant NAD(P)H peak, which increases the signal-to-noise ratio and affects statistical significance. The signal consists of a reduction in fluorescence intensity for the first few days of the cycle. This analysis indicates that hemoglobin absorption is the main menstrual-cycle effect on the use of fluorescence spectroscopy on the cervix. The effect is confined to a small set of excitation/emission wavelengths and to approximately the first 8 days of the cycle. This suggests that any problems from the menstrual cycle can be avoided with a simple requirement that the device not be used during the period of menstrual bleeding.

Index Headings: **Fluorescence spectroscopy; Cervical intraepithelial neoplasia; Menstrual cycle; Analysis of variance; Randomization tests; Multiple comparisons.**

INTRODUCTION

Fluorescence spectroscopy (FS) has been shown to be effective for diagnosing cancerous and precancerous lesions.[1–7] Each value in an FS measurement is associated with an excitation wavelength that excites the fluorophores in the tissue and an emission wavelength corresponding to the emitted light. This gives a two-dimensional array of data known as an excitation emission matrix (EEM). Our research group has demonstrated that FS is a promising tool for detecting cervical intraepithelial neoplasia (CIN).[8–10] Because the biological basis for the success of FS in cancer detection is not well understood, there is a concern that natural biological variability may present difficulties for this technology. Indeed, it appears that age and perhaps other factors do have a large effect.[11] Of particular concern for the application of FS to the cervix are changes that take place during the menstrual cycle. To assess the effect of the menstrual cycle on FS, ten patients were measured daily throughout approximately 30 consecutive days of their cycle.

These data have been analyzed in a separate manuscript.[12] There, a principal component was derived that appeared to show a strong correlation with the menstrual cycle. Also, the redox ratio (a ratio of FS intensities at two wavelength combinations) seemed to have a relationship with the menstrual cycle. However, these results were not tested for statistical significance.

In this article we subject the data to a rigorous analysis to determine if there is a statistically significant effect. The primary analysis is on the individual fluorescence intensity measurements at each excitation/emission wavelength combination, although the redox ratio and overall intensity are analyzed as well. A nonparametric technique was used and corrections for the multiple comparisons were included. Examination of the wavelength combinations where a significant signal was found revealed that they formed a region of contiguous pixels in the EEM. This region is near a hemoglobin absorption band, and the estimated signal in the region showed a lower intensity during the early part of the menstrual cycle. This suggests that the predominant signal of the menstrual cycle is associated with hemoglobin absorption during the first few days of the cycle when the patient has blood present on the cervix. This is potentially an important result for the medical application of the device, since any adverse effect this may have on using FS for diagnostic purposes can be eliminated by specifying that FS not be used when the patient may have menstrual bleeding.

MATERIALS AND METHODS

Subjects. The Institutional Review Board at the University of Texas M.D. Anderson Cancer Center and the University of Texas at Austin reviewed and approved the study protocol. Eligibility was limited to patients over the age of 18 who were not pregnant, had a history of normal menstrual cycles, and no history of abnormal Pap smears. Patients underwent a number of interviews and tests detailed elsewhere[12] and were asked particularly about their menstrual history and the date of their last menstrual period.

One of the investigators, Dr. Michele Follen, chose three colposcopically normal sites on each patient, and fluorescence EEMs were measured daily on these three sites for at least one full menstrual cycle. Both squamous and columnar epithelia were measured. But since relatively few columnar sites were measured (three, one each on patients 2, 6, and 10)—and since they react differently to FS and the hormonal changes of the menstrual cycle—

Received 10 October 2001; accepted 28 August 2002.
* Author to whom correspondence should be sent.

0003-7028/03/5701-0067$2.00/0

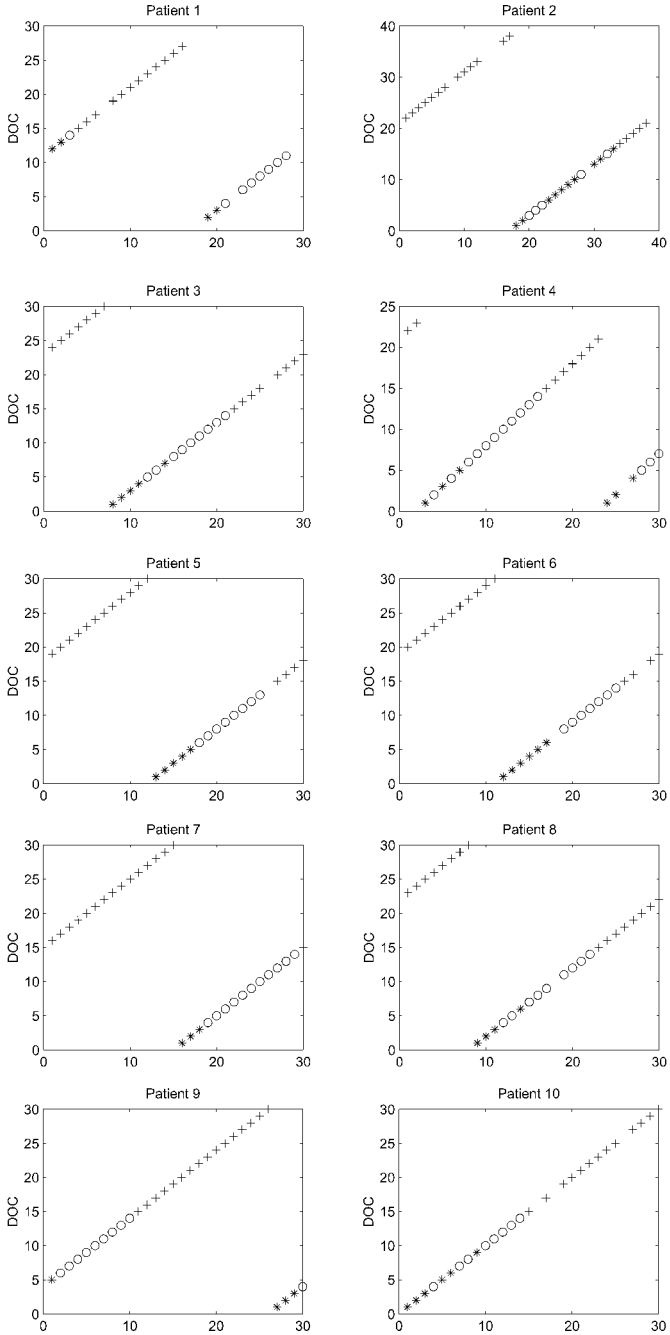

FIG. 1. Plot of day of cycle (DOC) by day of measurement for patients 1 through 10. The proliferative phase of the menstrual cycle is indicated by a circle (non-bleeding day) or an asterisk (a day when bleeding was observed) and the secretory phase by a cross. For instance, the first measurement day of patient 1 was the eleventh DOC, which was in the proliferative phase and there was bleeding, and a new cycle began on the nineteenth measurement day.

they were deleted from the analysis. A research nurse asked each patient daily whether she had any menstrual bleeding, and the nurse practitioner or physician making the measurements verified the patients' answers. Each date was assigned to a menstrual day of the cycle (DOC). Ovulation was assumed to occur at DOC 14; preceding days were assigned to the proliferative phase and following days to the secretory phase of the cycle. A graphical summary of the menstrual cycles by patient is shown in Fig. 1. The mean length of cycle observed for the ten patients was 29.8 days, with a minimum of 23 days and a maximum of 38 days. Seven patients had 30-day cycles and one had a 27-day cycle. The patient with the 23-day cycle was determined to be anovulatory.

Instrumentation and Measurements. The spectroscopic system used to measure fluorescence EEMs is described in more detail elsewhere.[12] Briefly, the system measures fluorescence emission spectra with a resolution of 5 nm at 16 excitation wavelengths, ranging from 330 to 480 nm in 10 nm increments. The system incorporates a fiber optic probe, a Xenon arc lamp coupled to a monochromator to provide excitation light, and a polychromator and thermo-electrically cooled charge-coupled device (CCD) camera to record fluorescence intensity as a function of emission wavelength. The initial and final emission wavelengths varied across the excitation wavelengths, and each spectrum had a total of 703 excitation/emission wavelength combinations where measurements were made.

Between uses, the probe was sterilized for 25 min in a Cydex solution, rinsed with water, and wiped with a sterile alcohol wipe. The probe was guided into the vagina and its tip was positioned flush with the cervical epithelium. Acetic acid is normally used during a colposcopic examination because it enhances the optical differences between normal and dysplastic tissue.[14] But an initial investigation determined it would be too irritating to the patients for daily application, and it was not used in this study.

The measured EEMs were subjected to a rigorous quality assurance inspection by three separate investigators. Only those EEMs determined to be the result of a faulty measurement process were excluded. The primary cause of a faulty measurement process is probe slippage. This is frequently detected by the provider making the measurements, and those observations were redone at the time. However, there were a number of EEMs where the provider did not note slippage, but it looked as if slippage had occurred. The typical feature of an EEM where slippage has occurred consists of a clearly anomalous emission spectrum for 1 or 2 contiguous excitation wavelengths. There were a very small number of EEMs that were clearly anomalous for unknown reasons, and they were deleted. Also, 83 EEMs that were from columnar tissue sites were deleted from the analysis. A total of 765 EEMs were included in the analysis presented here. The median EEM (i.e., the median fluorescence intensity at each excitation/emission wavelength combination) is shown in Fig. 2.

Data Analysis. The data were first preprocessed and then statistical inference was performed. The preprocessing incorporated the following steps:

(1) A few EEMs in the analysis had exact zeros. As we intended to take logarithms, these were removed by replacing a zero with one half of the smallest observed positive value of intensity.
(2) Each fluorescence intensity was normalized by dividing by an estimate of the peak value for that individual EEM. Normalization has several advantages, such as correcting for the tendency of fluorescence intensity to increase with patient age as previously observed.[11] We have found that such

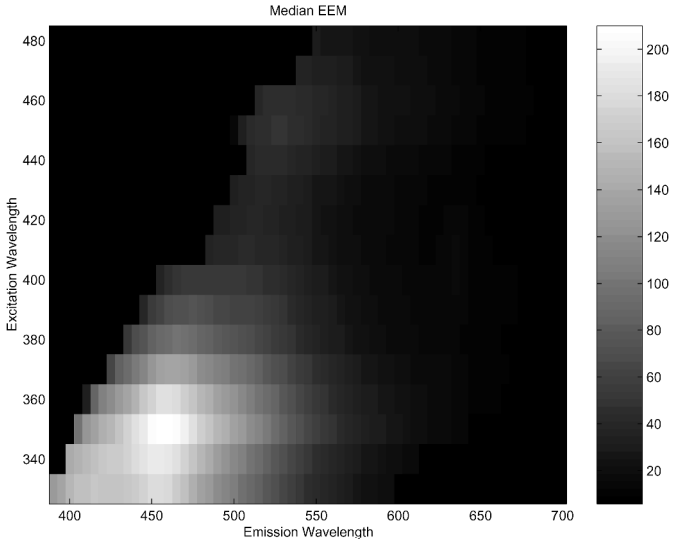

FIG. 2. Plot of the median EEM. The x-axis is emission wavelength in nanometers and the y-axis is excitation wavelength in nanometers. The high peak in the lower left is the result of the fluorophore NAD(P)H and collagen crosslinks.

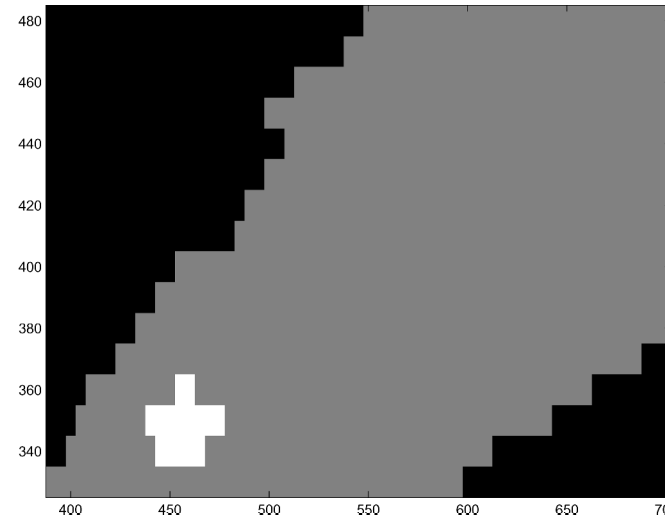

FIG. 3. Plot of the peak region used for normalization. The x-axis is emission wavelength in nanometers and the y-axis is excitation wavelength in nanometers. The area shown in white corresponds to the median EEM values above 190 in Fig. 2. The areas shown in white or gray correspond to the excitation/emission wavelength combinations actually used in the measured EEMs. The black area is not part of the measured EEM.

normalization is useful and apply it on a routine basis. The nominal peak value for normalization was taken as the average of the 19 intensity values at the excitation/emission wavelength combinations where the median intensity in Fig. 2 exceeded 190. A picture of this peak area of 19 excitation/emission wavelength combinations is shown in Fig. 3. There were many EEMs where the actual peak was not in this region. Since the normalizing value is the average of 19 intensities, it would never be the actual peak value in a given EEM (the maximum intensity will be larger than an average intensity), but it does reduce the noise somewhat to take an average.

(3) This normalizing value (the average of the intensities at the 19 excitation/emission wavelength combinations shown in Fig. 3) was kept as one of the variables of the analysis. One other variable that was added was the redox ratio.[12] The redox ratio R is the ratio of fluorescence intensity at 370 nm excitation, 530 nm emission (the approximate location of a peak associated with the fluorophore FAD) to the sum of intensities at 370 nm excitation, 530 nm emission (FAD) and 370 nm excitation, 450 nm emission (the approximate location of the NAD(P)H peak). In symbols, R = FAD/(FAD + NAD(P)H).

(4) After normalization, logarithms were taken. The logarithmic transformation removes some of the strong skewing in the histograms of intensities at a given excitation/emission wavelength combination. Other transformations (square root, reciprocal, and no transformation) were tried, but we judged that the logarithm gave the best results. We denote the values used in the analysis by:

x_{wtij} = logarithm of value (normalized spectrum value, normalization value, or redox ratio);
w = index for excitation/emission wavelength combination ($w < 704$) or normalizing value ($w = 704$), or the redox ratio ($w = 705$);
t = measurement day number, $1 \leq t \leq T_i$, where T_i is the number of measurements for patient i;
i = index for patient ($1 \leq i \leq 10$); and
j = index for site within patient ($1 \leq j \leq 3$).

(5) The next step is to center the data by the mean for each patient/site. This removes variability between patients and between sites within patients but leaves intact variability across time. Removing this variability will give more power to detect any signal associated with the DOC. Thus, we work with

$$y_{wtij} = x_{wtij} - x_{w.ij},$$

where the dot in the subscript $x_{w.ij}$ indicates that an average has been taken over the particular subscript t.

(6) Next, we convert measurement day to day in cycle. It is necessary to put all patients on some kind of standard "clock" for their cycle. As mentioned above, a day in cycle (DOC) was assigned to each measurement day for each patient. The majority of patients had 30-day cycles. To simplify matters, the total cycle for each patient was binned into 15 equal width intervals. Let D_{ki} denote the collection of measurement days for patient i belonging to interval k of the DOC binning. Note that there are typically two measurement days in each D_{ki}, but this varies because of some variability in the length of cycle for different patients. Also, patient 4 had two cycles beginning during the period of observation (see Fig. 1), so some of the DOC bins would contain days from two cycles. Mathematically, the assignment of measurement days t to D_{ki} was done by taking the DOC assigned to that t, dividing by the total length of cycle for that patient, multiplying by 15, and rounding to the near-

est integer. The number of bins is a more or less arbitrary choice. Too few bins makes it more difficult to detect rapid changes (except at the boundary points between bins) but means more observations are averaged within bins to compute the ANOVA test statistic described below, thus reducing signal-to-noise ratio. Too many bins means that there will not be enough observations averaged—one could even have bins without any observations. We felt that bins of a width of about 2 days of a cycle provided enough resolution and there would be sufficient observations within each bin. The number of observations within bins ranged from 51 to 64.

To assess the statistical significance of any effect associated with the variable k (DOC bin), we computed a standard one-way analysis of variance (ANOVA) F-Statistic[15] for each value of w (defined in step 4 above). In the ANOVA, we treat the value y_{wtij} as the response variable and the associated value k (binned DOC) as the ''treatment'' or explanatory variable. To describe the null hypotheses being tested, we define μ_{wk} to be the (theoretical) mean of the variables y_{wtij} averaged over the population of patients i, sites j, and day t for t in DOC bin D_{ik}. For each w, we test the null hypothesis H_0: $\mu_{w1} = \mu_{w2} = \ldots = \mu_{w15}$. If the null hypothesis is true for a given w, then there is no change in the mean associated with binned DOC. Of course, we expect that there is correlation between observations over time within the same site and between sites within the same patient, so the assumptions of ANOVA are seriously violated and one cannot use the standard tables to convert the F-statistic to a P-value in order to assess statistical significance. To overcome this difficulty, we used a randomization procedure. The null hypothesis (for each value of the index w) is that there is not systematic variation associated with the binned DOC. If this is true, then we can randomly vary the starting DOC for each patient and obtain a data set statistically equivalent to the original data set. Varying the binned DOC means adding a random quantity and taking remainder modulo 15 (since we binned the DOC values to 15 bins). This randomization is done independently across patients but is the same for sites within a patient so that any time drift associated with a particular patient is not altered. Using this randomization procedure, we generated 10 000 values of the F-statistics for each w, giving a 10 000 by 705 matrix of randomized F-statistic values. These are indeed generated under the null hypothesis of no effect from binned DOC since by randomizing the starting value within a patient, any systematic effect associated with DOC will be lost. This initial set of randomized F-statistic values will be referred to as the reference values. The 705 F-statistic values from the original observed data set were compared with these reference values and converted to P-values by computing the proportion of randomized reference F-statistic values greater than the observed F-statistic. Referring to the 10 000 by 705 matrix of randomized F-statistic values, we see that the P-value for variable w is the proportion of values in column w greater than the F-statistic for the actual data set.

Now we consider the correction for multiple comparisons. We are testing 705 null hypotheses of no menstrual signal at each excitation/emission wavelength combination (or the peak value used for normalization when $w = 704$ or redox ratio when $w = 705$). If all the 705 null hypotheses are true (i.e., absolutely no menstrual signal anywhere in the EEM, which is consistent across patients), and if we use the standard 0.05 level of significance, we would expect to (falsely) reject about $0.05 \times 705 \approx 35$ of the null hypotheses. In order to correct for the multiple comparisons problem, we applied the procedure of Westfall and Young.[16] This procedure produces corrected P-values, which one can compare with a desired level of significance and use to control the family-wise error rate (FWER), sometimes called the experiment-wise error rate. The FWER is the probability of rejecting a single true null hypothesis no matter what subset of null hypotheses is true. To apply the Westfall and Young method, we generated a second matrix of 10 000 F-statistic values from data with randomized starting values for the (binned) DOC. This second set of randomized F-statistics is referred to as the correction set. The correction set of F-statistic values was converted to a correction set of P-values by comparing each F-statistic in the correction set with the 10 000 corresponding F-statistic values in the reference set, i.e., by repeating exactly the calculation that converted the F-statistic for the observed data into a P-value. Given the reference P-values, the Westfall and Young procedure can then be applied to obtain corrected P-values for the observed data. One rejects all null hypotheses for which the corrected P-value is below the desired FWER, which we took as 0.05.

RESULTS

The ages of the ten patients who completed the study ranged from 26 to 45. All ten patients were pre-menopausal as determined by FSH levels, and nine of ten had ovulatory bleeding by medical history and estradiol and progesterone levels. One patient (patient 2) had an anovulatory cycle. Nine of the ten patients had normal results from the Pap smears taken as part of the study, and the other patient (patient 10) had CIN I/HPV-associated atypia. Eight of the ten patients participated in a second study after the menstrual-cycle study and underwent colposcopically directed biopsies on two of the measured sites. One of these patients displayed colposcopic abnormality and biopsies of all 3 measured sites were obtained at the end of the trial; the remaining seven patients had normal cervical biopsies. The patient with CIN I/HPV-associated atypia diagnosed from the Pap smear had a diagnosis of low-grade SIL (LG-SIL) confirmed in all three biopsies.

The results from the significance tests are shown in Fig. 4. The dark gray pixels were not statistically significant. The lighter gray areas represent excitation/emission wavelength combinations that were statistically significant at the 0.05 level without applying the correction for multiple comparisons. There were 51 such wavelengths. The normalizing (peak) value and the redox ratio did not have significant uncorrected P-values. The white area in Fig. 4 indicates the 13 wavelength combinations that were statistically significant at the 0.05 level after apply-

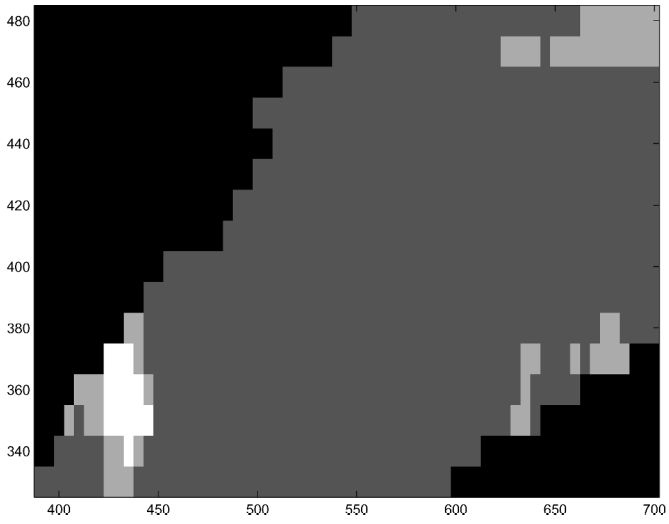

FIG. 4. Regions of the EEM that showed a statistically significant signal from the menstrual cycle. The x-axis is emission wavelength in nanometers and the y-axis is excitation wavelength in nanometers. The lighter gray areas show the excitation/emission wavelength combinations that were significant without the correction for multiple comparisons but not significant after the correction. The white areas show the wavelength combinations that were significant both before and after the correction. The darker gray area shows the remainder of the measured EEM. All significance levels are 0.05.

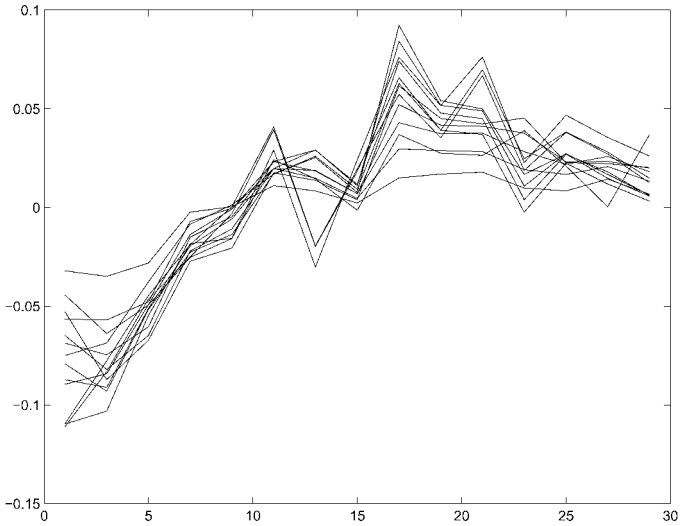

FIG. 5. Plot of the signal for the 13 significant wavelengths (shown in white in Fig. 4) after correction for multiple comparisons. The x-axis is day of cycle scaled to a 30-day cycle. The y-axis is the average (across patients and sites for a given DOC) relative change from the mean (across time for a given site within a patient) for that excitation/emission wavelength combination. Thus, on the first day of the cycle, the values tend to be about 8% below their mean value and on the twentieth day of the cycle, about 3% above their mean value.

ing the correction for multiple comparisons. The 13 excitation/emission pairs are as follows: at excitation 340, only emission 435; at excitation 350, emissions 425 through 445; at excitation 360, emissions 425 through 440; and at excitation 370, emissions 425 through 435.

To further investigate the nature of the menstrual signal, we have plotted the relative deviation from the mean of the (logarithmically transformed) spectral values at the 13 wavelengths found significant after correction. This appears in Fig. 5. Basically, we see low values near the first few days of the cycle; there is approximately an 8% decrease from the average over the entire cycle right at the start. The values tend to increase until about DOC 15 when the values range from 1 to 9% above the mean. It is interesting that there is a jump discontinuity from the end of the cycle to the beginning, indicating that the effect of lower intensity occurs very suddenly at the beginning of the cycle and then trails off gradually.

CONCLUSION

The evidence strongly suggests that hemoglobin absorption at the start of menstruation and for a few days thereafter is the dominant effect of the menstrual cycle on fluorescence spectroscopy of the cervix. The Soret band centered at 420 nm is the strongest absorption band of hemoglobin.[17] The emission wavelengths where there is a statistically significant effect from the day of cycle (Fig. 4) are between 425 and 445 nm inclusive. This is slightly displaced from the main hemoglobin absorption band at 420 nm. But the statistically significant emission wavelengths tend to be at a higher intensity because they are nearer the dominant NAD(P)H peak (Figs. 2 and 3). Statistical significance is heavily dependent on the signal-to-noise ratio, and we hypothesize that the higher signal values near the NAD(P)H peak cause a shift in the significant emission wavelengths. If one includes the region of uncorrected *P*-values below 0.05 (light gray pixels in Fig. 4), there is a fairly distinct band of absorption wavelengths centered around 425 nm where there appears to be an effect associated with the menstrual cycle. No significant effect is observed in excitation wavelengths around 420 nm, but this is a fairly low-intensity part of the spectrum (refer to Fig. 2), so the low signal-to-noise ratio may make it difficult to find an effect.

Furthermore, the detailed analysis of the significant values in the EEM (Fig. 5) shows that there is a reduction in intensity early in the cycle at these wavelength combinations. Because of absorption by the hemoglobin in menstrual blood, this effect is expected if hemoglobin absorption is the dominant effect. The plots in Fig. 5 are also marked by an apparent discontinuity from the last day of the cycle to the first, which is consistent with the sudden onset of menstrual bleeding.

In a previous paper, it was noted that the redox ratio displays a similar pattern. Our failure to find significance for this hypothesis may result from the small sample of 10 patients. A larger sample might reveal other effects.

If our hypothesis that the predominant effect is from hemoglobin absorption from menstrual bleeding is correct, then any adverse effect this may have on the diagnostic use of FS can be ameliorated simply by requiring that the procedure only be performed when the patient is unlikely to be bleeding. This is an important finding for the intended medical applications of fluorescence spectroscopy.

ACKNOWLEDGMENT

Financial support provided by NIH Program Project Award CA82710.

1. R. R. Alfano, G. C. Tang, A. Pradham, W. Lam, D. S. J. Choy, and E. Opher, IEEE J. Quantum Electron. **23,** 1806 (1987).

2. J. Hung, S. Lam, J. C. LeRiche, and B. Palcic, Lasers Surg. Med. **11,** 99 (1991).
3. M. A. Mycek, K. T. Schomaker, and N. S. Nishioka, Gastroint. Endosc. **48,** 390 (1998).
4. K. T. Schomacker, J. K. Frisoli, and C. Compton, Lasers Surg. Med. **12,** 63 (1992).
5. R. M. Cothren, M. V. Sivak, J. Van Dam, R. E. Petras, M. Fitzmaurice, J. M. Crawford, J. Wu, J. F. Brennan, R. P. Rava, R. Manoharan, and M. S. Feld, Gastroint. Endosc. **44,** 168 (1996).
6. I. Georgakoudi, B. C. Jacobson, J. Van Dam, V. Backman, M. B. Wallace, M. G. Muller, Q. Zhang, K. Badizadegan, D. Sun, G. A. Thomas, L. T. Perelman, and M. S. Feld, Gastroenterology **120,** 1620 (2001).
7. J. K. Dhingra, D. F. Perrault, Jr., K. McMillan, E. E. Rebeiz, S. Kabani, R. Manoharan, I. Itzkan, M. S. Feld, and S. M. Shapshay, Arch. Otolaryng. Head Neck Surg. **122,** 1181 (1996).
8. N. Ramanujam, M. Follen Mitchell, A. Mahadevan, S. Thomsen, A. Malpica, T. Wright, N. Atkinson, and R. Richards-Kortum, Lasers Surg. Med. **19,** 46 (1996).
9. K. Tumer, N. Ramanujam, J. Ghosh, and R. Richards-Kortum, IEEE Trans. Biomed. Eng. **45,** 953 (1998).
10. M. F. Mitchell, S. B. Cantor, C. Brookner, U. Utzinger, D. Schottenfeld, and R. Richards-Kortum, Obstet. Gynecol. **94,** 889 (1999).
11. C. Brookner, U. Utzinger, M. Follen, R. Richards-Kortum, and N. Atkinson, ''Effects of Biographical Variables on Cervical Fluorescence'', unpublished data.
12. S. Chang, M. Dawood, G. Staerkel, U. Utzinger, R. Richards-Kortum, and M. Follen, ''Fluorescence Spectroscopy for Cervical Pre-Cancer Detection: Is There Variance Across the Menstrual Cycle?'', unpublished data.
13. D. L. Heintzelman, U. Utzinger, H. Fuchs, A. Zuluaga, K. Gossage, A. M. Gillenwater, R. Jacob, B. Kemp, and R. Richards-Kortum, Photochem. Photobiol. **72,** 103 (2001).
14. A. Agrawal, U. Utzinger, C. Brookner, C. Pitris, M. Follen Mitchell, and R. Richards-Kortum, Lasers Surg. Med. **25,** 237 (1999).
15. J. Devore, *Probability and Statistics for Engineering and the Sciences* (Duxbury, Pacific Grove, CA, 2000), 5th ed., Chap. 10, pp. 403–409.
16. P. Westfall and S. Young, *Resampling Based Multiple Testing* (John Wiley and Sons, New York, 1993), Chap. 2, pp. 66–67.
17. I. D. Campbell and R. A. Dwek, *Biological Spectroscopy* (The Benjamin Cummings Publishing Company, Menlo Park, CA, 1984), Chap. 4, p. 71.