Scleroderma: diagnosis and experimental therapy

Enomoto, D.N.H.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 5

Quantification of cutaneous sclerosis with a skin elasticity meter in patients with generalized scleroderma

Enomoto DNH, Mekkes JR, Bossuyt PMM, Hoekzema R, and Bos JD.

published in:

Chapter 5

SUMMARY

**Background:** The skin score, a subjective assessment of skin elasticity, is widely used in patients with systemic sclerosis. Although this scoring method is regarded as a validated and accepted tool, the interobserver and intra-observer reproducibility is relatively poor.

**Objective:** Our purpose was to investigate whether the recently developed SEM 474 cutometer, which exerts a controlled vacuum force to the skin, can measure skin elasticity more objectively than the skin score.

**Methods:** Skin elasticity was measured in 74 different body areas in patients with systemic sclerosis, and compared with the skin score obtained from the same areas.

**Results:** The cutometer produced quantitative and reproducible data. A large-diameter (8 mm) measuring probe was superior to a small probe. The interobserver intraclass correlation coefficient (ICC) was 0.92; the intraobserver ICC was 0.94. A linear correlation was found with the clinical skin score; the Spearman rank correlation test was 0.69.

**Conclusion:** The correlation with the skin score was reasonable, despite the observation that regional differences in skin elasticity were detected by the cutometer but not by the human observer, who automatically compensates for these factors and integrates them in the skin score. The high interobserver and intraobserver ICC makes the cutometer more suitable for quantifying changes in skin thickness than the subjective skin score.

INTRODUCTION

In patients with systemic sclerosis, the sclerotic skin changes correlate with the overall disease activity and prognosis. The best accepted and most widely used evaluation method for skin thickness is the 'skin score', which is based on subjective examination of the skin by a trained observer. Although the interobserver and intraobserver reliability of the skin score can be low, it is still regarded as a suitable primary outcome variable in clinical trials because it is easy to use and clinically useful alternative methods are lacking.

To reduce subjectivity and increase reliability, many investigators have tried to develop devices that can measure skin sclerosis objectively and quantitatively. Thickening of the skin in patients with scleroderma is caused by an increase in collagen formation in the dermis and possibly or temporarily by an increased amount of edema in the skin. The amount of collagen can be measured in
standard skin biopsy specimens by weighing, histometric methods, or biochemical assays.\textsuperscript{8,15,16} The thickness of the dermis can be measured with high-frequency ultrasound\textsuperscript{6,17-21} and possibly by nuclear magnetic resonance imaging.\textsuperscript{22}

As a result of the accumulation of collagen and fluid, the skin develops its thickened appearance. It becomes impossible to pinch skin into a normal skinfold. This phenomenon, known as 'hidebinding' or 'tethering', is the most impressive change in sclerotic skin and is the basis of the skin score.\textsuperscript{1} To quantify this fixation of the skin, several mechanical instruments have been developed that can exert a controlled physical force to the skin, such as impression by a durometer,\textsuperscript{4,5} linear extension with an elastometer,\textsuperscript{23} or rotation with a twistometer.\textsuperscript{3,24,25} We tested a new skin elasticity meter, the SEM 474 cutometer,\textsuperscript{26-29} which lifts the skin into a measurement chamber and therefore represents the closest possible mechanical imitation of pinching skin into a skinfold. The purpose of this study was to evaluate the interobserver and intraobserver reliability of the elasticity measurements and to assess the correlation between skin score and skin elasticity measurements.

**PATIENTS AND METHODS**

**Patients**

We evaluated the skin of 19 consecutive patients with systemic sclerosis. All patients fulfilled the criteria for the diagnosis of systemic sclerosis as defined by the American Rheumatism Association in 1980\textsuperscript{30} and had the more severe type of scleroderma, described as type II or III scleroderma.\textsuperscript{31-36} Type I is acroscleroderma, type II is acroscleroderma with progression of the sclerosis to proximal areas such as the arms and legs, and type III is diffuse scleroderma, usually starting on the trunk, with rapid progression to other areas including the extremities. In each patient 74 body areas were evaluated by means of the skin score and measured with the skin elasticity meter. In five of these patients, all women between 20 and 66 years of age (average, 45 years) with type III systemic sclerosis of recent onset (< 48 months; average duration, 26 months), the 74 areas were measured with 2 different probe sizes and by two independent observers.
Skin Score
Cutaneous sclerosis was assessed on a 0 to 3 scale by a trained observer who palpated the skin in 74 body areas (Fig 1). The modified Rodnan skin score was used (0 = normal skin thickness, 1 = mild skin thickness, 2 = moderate skin thickness, 3 = severe skin thickness with inability to pinch the skin into a fold). To obtain a more refined score, the number of evaluated body areas was increased to 74 instead of the original 26 or 17 described by Rodnan. This score can be easily translated into the 26 or 17 areas skin score used in previous publications.

Figure 1. Skin score was assessed in these 74 body areas with a 0-3 scale known as the modified Rodnan score (see text).
The SEM 474 Cutometer
We used a specially developed skin elasticity meter, the SEM 474 cutometer (Courage + Khazaka Electronic GmbH, Cologne, Germany). This cutometer meter, equipped with an 8 mm measuring probe, must be connected to a personal computer (Fig. 2). For measurements the skin is drawn into a low-pressure (500 mBar) chamber. The depth of penetration in the low pressure chamber is determined by a noncontact optical measuring system (Fig 3).

Figure 2. The SEM 474 cutometer system consists of a measurement unit containing a vacuum pump, a measurement probe with a fixed opening (2 to 10 mm), and a personal computer.

The SEM cutometer measurement protocols can be varied with the included software, which is flexible and easy to use. The pressure can be adjusted between 50 and 500 mBar, and can be built up immediately or gradually at a controlled rate. The suction time and relaxation time can be changed from 0.1 to 60 seconds, and the number of measurement cycles from 1-99. The software can deliver a graphic and numeric output in four different modes. Both skin elasticity and skin relaxation can be evaluated. Exchangeable measurement probes with different apertures (2, 4, 6, and 8 mm) are available. The measurement protocol consisted of suction at 500 mBar low pressure for 1 second, followed by 1 second of relaxation. This was repeated three times.
Figure 3. Schematic diagram of the measuring head of the SEM 474 cutometer. A spring ensures that the central part of the probe is pushed to the skin at constant pressure. The skin is drawn into a vacuum chamber (V) by a constant 500 mBar vacuum force. The extent at which the skin is drawn into the aperture is measured by the interruption of an infra-red light beam, which is emitted by a light-emitting diode (L), and guided through glass (G) and mirrors towards a photoelectric cell (P).

Fig. 4 shows two repeated measurements. The maximal or final deformation (called $U_f$ according to the nomenclature used by Agache et al.\textsuperscript{37}) depends on the skin thickness. It consists of a linear elastic part ($U_e$) and a nonlinear viscoelastic part ($U_v$). The presence of fluid in the skin reduces the skin’s ability to recover to its initial position after deformation. As a result, the difference between the skin level after one suction cycle and the initial position is higher in edematous skin. This difference is called $R_4$ (result 4) in the graphic output from the cutometer (Fig. 4). On a theoretic basis, the final extension ($U_f$) after the first suction, as well as the maximum height after the second suction cycle ($R_3$) minus $R_4$ could be good indicators of skin sclerosis. Therefore both $U_f$ and $R_3-R_4$ were used in the data analysis.
Quantification of sclerosis with an elasticity meter

Figure 4. Graphic output from the SEM 474 cutometer. The lower line is obtained from a sclerotic skin area on the lower arm with skin score 3; the upper line is from a skin area on the upper arm with skin score 0 (normal skin). $U_e$ represents the elastic trajectory of the final skin extension ($U_f$), and $U_v$ the viscoelastic part. $R_4$ corresponds with the skin's inability to recover to its initial position.

To assess intraobserver reliability, 74 skin areas of a normal control subject were evaluated twice by the same observer (two repeated measurements within 1.5 to 2 hours). To assess the interobserver reliability, 74 areas of 5 patients were evaluated independently by two observers (within 1.5 to 2 hours). To obtain information about the intrinsic measurement errors of the equipment, 10 repeated measurements of the same skin area were done at a 5 minute interval. All measurements were performed with the patients in supine position, in an air conditioned room (temperature 20-21 °C, air humidity 55-60%). It takes about 45 minutes to measure 74 body areas.
Chapter 5

Statistical analysis

Important questions for the overall judgement of the skin elasticity meter are whether the results are reliable (reproducible), valid, and accurate.\textsuperscript{38,39} The intraobserver agreement was analyzed by calculating the intraclass correlation coefficient (ICC) of repeated measurements by the same observer.\textsuperscript{40-42} The interobserver agreement was analyzed by calculating the ICC of measurements by two independent observers. The ICC can be used to express levels of agreement between and within observers as a figure between 0 and 1. An ICC above 0.75 indicates an acceptable level of concordance.\textsuperscript{42} The validity was assessed by comparing cutometer measurements (the mean of two measurements performed by two independent observers) with the skin scores given to the same areas, by means of the Spearman rank correlation test. Student's $t$ tests were used to compare results between normal skin and the three gradations of sclerotic skin. Statistical significance was defined as $p < 0.05$.

RESULTS

All patients were initially measured with the small 2 mm probe which is recommended for studies on epidermal elasticity.\textsuperscript{27} After these data were analyzed, it appeared that the correlation between skin score and skin elasticity measured with the 2 mm probe was below 0.6 (Spearman correlation test), and that the correlation improved with the use of a larger probe. Therefore 5 patients with type III systemic sclerosis were measured again with the largest probe (8 mm). All further data in this report were obtained with the 8 mm probe in these five patients. Patient A had 36 body areas with skin score 0, 15 areas with score 1, 6 with score 2, and 17 with score 3. Patient B had 19 with score 0, 44 with score 1, 10 with score 2, and 1 with score 3. Patient C had 35 with score 0, 26 with score 1, and 13 with score 2. Patient D had 37 with score 0, 17 with score 1, 14 with score 2, and 6 with score 3, and patient E had 22 with score 0, 34 with score 1, 16 with score 2, and 2 with score 3.

The intraobserver agreement, expressed as the ICC calculated from two repeated measurements of 74 body areas by the same observer, was 0.94 for $U_f$, the 95% confidence interval (95% CI) was 0.92 to 1.00. The interobserver agreement, expressed as the ICC calculated from the skin elasticity measured by two independent observers in 74 different body areas of 5 patients with diffuse systemic sclerosis, was 0.92 (95% CI, 0.89 to 1.00) for $U_f$ (Fig. 5) and 0.89 (95% CI, 0.84 to 1.00) for $R_f-R_s$. 

112
Quantification of sclerosis with an elasticity meter

Ten repeated measurements of the same skin area on the thigh of a normal subject (score 0), performed with a 5-minute interval to allow for skin relaxation, showed an average distension of 1.25 mm with a standard deviation of 0.029 mm (95% CI, 1.23 to 1.27 mm).

The cutometer measurements were also compared with the skin score of the same areas, assessed by an experienced observer. In general, the results were comparable. The Spearman rank correlation coefficient was 0.67 for the final deformation \( U_0 \) and 0.63 for \( R_3-R_4 \). Fig. 6 shows the relation between the cutometer measurements and the different skin scores. As expected, low values for skin extensibility were found in skin with severe sclerosis, and high values were found in normal skin. The average skin extension was 1.29 (95% CI, 1.22 to 1.35 mm) in normal control skin, 1.11 mm (95% CI, 1.06 to 1.16 mm) for clinically normal skin (score 0) in the patients with scleroderma, 0.78 mm (95% CI, 0.75 to 0.82 mm) for score 1, 0.62 mm (95% CI, 0.57 to 0.66 mm) for score 2, and 0.41 mm (95% CI, 0.37 to 0.46 mm) for score 3. All differences between these scores were statistically significant (Student's \( t \) test, \( p < 0.001 \)).
Considerable differences in skin elasticity could be found between different anatomic regions, although these areas received the same skin score. For example, the mean distension (mean ± standard error of the mean) measured in skin areas judged as normal skin (score 0) by the human observer, was low at the dorsal side of the foot (0.83 ± 0.069 mm) and the leg (0.72 ± 0.023 mm), but high in the umbilical area (1.07 ± 0.033 mm), the gluteal region (1.35 ± 0.064 mm), and in the neck (1.44 ± 0.070 mm). Skin areas with skin score 1 in the face, showed lower distensions (0.68 ± 0.017 mm) than areas with the same score 1 at the inner arm (1.01 ± 0.037 mm).
DISCUSSION

The measurements with the SEM 474 cutometer were highly reliable. The standard deviation of repeated measurements was low and the intra- and interobserver agreement was good. These results were obtained despite the fact that the total number of patients was small. The reason is that the number of sample areas per patient was high (74), and all patients had type III scleroderma, therefore sufficient areas with skin scores of 0, 1, 2 and 3 were present. The reproducibility of the traditional skin score methods has been studied extensively.\textsuperscript{12-14} These studies showed that the interobserver ICC of the Rodnan skin score (0.43) and the modified Rodnan skin score (0.53) are considerably lower than the interobserver ICC of the cutometer (0.92).\textsuperscript{12} Also the intraobserver ICC of the Rodnan skin score is low (0.55 versus 0.94 with the cutometer).\textsuperscript{12} Some studies suggest that the skin score can be simplified by reducing the number of measurement sites per patients, but this can result in the loss of the ability to detect change in a patient.\textsuperscript{13} For this reason we prefer the combination of an increased number of body areas and a more accurate and reproducible measuring technique.

Despite the low inter- and intraobserver ICC of the traditional skin score, this method is still regarded as the best available tool until now to evaluate systemic sclerosis, because no other methods were available that are easy to use, reproducible, and clinically useful.\textsuperscript{14} We believe that the cutometer method could be an alternative for the skin score, especially in multicenter trials, and we suspect that the value of this equipment has been under-estimated by investigators who have tried it because the instrument in its standard configuration is not suitable for measuring scleroderma.

In its standard configuration, the cutometer is usually equipped with a measurement probe with a small aperture (2 mm) because the device was originally designed to measure elastic properties of the epidermis. It has been used successfully in studies of regional differences in skin thickness and on aging and photoaging of the skin.\textsuperscript{26-29} For the evaluation of scleroderma, however, the large probe (8 mm) gave the best results because with a smaller probe the contribution of the dermal component to the total measurement is relatively small.\textsuperscript{27} Increasing the probe size further would make the probe unsuitable for certain convex areas such as the dorsal surfaces of the fingers and toes. The viscoelastic properties, determined by matrix proteins and the presence and properties of fluid in the skin, are also important in the overall evaluation of skin sclerosis. For this reason, the total skin deformation ($U_f$ in Fig. 4) gives better results than $U_e$ (the linear elastic part of the deformation) or
R₃-R₄ (the skin deformation after subtracting the part of skin deformation which is not immediately reversible because of tissue edema).

Accuracy is difficult to assess, because there are no other objective methods available that can serve as a standard, except perhaps quantification of the amount of collagen in skin biopsy specimens from all measurement sites. This is, of course, impossible for ethical and practical reasons. The second best standard to compare it with is the skin score. However, there are several reasons that the skin score and the cutometer measurements may not be comparable. The skin score is subjective and therefore not free of observer bias. It is also not known whether the skin scores 0, 1, 2, and 3 are equidistant, or whether there is a linear relation between the skin score and the collagen content of the skin. Neither has it been shown that the scores are absolute, which means that score 2 in one patient is the same as score 2 in another patient. In fact, the skin score system could be partly a relational system, in which score 0 is normal skin, score 1 is not quite normal, score 3 is clearly abnormal, hidebound hard skin, and score 2 is somewhere in between. Furthermore, the human observer is capable of incorporating his or her knowledge about the normal regional differences in skin texture between, for instance, palm, inner arm, forehead, abdomen or back, into the overall judgement. Therefore, although the elasticity measured with the cutometer of normal abdominal skin is approximately twice as high as normal palm skin, both receive score 0 from the human observer. Finally, the palpating fingers of the human observer are not only assessing the inability to lift a skin fold, but also the induration of the skin, the resistance to pressure and lateral movement, and the presence of edema.

Our data indicate that the human observer using the subjective skin score is integrating all the information on the normal properties of skin in a certain area into the final score. This flexibility cannot be expected from a machine. The cutometer measures the absolute skin elasticity, which depends on the collagen amount but also on several other factors such as anatomical site, age, sex, genetic factors determining skin texture, habits (work, sun exposure), concomitant or previous skin diseases, or the presence of edema. In theory there can be large differences between patients, which makes the method more suitable for intraindividual comparisons such as follow-up during treatment. In practice, the skin score and cutometer measurements correlated well. Because of the differences caused by the anatomic localization, it is not recommended to sum all individual measurements of the different body areas to one total score, because then the distribution of the sclerosis in the different body areas would influence the total score. We recommend that for follow-up purposes all data are recalculated to a percentage of the baseline measurement for each body region.
The induration of the skin can also be measured by other equipment such as the recently described durometer, a simple and easily used device. The only disadvantage of the durometer appears to be that the method cannot be used over bony surfaces and that induration is not exactly the same as tethering. The SEM 474 cutometer measurement method is the closest imitation of lifting the skin between the fingers to assess hidebinding.

REFERENCES

Chapter 5

Quantification of sclerosis with an elasticity meter

40. Fleis JL. The design and analysis of clinical experiments. New York, Wiley. 1986