Assessing breast tissue density by transillumination breast spectroscopy (TIBS): an intermediate indicator of cancer risk

1K M BLACKMORE, MSC, 2J A KNIGHT, PHD, 3R JONG, MD and 1,4L LILGE, PHD

1Ontario Cancer Institute, University Health Network, Toronto, Ontario, Canada M5G 2M9, 2Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada M5G 1X5, 3Sunnybrook and Women’s College Health Science Centre, Toronto, Ontario, Canada M4N 3M5 and 4Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada M5G 2M9

ABSTRACT. Risk assessment by parenchymal density pattern, a strong physical indicator of future breast cancer risk, is available with the onset of mammographic screening programmes. However, due to the use of ionizing radiation, mammography is not recommended for use in younger women, thereby rendering risk assessment unattainable at an earlier age. Visible and near infrared light was used on 292 women with radiological normal mammograms to determine whether transillumination breast spectroscopy (TIBS) can identify women with a high parenchymal density pattern as an intermediate indicator of breast cancer risk. Principal component analysis (PCA) was used to reduce the spectral data and generate density scores for each woman. To assess the accuracy of TIBS, logistic regression was used to calculate crude and adjusted odds ratios (OR) and 95% confidence intervals (CI) for each score. Receiver operator characteristic (ROC) curves and area under the curve (AUC) were also calculated for the crude and adjusted logistic models. Optical information relating to tissue chromophores, such as water, lipid and haemoglobin content, was sufficient to identify women with high parenchymal density. The resulting AUC for the final and most parsimonious multivariate logistic model was 0.922 (95% CI 0.878–0.967). TIBS provides information correlating to high parenchymal density and is a promising tool for risk assessment, particularly for younger women.

Introduction

Breast cancer is the most commonly occurring cancer in women [1, 2]. While screening programmes have resulted in decreased mortality rates as a result of the detection of early stage cancers, the overall incidence of breast cancer is still increasing [1–3]. Consequently, within the field of preventive oncology, intervention (i.e. risk-reducing) strategies are being considered with the primary goal of decreasing breast cancer incidence rates and thereby maintain the health of the female population at a high level [4–8]. However, to attain such a goal, accurate and generalizable techniques are required to identify all women at greatest risk who would benefit from risk-reducing intervention programmes [4–8]. Therefore, the most useful risk assessment technique would also rely on an identifier demonstrating a large relative risk and be applicable to all women, unlike, for example, genetic markers such as BRCA1 and BRCA2 mutations, which occur only in a very small proportion of the female population [9].

Research into breast cancer aetiology has shown that the development of the disease is a slow process following initial transformation of the breast tissue, and that events early in life may be important in establishing future breast cancer risk [10, 11]. Moreover, initiating risk assessment in young women may have the added benefit that potentially less drastic intervention strategies (i.e. diet, exercise and lifestyle changes) acting over a longer period of time could be sufficient to achieve an adequate reduction in risk [12, 13].

One physical method of evaluating breast tissue risk that is relatively well established is the X-ray dense tissue content of the breast as obtained by standard mammography. Parenchymal density patterns assessed using quantitative methods have been consistently related to breast cancer risk [14, 15]. Several studies have shown that women with dense tissue occupying 75% or more of the total breast area compared with those with low density (< 25%) are four to six times more likely to develop breast cancer in the next decade [16–26]. Given that the extent of mammographic density is influenced by hormonal exposure (such as during the menstrual cycle and pregnancy), Boyd et al [14] have argued that mammographic density is a marker of susceptibility to breast cancer consistent with the concept of rate of breast tissue ageing introduced by Pike et al [10]. The main drawback to the use of mammography as an indicator of breast cancer risk is the required exposure to ionizing radiation, and hence there are concerns regarding its use in young women for regular screening (less than 40 years of age in the USA and younger than 50 years in Canada and the UK) [27, 28]. The technique also requires compression of the breast, which causes discomfort in some women.
Alternative methods that have been proposed for assessing breast tissue density include MRI and ultrasound. Preliminary work has demonstrated a correlation between relative water content derived with MRI and mammographic density [29, 30] and the extent of echogenic areas on ultrasound with mammographic parameters [31, 32]. Although MRI does not employ ionizing radiation, the technique is expensive and in high demand for other clinical applications, and being in an enclosed space can cause discomfort in some individuals. A more recently proposed method of assessing the state of breast tissue is to directly measure biomarkers in nipple aspirate fluid. There is some evidence that epithelial hyperplasia and atypical hyperplasia detected in this fluid are associated with increased breast cancer risk [33]. However, it is an invasive procedure, and nipple aspirate fluid was unobtainable in 40% of the cohort in whom breast cancer risk was assessed [33].

Near infrared (NIR) transillumination breast spectroscopy (TIBS) is a non-imaging, non-invasive technique that provides information about bulk tissue properties through the spectral dependency of photons that have passed through the breast tissue [34–36]. In contrast to mammography, TIBS uses non-ionizing visible and NIR light and can therefore be used on younger women, theoretically beginning at puberty. Plate compression of the breast tissue is not necessary, and each measurement takes no more than a few seconds. Additionally, no special training is required for its use, and the device and examination can be made available at costs at least an order of magnitude lower than mammography. Breast tissue is a highly scattering medium with relatively low absorption in the red and NIR wavelength ranges, permitting sufficient light penetration to detect signals through up to 7 cm of tissue, while maintaining the incidence power below government guidelines for light exposure to skin [37].

Multiple studies using NIR technologies to examine breast tissue have been published, and the findings for healthy breast tissue composition are fairly consistent [38–47]. In the NIR spectrum, the main absorbers of photons (i.e. chromophores) are water, lipids, collagen and haemoglobins (oxy- and deoxy-). In breast tissue, fibroglandular tissue results in increased water and concomitant decreased lipid-associated absorption. Stromal tissue, which contributes to high parenchymal density patterns, further suggests high collagen content. Larger total haemoglobin content and a trend towards lower oxygen saturation are also anticipated in high-density tissue compared with fatty tissue because of increased tissue vascularization and cellular proliferation, and thus increased metabolism. Variations in tissue scattering particle density (i.e. cells and intracellular organelles, collagen) affecting optical path length also will result in increased scattering and, conversely, in increased attenuation.

The study presented here is an extension of earlier publications by our group [34–36] and includes an analysis of the complete spectral data set obtained from a cross-sectional study comprising 292 pre- and post-menopausal women without radiologically suspicious lesions. The primary aim of the study was to evaluate the feasibility of identifying all women with very high breast tissue density (>75% dense tissue) in vivo using TIBS. The underlying hypothesis is that TIBS provides information consistent with conventional mammography in identifying breast tissue density and hence is a potential intermediate indicator of breast cancer risk. Mammographic density was selected as the gold standard because it is a physical assessment of breast tissue applicable to the entire female population over 40 years of age, and carries the highest odds ratio (OR) [4–6] with respect to other known non-genetic-based risk factors.

### Materials and methods

#### Study population

Three hundred participants were recruited among women having a standard X-ray mammography at the Marcelle Koffler Breast Centre in Mount Sinai Hospital, Toronto, Ontario (Table 1). Eligible women had a screening mammogram (233 film and 67 digital) within 12 months prior to recruitment (between 1 March 2000

<table>
<thead>
<tr>
<th>Total usable mammograms</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analogue</td>
<td>233</td>
</tr>
<tr>
<td>Digital</td>
<td>67</td>
</tr>
<tr>
<td>Missing demographic information</td>
<td>8</td>
</tr>
<tr>
<td>Included in final data analysis</td>
<td>292</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Breast density</th>
<th>Pre-menopausal&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Post-menopausal&lt;sup&gt;b&lt;/sup&gt;</th>
<th>All women&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (&lt; 25%)</td>
<td>33 (23.9%)</td>
<td>70 (45.5%)</td>
<td>103 (35.3%)</td>
</tr>
<tr>
<td>Medium (25–&lt; 75%)</td>
<td>78 (56.5%)</td>
<td>69 (44.8%)</td>
<td>147 (50.3%)</td>
</tr>
<tr>
<td>High (≥ 75%)</td>
<td>27 (19.6%)</td>
<td>15 (9.7%)</td>
<td>42 (14.4%)</td>
</tr>
<tr>
<td>Total</td>
<td>138</td>
<td>154</td>
<td>292</td>
</tr>
<tr>
<td>Age (years) mean (± SD) (range)</td>
<td>45.9 ± 4.3 (34–59)</td>
<td>55.4 ± 6.4 (41–77)</td>
<td>50.9 ± 7.2 (34–77)</td>
</tr>
<tr>
<td>BMI (kg m&lt;sup&gt;-2&lt;/sup&gt;) mean (± SD) (range)</td>
<td>25.3 ± 5.9 (18.0–54.9)</td>
<td>26.2 ± 4.9 (16.6–43.9)</td>
<td>25.8 ± 5.5 (16.6–54.9)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Pre-menopausal women had their last menstrual period less than 1 month before mammography and transillumination breast spectroscopy (TIBS) measurements.
<br>

<sup>b</sup>Post-menopausal women had their last menstrual period at least 12 months before mammography and TIBS measurements.
<br>

<sup>c</sup>Proportions from Canadian National Breast Screening Study: low = 37%, medium = 49%, high = 14%.
and 30 September 2004) showing no radiologically suspicious lesions. Women did not have previous surgery to the breast tissue, including reduction or augmentation. Women displaying variations in parenchymal density between breasts were excluded as their correlation with cancer risk is weaker. Information concerning participants’ age, menopausal status (pre-menopausal vs post-menopausal), height and weight were collected by means of a self-administered questionnaire. Participants were reimbursed for travel expenses only. Post-menopausal status was defined as having had no menstrual period for at least 12 months. Height and weight were used to calculate body mass index (BMI) defined as weight in kilograms divided by the square of the height in metres. This study was approved by the Institutional Review Boards (IRBs) of the University of Toronto, Mount Sinai Hospital and the University Health Network.

_quantification of parenchymal density from mammograms_

All usable mammograms (300) were classified on an ordinal scale by an expert radiologist (RJ) into low (< 25%), medium (25% to < 75%) or high (> 75% dense tissue area) density categories for analysis using ordinal data. Boyd et al [24] and Jong et al [48] have demonstrated a high level of inter-radiologist agreement for density classification (intraclass correlation coefficients of 0.94 and 0.89, respectively), indicating that the validity of the study was not negatively affected by using one radiologist. Furthermore, the radiologist used in the present study was very experienced and has participated in previous studies requiring classification of mammograms on an ordinal scale [24, 26, 48].

_optical set-up and procedure_

All optical measurements were collected prior to quantification of the participant’s tissue density class. The instrumentation used to gather transillumination spectra has been described in detail previously [34–36]. A 50 W halogen lamp served as the broadband light source. Ultraviolet, part of the visible spectrum and mid-infrared radiation were eliminated using a cut-on (λ > 550 nm) and a heat rejection filter, respectively. The remaining light was coupled into a 5 mm diameter liquid light guide (Fibre Guide, Bridgeport, CT) placed in contact with the skin on top of the breast. A total power of 0.25 W, covering the 550–1300 nm bandwidth, was delivered to the skin. Transmitted light was collected via a 7 mm diameter optical fibre bundle (140 Si/Si fibres, 200 μm core diameter, numerical aperture: 0.36; P & P Optica, Kitchener, Canada). The source and detector fibre bundles (optodes) were held coaxially, pointing towards each other, by a calliper attached to the resting platform, also providing the interoptode distance. The source fibre was placed against the skin on the top surface of the breast with minimal compression. Wavelength-dependent detection in the visible and NIR was achieved using a spectro-photometer (Kaiser, CA) with holographic transillumination grating (15.7 rules mm\(^{-1}\) blazed at 850 nm) and a two-dimensional cryogenically cooled silicon charge-coupled device (CCD; Photometrics, NJ) at a spectral resolution of better than 3 nm between 625 and 1060 nm, achieved by using a 0.5 mm entrance slit to the spectrophotometer. Total data acquisition times for all required measurements ranged from 160 to 200 s. Considering typical tissue optical properties and tissue thickness ranging from 2.5 to 7 cm, ovoid-shaped tissue volumes of 12–54 cm\(^3\) were interrogated per spectral measurement. Hospital Grade Canada Standards Association (CSA) certification and Health Canada Investigational New Device Class II approval were obtained.

All measurements were taken in the dark, with the participant seated comfortably in an upright position and the breast resting on the support platform. A total of eight measurements in craniocaudal projection were taken per individual, four per breast (centre; midline close to the pectoral muscle, medial; 2 cm from the inner edge, distal; 2 cm behind the nipple, and lateral; 2 cm from the outer edge), resulting in optical interrogation of different anatomical regions of the breast [34–36]. Temporal and spatial reproducibility of the optical measurement is good, as addressed previously [34, 35].

_Preparation of spectra for data analysis_

Spectra were corrected for daily variations in the wavelength-dependent signal transfer function of the optical system (< 1% day by day) and the thickness of the interrogated tissue, given by the interoptode distance. To correct for the signal transfer function, spectra were referenced to a transmission standard made of 1 cm thick ultrahigh-density polyurethane (Gigahertz Optics, Munich, Germany). The optical properties (OD cm\(^{-1}\) ~ 1.8–2.3 over the wavelength range of interest) of the polyurethane standard were measured in a separate experiment using an integrating sphere diffuse reflectance set-up [49]. Hence, spectra used in further data processing are independent of the instrument and the interoptode distance, and are expressed in units of optical density per centimetre (OD cm\(^{-1}\)), calculated using the negative log of the raw data spectrum, the reference spectrum of the polyurethane standard and the interoptode distance [34–36].

_Principal component analysis (PCA)_

To establish a correlation between the obtained transillumination spectra, here considered vectors (OD cm\(^{-1}\) vs wavelengths from 625 to 1060 nm) and a target (breast tissue density on an ordinal scale), PCA was used [48, 49]. PCA is a commonly used data analysis technique in the field of chemometrics and for spectroscopic analysis in medical applications [51, 52]. For PCA implementation, spectra from all usable high, medium and low tissue density women (Table 1; 300 women × 8 measurements, n = 2400) were employed. PCA was executed using Matlab 12.1 (The MathWorks Inc., MI). All other statistical analyses were carried out using SPSS (Statistical Packages for the Social Sciences, SPSS Inc., Chicago, IL), version 11.0 and SAS (Statistical Analysis Systems; SAS Institutes Inc., USA), version 9.1.
Prior to PCA implementation, the mean spectrum ($\bar{S}$) of all spectra was calculated and subtracted from each individual spectrum ($S_i$), resulting in $S_i = S_i - \bar{S}$, in order to derive mutually orthogonal principal components, and hence independent principal component scores. PCA derives a minimum number of representative spectra, the principal components ($p_a$), accounting for the majority of the variance seen in the entire mean-centred spectral data set. All principal component spectra contain a varying amount of metabolic and structural information from the interrogated tissue (tissue scattering by cellular or structural components, lipid and water content, deoxy- and oxyhaemoglobin content) [34–36]. Once the principal components were derived, scalar coefficients (i.e., scores, $t_{i,a}$) were assigned to each individual mean-centred spectrum ($S_i$) measured at each position for each woman, indicating the contribution of each principal component, and hence the optically interrogated chromophores, to that spectrum. In the present study, each individual spectrum is a linear combination of four principal component spectra ($p_a$) multiplied by the respective scalar coefficient or score ($t_{i,a}$), such that: $S_i = t_{i1}p_1 + t_{i2}p_2 + t_{i3}p_3 - t_{i4}p_4 + \epsilon$, where $\epsilon$ represents the residual error. (Note: while $t_1$ to $t_3$ showed an inverse relationship with mammographic density, $t_4$ showed a direct relationship.)

Bilateral symmetry in the spectra at corresponding quadrants was demonstrated previously [34, 35] in women without variations in parenchymal pattern. Consequently, the resulting principal component scores ($t_{i,a}$), herein referred to as ‘density’ scores, were averaged over all measurement positions on both breasts for each woman prior to subsequent statistical analysis. Averaging scores post PCA compared with averaging spectra prior to PCA results in a global optical assessment of the tissue, similar to mammographic density class assignment, while permitting assessment of intraperson variability. Descriptive statistics [median ± interquartile range (IQR)] were determined for the derived density scores ($t_{i,a}$) for the high ($\geq 75\%$) and combined low and medium density categories ($< 75\%$) for all women and by menopausal status (pre-menopausal vs post-menopausal). The analysis of high vs combined low and medium density best approximates the clinical decision-making process, requiring high sensitivity and specificity for the identification of only women at risk from among the entire female population. As the scores were generally not normally distributed, score differences between $< 75\%$ and $\geq 75\%$ tissue densities were tested by non-parametric methods using the Mann–Whitney $U$-test.

For all analyses, $p$-values $\leq 0.05$ were considered to be statistically significant. As eight women were missing information on height and weight, their BMI could not be calculated. For consistency, these women were excluded from further statistical analyses, and the data presented here are for $n = 292$ women with complete spectral and demographic information.

**Logistic regression using PCA scores**

To measure the association between the derived density scores ($t_{i,a}$) and breast tissue density, univariate and multivariate logistic regression was used to estimate both crude (i.e., unadjusted) and adjusted odds ratios (OR) and 95% confidence intervals (CI), respectively, with mammographic density as the outcome. For the analysis presented here, breast density was treated as a dichotomous variable, by comparing women with very dense breasts ($\geq 75\%$ dense tissue) with those with $< 75\%$ dense tissue. All models were fit with the density scores treated on a continuous scale.

To identify potentially confounding variables for inclusion in the multivariate models, univariate logistic regression analysis was also used to estimate crude OR and 95% CI for age, BMI and menopausal status with high mammographic density ($\geq 75\%$ dense tissue) as the outcome. For this analysis, age and BMI were treated on a continuous scale, while menopausal status was treated as a dichotomous variable (pre- vs post-menopausal status). The strength of the association of each density score with mammographic density as a function of menopausal status was assessed by creating an interaction term between the variable menopausal status and each density score and including these terms in the logistic models examined. For all logistic models, covariates with $p$-values $\leq 0.05$ were considered to be statistically significant. Furthermore, for both the univariate and the multivariate logistic models, the calculated individual probabilities of having density $\geq 75\%$ were used to derive receiver operator characteristic (ROC) curves and area under the curve (AUC) in order to evaluate how well each model predicts the outcome.

Correlation analysis was also executed between the scores ($t_{i,a}$), age and BMI to examine the relationship among the dependent variables.

**Results**

**Study population**

Table 1 lists the mammographic, demographic, reproductive and anthropometric information for all eligible women used in the data analysis ($n = 292$). The overall study population proportions are comparable to those seen in the Canadian National Breast Screening Study (CNBSS age range 40–59 years) [24].

**PCA of tissue density categories**

The first four principal components resulting from PCA (Figure 1) capture 96.41%, 1.84%, 1.06% and 0.35% of the variance, respectively, yielding a combined total of 99.66% for the data set comprising spectra from all three tissue density classes. Principal component $p_1$ carries information on the overall light attenuation of the interrogated tissue due to differential optical path length resulting from light scattering and losses at the boundary, while components $p_2$ to $p_4$ contain information about the water, lipid and the oxy- and deoxyhaemoglobin content of the tissue [34–36].

Box plots of the derived density scores $t_1$ to $t_4$ according to risk classification (low $< 75\%$ vs high $\geq 75\%$ dense tissue) are shown in Figure 2 for all women ($n = 292$). The median values and IQR for each of the
The first four PCA scores $t_1$ to $t_4$ for women with very dense tissue ($\geq 75\%$) vs women with $< 75\%$ dense tissue are shown in Table 2 for all women and by menopausal status. A Mann–Whitney $U$-test demonstrated that $t_1$, $t_3$ and $t_4$ were each significantly different between $< 75\%$ and $\geq 75\%$ tissue densities among all women and post-menopausal women, while only $t_1$ and $t_3$ retained significance among the pre-menopausal group. Density score $t_2$ was not significantly different between high and low risk tissue density for all women, or when examined by menopausal status (Table 2).

**Logistic regression using density scores**

Table 3 presents the unadjusted OR and accompanying 95% CI for the OR for each density score $t_1$ to $t_4$, age, BMI and menopausal status (pre- vs post-menopausal). The OR and accompanying 95% CI for the OR are expressed over the IQR of the respective density score $t_{\text{IQR}}$, age or BMI for $n = 292$ and represent the odds of high breast tissue density ($\geq 75\%$) vs the odds of $< 75\%$ dense tissue for an increase in the respective density score, age or BMI over the IQR [i.e. from the first quartile (Q1) to the third quartile (Q3)].

The odds of having $\geq 75\%$ tissue density was significantly and inversely associated with the values of $t_1$, $t_3$ and $t_4$ while, for density score $t_2$, the OR was not significant. BMI was strongly and inversely associated with high breast tissue density, while the odds of $\geq 75\%$ dense tissue among pre-menopausal women was 2.25 times the odds among post-menopausal women. The association between age and high breast tissue density was not significant. Among the univariate logistic models, the calculated AUC was highest for density score $t_3$.

Correlation analysis demonstrated a strong positive association between $t_1$ and BMI ($r = 0.663$, $p < 0.001$), and between $t_1$, $t_2$ and $t_4$ and age ($r = 0.248$, $r = 0.298$ and $r = 0.391$, respectively; all at $p < 0.001$).

Table 4 shows OR and accompanying 95% CI for the OR for each density score $t_1$ to $t_4$ adjusted for age, BMI and menopausal status (models I to IV). In the adjusted models, density score $t_4$ was no longer significantly associated with high breast tissue density. As in previous analyses, $t_2$ was not significantly associated with...
mammographic density, while the effect of $t_3$ did not change (Table 3 vs Table 4). However, unlike the other density scores, the strength of the inverse association of density score $t_4$ with high breast tissue density varied according to menopausal status (model IV, Table 4).

Figure 3 displays the estimated odds of high breast tissue density as a function of $t_4$ separately for pre- and post-menopausal women while adjusting for mean age (50.9 years) and mean BMI (25.8). Although the odds of high breast tissue density were higher among pre-

Table 2. Median, interquartile range (IQR) and results of Mann–Whitney U-test for derived scores $t_1$ to $t_4$ for all women ($n = 292$) and by menopausal status

<table>
<thead>
<tr>
<th></th>
<th>All women</th>
<th>Pre-menopausal</th>
<th>Post-menopausal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$&lt; 75%$</td>
<td>$&gt; 75%$</td>
<td>$&lt; 75%$</td>
</tr>
<tr>
<td>$n$</td>
<td>$n = 250$</td>
<td>$n = 42$</td>
<td>$n = 111$</td>
</tr>
<tr>
<td>$t_1$</td>
<td>0.91 (4.36)</td>
<td>-1.26 (4.72)</td>
<td>0.01 (4.73)</td>
</tr>
<tr>
<td>$t_2$</td>
<td>0.09 (0.55)</td>
<td>0.02 (0.87)</td>
<td>0.09 (0.56)</td>
</tr>
<tr>
<td>$t_3$</td>
<td>0.08 (0.29)</td>
<td>-0.31 (0.46)</td>
<td>0.10 (0.27)</td>
</tr>
<tr>
<td>$t_4$</td>
<td>-0.02 (0.25)</td>
<td>-0.09 (0.19)</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Figure 2. Box plots of derived scores $t_1$ to $t_4$ (a to d) by tissue density classification ($< 75\%$ vs $> 75\%$) for all women ($n = 292$).
menopausal women, this varied only slightly across values of \( t_4 \), while among post-menopausal women, the odds showed a strong inverse association as \( t_4 \) increased.

Table 5 displays the OR and 95% CI for a model including density scores \( t_3 \) and \( t_4 \), menopausal status, BMI and the interaction between \( t_4 \) and menopausal status. Because \( t_1 \), \( t_2 \) and age were not significantly associated with high breast tissue density, either independently (Table 3) or once adjusted for other covariates (Table 4), they were excluded from further analyses in order to attain the most parsimonious model. As can be seen from Table 5, the inclusion of both \( t_3 \) and \( t_4 \) in a single model resulted in a decrease in the OR (i.e. stronger) for \( t_3 \) for all women and for \( t_4 \) among post-menopausal women, and both remained significantly and inversely associated with high breast tissue density. BMI also remained strongly and inversely associated with the outcome. The AUC for the final model was 0.922 (95% CI 0.876–0.968) (Figure 4).

### Discussion

This study extends earlier publications by our group correlating TIBS with density classification on an ordinal scale [35, 36] to include the complete spectral data set of 292 pre- and post-menopausal women with radiologically normal breast tissue. While previous studies covered only high (>75%) vs low (<25%) mammographic tissue density, the main purpose of the analysis presented here was to determine the ability of TIBS to discriminate women with >75% dense tissue from the remainder of the total study population in order to establish the utility of TIBS as a potential assessment tool for breast cancer risk.

### Table 3. Results of univariate logistic regression analysis for each density score \( t_1 \) to \( t_4 \), age, body mass index (BMI) and menopausal status (n = 292)

<table>
<thead>
<tr>
<th>Density</th>
<th>OR ( a )</th>
<th>95% CI ( a )</th>
<th>IQR</th>
<th>( p )-value</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_1 )</td>
<td>0.43</td>
<td>0.29</td>
<td>0.64</td>
<td>4.31</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>( t_2 )</td>
<td>0.81</td>
<td>0.54</td>
<td>1.22</td>
<td>0.59</td>
<td>0.32</td>
</tr>
<tr>
<td>( t_3 )</td>
<td>0.25</td>
<td>0.16</td>
<td>0.38</td>
<td>0.37</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>( t_4 )</td>
<td>0.59</td>
<td>0.38</td>
<td>0.93</td>
<td>0.24</td>
<td>0.02</td>
</tr>
<tr>
<td>Age</td>
<td>0.70</td>
<td>0.49</td>
<td>1.11</td>
<td>9.00</td>
<td>0.14</td>
</tr>
<tr>
<td>BMI</td>
<td>0.25</td>
<td>0.14</td>
<td>0.53</td>
<td>6.25</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Pre-menopausal</td>
<td>2.25</td>
<td>1.14</td>
<td>4.44</td>
<td>4.44</td>
<td>0.02</td>
</tr>
</tbody>
</table>

AUC, area under the curve.

\( \text{OR} a \) Odds ratio (OR) and 95% confidence interval (CI) for OR calculated over an interquartile range (IQR) of the respective density score, age and BMI; IQR calculated for all women (n = 292). OR and 95% CI represent the odds of high breast tissue density (≥75%) vs the odds of <75% tissue density associated with an increase in the respective density score, age and BMI over the IQR (from Q1 to Q3).

### Table 4. Results of multivariate logistic regression analysis for each density score \( t_1 \) to \( t_4 \) adjusted for age, body mass index (BMI) and menopausal status (n = 292)

<table>
<thead>
<tr>
<th>Model</th>
<th>OR ( a )</th>
<th>95% CI ( a )</th>
<th>IQR</th>
<th>( p )-value</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.748</td>
</tr>
<tr>
<td>( t_1 )</td>
<td>0.68</td>
<td>0.41</td>
<td>1.13</td>
<td>4.31</td>
<td>0.14</td>
</tr>
<tr>
<td>Pre-menopausal</td>
<td>1.94</td>
<td>0.75</td>
<td>5.00</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.09</td>
<td>0.64</td>
<td>1.86</td>
<td>9.00</td>
<td>0.67</td>
</tr>
<tr>
<td>BMI</td>
<td>0.39</td>
<td>0.17</td>
<td>0.92</td>
<td>6.25</td>
<td>0.02</td>
</tr>
<tr>
<td>Model II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.739</td>
</tr>
<tr>
<td>( t_2 )</td>
<td>0.84</td>
<td>0.54</td>
<td>1.30</td>
<td>0.59</td>
<td>0.42</td>
</tr>
<tr>
<td>Pre-menopausal</td>
<td>1.89</td>
<td>0.73</td>
<td>4.88</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.09</td>
<td>0.64</td>
<td>1.86</td>
<td>9.00</td>
<td>0.77</td>
</tr>
<tr>
<td>BMI</td>
<td>0.27</td>
<td>0.13</td>
<td>0.56</td>
<td>6.25</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Model III</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.866</td>
</tr>
<tr>
<td>( t_3 )</td>
<td>0.25</td>
<td>0.16</td>
<td>0.40</td>
<td>0.37</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Pre-menopausal</td>
<td>3.16</td>
<td>1.07</td>
<td>9.36</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.09</td>
<td>0.54</td>
<td>2.22</td>
<td>9.00</td>
<td>0.82</td>
</tr>
<tr>
<td>BMI</td>
<td>0.35</td>
<td>0.17</td>
<td>0.72</td>
<td>6.25</td>
<td>0.002</td>
</tr>
<tr>
<td>Model IV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.788</td>
</tr>
<tr>
<td>( t_4 ) × Menopausal interaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.02</td>
</tr>
<tr>
<td>Pre-menopausal</td>
<td>1.02</td>
<td>0.63</td>
<td>1.66</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>Post-menopausal</td>
<td>0.27</td>
<td>0.10</td>
<td>0.71</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.43</td>
<td>0.71</td>
<td>2.90</td>
<td>9.00</td>
<td>0.33</td>
</tr>
<tr>
<td>BMI</td>
<td>0.30</td>
<td>0.15</td>
<td>0.64</td>
<td>6.25</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

AUC, area under the curve.

\( \text{OR} a \) Odds ratio (OR) and 95% confidence interval (CI) for OR calculated over the interquartile range (IQR) of the respective density score, age and BMI; IQR calculated for all women (n = 292). OR and 95% CI represent the odds of high breast tissue density (≥75%) vs the odds of <75% tissue density associated with an increase in the respective density score, age and BMI over the IQR (from Q1 to Q3).
to identify those women at greatest risk for the development of breast cancer and who would benefit most from risk-reducing interventions. Because risk assessment is not commonly available to women until midlife (i.e. 40 years or older), valuable years are lost for risk reduction interventions to exert their influence. Therefore, the ability to initiate risk assessment using a non-invasive pre-screening technique such as TIBS at a younger age (i.e. twenties to early thirties or even during puberty), when breast tissue is more vulnerable to hormone exposure and carcinogenic insults [10, 11], is very desirable.

Characterization of tissue density by TIBS

In previous work by our group [34–36], PCA was executed on non-mean-centred spectra, and $t_1$ was negative for all tissue density classes, thus representing the average attenuation for the tissue. In the current analysis using mean-centred spectra, information about the average attenuation is removed, thereby augmenting differences between tissue density groups (i.e. negative vs positive scores).

Table 5. Final logistic regression model including density scores $t_3$ and $t_4$, menopausal status, body mass index (BMI) and the interaction between density score $t_4$ and menopausal status

<table>
<thead>
<tr>
<th></th>
<th>OR*</th>
<th>95% CI</th>
<th>IQR</th>
<th>p-value</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_3$</td>
<td>0.17</td>
<td>0.10–0.31</td>
<td>0.37</td>
<td>&lt; 0.001</td>
<td>0.922</td>
</tr>
<tr>
<td>$t_4$</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_4 \times$ Menopausal interaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-menopausal</td>
<td>1.00</td>
<td>0.53–1.91</td>
<td>0.24</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>Post-menopausal</td>
<td>0.09</td>
<td>0.02–0.34</td>
<td>0.24</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>0.35</td>
<td>0.17–0.72</td>
<td>0.25</td>
<td>&lt; 0.001</td>
<td></td>
</tr>
</tbody>
</table>

AUC, area under the curve.

*Odds ratio (OR) and 95% confidence interval (CI) for OR calculated over the interquartile range (IQR) of the respective density score and BMI; IQR calculated for all women ($n = 292$). OR and 95% CI represent the odds of high breast tissue density (> 75%) vs the odds of < 75% tissue density associated with an increase in the respective density score and BMI over the IQR (from Q1 to Q3).

Similar to previous work, the obtained principal component spectra identify some metabolic and structural properties of breast tissue that are important for tissue density classification, including absorption by water, lipids and the haemoglobins (oxy and deoxy), as well as changes in photon path length due to light scattering and boundary losses [34–36]. In the analysis presented here, the sign and the magnitude of the density scores $t_3$, $t_4$ and $t_4$ associated with principal component spectra $p_1$, $p_2$ and $p_4$, are important in differentiating spectra as originating from women with high tissue densities (> 75%) vs those with < 75% dense tissue. Alternatively, as the spectrum of $p_2$ is flat between 62 nm and 875 nm (the primary region of haemoglobin absorption), and the only notable feature is a reduced lipid peak at 925 nm (Figure 1b), the scores $t_4$ associated with principal component $p_2$ were not significantly different between women with < 75% vs women with > 75% dense tissue.

Principal component $p_1$ carries information on the light scattering properties of the interrogated tissue, which varies in part with wavelength [34–36]. Scattering increases the path length travelled by light, resulting in a greater probability of absorption and thus a smaller probability of traversing the tissue. Additionally, it also increases the losses due to diffuse reflection at the tissue surface. Fibroglandular tissue has a larger scattering coefficient than fatty tissue because of increased cellular content [39, 40] and structural support tissues, such as the collagen matrix [53, 54].

In the analysis presented here, the median value for $t_3$ was positive for < 75% dense tissue and negative for women with > 75% tissue densities (Table 2). As depicted in Figure 1a, a smaller (i.e. negative) $t_4$ suggests more attenuation relative to a larger (i.e. positive) density score, as would be expected for dense tissue. This finding was confirmed by other groups who demonstrated a similar relationship between increased tissue density and increased light scattering and hence attenuation [39–47].

Adipose tissue predominantly characterizes lower density tissue [29] by its absorption maximum at 925 nm, while high water content with absorption at 970–975 nm was shown to correlate with higher density tissue [29, 30]. Larger total haemoglobin content and a trend towards lower oxygen saturation (indicated by more deoxy- vs oxyhaemoglobin) are anticipated in...
higher density tissue compared with fatty tissue on account of increased cellular proliferation and metabolism [43] and tissue vascularization [47]. These spectral features are also identifiable in the principal component spectra of \( p_3 \) and \( p_4 \) (Figure 1c and d), which contain information about the water and lipid content of the tissue, as well as the deoxy- and oxyhaemoglobin content of the tissue [34–36].

In \( \geq 75\% \) dense tissue, smaller (i.e. negative) \( t_3 \) and \( t_4 \) (Table 2) indicate greater contributions from water at 970–975 nm (\( p_3 \) and \( p_4 \)) and smaller contributions between 775 and 875 nm (\( p_4 \)) (Figure 1c and d). Underlying contributions from deoxyhaemoglobin between 625 and 725 nm (\( p_3 \)) (Figure 1c) and from oxyhaemoglobin from 750 to 900 nm (\( p_4 \)) are also present (Figure 1d). Conversely, in \(< 75\% \) density tissue, larger \( t_3 \) and \( t_4 \) (Table 2) suggest contributions from lipids at 925 nm (\( p_3 \) and \( p_4 \)) and smaller contributions at 825 nm (\( p_3 \)) (Figure 1c); contributions from deoxyhaemoglobin between 625 and 725 nm (\( p_4 \)) (Figure 1d) are also evident. The combined contribution (OD cm\(^{-1}\)) of both oxy- and deoxyhaemoglobin (i.e. THC) was greater in \( \geq 75\% \) density tissue compared with \(< 75\% \) dense tissue [34–36]. Furthermore, among high density tissue, the relative contribution of deoxyhaemoglobin to oxyhaemoglobin was larger [34–36].

The information captured in each principal component spectrum \( p_3 \) to \( p_4 \) with respect to tissue density is, as anticipated, according to the known anatomical and physiological properties of healthy, potentially at risk, breast tissue [34–36, 38–47].

**Discrimination of high breast tissue density by TIBS**

The observed inverse associations for each density score \( t_3 \), \( t_2 \) and \( t_4 \) in the logistic regression models with high breast tissue density are consistent with the spectral features captured by each principal component spectrum (Tables 3–5). As in other analyses (see Table 2), \( t_5 \) was not associated with high breast tissue density. In the crude models (Table 3), density score \( t_3 \) carried the strongest inverse association and highest level of accuracy, as indicated by the individual OR and AUC, followed by \( t_4 \), then \( t_2 \). The finding that density score \( t_3 \), a measure of overall light attenuation caused by the same structural components contributing to the parenchymal density pattern of the breast (i.e. relative amounts of connective and epithelial tissue and fat), resulted in lower AUC compared with \( t_3 \) suggests that TIBS through \( t_3 \) provides additional information to X-ray derived mammographic density. Specifically, information about the water to lipid ratio and deoxyhaemoglobin content of the tissue captured by \( p_4 \) is more accurate in discriminating between \(< 75\% \) and \( \geq 75\% \) density tissue.

Similar to previous studies, BMI showed a significant inverse association with high mammographic density, and the risk of high density parenchymal patterns was higher among pre-menopausal women (Table 3) [55, 56]. However, in contrast to other studies, age at TIBS was not significantly associated with high mammographic density. BMI and age also showed significant positive correlations with some of the density scores, namely \( t_1 \) (BMI and age) and \( t_2 \) (age only) and \( t_4 \) (age only). There was also some suggestion that discrimination of high mammographic density using density scores differed between pre- and post-menopausal women (Table 2). Based on these findings, all three demographic variables were retained in the multivariate logistic models as potential confounders in the association between the derived scores and mammographic density.

After adjusting for age, BMI and menopausal status (Table 4), the association of \( t_3 \) with high breast tissue density was no longer significant. This reduced significance is most probably explained by the inclusion of BMI in the same model as \( t_3 \) as both variables were strongly and positively correlated (\( r = 0.663 \)). This is expected as a higher BMI means more adipose tissue overall, more fatty replacement in the breast [55], and a larger \( t_1 \) (i.e. more positive) reflects reduced scattering and less attenuation due to smaller amounts of connective (collagen) and epithelial tissue relative to fatty tissue [34–36]. However, the fact that BMI retained significance after adjustment for \( t_3 \) indicates that, of the two measures, BMI is the more relevant independent predictor of high breast tissue density in this study. Consequently, BMI was retained in further logistic models in lieu of \( t_3 \). In terms of developing a predictive model, BMI is an easily and readily obtainable demographic variable.

In contrast to \( t_4 \), density score \( t_4 \) remained significantly and independently associated with high mammographic density among all women even after adjustment for BMI and other demographic covariates. The OR associated with density score \( t_4 \) also retained significance after adjustment for other indicators of mammographic density; however, this association was only significant among post-menopausal women. This latter finding suggests that the additional optical information captured by component spectrum \( p_4 \) relative to \( p_3 \), namely contributions from deoxyhaemoglobin between 625 and 725 nm in \(< 75\% \) dense tissue, and from oxyhaemoglobin and water...
between 775 and 875 nm and 750 and 900 nm in ≥ 75% dense tissue, respectively, are necessary to discriminate high from low density tissue among post-menopausal women, but not among pre-menopausal women. The use of menopause-specific PCA models should be explored; however, because of the limited number of women, and hence spectra, in each group in the current study, menopause-specific PCA was not feasible here.

In the final multivariate model, which included only those variables significantly associated with high mammographic density in both the crude and the adjusted analyses, the inclusion of both $t_3$ and $t_4$ in a single model resulted in a stronger inverse association of each score with mammographic density (post-menopausal women only for $t_4$) (Table 4 vs Table 5). A change in the OR associated with each score is likely, as all derived density scores were originally based on perpendicular mutually exclusive principal components and, even though the scores were independent of one another (i.e. not correlated), each optical spectrum is a linear combination of the mean spectrum of all four components $p_1$ to $p_4$ and all four density scores $t_1$ to $t_4$. However, the fact that, of the four density scores, only $t_3$ and $t_4$ (post-menopausal women only) remained significantly associated with high tissue density indicates that optical information relating to water, lipid and haemoglobin content is sufficient to identify women with high parenchymal density, after adjusting for BMI and menopausal status.

The high AUC associated with the final multivariate logistic model suggests that TIBS is a potential prescreening tool for preventive oncology. However, as all 292 women were considered in obtaining the AUC, the authors acknowledge that the value of 0.922 is likely to be an overestimate of the technique’s accuracy in identifying women with high breast tissue density (≥ 75%). Future work will focus on validating the predictive ability of TIBS using a group of women soon to be enrolled in a recently funded study for which both mammographic density and TIBS measurements will be available.

**Summary and conclusions**

This study demonstrated that in vivo TIBS is a physical method of assessing breast tissue composition and is thus a promising tool for breast cancer risk assessment. Specifically, the derived component scores ($t_{na}$) are analogous to “density scores” assigned to each woman, which identify her breast tissue as either low or high density and hence, by proxy, as being at low or high risk of future breast cancer, respectively.

Unlike X-ray mammography, TIBS does not exploit the atomic composition of the breast tissue, but rather some of its biomolecular markers, and does not carry a dose penalty. Specifically, contributions from water, lipids and haemoglobins were sufficient to discriminate between women with low and very high tissue density. This suggests that TIBS provides complementary information to tissue X-ray attenuation. An added advantage of TIBS over current imaging modalities is the fact that the results are derived from mathematical models; hence highly trained personnel are not required for image interpretation or assessment. The inherent safety (i.e. use of non-ionizing visible and NIR light) and comfort of this method will also permit risk assessment in women outside the recommended age range for X-ray mammography (<40 years of age), in whom risk assessment is currently not obtainable, as well as in those individuals who are non-compliant with standard screening.

Future investigation of menopause-specific PCA models (pre- vs post-menopausal) using a larger sample of women is warranted. Validation of the predictive ability of the model presented here on a new group of women will also be carried out to assess the ability of TIBS to discriminate high tissue density. Work is currently under way to study directly the relationship between TIBS and breast cancer incidence.

**Acknowledgments**

The authors wish to express their gratitude to all participants in this study. The authors would also like to thank Gina Lockwood for her statistical advice during the preparation of this paper. This work was funded by CDMRP (Congress Directed Medical Research Plan), under DAMD 17-00-1-0393, and by CIPI (Canadian Institute of Photonics Innovation).

**References**

Density assessment by TIBS


34. Simick M. Near infrared transillumination spectroscopy of breast tissue for correlation with mammographic density. Master’s thesis, Department of Medical Biophysics, University of Toronto, 2002.


