Diagnosing breast cancer using Raman spectroscopy: prospective analysis

Abigail S. Haka[*,†]
Zoya Volynskaya
Joseph A. Gardecki[‡]
Jon Nazemi[§]
Massachusetts Institute of Technology
George R. Harrison Spectroscopy Laboratory
77 Massachusetts Avenue
Cambridge, Massachusetts 02139

Robert Shenk
Nancy Wang
University Hospitals Case Medical Center
    and
Case Western Reserve
11100 Euclid Avenue
Cleveland Ohio 44106

Ramachandra R. Dasari
Massachusetts Institute of Technology
George R. Harrison Spectroscopy Laboratory
77 Massachusetts Avenue
Cambridge, Massachusetts 02139

Maryann Fitzmaurice
University Hospitals Case Medical Center
    and
Case Western Reserve
11100 Euclid Avenue
Cleveland Ohio 44106

Michael S. Feld
Massachusetts Institute of Technology
George R. Harrison Spectroscopy Laboratory
77 Massachusetts Avenue
Cambridge, Massachusetts 02139

Abstract. We present the first prospective test of Raman spectroscopy in diagnosing normal, benign, and malignant human breast tissues. Prospective testing of spectral diagnostic algorithms allows clinicians to accurately assess the diagnostic information contained in, and any bias of, the spectroscopic measurement. In previous work, we developed an accurate, internally validated algorithm for breast cancer diagnosis based on analysis of Raman spectra acquired from fresh-frozen *in vitro* tissue samples. We currently evaluate the performance of this algorithm prospectively on a large *ex vivo* clinical data set that closely mimics the *in vivo* environment. Spectroscopic data were collected from freshly excised surgical specimens, and 129 tissue sites from 21 patients were examined. Prospective application of the algorithm to the clinical data set resulted in a sensitivity of 83%, a specificity of 93%, a positive predictive value of 36%, and a negative predictive value of 99% for distinguishing cancerous from normal and benign tissues. The performance of the algorithm in different patient populations is discussed. Sources of bias in the *in vitro* calibration and *ex vivo* prospective data sets, including disease prevalence and disease spectrum, are examined and analytical methods for comparison provided. © *2009 Society of Photo-Optical Instrumentation Engineers.* [DOI: 10.1117/1.3247154]

1 Introduction

Breast cancer is the most common female cancer in the United States. Recent data indicate that a woman harbors a one-in-eight lifetime probability of developing breast cancer.[1] As such, significant research efforts have focused on breast cancer diagnosis, surgery, and management, resulting in dramatic improvements in outcome over the past several decades. Currently, a variety of optical imaging and spectroscopic techniques are being explored to improve breast cancer diagnosis and treatment.[2–12] They employ visible or near-infrared light, have the potential to provide chemical as well as morphologic information, and are less invasive than current diagnostic procedures.

Raman spectroscopy is a spectroscopic modality capable of providing detailed quantitative chemical/morphological information about tissue. It is an inelastic scattering process in which photons incident on tissue transfer energy to or from molecular vibrational modes.[13] This gives rise to a change in frequency (i.e., energy) of the emitted photon, hence the term "inelastic." Because the energy levels are unique for every molecule, Raman spectra are chemical specific. Raman spectroscopy is particularly amenable to *in vivo* measurements, because the excitation wavelengths and laser fluences used are nondestructive to the tissue and have a relatively large penetration depth.[14]

We present the first prospective test of Raman spectroscopy in diagnosing normal, benign, and malignant human

*Address all correspondence to: Abigail Haka, PhD, Department of Biochemistry, Weill Cornell Medical College, 1300 York Avenue, E-209, New York, NY. Tel: 212-746-8875; Fax: 212-746-8875; E-mail: abh2005@med.cornell.edu
†Abigail Haka is presently with Department of Biochemistry, Weill Cornell Medical College, New York NY.
‡Joseph Gardecki is presently with Wellman Center for Photomedicine, Massachusetts General Hospital, Boston MA.
§Jon Nazemi is presently with Prescient Medical, Incorporated, Doylestown, PA.

breast tissue in freshly excised surgical specimens. Several laboratories, including ours, have investigated the use of Raman spectroscopy for the examination of breast disease.[15–23] There are a number of ways in which Raman spectroscopy could aid in breast cancer diagnosis and treatment. A spectroscopic transdermal has the advantage of providing immediate diagnosis. As a result, the technique has the potential to reduce both the likelihood of a nondiagnostic biopsy that would require repeat needle or surgical biopsy, and patient anxiety by eliminating the currently unavoidable wait for a histopathology diagnosis. Because Raman spectroscopy provides immediate diagnosis, it may also aid in real-time intraoperative margin assessment. Accurate intraoperative margin assessment using Raman spectroscopy would lessen the need for reexcision surgeries resulting from positive margins, and thereby reduce the recurrence rate of breast cancer following partial mastectomy surgeries.

Our previous research demonstrated the ability of Raman spectroscopy to accurately diagnose normal, benign, and malignant tissues of the breast with high sensitivities and specificities.[21] This study examined *in vitro* fresh-frozen tissues in a laboratory setting. Four types of tissue, normal breast tissue, fibrocystic change, fibroadenoma, and invasive carcinoma, were studied in 126 sites from 58 patients. To extract information from the Raman spectra, we employ a spectroscopic model.[20] Modeling of the Raman spectrum yields fit coefficients that reflect the chemical makeup of the lesion, which is in turn associated with morphologic changes that pathologists routinely rely on to diagnose disease. Tissue composition extracted through modeling was used as the basis of a diagnostic algorithm. The fit coefficients for fat and collagen were found to be the key diagnostic parameters in distinguishing pathologies. These parameters were used to form the basis of a binary diagnostic algorithm, an x-y plot in which the fit coefficient for collagen is plotted on the y-axis, the fit coefficient for fat on the x-axis, and decision lines define the four classification regions. The resulting diagnostic algorithm, which classifies tissues, not just as benign or malignant, but according to specific pathological diagnoses, achieved a sensitivity (SE) of 94%, a specificity (SP) of 96%, and a total test efficiency of 95% for the diagnosis of cancer. We shall refer to this as the calibration analysis. This algorithm employed internal validation (leave-one-out cross validation); hence, its prospective capability had not been tested.

The excellent results of our *in vitro* calibration analysis supported testing the efficacy of the algorithm in a clinical setting. The current study prospectively evaluates the performance of Raman spectral diagnosis with this algorithm on freshly excised surgical specimens from a series of women undergoing needle localization breast biopsy or partial mastectomy. Sources of bias in the calibration and prospective data sets are examined, and analytical methods for comparison provided. Changes in performance of the Raman diagnostic algorithm can largely be explained by differences in the two patient populations in the calibration and prospective studies with respect to disease spectrum and cancer prevalence. Prospective testing of spectral diagnostic algorithms is of crucial importance because it allows clinicians to accurately assess the diagnostic information contained in, and any bias of, the spectroscopic measurement.

2 Materials and Methods

2.1 *Patient Population*

Breast tissue was obtained from a series of 28 consecutive patients undergoing excisional breast biopsy ($n=20$), partial mastectomy (lumpectomy, $n=7$), or simple mastectomy ($n=1$) at University Hospitals–Case Medical Center. Patient age averaged 52.5 years (range 35–80 years). All studies involving human tissue were approved by the University Hospitals–Case Medical Center Institutional Review Board and the Massachusetts Institute of Technology Committee On the Use of Humans as Experimental Subjects. Informed consent was obtained from all subjects prior to their surgical procedures.

2.2 *Tissue Preparation*

Data were collected from freshly excised surgical specimens in the pathology suite adjacent to the operating rooms, typically within 30 min of tissue excision. On removal, the outer surface of the surgical specimen was inked for identification of margins following standard pathology protocol. Each specimen was then sectioned and Raman spectra acquired from the uninked cut surface of the breast tissue from sites chosen by the pathologist. The number of spectra taken per patient varied, depending on the number and types of grossly visible breast tissue lesions. To reduce background, the breast specimens were placed in a light tight box for collection of Raman spectra. Following spectral acquisition, the breast specimens were marked with multicolored colloidal ink to uniquely identify each site sampled and fixed in 10% neutral buffered formalin. The fixed tissue samples were routinely processed, paraffin embedded, cut through the marked locations in 5-μm-thick sections, and stained with H&E. The histology slides were evaluated by an experienced breast pathologist who was blinded to the outcome of the Raman spectroscopy analysis. The pathology results served as the gold standard against which the Raman spectral diagnoses were compared. A total of 220 Raman spectra from the 28 patients were collected. Of these, 129 Raman spectra from 21 patients were appropriate for prospective analysis: 41 spectra from normal breast tissue from 18 patients, 82 from benign lesions, and six from malignant lesions. Benign lesions consisted of 73 regions of fibrocystic change from 16 patients and nine fibroadenomas from two patients. Malignant lesions consisted of 6 infiltrating ductal carcinomas (IDC) from five patients. Because multiple spectra were collected from each patient, some tissue samples are included in both the normal and diseased categories, depending on the pathology underlying the exact region of data collection. Ten spectra were excluded from the analysis due to excessive light contamination (six normal, one fibrocystic change, three IDC). Twenty spectra acquired from specimens diagnosed as ductal carcinoma *in situ* (DCIS) were also excluded, because this pathology was not encountered in the calibration data set used for diagnostic algorithm development. Sixty-one spectra acquired from two patients with breast cancer who had undergone preoperative chemotherapy (16 normal, nine fibrocystic change, three IDC), and five patients undergoing reexcision surgery to insure negative margins (14 normal, 17 fibrocystic change, two IDC) were also excluded from prospective analysis. Breast tissues from such patients exhibit histologic tissue changes not encountered in the calibration data set and, therefore, not
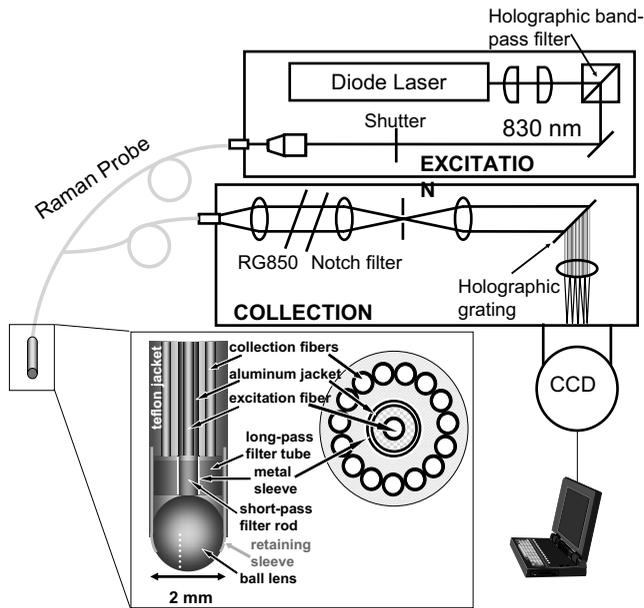
Fig. 1 Schematic of the clinical Raman system and optical fiber Raman probe. Light from an 830-nm diode laser is focused into the Raman probe's excitation fiber. The 4-m-long probe is <2 mm in diameter and consists of a single central excitation fiber surrounded by 15 collection fibers. The distal tip of the probe is terminated with a sapphire ball lens, which focuses the excitation light and efficiently couples the Raman scattered light from the tissue into the collection fibers. The linear array of collection fibers at the proximal end is coupled to a spectrograph for dispersion onto a CCD detector.

addressed in the original diagnostic algorithm development.[24,25]

2.3 Raman Spectroscopic Measurements

Data were acquired using the clinical Raman system and Raman optical fiber probe shown in Fig. 1 and described in detail elsewhere.[26,27] Briefly, light from an 830-nm-diode laser is collimated and then bandpass filtered before being focused into the Raman probe's excitation fiber. The 4-m-long probe is <2 mm in diameter and consists of a single central excitation fiber surrounded by 15 collection fibers. All fibers are low-OH fused silica and have a 200-μm core diameter. The probe's distal end is registered with a dual-filter module that rejects the intense interfering signals generated in the fibers. The distal tip of the probe is terminated with a sapphire ball lens, which focuses the excitation light and efficiently gathers and couples the Raman scattered light from the tissue into the collection fibers. The linear array of collection fibers at the proximal end is coupled to an f/1.8 spectrograph for dispersion onto a liquid-nitrogen-cooled, back-illuminated, deep-depletion CCD detector. Raman spectra in this study were acquired with a 10-to-30-s integration time, depending on signal intensity, and a spectral resolution of 8 cm^{-1}. The average laser excitation power varied between 100 and 150 mW. No tissue damage was observed, either grossly or on histological review.

2.4 Data Processing

Prior to data collection, calibration spectra (not to be confused with the Raman data set used for algorithm development) were collected for spectral corrections. Wave-number calibration was established with a spectrum of 4-acetamidophenol. Chromatic intensity variations were corrected by collecting the spectrum of a tungsten white light source diffusely scatted by a reflectance standard ($BaSO_4$). The remaining probe background generated in the optical fibers was characterized by collecting the excitation light scattered from a roughened aluminum surface. This background was optimally subtracted from the data in an iterative loop by using a scaling factor related to the tissue's optical properties.[27] The tissue fluorescence background was modeled and removed with a sixth-order polynomial. Following spectral correction, the basis spectra of the Raman spectral model were fit to the spectrum obtained from the breast tissue via non-negativity constrained least-squares minimization.[20] Raman spectra of epoxy and sapphire, two probe components, were added to the model, and a background spectrum acquired in the light tight box with no sample present was included to account for light contamination. Following the procedure used in the calibration study, the fit coefficients for each Raman spectrum of the data set were normalized to sum to unity. The microcalcification spectra were excluded from this normalization, as these species were not present in the tissue samples used for diagnostic algorithm development; their diagnostic significance will be considered elsewhere. We also observed increased contributions from cholesterol-like lipid deposits in this data set, and thus, cholesterol-like was excluded from normalization. Possible reasons for the increased contribution from cholesterol-like lipid deposits seen in the clinical data set are discussed below.

Because of the more realistic conditions encountered in fresh tissue, the signal-to-noise ratio (SNR) was lower for this *ex vivo* clinical data set than for the *in vitro* calibration data set. Therefore, we determined the error in each fit coefficient, excluded model components with fit coefficients less than twice these errors, and renormalized the remaining fit coefficients. χ^2 analysis was used to calculate the goodness of fit and the error associated with model fitting.[28] The Raman spectra in each diagnostic category have different SNRs; thus, mean errors are reported for each. Fitting errors for the two diagnostic model components, fat and collagen, are 0.003 and 0.001 for normal breast tissue, 0.025 and 0.014 for fibrocystic change, and 0.021 and 0.011 for fibroadenoma and IDC, respectively. Errors are slightly larger for fat than for collagen because the Raman spectrum of fat has more similarity to other model components than that of collagen. A cutoff of twice the error was employed in this analysis because this is the degree of agreement observed between experimental Raman data and errors determined via χ^2 analysis.[29]

3 Results

3.1 Prospective Analysis

Each of the 129 Raman spectra in the prospective data set was analyzed as described above to obtain the normalized fit coefficients of collagen and fat. The previously developed diagnostic algorithm was then applied prospectively to obtain the
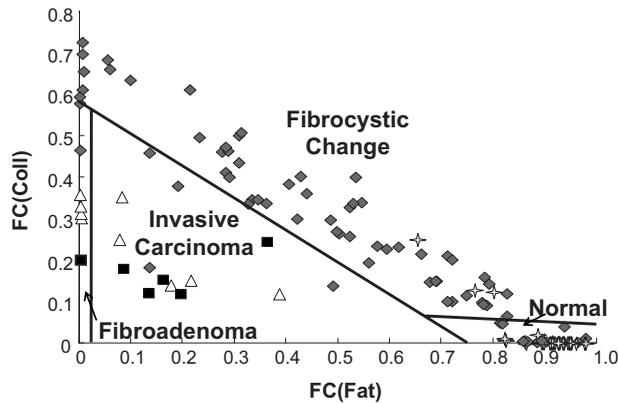
Fig. 2 Prospective application of the diagnostic algorithm developed *in vitro* on fresh-frozen tissues to an *ex vivo* clinical data set of freshly excised tissues which closely mimics the *in vivo* environment. Normal (stars), fibrocystic change (diamonds), fibroadenoma (triangles), and invasive ductal carcinoma (squares).

Raman spectral diagnoses, which were subsequently compared to traditional histopathology diagnoses. The results are shown in Fig. 2. Prospective application of this algorithm resulted in correct diagnosis of five of six cancerous sites (IDCs) and 114 of 123 noncancerous sites (normal breast tissues and benign lesions). This corresponds to a SE of 83% and a SP of 93%, giving a total test efficiency of 92% for the diagnosis of cancer. The overall accuracy of correctly classifying each of the four tissue types individually is 78% (101/129). The Raman spectral diagnoses and the histopathology diagnoses are compared in Table 1. We note that although there are only six cancerous specimens in the present data set, the diagnostic algorithm has previously been validated with 31 cancerous specimens. For reference, the results of the internally validated calibration set are SE 94% (29/31), SP 96% (91/95), total test efficiency 95%, and overall accuracy of 86% (108/126).[21]

Table 1 Comparison of pathologic diagnosis with that of the Raman diagnostic algorithm. Prospective application of the Raman diagnostic algorithm results in a sensitivity of 83%, a specificity of 93% and a negative predictive value of 99% for distinguishing cancerous from normal and benign tissues.

Raman diagnosis	Pathology diagnosis			
	Normal 41 spectra 18 patients	Fibrocystic change 73 spectra 16 patients	Fibroadenoma 9 spectra 2 patients	Invasive carcinoma 6 spectra 5 patients
Normal	38	13	0	0
Fibrocystic change	3	54	0	0
Fibroadenoma	0	2	4	1
Invasive carcinoma	0	4	5	5

Table 2 Summary of results from the *in vitro* calibration and *ex vivo* prospective data sets.

	In vitro calibration data set (%)	*Ex vivo* prospective data set (%)
SE	94	83
SP	96	93
PPV	88	36
NPV	98	99

The predictive values of the prospective and calibration data can also be computed. Unlike SE and SP, these depend on the prevalence of disease in the respective data sets and must be considered carefully in this light. The positive predictive value (PPV) and negative predictive value (NPV) of the prospective data set for the diagnosis of IDC are 36 and 99%, respectively. The corresponding values for the calibration set are 88 and 98%. These results are summarized in Table 2, and their significance examined in Sec. 4.

It is clear that any clinically accepted implementation of Raman spectroscopy to breast cancer diagnosis must encompass DCIS. Because of its diagnostic importance, breast tissue harboring DCIS was not routinely available in the case-control calibration study used for algorithm development, and thus, these specimens were excluded from prospective analysis. Although the algorithm was not developed to examine DCIS, it is of interest to observe where the samples diagnosed as DCIS fall on the diagnostic plane. Using our current algorithm based on the fit coefficients for fat and collagen, 5 of 20 DCIS specimens were correctly diagnosed as cancerous. The remaining 15 DCIS specimens were diagnosed as noncancerous (seven normal breast tissues, seven fibrocystic change, and one fibroadenoma). We note that eight of these spectra were obtained from patients who had undergone preoperative chemotherapy. All eight of these spectra were incorrectly diagnosed. It is clear that other fit coefficients must be incorporated into our algorithm to correctly diagnose DCIS. Studies are currently underway to expand our diagnostic algorithm to incorporate DCIS.

One area for algorithm refinement may be the incorporation of the nuclear-to-cytoplasm (N/C) ratio. Enlargement of cell nuclei is a hallmark of cancer.[30,31] In our studies, the spectroscopic parameter characterizing the N/C ratio is obtained by dividing the fit coefficient of the cell nucleus basis spectrum by that of the epithelial cell cytoplasm basis spectrum. Fibrocystic change and fibroadenoma have mean N/C parameters of 0.02 and 0.01, respectively, whereas infiltrating carcinoma has a much higher mean N/C parameter of 0.06. The mean N/C parameter of the samples diagnosed as DCIS is 0.05, intermediate between that of benign breast conditions and IDC, indicating the potential for detecting DCIS. Similar trends were seen in the mean N/C values in the calibration study, but the N/C parameter was not diagnostic because there was significant variability within pathologies.

It is also of interest to examine where the samples from patients who had undergone preoperative chemotherapy or were undergoing reexcision surgery fall on the diagnostic plane. Breast tissues from these patients demonstrate a fibrous healing reaction to surgery- or chemotherapy-induced tissue injury.[24,25] Consequently, the spectra display an increase in the fit coefficient for collagen and the majority of these data points were diagnosed as fibrocystic change. For instance, the mean collagen fit coefficient for samples diagnosed as IDC increased by 35% in patients who had received preoperative chemotherapy and 64% in patients who were undergoing re-excision surgery. Upcoming studies will focus on development of distinct diagnostic algorithms for patients who have undergone a prior surgery or preoperative chemotherapy.

4 Discussion

The performance of the diagnostic algorithm in this *ex vivo* prospective clinical study, summarized in Table 1, is generally quite good. In particular, the NPV of 99% is excellent, and as discussed below, NPV is the key parameter for making clinical decisions in either of our proposed applications. As expected, the SE and SP of the algorithm, applied prospectively, are lower than for the calibration data set. The NPV remains very high, while the PPV is greatly reduced.

In comparing the calibration and validation study, we have identified four factors that influence the performance of the diagnostic algorithm. (*i*) The use of prospective data rather than internally validated calibration data. The latter is always expected to give "more efficacious" results because the cross-validation employed to develop the algorithm gives internally consistent results. We note that some investigators regard such internal-validation techniques to be prospective.[32–34] (*ii*) Differences in the patient populations studied (cohort versus case control).[34] As discussed below, potential sources of bias in the patient populations must be properly taken into account. (*iii*) The use of freshly excised surgical samples as opposed to fresh-frozen samples. Past studies have shown that the Raman spectral line shape is equivalent in the two types of samples with regard to the moieties that comprise our model, since they are structural rather than metabolic in nature.[35] However, we did witness changes in the contribution of particular model components to the bulk Raman spectrum between the calibration and prospective studies. This may be due to changes in disease spectrum, as discussed below, or alternatively to changes in the tissue density/scattering between the fresh (*ex vivo*, prospective) and fresh-frozen (*in vitro*, calibration) data sets. Further experiments are needed to investigate the basis of these changes. (*iv*) Instrumentation factors, particularly SNR.[36]

4.1 Cohort versus Case-Control Study

Tissue samples were obtained by different methods in the calibration and prospective studies. Our initial calibration study was a case-control study, in which tissues were procured from 58 patients several different hospitals via the Co-operative Human Tissue Network (CHTN), snap frozen, and shipped to MIT for spectral interrogation. CHTN was asked to provide samples with grossly visible/palpable lesions paired with normal control tissues whenever possible. This artificially increased the prevalence of cancer (31/126; 25%) in

the calibration data set. The current prospective study was a cohort study, in which we examined freshly excised breast tissue from a series of 28 consecutive patients at a single hospital undergoing excisional biopsy, lumpectomy, or simple mastectomy. Spectra were collected from all types of breast tissue present in the surgical specimens and therefore included a much broader range of lesions. Only six of these patients had grossly visible/palpable lesions; the remainder had mammographically suspect/nonpalpable lesions, the majority of which were noncancers. This significantly reduced the prevalence of cancer (IDC) in the prospective study. However, the decreased cancer prevalence in the cohort data set better represents that expected in the target patient population for clinical spectroscopic diagnosis.[37] The decreased disease prevalence in the *ex vivo* prospective data set adversely affected the PPV of the diagnostic algorithm.

4.2 Disease Prevalence

Disease prevalence (or prior probability) refers to the proportion of individuals with a disease in a given population. The predictive value of a diagnostic test is influenced by disease prevalence. There are two types of predictive value. PPV gives the probability of disease if a test result is positive, while NPV gives the probability of no disease if a test result is negative. As the disease prevalence in a particular data set decreases, PPV decreases and NPV increases. Thus, the lower the disease prevalence is, the less the probability that a positive result will be "correct," regardless of other parameters of test performance (SE and SP).

The cancer prevalence in the calibration study was 25%, while in the prospective study it fell to 5%. As a result, PPV decreased and NPV increased in the prospective study. To examine the effect of disease prevalence on algorithm performance, we calculated predictive values for the calibration data set as a function of disease prevalence. According to Bayes' theorem,

$$PPV = pSE/[pSE + (1 - p)(1 - SP)],$$

$$NPV = (1 - p)SP/[(1 - p)SP + p(1 - SE)],$$

where p is prior probability of encountering disease (disease prevalence).[38] These equations allow the effectiveness of the algorithm to be extrapolated to data sets that are comprised of largely different proportions of normal and diseased tissues. As shown in Fig. 3, with a drop in the prevalence of invasive cancer from 25 to 5%, the PPV is expected to decrease from 88 to 53% while the NPV is expected to increase from 98 to 99%. Thus, the decrease in PPV seen in the prospective study (88–36%) is due, in large part, to the decrease in cancer prevalence in the patient cohort.

4.3 Disease Spectrum

SE, the probability of a positive test result among patients with the disease, and SP, the probability of a negative test result in a population without the disease, are not affected by disease prevalence. This is because SE depends on the distribution of positives and SP on the distribution of negatives, but both are independent of the relative number of positives and negatives in the data set. However, SE and SP are affected by
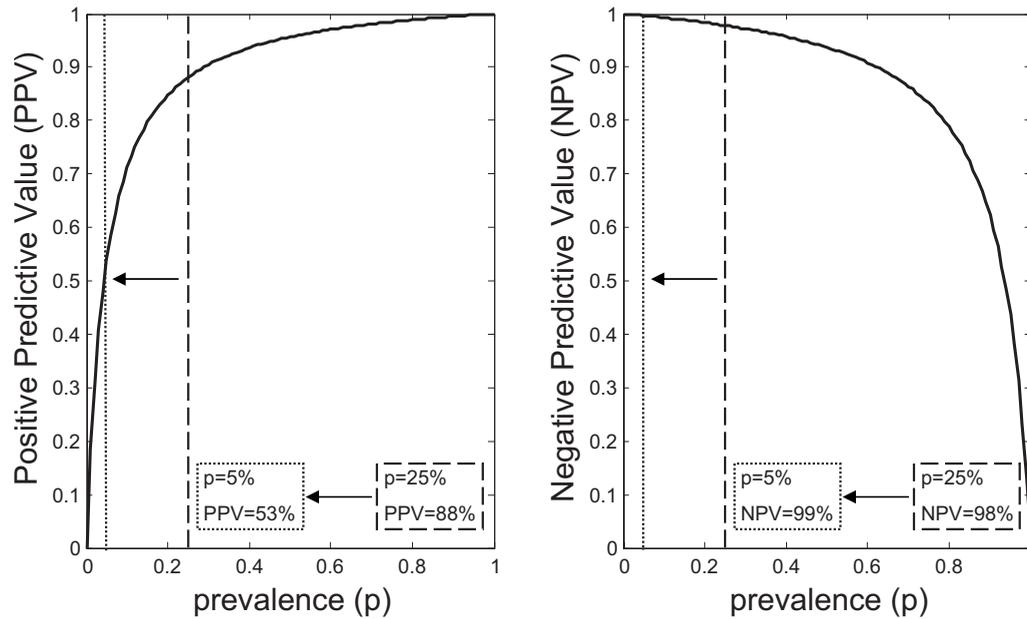
Fig. 3 Positive and negative predictive value of the diagnostic algorithm as a function of disease prevalence for the calibration data set. With a drop in the prevalence of invasive cancer from 25 to 5%, the PPV is predicted to fall from 88 to 54% while the NPV increases from 98 to 99%. The experimental PPV and NPV of the diagnostic algorithm in the prospective data set, which had a prevalence of invasive cancer of 5%, were 36 and 99%, respectively.

disease spectrum or variability in severity of disease in the study population.[39] Note that disease spectrum discussed in this section refers to disease variation and not Raman spectral variation. SE and SP will be high if the control tissues (non-cancers) are distinctly different from the target tissues (cancers). SE and SP both fall if the spectrum of disease in the control population changes such that the difference from the target population is less marked.[39]

This was the case for the fibrocystic change subset in the prospective study, where a broader range of histologic manifestations was encountered than in the calibration study. In the calibration study, the predominant manifestation of fibrocystic change seen was stromal fibrosis, a pathology characterized by collagen accumulation. In the prospective study, all three manifestations of fibrocystic change were observed, stromal fibrosis, cyst formation, and adenosis (gland proliferation). Representative histopathologic images illustrating the disease spectrum seen in fibrocystic change in the calibration and prospective data set are shown in Fig. 4. In addition to the pres-

ence of duct cysts and adenosis, the fibrocystic change specimens show a marked increase in the amount of fat relative to stromal fibrosis. Consistent with histopathology, within the fibrocystic change specimens, we observed an increase in the mean fit coefficient of fat from 0.32 to 0.50 and a decrease in the mean fit coefficient of collagen from 0.38 to 0.29 from the calibration to the prospective data set.

The change in the disease spectrum of the fibrocystic change control population can also be examined in the context of the entire Raman spectrum. Correlation coefficients, a measure of Raman spectral similarity, were calculated for the mean spectra in each diagnostic category and are shown in Table 3. The spectral correlation coefficient for fibrocystic change and IDC increased from 0.95 in the calibration study to 0.99 in the prospective study, indicating that the Raman spectra of the control tissues (fibrocystic change) and the target tissues (IDC) are more similar in the prospective study. This change in disease spectrum likely contributes to the decrease in both SE and SP in the prospective study.
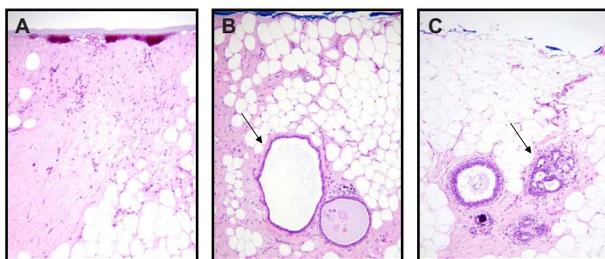
Fig. 4 Histopathologic photomicrographs (H&E; 10X) illustrating the disease spectrum seen in fibrocystic change in the calibration (A only) and prospective (A, B, and C) data sets: (A) fibrosis; (B) cyst formation (arrow), (C) adenosis (arrow).

Table 3 Raman spectral correlation coefficients for the mean spectra in each diagnostic category.

| | Correlation coefficient | |
	Calibration	Prospective
IDC versus fibrocystic change	0.95	0.99
IDC versus fibroadenoma	0.97	0.94
IDC versus normal	0.80	0.87

We also encountered an increase in the rate of misdiagnosis between fibroadenoma and IDC in the prospective data set. However, unlike fibrocystic change, the spectral correlation coefficient for fibroadenoma and IDC decreased from 0.97 in the calibration study to 0.94 in the prospective study, indicating that, in this case, there is more difference between the control tissues (fibroadenoma) and the target tissues (IDC) in the prospective study. This may be the result of spectral changes unrelated to fat and collagen content and thus may not play a role in prospective application of a diagnostic algorithm based on these two parameters. This conjecture is bolstered by examination of the root mean square (rms) of the mean residual (the difference between the spectrum and model fit) in each diagnostic category. The rms is a frequently used measure of the difference between values predicted by a model and the values actually observed. For samples diagnosed as fibroadenoma and IDC, the rms doubled between the calibration and validation data, while the rms remained relatively constant for samples diagnosed as normal and fibrocystic change. This indicates that there may be new components present in the fibroadenoma and IDC tissue samples examined in the prospective study. However, all of the rms values were small and the model was able to account for the majority of the spectral features observed. Thus, the new components are present in low concentrations or have relatively small Raman scattering cross sections.

The reason for the increase in the rate of misdiagnosis of fibroadenoma and IDC is currently unclear. Although we have minimized the effects of SNR through our spectral error analysis, we may not have completely nullified all SNR effects. Any remaining SNR effects would likely have the greatest impact on the diagnosis of fibroadenoma, as this diagnostic category occupies the smallest area in our diagnostic plane (Fig. 2). Nevertheless, the fact that the correlation coefficient for fibroadenoma and IDC decreased in the prospective data set is encouraging for future algorithm refinement.

Although it is difficult to quantify the effects of changes in disease spectrum on algorithm performance, it is clear that the decrease in SE and SP in the prospective study results, at least in part, from changes in disease spectrum.[39] Changes in disease spectrum also affect PPV and NPV, which depend not only on disease prevalence but additionally on SE and SP. Thus, the change in performance of the Raman diagnostic algorithm can be largely explained by differences in disease spectrum and cancer prevalence in the calibration and prospective studies.

4.4 *Clinical Significance of Positive and Negative Predictive Value*

Overall, the prospective PPV of our Raman diagnostic algorithm was more adversely affected than the NPV. However, in either of our proposed clinical applications, NPV is the statistic clinicians would rely on for decision making. In the case of transdermal spectral diagnosis via needle at mammography, the goal is to identify lesions that need to be biopsied or excised. A high NPV, such as in our prospective study, would allow the clinician to decide that a benign spectral diagnosis is sufficient to leave a lesion unbiopsied, while a low PPV might lead a clinician to unnecessarily biopsy a benign lesion. In this scenario, there is less risk to the patient in biopsying a benign lesion than in leaving a malignant lesion unbiopsied. The same is true in intraoperative margin assessment, where a high NPV would allow the surgeon to accept a margin as negative and not excise more tissue. In other words, a high NPV represents a high likelihood that the lesion/margin is not cancer.

Nevertheless, a higher PPV would result in a more robust overall diagnostic algorithm. In order to improve PPV, a new diagnostic algorithm will be devised using data sets with a sufficient number of samples in each diagnostic category that better represent the target patient population of our proposed applications. An iterative process of devising and prospectively validating new diagnostic algorithms in progressively larger *ex vivo* and *in vivo* clinical studies is needed to realize the full potential of Raman spectroscopy for breast cancer diagnosis.

5 Conclusions

The current study has validated a Raman spectroscopic algorithm for the diagnosis of breast cancer that was developed *in vitro,* on a large prospective *ex vivo* data set that closely mimics the target patient population of our anticipated *in vivo* clinical applications. It is the first prospective application of Raman spectroscopy in diagnosing normal, benign, and malignant breast tissue. The NPV of our diagnostic algorithm in this prospective study was excellent. Effects on the PPV, SE, and SP of the diagnostic algorithm were seen largely due to changes in disease prevalence and disease spectrum in the prospective data set. As the application of optical techniques to medicine and breast cancer diagnosis matures, algorithms must be tested prospectively to allow clinicians to accurately assess the diagnostic information contained in and any bias of the measurements. Although these preliminary results are promising, in order to fully assess the potential of Raman spectroscopy for breast cancer diagnosis, further algorithm development and larger scale *ex vivo* and *in vivo* studies are needed.

References

1. *American Cancer Society Breast Cancer Facts and Figures* (2007–2008).
2. M. Guven, B. Yazici, X. Intes, and B. Chance, "Diffuse optical tomography with a priori anatomical information," *Phys. Med. Biol.* **50**(12), 2837–2858 (2005).
3. G. Boverman, E. L. Miller, A. Li, Q. Zhang, T. Chaves, D. H. Brooks, and D. A. Boas, "Quantitative spectroscopic diffuse optical tomography of the breast guided by imperfect *a priori* structural information," *Phys. Med. Biol.* **50**(17), 3941–3956 (2005).
4. X. Cheng, J. M. Mao, R. Bush, D. B. Kopans, R. H. Moore, and M. Chorlton, "Breast cancer detection by mapping hemoglobin concentration and oxygen saturation," *Appl. Opt.* **42**(31), 6412–6421 (2003).
5. J. C. Hebden, H. Veenstra, H. Dehghani, E. M. Hillman, M. Schweiger, S. R. Arridge, and D. T. Delpy, "Three-dimensional time-resolved optical tomography of a conical breast phantom," *Appl. Opt.*

40(19), 3278–3287 (2001).

6. N. Shah, A. Cerussi, C. Eker, J. Espinoza, J. Butler, J. Fishkin, R. Hornung, and B. Tromberg, "Noninvasive functional optical spectroscopy of human breast tissue," *Proc. Natl. Acad. Sci. U.S.A.* **98**(8), 4420–4425 (2001).

7. V. Quaresima, S. J. Matcher, and M. Ferrari, "Identification and quantification of intrinsic optical contrast for near-infrared mammography," *Photochem. Photobiol.* **67**(1), 4–14 (1998).

8. C. Zhu, G. M. Palmer, T. M. Breslin, J. Harter, and N. Ramanujam, "Diagnosis of breast cancer using diffuse reflectance spectroscopy: comparison of a Monte Carlo versus partial least squares analysis based feature extraction technique," *Lasers Surg. Med.* **38**(7), 714–724 (2006).

9. S. K. Majumder, P. K. Gupta, B. Jain, and A. Uppal, "UV excited autofluorescence spectroscopy of human breast tissues for discriminating cancerous tissue from benign tumor and normal tissue," *Lasers Life Sci.* **8**, 249–264 (1998).

10. P. K. Gupta, S. K. Majumder, and A. Uppal, "Breast cancer diagnosis using N2 laser excited autofluorescence spectroscopy," *Lasers Surg. Med.* **21**(5), 417–422 (1997).

11. Y. Yang, A. Katz, E. Z. Celmer, M. Zurawska-Szczepaniak, and R. R. Alfano, "Excitation spectrum of malignant and benign breast tissues: a potential optical biopsy approach," *Lasers Life Sci.* **7**, 115–127 (1997).

12. I. J. Bigio, S. G. Bown, G. Briggs, C. Kelley, S. Lakhani, D. Pickard, P. M. Ripley, I. G. Rose, and C. Saunders, "Diagnosis of breast cancer using elastic-scattering spectroscopy: preliminary clinical results," *J. Biomed. Opt.* **5**(2), 221–228 (2000).

13. C. V. Raman and K. S. Krishnan, "A new type of secondary radiation," *Nature (London)* **121**, 501–502 (1928).

14. E. B. Hanlon, R. Manoharan, T. W. Koo, K. E. Shafer, J. T. Motz, M. Fitzmaurice, J. R. Kramer, I. Itzkan, R. R. Dasari, and M. S. Feld, "Prospects for *in vivo* Raman spectroscopy," *Phys. Med. Biol.* **45**(2), R1–59 (2000).

15. R. R. Alfano, C. H. Liu, and W. Sha, "Human breast tissues studied by IR Fourier transform Raman spectroscopy," *Lasers Life Sci.* **4**, 23–28 (1991).

16. D. Redd, Z. Feng, K. Yue, and T. Gansler, "Raman spectroscopic characterization of human breast tissues: implications for breast cancer diagnosis," *Appl. Spectrosc.* **47**, 787–791 (1993).

17. C. J. Frank, D. C. Redd, T. S. Gansler, and R. L. McCreery, "Characterization of human breast biopsy specimens with near-IR Raman spectroscopy," *Anal. Chem.* **66**(3), 319–326 (1994).

18. C. J. Frank, R. L. McCreery, and D. C. Redd, "Raman spectroscopy of normal and diseased human breast tissues," *Anal. Chem.* **67**(5), 777–783 (1995).

19. R. Manoharan, K. Shafer, L. Perelman, J. Wu, K. Chen, G. Deinum, M. Fitzmaurice, J. Myles, J. Crowe, R. R. Dasari, and M. S. Feld, "Raman spectroscopy and fluorescence photon migration for breast cancer diagnosis and imaging," *Photochem. Photobiol.* **67**(1), 15–22 (1998).

20. K. E. Shafer-Peltier, A. S. Haka, J. T. Motz, M. Fitzmaurice, R. R. Dasari, and M. S. Feld, "Model-based biological Raman spectral imaging," *J. Cell Biochem. Suppl.* **39**, 125–137 (2002).

21. A. S. Haka, K. E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R. R. Dasari, and M. S. Feld, "Diagnosing breast cancer by using Raman spectroscopy," *Proc. Natl. Acad. Sci. U.S.A.* **102**(35), 12371–12376 (2005).

22. P. Matousek and N. Stone, "Prospects for the diagnosis of breast cancer by noninvasive probing of calcifications using transmission Raman spectroscopy," *J. Biomed. Opt.* **12**(2), 024008 (2007).

23. N. Stone and P. Matousek, "Advanced transmission Raman spectroscopy: a promising tool for breast disease diagnosis," *Cancer Res.* **68**(11), 4424–4430 (2008).

24. A. A. Tardivon, J. M. Guinebretiere, C. Dromain, M. Deghaye, H. Caillet, and V. Georgin, "Histological findings in surgical specimens after core biopsy of the breast," *Eur. J. Radiol.* **42**(1), 40–51 (2002).

25. A. Moreno, A. Escobedo, E. Benito, J. M. Serra, A. Guma, and F. Riu, "Pathologic changes related to CMF primary chemotherapy in breast cancer: pathological evaluation of response predicts clinical outcome," *Breast Cancer Res. Treat.* **75**(2), 119–125 (2002).

26. J. T. Motz, S. J. Gandhi, O. R. Scepanovic, A. S. Haka, J. R. Kramer, R. R. Dasari, and M. S. Feld, "Real-time Raman system for in vivo disease diagnosis," *J. Biomed. Opt.* **10**(3), 031113 (2005).

27. J. T. Motz, M. Hunter, L. H. Galindo, J. A. Gardecki, J. R. Kramer, R. R. Dasari, and M. S. Feld, "Optical fiber probe for biomedical Raman spectroscopy," *Appl. Opt.* **43**(3), 542–554 (2004).

28. J. F. Kenney and E. S. Keeping, *Mathematics of Statistics*, Pt. 2, Van Nostrand, Princeton (1951).

29. O. R. Scepanovic, K. L. Bechtel, A. S. Haka, W. C. Shih, T. W. Koo, A. J. Berger, and M. S. Feld, "Determination of uncertainty in parameters extracted from single spectroscopic measurements," *J. Biomed. Opt.* **12**(6), 064012 (2007).

30. C. W. Elston and I. O. Ellis, "Pathological prognostic factors in breast cancer. I. the value of histological grade in breast cancer: experience from a large study with long-term follow-up," *Histopathology* **19**(5), 403–410 (1991).

31. S. A. Hoda and P. P. Rosen, "Practical considerations in the pathologic diagnosis of needle core biopsies of breast," *Am. J. Clin. Pathol.* **118**(1), 101–108 (2002).

32. N. Ramanujam, M. F. Mitchell, A. Mahadevan, S. Thomsen, A. Malpica, T. Wright, N. Atkinson, and R. Richards-Kortum, "Development of a multivariate statistical algorithm to analyze human cervical tissue fluorescence spectra acquired *in vivo*," *Lasers Surg. Med.* **19**(1), 46–62 (1996).

33. N. Ramanujam, M. F. Mitchell, A. Mahadevan, S. Thomsen, A. Malpica, T. Wright, N. Atkinson, and R. Richards-Kortum, "Spectroscopic diagnosis of cervical intraepithelial neoplasia (CIN) *in vivo* using laser-induced fluorescence spectra at multiple excitation wavelengths," *Lasers Surg. Med.* **19**(1), 63–74 (1996).

34. R. S. Galen and S. R. Gambino, *Beyond Normality: The Predictive Value and Efficienty of Medical Diagnoses*, Wiley, Hoboken, NJ (1975).

35. J. F. Brennan, "Near infrared Raman spectroscopy for human artery histochemistry and histopathology," Ph.D. thesis, Department of Electrical Engineering, Massachusetts Institute of Technology, Cambridge (1995).

36. U. Utzinger, E. V. Trujillo, E. N. Atkinson, M. F. Mitchell, S. B. Cantor, and R. Richards-Kortum, "Performance estimation of diagnostic tests for cervical precancer based on fluorescence spectroscopy: Effects of tissue type, sample size, population, and signal-to-noise ratio." *IEEE Trans. Biomed. Eng.* **46**(11), 1293–1303 (1999).

37. J. J. Jobsen, J. Van Der Palen, F. Ong, and J. H. Meerwaldt, "Differences in outcome for positive margins in a large cohort of breast cancer patients treated with breast-conserving therapy," *Acta Oncol.* **46**(2), 172–180 (2007).

38. E. K. Harris and A. Albert, *Multivariate Interpretation of Clinical Laboratory Data*, Decker, New York (1987).

39. S. A. Mulherin and W. C. Miller, "Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation." *Ann. Intern Med.* **137**(7), 598–602 (2002).