Journal of Pain Research

Dovepress
open access to scientific and medical research

Open Access Full Text Article

METHODOLOGY

# Computer-aided surface estimation of pain drawings – intra- and inter-rater reliability

This article was published in the following Dove Press journal:
Journal of Pain Research
13 May 2011
Number of times this article has been viewed



Ann L Persson[1]
Sofia Garametsos[2,*]
Jonna Pedersen[2,*]

[1]Rehabilitation and Research Centre for Torture Victims, Copenhagen K, Denmark; [2]Department of Health Sciences, Division of Physiotherapy, Lund University, Lund, Sweden

*These authors contributed equally to this study

**Abstract:** Pain drawings are often utilized in the documentation of pain conditions. The aim here was to investigate intra- and inter-rater reliability of area measurements performed on pain drawings consecutively, using the computer program Quantify One. Forty-eight patients with chronic nonmalignant pain had shaded in their experienced pain on the front and back views of a pain drawing. The templates were scanned and displayed on a 17-inch computer screen. Two independent examiners systematically encircled the shaded-in areas of the pain drawings with help of a computer mouse, twice each on two separate days, respectively. With this method it is possible to encircle each marked area and to obtain immediate details of its size. The total surface area ($mm^2$) was calculated for each pain drawing measurement. Each examiner measured about 2400 areas, and as a whole, the number of areas measured varied only by 3%. The intra-rater reliability was high with intraclass correlation coefficients 0.992 in Examiner A and 0.998 in Examiner B. The intra-individual absolute differences were small within patients within one examiner as well as between the two examiners. The inter-rater reliability was also high. Still, significant differences in the absolute mean areas (13%) were seen between the two examiners in the second to fourth measurement sessions, indicating that one of the examiners measured systematically less. The measurement error was ≤10%, indicating that use of the program would be advantageous both in clinical practice and in research, but if repeated, preferably with the same examiner. Since pain drawings with this method are digitized, high quality data without loss of information is possible to store in electronic medical records for later analysis, both regarding precise location and size of pain area. We conclude that the computer program Quantify One is a reliable method to calculate the areas of pain drawings.

**Keywords:** area calculation, assessment, chronic pain, digitized pain drawing

## Introduction

Pain drawings are often utilized in the assessment, diagnosis, and documentation of pain conditions both regarding physical and psychological components,[1–11] or evaluating a particular treatment.[12] However, further information such as taking a history and performing a physical examination,[5,13] is usually necessary and must be added.

Pain drawings vary in design, but are all based on the same principle: a human line drawing usually anterior and posterior views. Patients are asked to shade-in the body areas where they experience pain using a pencil,[5,6] or to mark the areas with different colors[7] or symbols for pain.[10,14] To assess exactly how large an area is pain-affected and to see the precise delineations can be difficult.

There are different methods of rating a pain drawing.[15] One of the first assesses the patient according to a point scale where 0–2 points is deemed normal,

Correspondence: Ann L Persson
Rehabilitation and Research Centre for Torture Victims, PO 2107, DK-1014 Copenhagen K, Denmark
Tel +45 33 760 600
Fax +45 33 760 510
Email ap@rct.dk

and above 2 indicates an abnormal or excessive pain manifestation.[10] Margolis and coworkers,[6] split up the drawing into 45 areas and assess which and how many are marked. Another method places a squared overlay on the drawing and assesses the number of shaded-in squares,[16] and thereby the extent of the painful area. These manual methods have demonstrated a high intra- and inter-rater reliability, also in a test–retest situation.[14,17] However, the described methods are complicated and time consuming, and the calculated area is not precise. In spite of this, many pain clinics use qualitative assessments of pain drawings in their clinical work, and if using quantitative assessments these are usually backed by manikin templates.[18]

There are indications that computer programs rate pain drawings with higher accuracy than a subjectively performed manual rating. For this reason, a number of different software programs have been developed that can help with the evaluation of pain drawings.[5,13,19] Quantify One[20] is a computerized measuring instrument, developed to use in the rating of pain drawings to interpret research results and in clinical practice.[21] Pain drawings may well be evaluated by a number of different examiners and at varying intervals; it is thus extremely important that the rating be reliable.

To our knowledge, no reliability tests of this program have been performed. The aim of the present study was to investigate the intra- and inter-rater reliability of measurements performed on pain drawings, using the computer program Quantify One on two different occasions by two examiners. If its reliability was found to be high, and the time consumption was limited, the method would facilitate the registration of pain drawings in the clinic and increase the possibility of documenting and communicating results quantitatively and electronically.

# Material and methods

## Pain drawings

The material consisted of 48 pain drawings filled in by consecutive patients referred to a specialist university pain clinic for evaluation of chronic nonmalignant pain conditions. The patients had participated in an earlier study on pain outcome measures. The pain diagnoses were categorized as: myalgia, psychalgia, neuralgia, lumbago, thoracalgia, cervicalgia, and other. The painful shaded-in areas were assessed by ALP, a physiotherapist experienced in pain management, who also clarified the areas that were not distinctly demarcated. The main purpose was to create a set of pain drawings from clinical data testable for computerized surface area estimation.

## Computer program Quantify One

Quantify One[20] is a computer program, developed for the purpose of performing fast and repeated calculations of area, length, and perimeter on scanned surfaces of lab results by moving a cursor on the computer screen with the help of a computer mouse. The program runs on a standard personal computer with at least an 80 GB hard disc, processor speed Pentium 4, and Microsoft Windows XP Professional.

## Computer settings

The shaded-in A4 size pain drawings were scanned[22] (HP Scanjet 5530, screen resolution 2400 dpi × 4800 dpi) into the Quantify One program. The pain drawing (Figure 1) outline was adjusted to be displayed in whole at maximum screen size (17-inch screen), which made it 86% of the original outline size (14.5 cm). The scanning resolution was set to 100 dpi, and the scale to body height 175 cm equaled 14.5 cm. To facilitate control of the mouse's movements, the speed level was specified by choosing speed level 4 out of 10 possible levels. The mouse hardware was a Microsoft PS/2 HID-compliant mouse, and a plastic blue desk pad size 50 cm × 65 cm was used as a mouse-pad. There was a one to one relation (standard) between the movement of the mouse and the cursor movement on the screen. We chose the mode "continuous area" to mark the shaded areas.

## Examiner training session

All measurements were carried out by two physiotherapy students (Examiner A and Examiner B), both of whom took

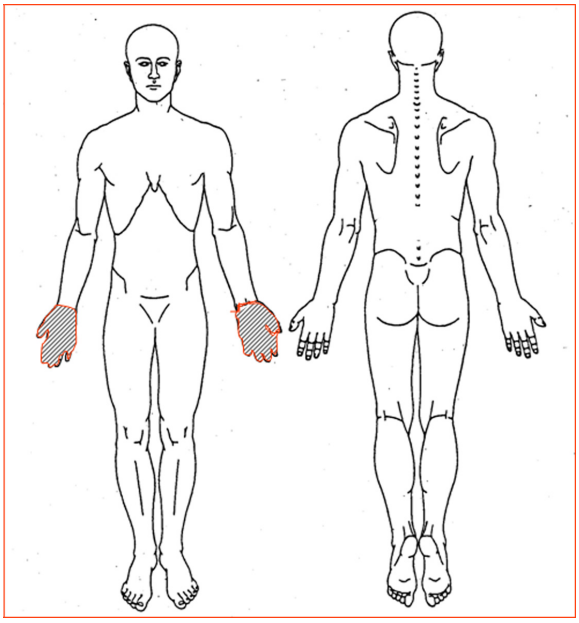

**Figure 1** The pain drawing used for this study.

part in a preliminary training session to practice using the method and to familiarize themselves with the apparatus. On this occasion, five pain drawings were marked, and calculations showed that it took about 1 minute per drawing to perform a measurement.

## Protocol

Before the project measurement sessions were started, it was agreed that the examiners should encircle the shaded areas systematically from the top downwards on the screen, from head to feet and from left to right on the figure. Once a drawing had been marked, the next one was brought up and dealt with in the same way. After marking five drawings, the examiner took a short break before dealing with the next five, and continued in this way until all 48 drawings had been dealt with. On day 1, Examiner A first dealt with all 48 pain drawings alone and unassisted. She brought up the drawings on the screen, one at a time, and encircled the shaded-in areas. With the help of the mouse, each shaded part was then encircled, and its area in $mm^2$ was displayed on the screen and saved in a Microsoft Excel file. Examiner B then took over and performed the same procedure, also alone and unassisted. The total shaded-in area for each pain drawing was used for presenting the results. The software has an option where it is possible to correct the first lines drawn with the mouse in an enlarged correction mode, but that was not used here due to time constraints. Four measurement sessions were then performed on the same day, twice by each examiner alternately, ie, ABAB. This procedure was repeated 1 week later in reverse order, giving in total eight measurement sessions.

## Statistical analysis

The areas of the shaded-in regions on the pain drawings were calculated by the computer program and recorded in $mm^2$. Mean values and 95% confidence intervals (95% CI) were calculated. Student's paired *t*-test was used for group comparisons. The *P*-value was set to 0.01.

In research, the term reliability means 'repeatability' or 'consistency'. We estimated the reliability as the correlation between four observations of the same measure. The intraclass correlation coefficient (ICC) with 95% CI was used to measure test–retest reliability between each examiner's four measurements and inter-rater reliability between the two examiners' measurements. The measurement error was analyzed and then presented as the absolute maximal difference between the mean values for examiners A and B together with ±1 standard deviation (SD) for all four sessions. To compensate for possible learning effects, calculations not including the first set of measurements were also made. The data were analyzed with the help of Statistical Package for the Social Sciences (SPSS, version 15).

# Results

## Pain drawings

The number of areas that had been shaded-in on a single pain drawing varied between 1 and 26 areas. Each of the two examiners performed four measurement sessions on the 48 drawings. In total, 384 drawings were measured, and each examiner measured about 2400 areas. Hence each drawing contained on the average 6.25 separate areas. However, a control inspection of the pain drawings after all of the measurements had been carried out revealed that the two examiners had arrived at a somewhat different number of areas on a few of the drawings, considering two adjacent areas as one, but as a whole, the number of areas measured varied by only 3%.

## Individual pain areas

The individual areas, calculated in $mm^2$ of the 48 patients' pain drawings, are shown in Figures 2A and 2B for the two examiners. Each column represents a measurement of the shaded-in areas on the same pain drawing as numbered on the x-axis and contains the four measurements for Examiner A (Figure 2A) and Examiner B (Figure 2B). From the individual graphs, it can be seen that the intra-individual absolute differences were small within patients for each examiner, respectively. When comparing the two examiners, it can be seen that Examiner A's values are often somewhat lower.

## Intra-rater reliability

In Tables 1A and 1B, the ICC with a 95%CI, the absolute mean maximal differences, and ±1 SD are presented. When all four sessions were included in the analysis, a high intra-rater reliability was found for both examiners (ICC = 0.992 and 0.998), respectively (Table 1A). The absolute mean difference in $mm^2$ for Examiner A was almost double the size of Examiner B.

When the first session was excluded (Table 1B), the intra-rater repeatability increased for Examiner A. The absolute mean difference decreased for both examiners, by 72 $mm^2$ for Examiner A and by 22 $mm^2$ for Examiner B. In view of the fact that the size of the mean pain areas varied between 1062 $mm^2$ and 1236 $mm^2$, it means that the method with one examiner has a measurement error of 6.5%–9.1%.

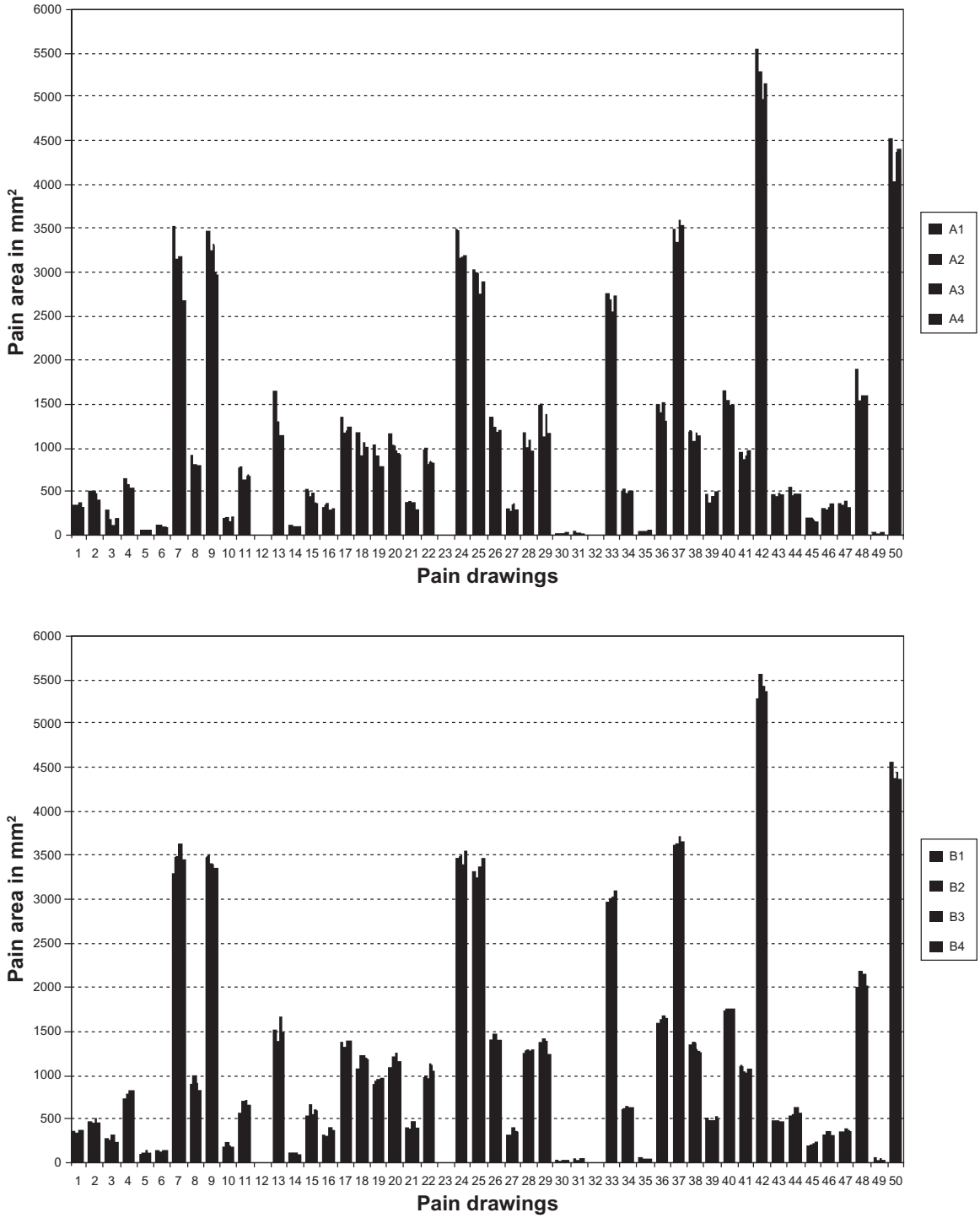


**Figure 2** All individual pain area measurements of the 48 patients' pain drawings: **A**) Examiner A; and **B**) Examiner B. Each cluster represents the same pain drawing as numbered on the x-axis and contains the four measurements.

## Inter-rater reliability

The inter-rater reliability was high in both analyses. The measurement error calculated as the absolute difference between the mean values for the two examiners in all four measurements was 10.1% (117 $mm^2$), and for measurements two, three, and four, it was 13% (151 $mm^2$).

Thus, it appears that a significant difference (13%) in the size of absolute mean areas was seen between the two examiners in the second, third, and fourth measurement sessions, indicating that even if the ICC was high, one examiner measured systematically lower values, as can be seen from the mean area calculations from the eight measuring sessions (Figure 3).

**Table 1** Intra- and inter-rater reliability presented as ICCs with 95% CIs, absolute mean maximal differences mm², and ±SD for "All 4 measurements" and for "Measurements 2, 3 and 4" (first measurement excluded) for Examiners A and B (n = 48)

| | ICC (95%CI) | Absolute mean maximal difference mm² (SD) |
|---|---|---|
| **A. All 4 measurements** | | |
| By Examiner A (intra) | 0.992 (0.985–0.996) | 178 (173) |
| By Examiner B (intra) | 0.998 (0.997–0.999) | 97 (78) |
| By Examiners A and B (inter) | 0.992 (0.923–0.998) | 117 (109) |
| **B. Measurements 2, 3, and 4** | | |
| By Examiner A (intra) | 0.996 (0.993–0.997) | 106 (105) |
| By Examiner B (intra) | 0.998 (0.998–0.999) | 75 (64) |
| By Examiners A and B (inter) | 0.987 (0.865–0.996) | 151 (139) |

**Abbreviations:** CI, confidence interval; ICC, intraclass correlation coefficient; SD, standard deviation.

# Discussion

## Reliability

Using the computer program Quantify One, we found high intra-rater repeatability (ICC) with a mean maximal difference of only 6%–9% of the measurements consecutively performed on 48 pain drawings on two different occasions by two examiners when the measurements from their first session were excluded from the calculations. Our interpretation is that the first session served as a learning session, since this was the case for both examiners, although to a somewhat higher degree for Examiner A. The inter-examiner difference was somewhat higher, 13% between examiners. Across a wide range of pain surface areas measured (from 100 mm² to 5500 mm²), the absolute mean maximal difference was close to 100 mm², ie, a measurement error of 10% as compared with the mean pain areas. The measurement error can therefore be considered highly acceptable.

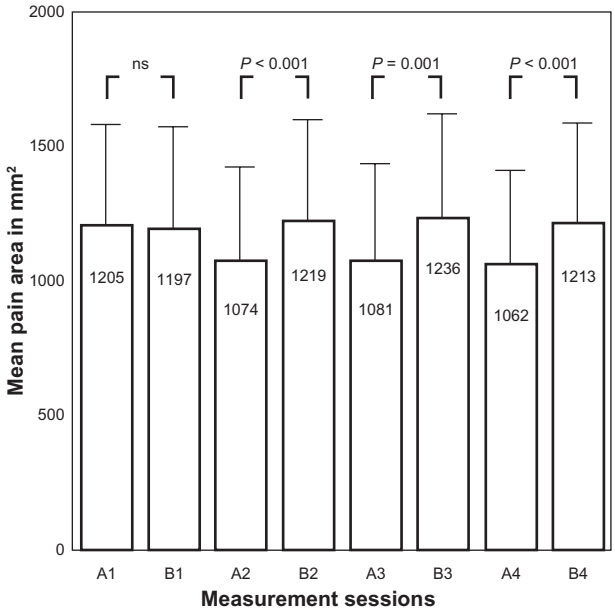


**Figure 3** Bars show the mean measured area in mm² for Examiners A and B on measuring sessions 1, 2, 3, and 4. Error bars show 95%CI of mean; (n = 48; Student's paired *t*-test).

That pain drawings are reliable instruments both for assessment and as test–retest situations has been found earlier,[14,16,17,23] but those studies have been performed with manual and not computer-aided interpretation methods. In a study by Finnerup et al,[24] however, two assessors blinded to patient history used Quantify One to calculate lesion areas shaded on MRI (magnetic resonance imaging) images of patients with spinal cord injuries. They also found a high correlation between assessors (Pearson's correlation coefficient $r = 0.98$, $P < 0.01$), but they did not analyze mean absolute values to support their results.

## Preciseness

A control inspection of the pain drawings after all of the measurements had been made revealed that the examiners had performed a somewhat different number of measurements on each of the four sessions, but the differences were very small, considering that, in all, each examiner measured about 2400 areas. Lacey et al,[23] studied the inter-rater reliability between eight evaluators who rated the number of areas (and thereby what percentage of body area) patients had shaded-in on 50 different pain drawings. The pain drawings were estimated to have an average of 10.2–11.3 shaded areas, and the inter-rater reliability was at least 78%. Since their pain drawings had 48 areas already marked, and all that was needed was to count the filled-in squares, these results cannot be directly compared with ours. They do show, however, that a factor of importance is whether the pain drawing is already squared or not (as in the present study).

Comparisons of the pain areas calculated from the individual pain drawings on the eight sessions revealed that they were not identical. In previous studies, when calculating surface area, a squared overlay has been placed on top of the filled-in pain drawing, and those squares containing shading have been included in the calculation.[6,16,25] Margolis et al,[6] made use of a so-called 'grid system' and divided the outline body into 45 sections. In certain cases, this sectioning

is not detailed enough to obtain a correct figure for the actual area of the body in which the patient experiences pain, especially where small areas are concerned, since even when a square is only partly shaded, it is rated as a whole square. For this reason, several grid systems have been developed, with a varying number of squares that range from 200[25] to 560[16] squares, which specify more exactly the size of the total shaded body pain area. Even so, it is not likely that calculations can ever be exact. Instead it has been shown that evaluators often overestimate the area in rating a pain drawing.[26] The present technique enables the possibility to express pain areas also in percentage of the total body area, which could be valuable when studying the effect of pain treatment in groups of individuals as well as clinically on single cases. The validity of quantifying pain drawings has been demonstrated earlier.[19,27]

Using Quantify One, it is possible to encircle more precisely each area marked by the patient, and to obtain immediate details of its size. It is evident that Quantify One, in comparison with grid systems, should make it possible to obtain far more exact information on how large an area of the body is experienced to be in pain. Wenngren and Stålnacke[21] recently published a pilot study using the first version of Quantify One to assess pain drawings in chronic pain. The clinical implications they suggested were that the method could be useful for systematic comparisons of repeated measures over time or when relating pain areas to other instruments.

On each measuring day, the 48 filled-in pain drawings were dealt with twice by each examiner (once in the morning, once in the afternoon). Since great concentration was required to measure a large number of drawings in succession, it is possible that individual measurements were affected by variations in the examiners' attention. Despite this possibility, there is a very high correlation between all the measurements. The pain drawing outline on the computer screen was adjusted to be 86% of its original size. With a larger format, less precision would perhaps have been demanded of the examiner, but then it would not have been possible to show the whole drawing on the screen. The whole procedure would, on the other hand have been more time consuming, and if a method is to be introduced in clinical practice it should be easy to use and feasible.

Since the pain drawings with this method are digitized, high quality data without loss of information is possible to store in electronic medical records for later analyses, both regarding precise location and size of pain area, whereas the template method has unsatisfactory preciseness with respect to pain location. It is also possible to analyze pain drawings from a multicenter study at the same occasion and in one center.

Since the dominating storage form of medical records today is currently electronic, the Quantify One method is a suitable technique to integrate scanned and quantified pain drawings into modern records, both for clinical and scientific use. As apart from applying manikins or written questions, the present technique offers both precise location for diagnosis and precise area quantification for therapeutic follow-up.

## Conclusion

Our results show that the computer program Quantify One is a reliable method to calculate marked areas of pain drawings. Both intra- and inter-rater reliability were high, and the measurement error was 10% or less, indicating that use of the program would be advantageous both in clinical practice and in research, but if repeated, preferably with the same examiner.

## Disclosure

The authors declare no conflict of interest in relation to this work.

## References


1. Bertilson B, Grunnesjö M, Johansson SE, Strender LE. Pain drawing in the assessment of neurogenic pain and dysfunction in the neck/shoulder region: inter-examiner reliability and concordance with clinical examination. *Pain Med*. 2007;8(2):134–146.
2. Dahl B, Gehrchen PM, Kiaer T, Blyme P, Tøndevold E, Bendix T. Nonorganic pain drawings are associated with low psychological scores on the preoperative SF-36 questionnaire in patients with chronic low back pain. *Eur Spine J*. 2001;10(3):211–214.
3. Gioia F, Gorga D, Nagler W. The value of pain drawings in the care of neck and back pain. *J Back Musculoskel*. 1997;8(3):209–214.
4. Hildebrandt J, Franz CE, Choroba-Mehnen B, Temme M. The use of pain drawings in screening for psychological involvement in complaints of low-back pain. *Spine*. 1988;13(6):681–685.
5. Jamison RN, Fanciullo GJ, Baird JC. Usefulness of pain drawings in identifying real or imagined pain: accuracy of pain professionals, nonprofessionals, and a decision model. *J Pain*. 2004;5(9):476–482.
6. Margolis RB, Tait RC, Krause SJ. A rating system for use with patient pain drawings. *Pain*. 1986;24(1):57–65.
7. Masferrer R, Prendergast V, Hagell P. Colored pain drawings: preliminary observations in a neurosurgical practice. *Eur J Pain*. 2003;7(3):213–217.
8. Ohnmeiss DD, Vanharanta H, Ekholm J. Relationship of pain drawings to invasive tests assessing intervertebral disc pathology. *Eur Spine J*. 1999;8(2):126–131.
9. Parker H, Wood PL, Main CJ. The use of the pain drawing as a screening measure to predict psychological distress in chronic low back pain. *Spine*. 1995;20(2):236–243.
10. Ransford AO, Cairns D, Mooney V. The pain drawing as an aid to the psychologic evaluation of patients with low-back pain. *Spine*. 1976;1(2):127–134.
11. Toomingas A. Characteristics of pain drawings in the neck-shoulder region among the working population. *Int Arch Occup Environ Health*. 1999;72(2):98–106.

12. Grunnesjö M, Bogefeldt J, Blomberg S, Delaney H, Svärdsudd K. The course of pain drawings during a 10-week treatment period in patients with acute and sub-acute low back pain. *BMC Musculoskelet Disord*. 2006;7:65.
13. Sanders NW, Mann NH 3rd. Automated scoring of patient pain drawings using artificial neural networks: efforts toward a low back pain triage application. *Comput Biol Med*. 2000;30(5):287–298.
14. Ohnmeiss DD. Repeatability of pain drawings in a low back pain population. *Spine (Phila Pa 1976)*. 2000;25(8):980–988.
15. Schott GD. The cartography of pain: the evolving contribution of pain maps. *Eur J Pain*. 2010;14(8):784–791.
16. Bryner P. Extent measurement in localised low-back pain: a comparison of four methods. *Pain*. 1994;59(2):281–285.
17. Margolis RB, Chibnall JT, Tait RC. Test-retest reliability of the pain drawing instrument. *Pain*. 1988;33(1):49–51.
18. Van den Hoven LH, Gorter KJ, Picavet HS. Measuring musculoskeletal pain by questionnaires: the manikin versus written questions. *Eur J Pain*. 2010;14(3):335–338.
19. Jamison RN, Fanciullo GJ, Baird JC. Computerized dynamic assessment of pain: comparison of chronic pain patients and healthy controls. *Pain Med*. 2004;5(2):168–177.
20. K:L:O:N:K Quantify Image [homepage on the internet] Available from: http://www.quantifyimage.net/1home.aspx. Accessed February 16, 2011.
21. Wenngren A, Stålnacke BM. Computerized assessment of pain drawing area: a pilot study. *Neuropsychiatr Dis Treat*. 2009;5:451–456.
22. Sanders NW, Mann NH 3rd, Spengler DM. Pain drawing scoring is not improved by inclusion of patient-reported pain sensation. *Spine*. 2006;31(23):2735–2741; discussion 2742–2743.
23. Lacey RJ, Lewis M, Jordan K, Jinks C, Sim J. Interrater reliability of scoring of pain drawings in a self-report health survey. *Spine*. 2005; 30(16):E455–E458.
24. Finnerup NB, Gyldensted C, Nielsen E, Kristensen AD, Bach FW, Jensen TS. MRI in chronic spinal cord injury patients with and without central pain. *Neurology*. 2003;61(11):1569–1575.
25. Gatchel RJ, Mayer TG, Capra P, Diamond P, Barnett J. Quantification of lumbar function. Part 6: The use of psychological measures in guiding physical functional restoration. *Spine*. 1986;11(1):36–42.
26. Ramsay B, Lawrence CM. Measurement of involved surface area in patients with psoriasis. *Br J Dermatol*. 1991;124(6):565–570.
27. Ohlund C, Eek C, Palmbald S, Areskoug B, Nachemson A. Quantified pain drawing in subacute low back pain. Validation in a nonselected out-patient industrial sample. *Spine (Phila Pa 1976)*. 1996;21(9):1021–1030; discussion 1031.