Published in final edited form as:
Am J Obstet Gynecol. 2009 May ; 200(5): 512.e1–512.e5. doi:10.1016/j.ajog.2008.11.024.

Effect of hormonal variation on Raman spectra for cervical disease detection

Elizabeth M. KANTER, PhD[1], **Shovan MAJUMDER, PhD**[1], **Gary J. KANTER, MD**[2], **Emily M. WOESTE, MD**[2], and **Anita MAHADEVAN-JANSEN, PhD**[1]

[1] *Department of Biomedical Engineering, Station B, Box 351631, Nashville, TN 37235, Phone: 615-343-4787, Fax: 615-343-7919, Anita.mahadevan-jansen@vanderbilt.edu* [2] *Tri-State Women's Health, Florence, KY*

Abstract

Objective(s)—To characterize the variations in normal cervical spectra due to menopausal status and location within the menstrual cycle. Using the information obtained, the accuracy of Raman Spectroscopy to diagnose low grade squamous intraepithelial lesion (LGSIL) will be improved.

Study Design—A total of 133 patients undergoing either colposcopy or Pap smear were recruited from either Vanderbilt University or Tri-State Women's Health. Raman spectra were collected from both normal and diseased areas. The data was processed and analyzed using a multi-class discrimination and classification algorithm to determine if the spectra can be correctly classified.

Results—Stratifying the data by menopausal state resulted in correctly classifying of LGSIL 97% of the time (from 74%).

Conclusion(s)—This study brings Raman Spectroscopy one step closer to clinical use by improving the sensitivity to differentiate LGSIL from normal.

Keywords

Cervical Dysplasia; Optical diagnosis; Raman spectroscopy

Introduction

Cervical cancer is the second most common malignancy among women worldwide, with more than 490,000 cases diagnosed and 274,000 deaths each year.[1] In the United States alone, it is estimated that in 2008, 3,870 deaths will occur from this disease, and 11,070 new cases of invasive cervical cancer will be diagnosed.[2] Several recent meta-analyses have reported low Pap smear sensitivities in the range of 20–50%.[3, 4] Most of these studies, however, indicate that the Pap smear is generally a very specific test, meaning that cytology correctly identifies most women who do not have high grade lesions or cancer. Although cervical cancer affects young women, many older women do not realize that the risk of developing cervical cancer is still present as they age. Slightly over 20% of women with cervical cancer are diagnosed when they are over 65.[5] Further, cervical cancer in Hispanic women occurs at a rate that is more than two times that of non-Hispanic white women. African-American women develop cervical

Correspondence to: Anita MAHADEVAN-JANSEN.

cancer about 50% more often than non-Hispanic white women.[5] Thus there is a continued need for an effective diagnostic and guidance tool.

Optical methods, including optical coherence tomography and fluorescence, reflectance, and Raman spectroscopy, have been investigated as new diagnostic tools. In particular, Raman spectroscopy has shown considerable promise in detecting and screening for cervical pre-cancers with improved sensitivity and specificity over more traditional methods. Current studies suggest that Raman spectroscopy can differentiate between normal tissue, metaplasia, low grade squamous intraepithelial lesion (LGSIL), and high grade squamous intraepithelial lesion (HGSIL) with an accuracy of 90%. The most difficult pathology to classify is LGSIL, which is often misclassified as normal.[6] Accordingly, one goal of this study was to improve the ability of Raman spectroscopy to differentiate between LGSIL and healthy squamous epithelium.

One method to improve the differentiation of LGSIL from healthy tissue is to determine a source of spectral variability of the normal cervix that may be used to stratify tissue samples. Many different physiological factors may affect the normal cervix, such as hormonal status and previous disease of the cervix. Several studies have shown that menopausal state and location in the menstrual cycle impact the optical properties. For example, a study by Gill *et. al*. showed that there was a statistical difference between fluorescence intensity in pre- and post-menopausal cervix: on average, post-menopausal women have a higher average fluorescence signal. This difference is likely due to changes in collagen cross-linking of the cervix that occur during menopause.[7] Cox *et. al.* found that the effects of the menstrual cycle on fluorescence measurements were confined to the presence of hemoglobin absorption; this was avoided when measurements were not taken during the first eight days of the cycle.[8] These previous studies used fluorescence spectroscopy. Other changes in the cervix may be detected by Raman since its biochemical composition changes during a menstrual cycle, but this has not been investigated.

The goal of this study is to improve the ability of Raman spectroscopy to diagnose low grade dysplasia in the cervix; this study looked at variations in normal cervical spectra due to menopausal status and time point in the menstrual cycle. Based on the significance of this result, the Raman spectra were stratified according to menopausal status and menstrual cycle prior to differentiation. Raman spectra were analyzed using multivariate statistical methods. From our results from the normal cervix, we found that data needs to be stratified according to location within the menstrual cycle. Adding this information to the diagnostic algorithm then results in a 97% classification rate of LGSIL.

Methods

Clinical Study Design-Pap Smear Patients

Ninety-one patients undergoing a routine annual Pap smear were recruited to participate in the study which was approved by the Vanderbilt and Copernicus Group Institutional Review Board (IRB). To be eligible for enrollment, the patient must be undergoing a routine Pap smear, be between the ages of 18–75, and have a cervix (no history of a hysterectomy). Informed consent was obtained from each patient prior to the procedure. The cervix was exposed, visually examined by the attending physician, and wiped clean with a dry cotton swab and then with a saline solution. Multiple Raman spectra (3–5) of normal appearing sites were measured *in vivo*, and tissue sites were recorded as squamous epithelium, columnar epithelium, or at the squamous-columnar junction as noted by the attending physician. The Pap procedure then proceeded according to standard clinical protocol. The acquired spectra were considered normal if the Pap smear was negative. The patient's age, last menstrual period, artificial hormones (including contraception), and menopausal status were all noted upon chart review.

Clinical Study Design- Colposcopy Patients

Thirty one patients undergoing colposcopy-guided biopsy were recruited to participate in the study as approved by the Vanderbilt and Copernicus Group IRBs. To be eligible for enrollment, the patient must be undergoing a colposcopy-guided biopsy, be between the ages of 18–75, and have a cervix (no history of a hysterectomy). Informed consent was obtained from each patient prior to the procedure. The cervix was exposed and visually examined by the attending physician. Acetic acid was applied to the cervix to turn abnormal areas white, enabling visualization of abnormal areas, and iodine was applied to clean the tissue and show the location of squamous epithelium. The abnormal tissue was removed and underwent pathology examination. Spectra were measured after the application of the acetic acid but before the application of the iodine (if needed). Spectra were acquired from each area where a biopsy was planned and one visually normal area. The patient's age, last menstrual period, abnormal Pap smear result, and menopausal status were noted upon chart review.

Data Acquisition

Raman spectra were collected from multiple sites *in vivo* using a portable Raman spectroscopy system consisting of a 785 nm diode laser (Process Instruments, Inc., Salt Lake City, UT), beam-steered fiber optic probe (Visionex Inc., Atlanta, GA), imaging spectrograph (Kaiser Optical Systems, Inc., Ann Arbor, MI), and back-illuminated, deep-depletion, thermoelectrically cooled charge coupled device (CCD) camera (Roper Scientific, Inc., Princeton, NJ), all controlled with a laptop computer. A photograph of the system can be seen in figure 1; more details of the system can be found elsewhere [9]. The fiber optic probe deliverers 80mW of incident light onto the tissue for 3 seconds. During each measurement, the overhead fluorescent lights were turned off.

Each day, spectral calibrations were preformed using a neon-argon lamp, naphthalene and acetaminophen standard to correct for system wavenumber, laser excitation, and throughput variations. The spectra were further corrected using a calibrated tungsten lamp to account for system variations. The spectra were processed for fluorescence subtraction and noise smoothing using a modified mean polynomial fitting approach [9]. Following data processing, each spectrum was normalized to its mean spectral intensity across all Raman bands to account for overall intensity variability. These normalized spectra were categorized according to histology (squamous epithelium, columnar epithelium, or at the junction) as determined by the attending physician and menopausal status and used for further comparison and analysis.

Statistical Analysis

The analysis technique used in this paper has been previously described [10]. The process consists of two steps - the first is extraction of diagnostic features from the spectra using nonlinear maximum representation and discrimination features (MRDF). The processed data set undergoes a two-step, non-linear transform to extract relevant features that provide the best class separation. The second step is a probabilistic classification scheme based on sparse linear multinomial logistic regression (SMLR) for classifying nonlinear features into corresponding tissue categories. SMLR is a probabilistic multi-class model based on a Bayesian machine-learning framework of statistical pattern recognition. The main focus of SMLR is to separate a set of labeled input data into its class by predicting the posterior probabilities of their class membership. An unbiased classification algorithm was developed using these two methods and was tested using leave-one-patient-out cross-validation.

Results

Spectra from a total of 122 patients were obtained and the results are discussed below.

Variations in Menopausal Status

The spectra in Figure 2 are shown stratified by menopausal status. The data is separated into one of 4 groups: pre-menopausal proliferative phase (days 1–14 of the menstrual cycle) or Pre-menopausal Before Ovulation (PBO); pre-menopausal secretory phase (days 15–28+ of the menstrual cycle) or Pre-menopausal After Ovulations (PAO); peri-menopausal (PERI); and post-menopausal (POST). Women who were on traditional oral contraceptives were placed in the category in which they best fit. Overall the spectra have subtle but consistent differences according to location within the menstrual cycle and menopausal status, especially at 1250cm^{-1} (typically associated with collagen) and 1300–1320cm^{-1} (typically associated with cellular features)[11]. These differences could be due to changes caused by varying levels of estrogen and progesterone during a women's lifetime. As seen in the confusion matrix shown as Table 1, the data classified with an overall accuracy of 98.2%. In the confusion matrix, the rows correspond to the histopathology, and the columns correspond to the result of the Raman algorithm. The number along the diagonal represents the number of spectra that classify correctly. In this example, only 3 spectra were classified incorrectly, and one of the misclassifications was a post-menopausal spectrum that was atrophic. This is relatively common in post-menopausal women and is clinically treated as normal, although an atrophic cervix has different biochemical components.

Dysplasia Study

Average LGSIL spectra and the average normal spectra are displayed in Figure 3. The largest difference occurs between 1230cm^{-1} and 1300cm^{-1}. A variety of biochemical changes occur as tissue becomes dysplastic. With LGSIL, these changes are very subtle because relatively fewer cells have undergone this transformation. The classification accuracy without menstrual cycle stratification to discriminate between LGSIL and normal tissue is 74% using MRDF-SMLR.[6]

Using the MRDF-SMLR algorithm stratified by menopausal status, LGSIL classification accuracy improved to 97%. The new, higher-class discrimination algorithm was developed using the information obtained from the menopausal/menstrual status study. PERI and POST spectra were not included in this analysis because there were not enough patients in the categories. In table 3, the classification using 4 classes: normal-PBO, normal-PAO, low grade-PBO, and low grade-PAO is shown. In this table, a total of 4 spectra (out of 60) classified incorrectly: two LGSIL and two normal. Overall, stratifying spectra based on menopausal state and menstrual cycle results in an improvement of classification accuracy.

Comment

Optical spectroscopy for diagnosing abnormal tissue has many advantages, such as real time monitoring, the ability to do "see and treat" procedures, a reduced need for biopsies, and the ability to monitor progression of disease. Raman spectroscopy provides molecularly specific vibrational signatures and more detailed information than other methods of optical spectroscopy. Although Raman signatures are significantly weaker than other forms of optical spectroscopy, the slightly longer acquisition time and data processing is negated by its increased performance. For example, using fluorescence spectroscopy, the prospective sensitivity and specificity of a paired multivariate algorithm for discriminating precancers from non-precancerous tissue is 82% and 68%[12], whereas Raman spectroscopy has been shown to discriminate these tissues with an accuracy of 97%.[6]

This study brings Raman spectroscopy one step closer to clinical use by improving the specificity in diagnosing dysplasia. This improvement was accomplished by incorporating variations in the normal cervix to differentiate LGSIL from normal. Changes due to menopausal state and menstrual cycle location in the normal cervix can be detected with Raman

spectroscopy. These stratifications need to be considered when using a classification algorithm to differentiate between normal and dysplastic tissue. When this information is incorporated into the algorithm, the classification accuracy improves from 74% to 97%, indicating the potential of Raman spectroscopy to diagnose low grade dysplasia. Previous studies have already suggested that Raman spectroscopy can differentiate HGSIL with high success.[13]

One limitation to stratifying data before analysis is that there are some women who do not fit into any category. (PBO, PAO, PERI or POST). This can be due to a form of birth control (Depo-Provera (Depo)) or to very irregular periods (amenorrhea). We found that the cervix of women who use Depo for either birth control or health reasons consistently classifies as PAO. Since Depo delivers a high level of progesterone, it stops the ovaries from releasing eggs causing cervical mucus to thicken and changes the uterine lining, similar to changes after ovulation. Normally, after ovulation, the corpus luteum produces high levels of progesterone, and this progesterone thickens the mucus in the cervix and acts on the lining of the uterus, therefore it is expected for the spectrum to be similar. The amenorrhea results were expected to be variable because these women have very inconsistent periods and therefore may fall at any point in the cycle. With more data, we may be able to ascertain a clearer way to stratify such patients.

Raman spectroscopy has the potential to detect small variations both in the normal and dysplastic cervix. [13, 14] The ectocervix consists of a dense fibromuscular stroma which is composed primarily of collagenous connective tissue and a ground substance of mucopolysaccharides. The connective tissue is approximately 15% smooth muscle with a small amount of elastic tissue. Hormonal changes, such as menopausal status and location in the menstrual cycle, change the composition of the ectocervix.[15] Therefore, Raman signatures vary significantly depending on location within the menstrual cycle and with the onset and completion of menopause. These differences are important when trying to correctly classify the tissue. Spectral differences are shown in figure 1; the most notable differences occur around $1250cm^{-1}$, $1300cm^{-1}$, and $1320cm^{-1}$, most likely due to changes in protein levels (especially elastin and collagen). During the menstrual cycle, the cervix becomes softer or more elastic as the level of estrogen increases. After ovulation, this process is reversed, and the cervix loses some of its elasticity. During peri-menopause, the layer of epithelial cells thins, and the vascularity and content of the cervix is erratic, but the spectra remain consistent. The most variable and therefore the hardest group to classify is the post-menopausal group. The absence of ovarian estrogen and progesterone causes the cervix to change, leading to both dryness and atrophy. These conditions are considered normal in a woman who has gone through menopause but may cause major differences in Raman spectra.

Changes due to dysplasia are different than those associated with hormonal variations. The largest change when comparing LGSIL Raman spectra to normal Raman spectra is at the 1230–1300 cm^{-1} peak. This peak range is usually associated with proteins (Amide III), DNA (Guanine) and lipids (CH transformations) [11]. It is expected that there will be variations in the protein and lipid content when dysplastic changes occur. Another expected change is a reduction in glycogen peaks[16] that occur around 1300 cm^{-1}. This difference is expected to be minimal in LGSIL because the disease only affects a small portion of the epithelium. As the disease progresses towards HGSIL, this drop in the glycogen peak is expected to be more drastic.

Stratification and classification with a non-linear multi-class algorithm yields a posterior probability of how likely the spectra are to fit in a particular category. This is a powerful tool because it quantifies the confidence with which a particular area is classified correctly with Raman spectroscopy. If the confidence value is low and the area is suspicious based on the doctor's observation, a biopsy could be taken as is the current clinical protocol. Our method

may greatly reduce the number of biopsies taken, even if they are not completely eliminated. Most diagnoses could be determined in the clinic during the visit instead of several days to a week later, reducing the stress women feel while waiting for results and the need for a follow up visit to discuss the biopsy results.

This method would be particularly useful in developing countries, where “see and treat” methods are optimal. One major problem with screening alone is poor follow-up testing among women with abnormal pap smears. Our method would allow a nurse to find suspicious areas, take Raman measurements, and report an accurate diagnosis. The nurse could then decide on the appropriate action: allow the women to go home untreated, treat the patient, or send the patient to a doctor. This process could reduce the number of patients who are treated unnecessarily and ensure patients receive an accurate diagnosis the same day in the clinic.

Overall, this is a promising method for LGSIL diagnosis that would greatly benefit developing countries, rural communities, and working individuals by limiting time spent at the doctor’s office.

Acknowledgments

Funding Source- NCI/NIH (R01-CA95405)

The authors were like to acknowledge the financial support of the NCI/NIH (R01-CA95405). We would also like to thank the doctors and staff at Tri-state Women’s Health and Vanderbilt University for all their help.

References

1. AMERICAN CANCER Society. Cervical Cancer Resource Center. 2007
2. Anderson GH, Boyes DA, Benedet JL, et al. Organization and Results of the Cervical Cytology Screening-Program in British-Columbia, 1955–85. British Medical Journal 1988;296:975–978. [PubMed: 3129115]
3. Fahey MT, Irwig L, Macaskill P. Metaanalysis of Pap Test Accuracy. American Journal of Epidemiology 1995;141:680–689. [PubMed: 7702044]
4. Nanda K, McCrory DC, Myers ER, et al. Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: A systematic review. Annals of Internal Medicine 2000;132:810–819. [PubMed: 10819705]
5. Parham GP, Sahasrabuddhe VV, Mwanahamuntu MH, et al. Prevalence and predictors of squamous intraepithelial lesions of the cervix in HIV-infected women in Lusaka, Zambia. Gynecologic Oncology 2006;103:1017–1022. [PubMed: 16875716]
6. Kanter E, Majumder SK, Vargis E, et al. Multi-class discrimination of cervical precancers using Raman spectroscopy. Journal of Raman Spectroscopy. 2008In review
7. Gill EM, Malpica A, Alford RE, et al. Relationship between collagen autofluorescence of the human cervix and menopausal status. Photochemistry and Photobiology 2003;77:653–658. [PubMed: 12870852]
8. Cox DD, Chang SK, Dawood MY, et al. Detecting the signal of the menstrual cycle in fluorescence spectroscopy of the cervix. Applied Spectroscopy 2003;57:67–72. [PubMed: 14610938]
9. Lieber CA, Mahadevan-Jansen A. Automated method for subtraction of fluorescence from biological Raman spectra. Applied Spectroscopy 2003;57:1363–1367. [PubMed: 14658149]
10. Majumder SK, Gebhart S, Johnson MD, Thompson R, Lin WC, Mahadevan-Jansen A. A probability-based spectroscopic diagnostic algorithm for simultaneous discrimination of brain tumor and tumor margins from normal brain tissue. Appl Spectrosc 2007;61:548–57. [PubMed: 17555625]
11. Bitar RA, Martinho Hda S, Tierra-Criollo CJ, Zambelli Ramalho LN, Netto MM, Martin AA. Biochemical analysis of human breast tissues using Fourier-transform Raman spectroscopy. J Biomed Opt 2006;11:054001. [PubMed: 17092150]

12. Ramanujam N, Mitchell MF, Mahadevan Jansen A, et al. Cervical precancer detection using a multivariate statistical algorithm based on laser-induced fluorescence spectra at multiple excitation wavelengths. Photochemistry and Photobiology 1996;64:720–735. [PubMed: 8863480]

13. Robichaux-Viehoever A, Kanter E, Shappell H, Billheimer D, Jones H, Mahadevan-Jansen A. Characterization of Raman spectra measured in vivo for the detection of cervical dysplasia. Applied Spectroscopy 2007;61:986–993. [PubMed: 17910796]

14. Majumder, SK.; Kanter, E.; Viehoever, AR.; Jones, H.; Mahadevan-Jansen, A. Near-infrared Raman spectroscopy for in-vivo diagnosis of cervical dysplasia: a probability-based multi-class diagnostic algorithm. In: Vo-Dinh, T.; Grundfest, WS.; Benaron, DA.; Cohn, GE.; Raghavachari, R., editors. Advanced Biomedical and Clinical Diagnostic Systems V. Vol. 6430. San Jose, CA, USA: SPIE; 2007.

15. Katz, VL.; Lobo, RA.; Lentz, G.; Gershenson, D. Katz: Comprehensive Gynecology. Vol. 5. Elsevier Inc; 2008.

16. Sellors, J.; Sankaranarayanan, R. Colpsocopy and Treatment of Cervical Intreaepithelial Neoplasia: A Beginners' Manual. Lyon: International Agency for Research on Cancer; 2003.

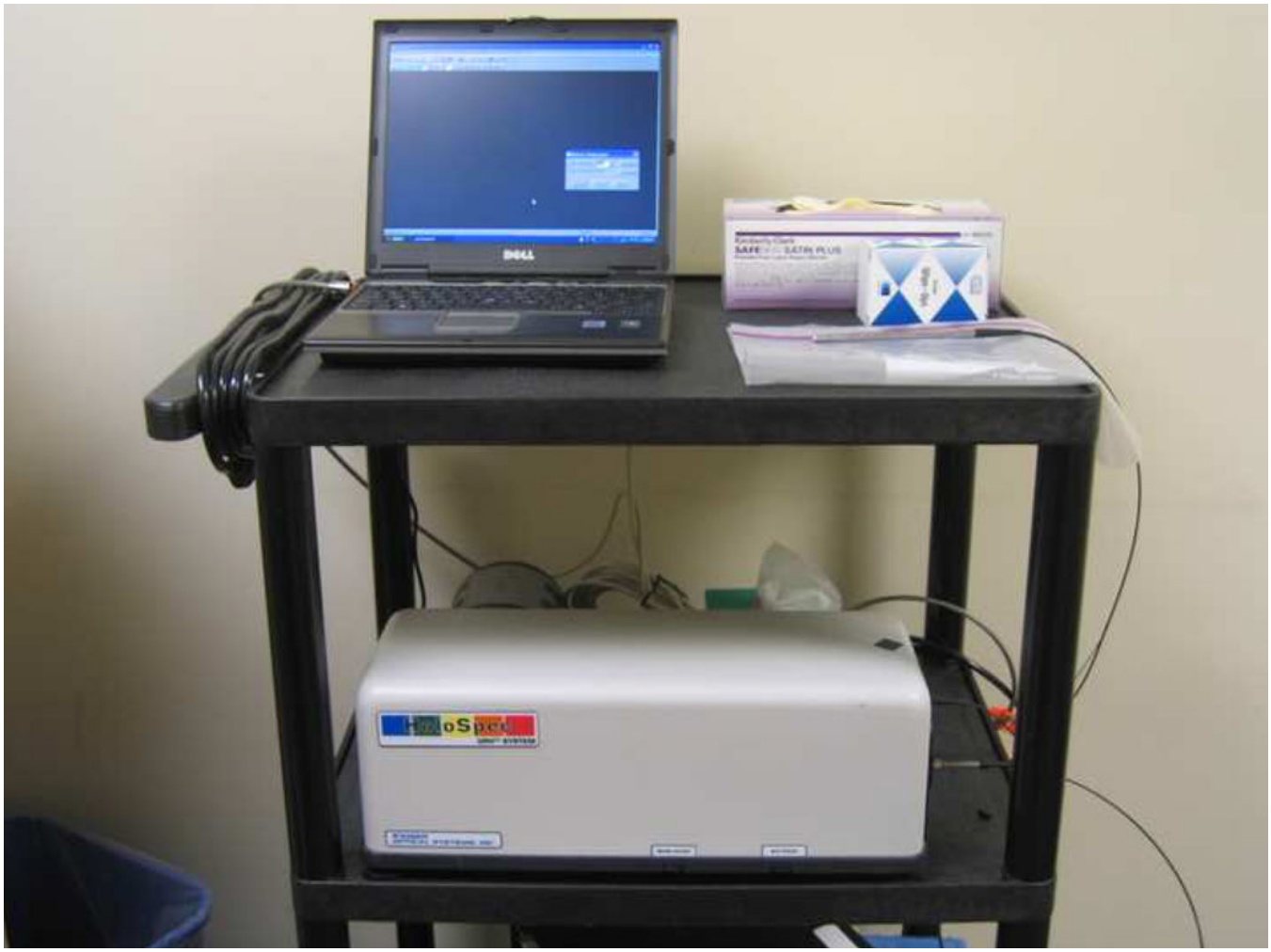

Figure 1.
Photograph of the system

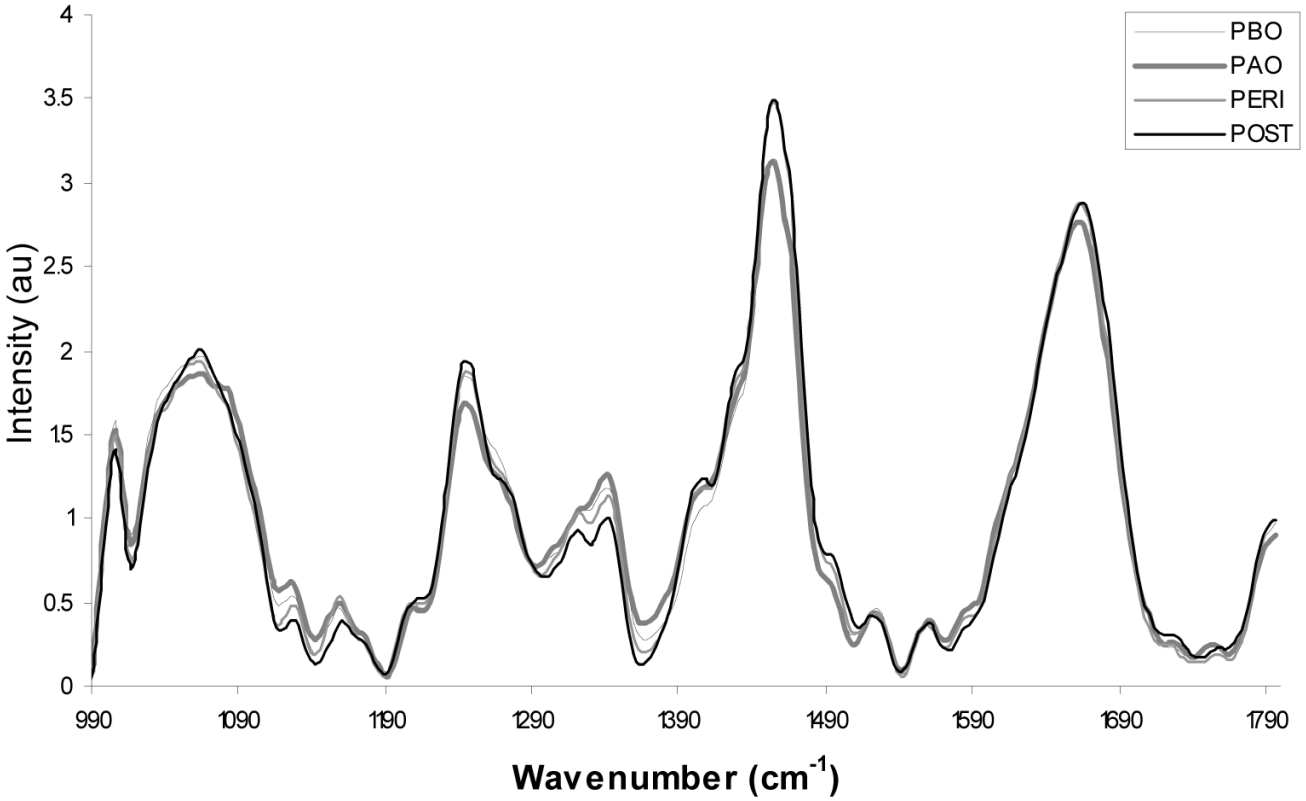

Figure 2.
Average Raman spectra for post menopausal normal cervix (POST-30), peri menopausal normal cervix (PERI-34), pre-menopausal after ovulation normal cervix (PAO-54) and pre-menopausal before ovulation normal cervix (PBO-47).

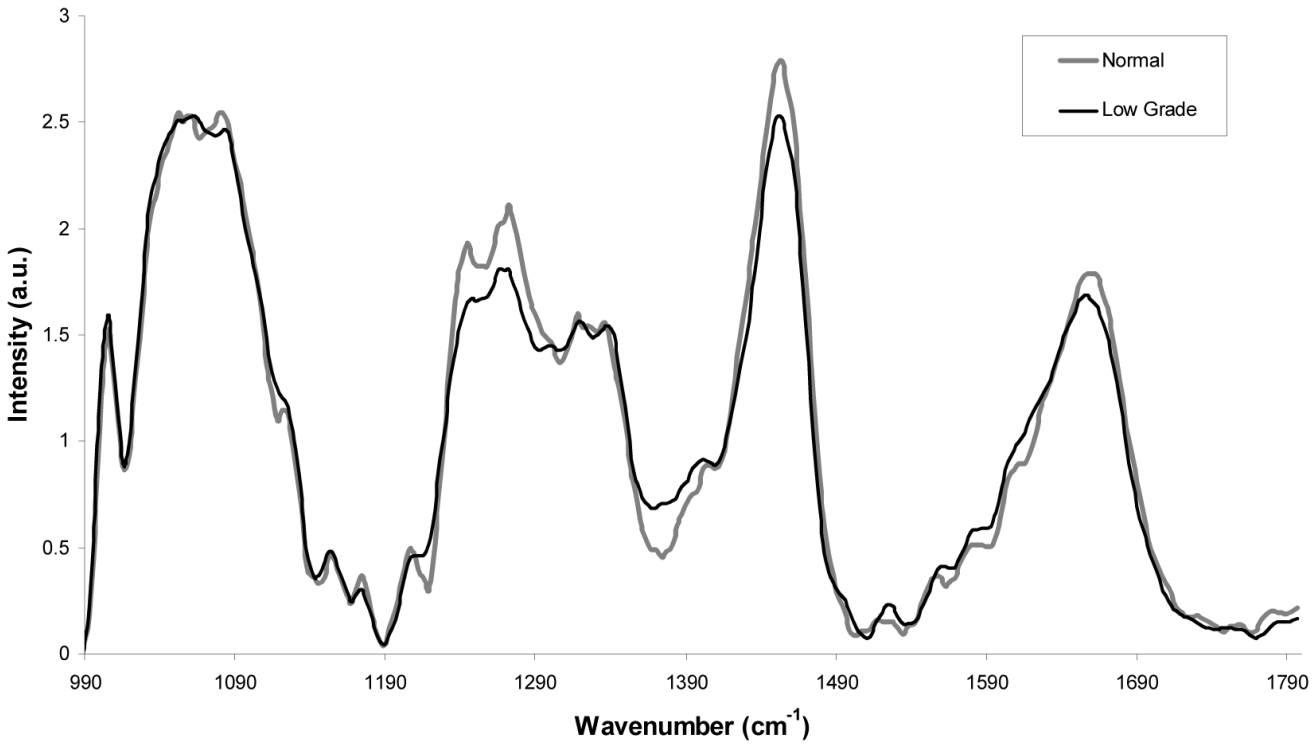

Figure 3.
Average Raman spectra for normal cervix (34) and low grade cervix (30).

Table 1

Confusion matrix from the menopausal status data

		Histopathology			
		PBO	**PAO**	**PERI**	**POST**
Raman Algorithm	PBO	47	0	0	0
	PAO	0	53	0	0
	PERI	0	0	33	1
	POST	0	1	1	28

Table 2

Classification of amenorrhea and depo data as classified by MRDF and SMLR

		Histopathology	
		Amenorrhea	**Depo**
Raman Algorithm	PBO	1	0
	PAO	2	4
	PERI	1	0
	POST	2	0

Table 3

Confusion matrix for LGSIL vs. Normal separated by location in the menstrual cycle

		Histopathology			
		Normal-PBO	**Normal-PAO**	**LGSIL-PBO**	**LGSIL-PAO**
Raman Algorithm	Normal-PBO	19	0	1	1
	Normal-PAO	0	13	0	0
	Low Grade-PBO	1	0	14	0
	Low Grade-PAO	1	0	0	14