Disease quantification in dermatology: *in vivo* near-infrared spectroscopy measures correlate strongly with the clinical assessment of psoriasis severity

Tanja Maria Greve
Søren Kamp
Gregor B. E. Jemec
Disease quantification in dermatology: in vivo near-infrared spectroscopy measures correlate strongly with the clinical assessment of psoriasis severity

Tanja Maria Greve,a Søren Kamp,b,c and Gregor B. E. Jemecb
aSpectroscopy and Physical Chemistry, LEO Pharma A/S, Ballerup, Denmark
bRoskilde Hospital, Department of Dermatology, Roskilde, Denmark
cUniversity of Copenhagen, Health Sciences Faculty, Copenhagen, Denmark

Abstract. Accurate documentation of disease severity is a prerequisite for clinical research and the practice of evidence-based medicine. The quantification of skin diseases such as psoriasis currently relies heavily on clinical scores. Although these clinical scoring methods are well established and very useful in quantifying disease severity, they require an extensive clinical experience and carry a risk of subjectivity. We explore the opportunity to use in vivo near-infrared (NIR) spectra as an objective and noninvasive method for local disease severity assessment in psoriasis patients. In whom selected plaques were scored clinically. A partial least squares (PLS) regression model was used to analyze and predict the severity scores on the NIR spectra of psoriatic and uninvolved skin. The correlation between predicted and clinically assigned scores was $R = 0.94$ (RMSE = 0.96), suggesting that in vivo NIR provides accurate clinical quantification of psoriatic plaques. Hence, NIR may be a practical solution to clinical severity assessment of psoriasis, providing a continuous, linear, numerical value of severity.

Key words: near-infrared; dermatology; psoriasis; severity assessment; partial least squares regression.

Paper 12738R received Nov. 15, 2012; revised manuscript received Jan. 30, 2013; accepted for publication Feb. 19, 2013; published online Mar. 18, 2013.

1 Introduction

Access to accurate, reliable, and appropriate methods for documenting disease severity and change in disease severity is a prerequisite for successful clinical research and for the practice of evidence-based medicine. In spite of their visibility and accessibility, the quantification of skin diseases such as psoriasis currently relies heavily on simple clinical scores. Although these clinical scoring methods are well established and very useful in quantifying disease severity, they require an extensive clinical experience and carry a risk of subjectivity. Hence, these scores are subject to inter- and intra-rater variation as well as the underlying variation of the disease.1,2 The development of new objective scores, therefore, is of considerable interest to further clinical development in dermatology.

Near-infrared (NIR) spectroscopy provides an objective, noninvasive, and relatively fast sample measure and can be used for in vivo as well as in vitro measurements. In dermatological research, NIR spectroscopy has predominantly been applied to the discipline of skin characterization with respect to water content.3–9 Other studies have used NIR spectroscopy for in vivo characterization of healthy skin with respect to age and gender,10 classification of diabetes patients,11 blood glucose measurement,12 and body fat measurement.13 Psoriasis is a common human skin disease that is seen worldwide, in all races and both sexes.14–17 It is an inflammatory disease characterized by the dysregulation of various inflammatory pathways, leading to a chronic disease with a characteristic relapsing/remitting course.18–20 The aim of the present exploratory study was to investigate the correlation between local clinical scores and in vivo NIR spectra of the skin in psoriasis patients. A further objective was to explore the possibility of building a predictive partial least squares (PLS) regression model for objective measurement of the severity of a given psoriatic plaque based on NIR spectroscopy.

2 Materials and Methods

2.1 Skin Samples

in vivo NIR spectra were recorded from a total of 31 psoriasis patients: 23 males of ages ranging from 18 to 69 years (average age 41 years) and 8 females of ages ranging from 35 to 65 years (average age 52 years). For each patient, two spectra were obtained from a psoriasis plaque and one spectrum was obtained from clinically uninvolved psoriatic skin that served as a regional control. Uninvolved psoriatic skin was regarded as apparently healthy-looking skin approximately 5 cm away from the periphery of the measured psoriatic lesion. Furthermore, each measured lesion was assigned a clinical severity score (see Sec. 2.3 for more details).

The study was approved by the Danish National Committee on Biomedical Research Ethics, and informed consent was obtained from all subjects.

2.2 Clinical Scores

Local clinical scores were made based an assessment of erythema, induration, and scaling. These components have
been used to score the severity of plaques with standard score systems such as the psoriasis area severity index (PASI) and have proved to be a good guide for clinical assessment of disease severity. However, established clinical scores such as PASI also take the affected body surface area and localization of the plaques into account to assess disease severity.

Each parameter was scored 0 to 4, as follows: 0 = none/not present, 1 = mild, 2 = moderate, 3 = severe, and 4 = very severe. In agreement with the PASI score, the scores were added to get a combined severity score for each assessed lesion, so the total score ranged from 0 to 12.

The scores were determined by a skilled physician with extensive dermatological experience. Since a spectrum corresponding to a clinical score of 0—from healthy-looking skin—was recorded for each patient, we endeavored to obtain as many spectra with high scores as possible in order to span the entire range. Hence, the decision of which psoriasis plaque to use for the measurements was based on the following criteria:

1. The area of the plaque should be large enough for the NIR measurements—i.e., an area of at least 3 mm in diameter.
2. The assigned clinical score should be the highest possible.
3. If the fulfillment of these two criteria left more than one area/plaque, the areas were prioritized by location as follows: arm > thigh > back > abdomen > buttock > ankle > shin > knee > elbow > palm > face.

Table 1 gives an overview of the number of patients measured for each body region. The prioritization of the body areas was based on ease of access and a wish to get the best possible reference (score 0) spectra—i.e., to avoid areas that are often more dry (knee and elbow), have a thickened stratum corneum (palms), or are classified as areas with thin skin (face). The vast majority of the patients were assigned a clinical score between 2 and 8 for the selected psoriatic plaque, with a fairly even distribution in the assignment of the different scores. Table 2 gives an overview of the number of subjects assigned a given clinical score.

<table>
<thead>
<tr>
<th>Score</th>
<th>Number of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2 Number of patients assigned a given score.

2.3 FT-NIR Spectroscopy

A Q-Interline Quant spectrometer (Q-Interline, Tølløse, Denmark) equipped with an InGaAs detector was used to obtain Fourier transform (FT)-NIR spectra in the region 15,000 to 4000 cm$^{-1}$. However, only the 9000 to 4200 cm$^{-1}$ spectral region was used for the analysis of the spectra, as for the skin, no spectral bands are found above 9000 cm$^{-1}$ and the signal-to-noise ratio is fairly reduced below 4200 cm$^{-1}$. The spectrometer was coupled to a fiberoptic probe that has a circular measurement area of approximately 3 mm in diameter. For each sample, 32 scans at 8 cm$^{-1}$ resolution were collected for approximately 30 s and averaged. The measurements are based on diffuse transfectance.

The exact depth of the measurement is not clear and varies in the literature. The measurement depth has been reported to be 150 to 200 μm. Using the 50% measurement depth, which is 120 μm at 6700 cm$^{-1}$ and 30 μm at 5250 cm$^{-1}$, the majority of the absorption observed in the spectra can be assigned to skin layers above this depth. Another report states that the penetration depth of the incident NIR light is up to 2 mm, but due to scattering, the majority of the light returning from the skin and contributing to the spectra will come from the upper layers of the skin, as that fraction of light will be the least attenuated. Thus, although the arguments differ and reports of the exact depth vary, it is agreed that the measurement depth is wavelength dependent—i.e., it varies across the different measurement points of the spectra—and that the majority of the signal observed in the spectra arise from the stratum corneum and epidermis of the skin.
Spectral analysis was performed using OPUS® version 6.5 (Bruker Optics, Ettlingen, Germany). Prior to comparison of the various spectra, they were all normalized using the vector normalization function in the region 9000 to 4000 cm⁻¹.

2.4 Chemometrics

PLS regression was used to make calibration models between the NIR data and the clinical scores for the patients. PLS can be used to disclose the structure of a data set in view of external information. In PLS, the measured data are represented in a matrix $X$ with $n$ rows (number of subjects) and $m$ columns (number of variables) and the external information—here, the clinical scores—are represented in a matrix $Y$. Then PLS considers the ability of the $X$ matrix components to optimally predict the $Y$ variables. Correlated predictor or response variables are combined into PLS components, which explains the variations of the data set and maximizes the orthogonality of the source contributions.

Based on these components, the model can be used to calculate a score by decomposing the $X$ block and building up the $Y$ block. Full cross-validation (leave one out) was used for the model. This provided a predicted score for each sample, which was compared to the reference value—i.e., the assigned clinical score.

The Unscrambler®X ver. 10.1 (CAMO, Oslo, Norway) was used for the PLS regression analysis. Prior to multivariate data analysis, all spectral data sets were pretreated with multiplicative scatter correction (MSC) to avoid differences in baseline offset. The Pearson correlation coefficient ($R$) and the root mean square error (RMSE) were used to evaluate the models.

3 Results and Discussion

3.1 Correlation Analysis

The regression results for the PLS model based on the NIR measurements on two psoriatic sites and one uninvolved skin site for all of the 31 psoriasis patients, along with the target line, are shown in Fig. 1. The model uses seven PLS regression components and yields a very good correlation between the clinically assigned and the predicted scores, with a correlation coefficient of 0.94 ($p < 0.0001$) and RMSE = 0.96. The offset of the regression line is 0.42, and the slope is 0.88. The average difference between the clinical scores and the predicted scores is 0.0, with a standard deviation of ±1.0. That is, for 37% of the recorded NIR spectra, the predicted scores equal the clinical scores; for 52% of the recorded NIR spectra, the predicted scores are 1 unit higher or lower than the clinical scores; and for 11% of the recorded NIR spectra, the predicted scores are 2 units higher or lower than the clinical scores. The 95% confidence interval and the 95% prediction interval are shown in Fig. 2.

For some patients, there is a discrepancy between the two measurements of the psoriatic plaque. This indicates either that the NIR measurement method is inaccurate, or, more likely, that the spectra have not been recorded in the exact same spot and psoriasis severity varies even within small areas of skin, as measured in the current study. Performing the PLS regression model by using an average of the two psoriatic plaque measurements—thus yielding one NIR spectrum of a psoriatic plaque and one NIR spectrum of uninvolved skin for each patient—gives a correlation coefficient of 0.96 and an RMSE of 0.87. An increased correlation coefficient could just be a natural consequence of the decreased number of data points, as would happen with any multilinear regression fitting. However, a random reduction in the number of data points would also increase the RMSE—e.g., performing PLS regressions on either the first or the second of the recorded spectra from each patient shows an increase in the correlation coefficient similar to what happens when using the average spectra, but the RMSEs also increase to 1.09 and 1.12, respectively. Hence, the decreased RMSE combined with the increased correlation coefficient show that the PLS regression model based on the averaged NIR psoriasis spectra really does improve the model considerably. This indicates that the measurement position is extremely important to obtain the best possible model. However, as the exact measurement positions were not noted and the variation with respect to severity scores within each psoriatic plaque was not analyzed, the remainder of the analysis will be based on the separated spectra.
Previously published spectra of healthy skin show some variation with age, gender, and body region.\textsuperscript{3,10,26} However, when dividing the NIR spectra into different groups based on gender or body region, each subset becomes too small for the results to be considered significant. Even though there is an indication that the correlation might be slightly improved when dividing the data into specified subsets, it is found that with a correlation coefficient of 0.89 for all regions and both genders, the possibly improved predictive ability of region- or gender- specific measurements is unlikely to provide a clinically significant improvement. It is further speculated that the added complexity will detract from the utility of the proposed method. Also, when evaluating the correlation between predicted and clinically assigned severity scores, one should take into account that even for experienced raters, the standard deviation for the PASI score (using a scale from 0 to 72) has been shown to be ±1.2 for each individual rater and ±8.1 when comparing different raters.\textsuperscript{3} In light of the significant uncertainty of the clinical severity estimates, scavenging for minute improvements in accuracy through subdivision of test subjects may be less clinically meaningful.

3.2 Spectral Analysis

The various peaks in the NIR spectrum can be assigned to generalized functional groups of the skin constituents: water, proteins, and lipids. However, it is generally not possible to assign the peaks to specific features, structures, or layers of the skin due to the very complex nature of the bands consisting of overtones and combinations of the fundamental vibrations seen in the mid-infrared (MIR) spectrum.

Representative NIR spectra of the uninvolved psoriatic skin and various psoriasis plaques are shown in Fig. 3. The dominant features of the uninvolved psoriatic skin spectrum are the two water peaks at 6900 and 5180 cm\(^{-1}\) assigned to the first overtone of the OH stretching vibration and the combination mode of the OH stretching and HOH bending vibrations.\textsuperscript{5} Weaker overlapping bands are found around 8550, 8410, 8220, 5925, 5780, 5675, 5600, 4880, and 4600 cm\(^{-1}\). The 5925-, 5780-, and 5675 cm\(^{-1}\) bands are assigned to CH groups in skin lipids,\textsuperscript{5,10} whereas the 5600 cm\(^{-1}\) band is assigned to CH groups in skin proteins.\textsuperscript{5} The two bands located around 4880 and 4600 cm\(^{-1}\) are due to the NH-stretching vibration and CONH\(_2\) groups of amide II, and the combinations of amide I and CONH and CONH\(_2\) of amide II\textsuperscript{27} in the skin proteins. These bands are in the NIR spectrum of uninvolved psoriatic skin covered by the broad and intense water band at 5180 cm\(^{-1}\), but they are very distinct in the spectrum of psoriatic skin, with a severity score of 10.

The most obvious differences when comparing the NIR spectrum of uninvolved psoriatic skin to those of the psoriatic plaques are the intensity decrease and the broadening of the 6900 cm\(^{-1}\) water peak and the intensity decrease of the 5180 cm\(^{-1}\) water peak. For both water bands, the intensity is decreased with the increased severity score, although the decrease is more pronounced in the low wavenumber band. This decrease in water content may be explained by a combination of several factors. First, the psoriatic epidermis has a reduced water-holding capacity,\textsuperscript{29} leading to a reduced amount of water in the skin. Second, the increased thickness of the psoriasis stratum corneum in the psoriatic scales\textsuperscript{29} means that the fraction of the incident NIR light that reaches the viable (and more water-containing) epidermis is decreased. Thus, the decreased water contribution in the spectra of severe psoriasis is caused by a combination of a change in the penetration pathway of the incident light and an actual decrease in the water content of psoriatic skin compared to uninvolved skin. Another variation in the spectra is found in the region of the CH stretching vibration bands of the skin lipids. In general, the bands are decreased and shifted toward lower wavenumbers when the severity scores increase. These shifts can be explained by previously reported findings that the content of the various lipid classes is changed in the psoriatic scales,\textsuperscript{30,31} or that in the psoriatic scales, there is a loss of structural integrity of the lipid chains as the lipids lose their crystalline structure and become more fluid.\textsuperscript{32,33}

When comparing spectra having similar clinical scores but different predicted scores, the spectral variations are in accordance with the abovementioned characteristics. Taking the 12 spectra (from six different patients) with the assigned clinical score of 5, 6 of these spectra have an accurate predicted score, 4 have a predicted score of 4 or 6 (i.e., the clinical score ±1), and 2 have a predicted score of 3 (i.e., the clinical score ±2). The spectra are shown in Fig. 4. It is obvious that the 2 spectra with predicted scores of 3 differ from the peer spectra as the intensity of the 5180 cm\(^{-1}\) water peak increases, the 6900 cm\(^{-1}\) water peak is narrowed, and in the 6000 to 5450 cm\(^{-1}\) region, the lipid peaks are shifted 5 to 10 cm\(^{-1}\) toward higher wavenumbers. These differences are consistent with the overall grouping of the spectra shown in Fig. 3. Hence, the observed spectral variations agree with the predicted scores from the model, suggesting some variations in, or deviations from, the clinical scores.

The differences between clinical scores and spectral characteristics are probably due to a combination of the following factors:

1. The NIR measurements are able to detect subtle differences that may be lost in the visually based clinical scoring.

2. Natural biological variations will be displayed as part of the NIR measurements. Hence, the NIR spectra
could display variations not relevant to the disease severity.

3. The clinical scores are the sum of three parameters, so identical summed scores can cover very different cases. For example, a patient with a high score on induration can have the same total score as a patient with a high score on scaling. The NIR spectra of such patients would probably not be identical, and hence, some variations with respect to prediction would arise. Still, even though the interpretation of the summed scores is complex, it reflects the real-life cases better than the single parameters in a subjective assessment. The NIR spectra are also the sum of several factors influencing the skin condition, and therefore, the summed scores are expected to give the best overall correlation to the NIR spectra in the model.

3.3 Opportunities and Limitations

The NIR measurements and the predicted scores from the developed PLS regression model does not provide information about the overall severity of the disease (e.g., the involved body area is not taken into account). Rather, the NIR measurements provide a measurement of local disease severity on selected plaques. Such a method could be a powerful tool in the drug development process, where the effect of a given drug is tested over time at specific plaques. Also, in clinical practice, a rapid and objective measure of local disease severity could be very useful in the evaluation of treatment response and progression of the involved plaques. But the method will need to be supported by other parameters, such as body surface area and position of the plaques, to evaluate the general severity of the disease.

For experienced operators, the intra-operator variation of the NIR measurements has been found to be negligible (healthy skin, not published). However, the inter-operator variation is not clarified. That is, the performance of the NIR measurements and the calculation of the predicted scores are unbiased, as one cannot influence the measurements toward a given score, either intentionally or unintentionally. But the measurement of the spectrum might still be slightly operator dependent. Hence, a comparison of measurements performed by different people could clarify the inter-operator variations and verify the precision of the model. Also, this study has shown that variations in severity can occur within a selected plaque. Therefore, it is very important for the accuracy of the model to have a common understanding and definition of where the measurements are performed.

The use of a single, expert clinical assessor as a reference provides both stability and a source of variation in the estimated accuracy. It is currently the gold standard for clinical assessment of disease severity, but is subject to inter- as well as intra-observer variation, which may explain some of the variation.

Another limitation in the developed model is that only very few of the measured patients were assigned clinical scores at the top of the scale. Hence, the next step toward a further strengthening of the model will be to include more patients with very severe psoriatic plaques to ensure that the model truly covers the entire scale.

4 Conclusions

NIR spectra of psoriasis closely reflect local clinical scores and provide automated, unbiased values. NIR spectra, therefore, may provide a rapid method to quantify several aspects of skin pathology. This trial suggests that NIR spectra correlate strongly with existing clinical psoriasis severity scores. Hence, NIR may be a practical solution to the clinical severity assessment of psoriasis, providing a continuous, linear, numerical value of severity.

While this finding requires confirmation, it is potentially very useful not only in pharmaceutical research, but also in clinical practice, as it may replace the considerable uncertainties inherent in the use of subjective scores.

References