Analyzing acute procedural pain in clinical trials

Elvira V. Lang[1], **Gabriel Tan**[2], **Ido Amihai**[2], and **Mark P. Jensen**[3]

[1]Hypnalgesics, LLC, Brookline, MA

[2]Department of Psychology, National University of Singapore

[3]Department of Rehabilitation Medicine, University of Washington

Abstract

Because acute procedural pain tends to increase with procedure time, assessments of pain management strategies must take that time relationship into account. Statistical time course analyses, however, are complex and require large patient numbers to detect differences. The current study evaluated the abilities of various single and simple composite measures such as averaged pain or individual patient pain slopes to detect treatment effects. Secondary analyses were performed with the data from 3 prospective randomized clinical trials that assessed the effect of a self-hypnotic relaxation intervention on procedural pain, measured every 10–15 min during vascular/renal interventions, breast biopsies, and tumor embolizations. Single point-in-time and maximal pain comparisons were poor in detecting treatment effects. Linear data sets of individual patient slopes yielded the same qualitative results as the more complex repeated measures analyses, allowing use of standard statistical approaches (e.g. Kruskal-Wallis), and promising analyses of smaller subgroups, which otherwise would be underpowered. With non-linear data, a simple averaged score was highly sensitive in detecting differences. Use of these two workable and relatively simple approaches may be a first step towards facilitating the development of data sets that could enable meta-analyses of data from acute pain trials.

Corresponding Author: Elvira V. Lang, MD, FSIR, FSCEH, Hypnalgesics, LLC, 157 Ivy Street, Brookline, MA 02446, Tel: 978 404 9724, Fax: 617 734 9087, drevlang@gmail.com.

1. Introduction

Contemporary healthcare strives to be evidence-based. While one properly designed prospective randomized trial may suffice to establish some confidence in the relative risks and benefits of a specific treatment, the highest level of evidence derives from concurrent results of several such trials [3]. The premise of this view is that the measures used in different trials are comparable and can easily be combined and entered in meta-analyses. With objective single-point outcome measures such as disease-free intervals or survival time, the task is relatively straightforward. However, when outcome measures are multidimensional, subjective, and have uncertain trajectories and time intervals across subjects–such as is the case for measures used in pain clinical trials–assessment methods become more complex [7].

The National Institute of Health initiated the Toolbox Project to provide a set of brief, validated outcome measures that can be used across diverse study designs. To assess pain, the Toolbox includes a 0–10 numeric intensity rating scale and a pain interference item bank [5]. Investigators still need to decide whether to choose single, multiple, averaged, or otherwise aggregated measures to reflect treatment effects [8].

Common approaches are point-in-time comparisons, use of averages [1; 9; 18; 19], and maximal pain measures [14; 15; 17]. Jensen and colleagues showed that in the assessment of chronic pain a single 24 hr recall rating can potentially be as valid (sensitive) for detecting treatment differences as are 9 individual measures combined;, allowing considerable savings in cost and burden of clinical trials [7] (but see also Stone et al [16]). Assessing the effect of interventions on stimulus-evoked or procedural acute pain, however, may not be as straightforward, because the time factor is a more critical element of analysis.

In a clinical trial of patients undergoing invasive vascular and renal procedures, patients' pain perception increased linearly over time under standard care conditions [10]. This phenomenon replicated in two subsequent studies [11; 12], indicating a need for time-sensitive methods of analysis. However, time series analyses require large sample sizes and complex statistical approaches. Moreover, with effective interventions, the appearance of zero-pain assessments can make transformation into normally-distributed data impossible (as occurred in two of the trials cited above). This factor makes statistical approaches even more demanding, exceeding the repertoire of many investigators and preventing inclusion of results in meta-analyses.

The purpose of this study was to evaluate the ability of various analytical approaches to detect treatment effects on acute pain. Examining data from three previously published trials, we were particularly interested in whether a single composite pain rating or a relatively straightforward measure, such as slope derived from a per-subject regression analysis, would be as valid as more complex approaches. Because such comparisons, to our knowledge, have not yet been performed, we did not have specific *a priori* hypotheses regarding which method would be superior. Nevertheless, in the event that one specific data treatment proved to be more valid, this could have significant implications for the design and analyses of acute pain clinical trials.

2 Materials and methods

2.1. Data Sets

We performed secondary analyses using the raw data sets of three original prospective randomized clinical trials that had been performed with IRB approval and HIPPAA compliance [10–12].

The trials compared measures of acute pain intensity and anxiety in patients undergoing invasive medical procedures. Patients were randomized to standard care, empathic attention control, or self-hypnotic relaxation groups. The three trials differed in terms of access to sedation, invasiveness of the procedures, as well as procedure risk and meaning for the patients. In the "Vascular/Renal Trial." 241 patients undergoing procedures to their blood vessels or kidneys had access to IV midazolam and fentanyl through a patient-controlled analgesia model, and had their puncture sites over the groin vessels or over the kidney anesthetized with lidocaine [10]. The "Breast Biopsy Trial" only allowed local anesthetic for biopsies with 8 or 14 gauge devices in a pure outpatient setting with 236 patients [11]. The "Tumor Embolization Trial" was a model of particular invasiveness and enrolled 201 patients with liver cancers or benign uterine fibroids. Patients had access to IV midazolam and fentanyl through a patient-controlled analgesia model, had local anesthetics and received intra-arterial infusion of particles with or without chemotherapy to promote organ ischemia which was expected to set in during the time of treatment with the associated potential discomfort [12].

In the standard care patient condition, personnel were instructed to abstain from suggesting or inducing imagery in the patients, but to otherwise behave naturally. In both the empathy and hypnosis conditions a research assistant displayed a set of standardized behaviors such as adapting to the patient's preferred mode of verbal and nonverbal communication, avoidance of negative suggestions, attentive listening, providing encouragement, but avoiding praise, and supporting the perception of control. In the hypnosis condition the research assistant also read a script that included a hypnotic induction followed by generic prophylactic suggestions for pain management.

Features of the primary trials and analyses used in the original trials are summarized in Table 1.

2.2. Pain Measures

Study participants were asked to rate their comfort level on a scale from 0 (no pain at all) to 10 (worst pain possible) and on a scale from 0 (no anxiety at all) to 10 (terrified) prior to the procedure (baseline); every 15 min in the Vascular/Renal and Embolization Trials, and every 10 min in the Breast Biopsy Trial following the first assessment; and at the end of the procedure. When participants indicated pain spontaneously outside the question intervals they were asked to rate that pain as well. The highest rating for each time interval was entered into analysis for the time interval in which it was voiced. Average baseline pain was 2.1 in the Vascular/Renal Trial, 0.94 in the Breast Biopsy Trial, and 1.6 in the Tumor Embolization Trial. Baseline pain was not entered into the slope nor other calculations.

2.3. Statistical Analyses

To assess how use of the various approaches of data analysis affected ability to detect the differences observed relative to the more complex statistics used in the primary analyses for these trials, we tested the following outcome variables:

1. Original single ratings at each time point for which they were collected;

2. Maximal pain ratings reported over 1–10 time intervals;

3. Average of ratings over a sequence of 2–10 time points, and

4. Slopes, derived from a per-subject regression analysis.

Of note is that with increasing procedure length fewer and fewer patients remained in each group with some differential based on treatment (see Table 1). The single time point ratings include only the patients still in the procedure room at these time-points.

Averaged and maximal pain scores at the later time points thus contained the total number of measures obtained per patient up to this point even though for many patients the procedures may have been completed at earlier time points. For example, in the Breast Biopsy Trial, more than half of the patients had their procedure completed at 70 min and more than four fifths of patients (80%) had left at 80 min (Table 1).

Analyses were performed using group averages of raw and logarithmically-transformed data which were separately subjected to ANOVAs (F-tests) and non-parametric Independent Kruskal-Wallis Tests. Significance was considered at $p < 0.05$ in two-sided tests.

For a test to be recommended for future analysis the following criteria were used:

1. The test should be at least as sensitive in detecting the differences as the more complex statistics used in the primary analyses

2. the test of choice should be the one showing the greatest sensitivity to detect differences between test and treatment groups.

3. Results

3.1. Overview

Tables 2–4 compare the results obtained with use of single time-point ratings, averaged pain ratings, individual patient slopes, and averaged maximal pain ratings encompassing various increasing time intervals for the Vascular/Renal Trial (Table 2), the Breast Biopsy Trial (Table 3), and the Tumor Embolization Trial (Table 4). The Tables contain averages and standard deviations, although it is important to keep in mind that not all of the variables are normally distributed. Logarithmic transformation of data enabled use of a normally-distributed data set in the Vascular/Renal Trial, which did not qualitatively alter the outcomes of the analyses (not shown). Of note is that linearity of data was well preserved in the Standard Care Groups of the 3 trials with relatively similar trajectories of the pain response despite vastly different invasiveness of the procedures (Fig. 1a). However, in the hypnosis groups of the Breast Biopsy and Tumor Embolization trials, after very low pain levels during the early procedural steps (e.g. lidocaine application, imaging with contrast

medium or breast compression, lesion access,) a secondary "hump" of the pain experience appeared in the Breast Biopsy Trial when the biopsy device was actually activated and in the Tumor Embolization Trial when particles with or without chemotherapy were infused intraarterially producing organ ischemia (Fig. 1b). Due to the presence of many zero values in the self-hypnotic relaxation group and a biphasic pain response in the tumor embolization group (Fig. 1), it was not possible, even with logarithmic transformations, to transform the variables from the Breast Biopsy and Tumor Embolization Trials into normally distributed data sets. Thus use of F-tests in the current analysis has to be viewed with the caveat that the prerequisite of ANOVA of a normal distribution was not fulfilled in the Breast Biopsy and Tumor Embolization Trials— identifying the Kruskal-Wallis Test as the more appropriate of approaches, in general. Also, the use of the nonparametric Kruskal-Wallis Test was clearly superior in Vascular/Renal Trial when averages of slopes were used.

3.2. Single time-point ratings

Use of single time-point ratings at any given time point would have missed differences at 10 out of 10 time points in the Vascular/Renal Trial; 7 out of 10 time points in the Breast Biopsy trial, and 6 out of 10 time points in the tumor embolization trial.

3.3. Averaged pain

The composite scores representing averaged pain yielded significant group differences for all time-points in the Breast Biopsy and Tumor Embolization Trials. Averaged pain scores, however, failed to show differences in the Vascular/Renal Trial except for sampling times 120 min when only 10 standard care patients, 3 empathy patients, and 2 hypnosis patients remained e.g. had not yet completed their procedures.

3.4. Individual slopes

Use of individual patient slopes showed significant differences in the Vascular/Renal Trial when 4 data points were used for slope calculations and the nonparametric Kruskal-Wallis Test was used. Conversely, use of individual patient slopes showed differences only for slopes based on the early 3–4 time data points in the Breast Biopsy Trial. No data point combination for the slope measures from the Tumor Embolization Trial yielded a statistically significant difference between-groups..

3.5. Maximal pain

Measures of maximal pain failed to show any significant treatment effects except when measurements extended to the very last time intervals in the Renal/Vascular and Breast Biopsy Trials when few patients remained.

4. Discussion

Measuring pain during a medical procedure at a single predetermined time-point may seem an attractive, effort-conscious, and scientifically rigorous premise, but it proved quite unreliable in identifying significant treatment effects across three studies. Single time point measurements missed differences entirely in the Vascular/Renal Trial. This happened at early time points with many patients still enrolled and at later time points with fewer patients

but greater expected differences, albeit also greater variability. In contradistinction, the original repeated measures analyses yielded highly significant differences among groups in the Renal/Vascular Trial. One explanation could be that potentially painful stimuli cannot be timed exactly in a clinical setting e.g. it may take more or less time to position the patient appropriately or compress the breast just right to show a lesion, navigate a difficult anatomy to get to the target area, and/or get the appropriate confirming images. Such factors would tend to average out in analyses using composite scores encompassing multiple time points. As a matter of fact, had we realized the potential of using averaged ratings in the Tumor Embolization Trial we would have used those as more sensitive measure.

Use of maximum pain scores, which are sometimes used as outcomes in pain trials [1; 2; 13–15; 17], did not yield significant treatment effects unless carried out to include the longest (i.e., toughest) cases in which fewer patients remain and which in itself may introduce bias towards treatment-accelerating interventions. Shortening of a procedure – as occurred in the Renal/Vascular Trial in the self-hypnotic relaxation group– may be an effect of the patient being calmer translating into the surgical team being able to focus better on the work at hand. Thus, unless there is a compelling reason to report maximum pain scores and use them as an outcome in an acute-pain clinical trial, our findings suggest that this use may result in an underestimation of treatment efficacy.

The effects of time have to be considered in analyses of acute pain experienced during medical procedures. Under standard care conditions, acute pain tends to increase over time; a finding evident in all three trials studied here. Assuming such a linear relationship for procedural pain, one would expect the total pain experience to decrease when the steepness of the time trend of increasing pain lessens, when it's shape changes while decreasing the area under the pain curve, and/or the procedure time (i.e., time of experienced pain) is shortened. Based on the results of these analyses, we postulate that the ability of the analytical approach in detecting treatment differences depends on whether or not linearity of the pain/time curve is maintained in the treatment condition.

In the Vascular Renal Trial where linearity of the pain/time curve was maintained in all three treatment conditions, the use of individual patient slopes was highly sensitive. The results concurred with the results obtained in the more complex primary analysis of the trial utilizing a repeated-measures analysis of pain responses [10]. The secondary analysis showed, as did the primary analyses, significant differences between the hypnosis and empathy groups and between the hypnosis and standard care groups, but no significant differences between the empathy and standard care groups. The primary trial analysis provided an opportunity to make some additional statements on the underlying theoretical framework of the pain response by determining that the pain scores increased linearly with procedure time in the standard group but remained flat in the hypnosis group.

Chapman et al. also found individual patient slopes (as used in our secondary analyses) as superior to single measures [4]: The same authors though also added a repeated measures analysis over the group means over time for additional information in support of a general theoretic framework of the pain response over time. Thus they were able to demonstrate that the decrease in pain over the 6 days after discharge from an emergency room indeed

followed a significant negative linear trend a finding they could not have stated using individual patient slopes alone. The ability to statistically assess whether a trend of pain over time is positive, negative, or flat (not significantly different from zero) by repeated measures analysis is instructive, but assessment of treatment effects and group comparisons by repeated measures analyses very quickly requires large numbers of patients (e.g., 70–80 per group based on the power calculations performed for the prior trials). The advantages of the use of individual patient slopes in a data set with linear pain trends include that it yields the same qualitative results as the more complex repeated measures analyses, allows use of standard statistical approaches (e.g., Kruskal-Wallis), and enables analyses of smaller subgroups, which otherwise would be underpowered.

Although use of patients' average individual slopes showed high sensitivity in demonstrating differences in the Vascular/Renal Trial (but for the nonparametric Kruskal-Wallis Test only; not for the ANOVAs), the approach failed partly in the Breast Biopsy Trial, and completely in the Tumor Embolization trial. The hallmark of these two latter trials is the loss of a linear response in the treatment condition. There appeared to be a bi-modal pain response in the Tumor Embolization group around 75–125 min into the procedure when organ ischemia is expected to cause potential discomfort superimposed on the general angiographic experience. In addition presence of many zeros made transformation into a normally distributed data set impossible. In this setting use of averaged pain scores proved highly sensitive in detecting treatment effects and greatly facilitated analysis.

Use of averaged pain data yielded qualitatively the same results as the highly complex logistic regression model used in the primary analysis of the Breast Biopsy Trial showing significant differences between hypnosis and standard care and between empathy and standard care [11]. The logistic regression model allowed information in characterizing the pain response over time in the various conditions (e.g. increasing under standard care conditions, remaining flat with empathic attention, and decreasing with self-hypnosis) that provides interesting insights but is difficult to integrate into multi-institutional data sets and meta-analyses. Ordinal logistic regression describes the likelihood of a pain perception at or above a series of given pain thresholds over time in logit slopes and can identify positive, negative, or flat trends. Pain likelihood above threshold increased over time in all three groups in the primary analyses. Pain rose more slowly with Empathy than Standard care, and pain rose more slowly with Hypnosis than Standard care. There was no evidence that rate of change in pain differed between hypnosis and empathy.

In the primary analysis of the Tumor Embolization Trial none of the regression models could be used and point-in-time comparisons among groups were performed [12]. If the results of the analyses performed in the current study had been available, the authors would have used averaged pain ratings as more encompassing and sensitive approach. This would have allowed a more comprehensive assessment of the entire procedure than just the single time points.

"Procedural pain" in this manuscript refers to interventions that include application of potentially painful stimuli to conscious individuals with or without additional sedative drugs and local anesthetics. This would cover in principle invasive medical procedures such a

heart catheterizations, surgeries performed without general anesthesia, steps preceding induction of anesthesia, dental work, needle placement with subsequent treatment such as chemotherapy or dialysis, any kind of biopsies, and experimental work in the laboratory with volunteers. How far the findings of this study will be generalizable requires further work in particular in the assessment of the postoperative pain experience.

Based on the results of these secondary analyses, it was not possible to identify a single unifying measure that would work reliably for both linear and non-linear relationships of acute pain and time. Reduction to two workable and relatively simple approaches based on these parameters however may be a first step towards facilitating data analyses of clinical trials without the need for more complex statistical approaches. Provision of such sets that would not only facilitate outcome analyses of individual trials but also comparison among different trials and enable "big data" sets. Simply plotting the data means or medians over time may allow a quick guide into which category the data falls such as done in Fig. 1. If a linear relationship is deemed to be unlikely based on visual inspection alone, the use of averaged pain ratings is recommended as measure of choice. Basic linearity can also be easily assessed with a standard regression or trend line approach, and if present the recommended mode of analysis is use of individual patient slopes and nonparametric statistics. More complex time-course analyses can be reserved for situations in which there is a strong empirical or theoretical reason to obtain additional information on the shape of the pain/time interrelationship, trends, thresholds, and/or intercepts. If no such reason exists and/or analyzes of smaller groups are desired, then the easier to analyze and interpret individual patient pain/time slopes are preferable. Even when more complex analyses are used providing these simpler measures will enable the researchers to have their data included in meta-analyses which may otherwise not happen.

Acknowledgments

References

1. Alstergren P, Ernberg M, Nilsson M, Hajati AK, Sessle BJ, Kopp S. Glutamate-induced temporomandibular joint pain in healthy individuals is partially mediated by peripheral NMDA receptors. Journal of orofacial pain. 2010; 24(2):172–180. [PubMed: 20401355]

2. Aubin M, Vezina L, Parent R, Fillion L, Allard P, Bergeron R, Dumont S, Giguere A. Impact of an educational program on pain management in patients with cancer living at home. Oncol Nurs Forum. 2006; 33(6):1183–1188. [PubMed: 17149401]

3. Burns PB, Rohrich RJ, Chung KC. The Levels of Evidence and their role in Evidence-Based Medicine. Plastic and reconstructive surgery. 2011; 128(1):305–310. [PubMed: 21701348]

4. Chapman CR, Fosnocht D, Donaldson GW. Resolution of acute pain following discharge from the emergency department: the acute pain trajectory. J Pain. 2012; 13(3):235–241. [PubMed: 22285610]

5. Cook KF, Dunn W, Griffith JW, Morrison MT, Tanquary J, Sabata D, Victorson D, Carey LM, Macdermid JC, Dudgeon BJ, Gershon RC. Pain assessment using the NIH Toolbox. Neurology. 2013; 80(11 Suppl 3):S49–53. [PubMed: 23479545]

6. Flory N, Lang EV. Distress in the radiology waiting room. Radiology. 2011; 260(1):166–173. [PubMed: 21474702]

7. Jensen MP, Hu X, Potts SL, Gould EM. Single vs composite measures of pain intensity: relative sensitivity for detecting treatment effects. Pain. 2013; 154(4):534–538. [PubMed: 23375512]

8. Jensen, MP.; Karoly, P. Self-report scales and procedures for assessing pain in adults. In: Turk, DC.; Melzack, R., editors. Handbook of pain assessment. 3. New York: Guilford Press; 2011. p. 19-44.

9. Kallmes DF, Comstock BA, Heagerty PJ, Turner JA, Wilson DJ, Diamond TH, Edwards R, Gray LA, Stout L, Owen S, Hollingworth W, Ghdoke B, Annesley-Williams DJ, Ralston SH, Jarvik JG. A randomized trial of vertebroplasty for osteoporotic spinal fractures. N Engl J Med. 2009; 361(6): 569–579. [PubMed: 19657122]

10. Lang EV, Benotsch EG, Fick LJ, Lutgendorf S, Berbaum ML, Berbaum KS, Logan H, Spiegel D. Adjunctive non-pharmacologic analgesia for invasive medical procedures: a randomized trial. Lancet. 2000; 355:1486–1490. [PubMed: 10801169]

11. Lang EV, Berbaum KS, Faintuch S, Hatsiopoulou O, Halsey N, Li X, Berbaum ML, Laser E, Baum J. Adjunctive self-hypnotic relaxation for outpatient medical procedures: A prospective randomized trial with women undergoing large core breast biopsy. Pain. 2006:155–164. [PubMed: 16959427]

12. Lang EV, Berbaum KS, Pauker SG, Faintuch S, Salazar GM, Lutgendorf S, Laser E, Logan H, Spiegel D. Beneficial effects of hypnosis and adverse effects of empathic attention during percutaneous tumor treatment: When being nice does not suffice. J Vasc Interv Radiol. 2008; 19(6):897–905. [PubMed: 18503905]

13. Logan HL, Sheffield D, Lutgendorf S, Lang E. Predictors of pain during invasive medical procedures. J Pain. 2002; 3(3):211–217. [PubMed: 14622775]

14. Pavlin DJ, Pavlin EG, Horvath KD, Amundsen LB, Flum DR, Roesen K. Perioperative rofecoxib plus local anesthetic field block diminishes pain and recovery time after outpatient inguinal hernia repair. Anesth Analg. 2005; 101(1):83–89. table of contents. [PubMed: 15976211]

15. Smith WR, Penberthy LT, Bovbjerg VE, McClish DK, Roberts JD, Dahman B, Aisiku IP, Levenson JL, Roseff SD. Daily assessment of pain in adults with sickle cell disease. Ann Intern Med. 2008; 148(2):94–101. [PubMed: 18195334]

16. Stone AA, Schneider S, Broderick JE, Schwartz JE. Single-day Pain Assessments as Clinical Outcomes: Not So Fast. Clin J Pain. 2013

17. Strigo IA, Duncan GH, Bushnell MC, Boivin M, Wainer I, Rodriguez Rosas ME, Persson J. The effects of racemic ketamine on painful stimulation of skin and viscera in human subjects. Pain. 2005; 113(3):255–264. [PubMed: 15661431]

18. Toms L, McQuay HJ, Derry S, Moore RA. Single dose oral paracetamol (acetaminophen) for postoperative pain in adults. Cochrane Database Syst Rev. 2008; (4):CD004602. [PubMed: 18843665]

19. Wernicke JF, Pritchett YL, D'Souza DN, Waninger A, Tran P, Iyengar S, Raskin J. A randomized controlled trial of duloxetine in diabetic peripheral neuropathic pain. Neurology. 2006; 67(8): 1411–1420. [PubMed: 17060567]

Summary sentence

Standard tests using individual patient slopes for linear data sets and averaged measures for others are highly sensitive in detecting treatment effects.
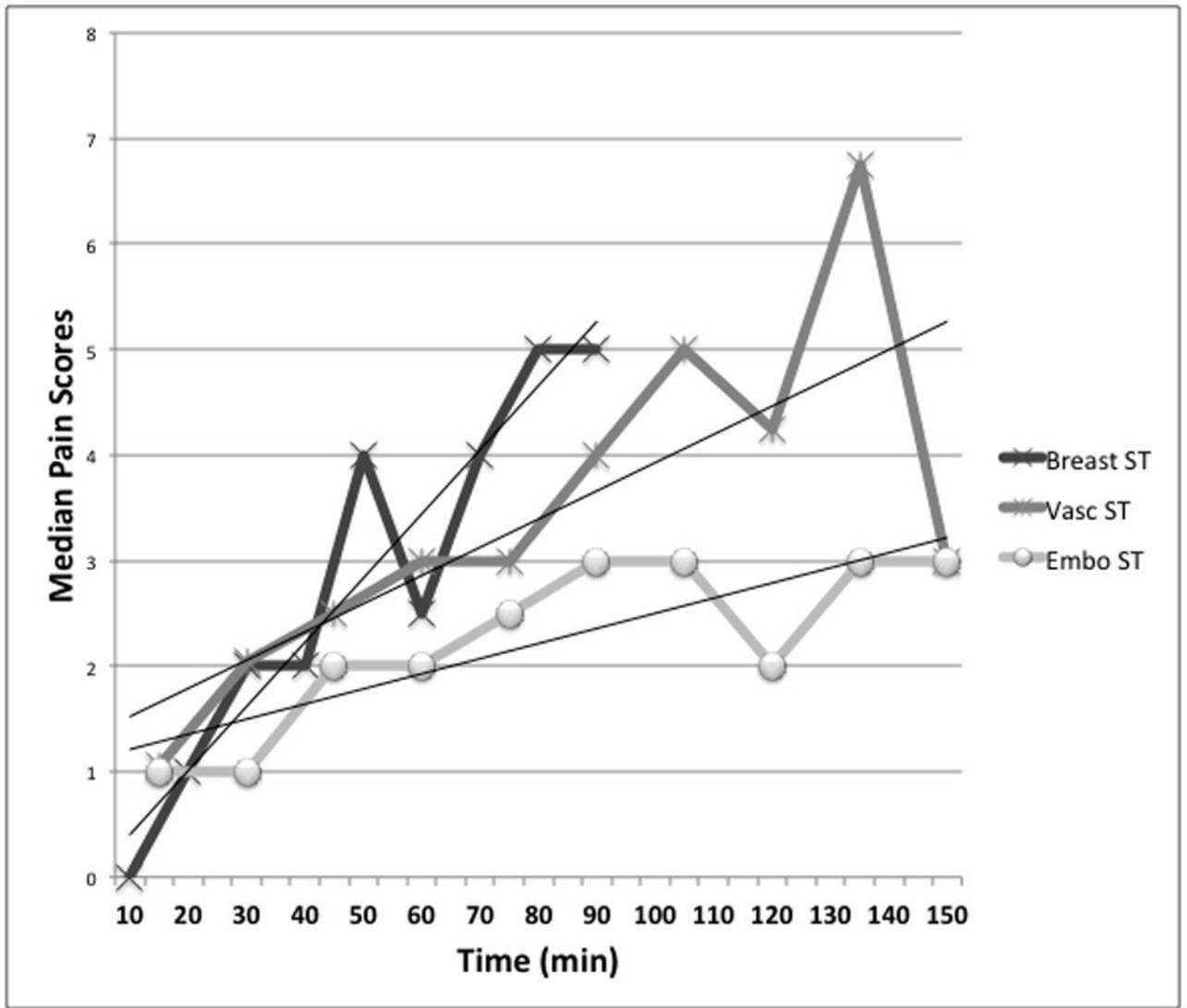
Fig. 1.
Figure 1a and b. Median pain scores under standard care conditions (ST) (Fig. 1a) and the self-hypnotic relaxation conditions (HYP) (Fig. 1b) in the 3 clinical trials. Breast = Breast Biopsy Trial, Vasc = Vascular/Renal Trial, Embo = Tumor Embolization Trial. Trend lines were placed for illustrative purposes only.

Table 1

Features and statistical analyses of the primary trials. All patients received local anesthetics.

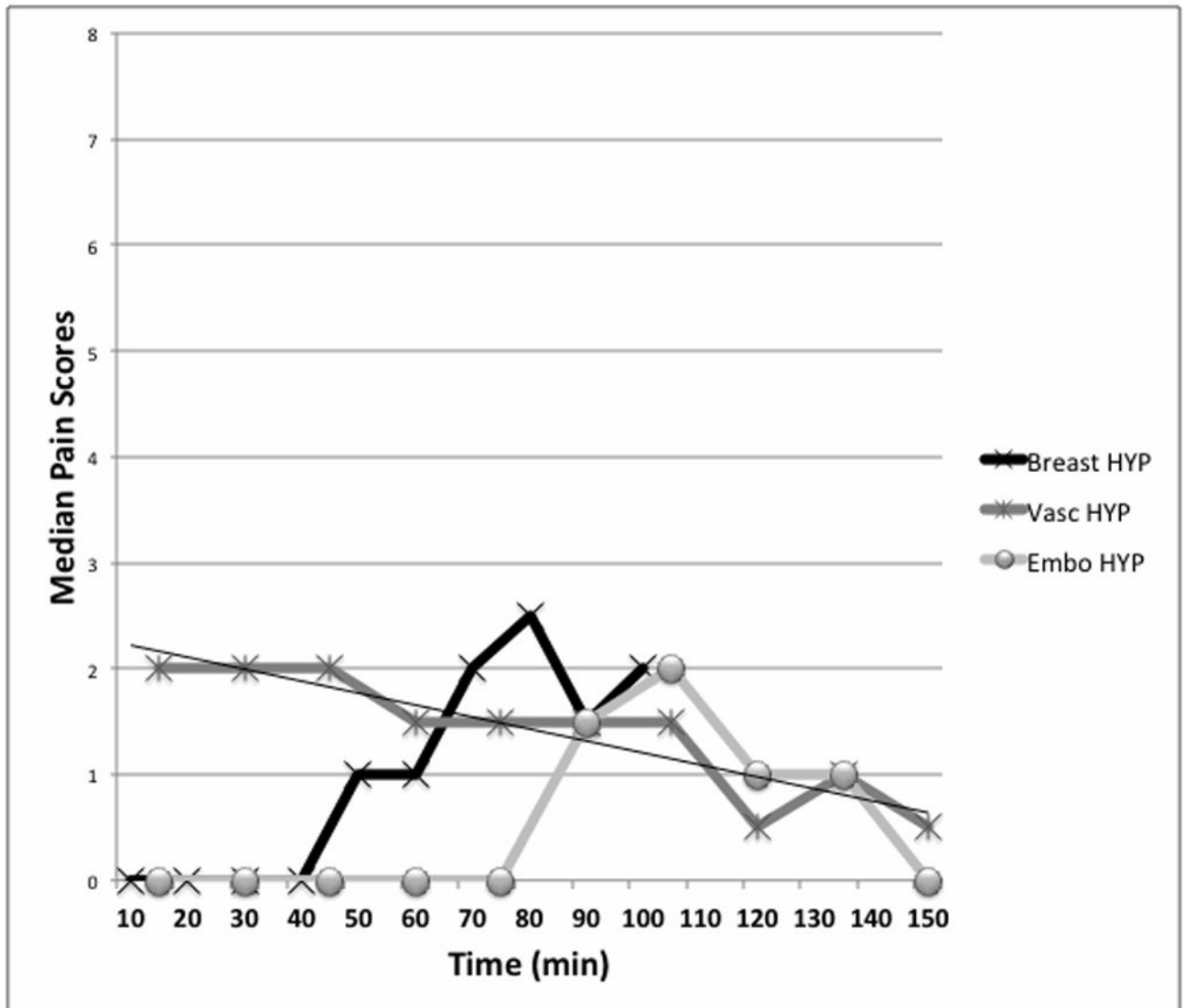| Trial	Procedure specifics	Mean Procedure time	Problems preventing choice of intended analysis	Methods used for published analyses	Results of published primary analyses
Vascular/Renal Trial [10] n=241	IV sedation	78 min S 67 min E 61 min H (p=0.0016)	None; logarithmic transformation allowed normalization of data	Repeated measures analysis with results presented in terms of time trends of pain ratings; linear mixed models estimated by restricted maximum likelihood in BMDP version 5 {Dixon, 1992 #274} to obtain unbiased estimates of intercepts and slopes; comparison of slopes by two-tailed Wald statistics.	Positive linear time trends for S (slope 0·09 in pain score/15 min, p<0·0001) and E (slope 0·04/15 min; p=0·0425); slight negative slope for H not different from zero (slope= –0·03, p=0·234). Trend in H significantly less than trend in S (p<0·0001) and E (p=0·0259).
Breast Biopsy Trial [11] N=236	Higher anxiety levels than other groups [6]	46 min S 43 min E 39 min H (p=0.18)	47% of pain responses were 0 making normalization impossible thus not permitting repeated measures analysis	Ordinal regression with results presented in terms of the slopes of the time course of pain on logit scales; data adaptation to assure meeting of the proportional odds assumption of ordinal regression (sparseness of data in the highest response ratings of 8, 9, and 10 requiring collapsing these into a single category).	Significant increase in pain over time in all three groups (S slope = 0.53, p < 0.001; E slope = 0.37, p < 0.001; H slope = 0.34, p < 0.001). Increase of slope E < slope S (p = 0.024); slope H<S (p = 0.018), no difference in slopes of E and H (p = 0.73).
Tumor Embolization Trial [12] N=201	IV sedation, end organ ischemia in later part of procedure; concurrent IA chemotherapy for 88 patients	110 min S 120 min E 110 min (p=0.77)	High adverse event rate in empathy group resulted in halting the trial with less than planned 350 patients; bimodal distribution of pain ratings (Fig. 1) did not allow normalization of data nor ordinal regression	Mann-Whitney rank-sum tests for individual time points with comparisons between H and E, H and S, and S and E.	Significant difference at 4 time points for less pain with H than E and S.

S – Standard Care Group, E = Empathic attention group, H – self-hypnotic relaxation group

Table 2

Vascular/Renal Trial.

Time Point (min)	Number of Ratings	Average (SD) of Measures			F(p) Value for Group Differences	χ^2(p) Kruskal-Wallis Test
		Standard	Empathy	Hypnosis		
		Average of Original (Single) Ratings at Each Time Point n= number of patients giving ratings at each time point				
15	1	2.19(2.69) n=79	2.08(2.36) n=80	2.34(2.54) n=82	0.200(0.819)	0.818(0.664)
30	1	2.61(2.45) n=78	2.6(2.54) n=79	2.68(2.79) n=79	0.023(0.978)	0.014(0.993)
45	1	3.05(2.72) n=75	2.67(2.8) n=72	2.60(2.82) n=68	0.527(0.591)	1.624(0.444)
60	1	3.25(2.71) n=62	2.53(2.74) n=56	2.39(2.74) n=45	1.476(0.232)	4.218(0.121)
75	1	3.21(2.48) n=57	2.81(2.79) n=37	2.75(2.43) n=24	0.327(0.722)	1.105(0.576)
90	1	4.17(3.20) n=50	3.44(3.38) n=18	3.73(2.86) n=13	0.286(0.753)	0.708(0.702)
105	1	4.16(3.14) n=25	3.27(2.21) n=13	2.33(2.34) n=7	1.123(0.337)	1.475(0.478)
120	1	4.04(2.85) n=16	2.64(1.55) n=8	1.13(1.93) n=5	2.414(0.115)	3.495(0.174)
135	1	5.40(2.87) n=12	4.13(2.10) n=4	2.00(3.37) n=5	2.064(0.162)	4.178(0.124)
150	1	4.50(3.02) n=10	3.17(2.84) n=3	1.00(1.41) n=4	2.173(0.165)	3.951(0.139)
		Average (SD) of Averaged Ratings				
15–30	2	2.19(2.69)	2.08(2.36)	2.34(2.54)	0.171(0.843)	0.261(0.878)
15–45	3	2.4(2.57)	2.34(2.45)	2.51(2.66)	0.222 (0.801)	0.315(0.854)
15–60	4	2.61(2.62)	2.44(2.56)	2.54(2.71)	0.846(0.430)	1.851(0.396)
15–75	5	2.74(2.63)	2.46(2.59)	2.51(2.71)	1.204(0.301)	3.016(0.221)
15–90	6	2.80(2.59)	2.50(2.7)	2.53(2.75)	1.730(0.178)	3.963(0.138)
15–105	7	2.90(2.73)	2.55(2.84)	2.58(2.73)	2.396(0.092)	4.914(0.086)
15–120	8	2.96(2.79)	2.58(2.79)	2.57(2.64)	**3.046(0.048)**	**6.140(0.046)**
15–135	9	2.99(2.84)	2.58(2.72)	2.55(2.48)	**3.946(0.020)**	**7.839(0.02)**

Time Point (min)	Number of Ratings	Average (SD) of Measures			F(p) Value for Group Differences	χ^2(p) Kruskal-Wallis Test
		Standard	Empathy	Hypnosis		
15–150	10	3.06(3.07)	2.60(2.59)	2.55(2.67)	**4.523(0.011)**	**9.038(0.011)**
Average (SD) of Individual Patient Slopes						
15–30	2	0.019(0.181)	0.033(0.183)	0.015(0.162)	0.221(0.802)	2.316(0.314)
15–45	3	0.020(0.116)	0.027(0.102)	0.002(0.110)	1.055(0.350)	5.885(0.053)
15–60	4	0.021(0.095)	0.026(0.083)	0.001(0.098)	1.658(0.193)	**8.788(0.012)**
15–75	5	0.019(0.093)	0.025(0.075)	0.002(0.094)	1.446(0.238)	**7.572(0.023)**
15–90	6	0.020(0.085)	0.026(0.074)	0.002(0.094)	1.624(0.199)	**9.088(0.011)**
15–105	7	0.020(0.085)	0.027(0.073)	0.001(0.093)	1.845(0.160)	**10.901(0.004)**
15–120	8	0.019(0.085)	0.026(0.073)	0.000(0.093)	1.850(0.160)	**11.009(0.004)**
15–135	9	0.020(0.084)	0.026(0.073)	0.000(0.093)	1.893(0.153)	**12.128(0.002)**
15–150	10	0.019(0.084)	0.026(0.073)	0.000(0.093)	1.892(0.153)	**11.469(0.003)**
Average Maximal Pain(SD)						
15	1	2.19(2.69)	2.08(2.36)	2.34(2.54)	-	-
15–30	2	3.34(2.65)	3.25(2.62)	3.4(2.76)	0.033(0.968)	1.143(0.565)
15–45	3	4.04(2.77)	3.9(2.89)	3.64(2.81)	0.013(0.987)	0.267(0.875)
15–60	4	4.53(2.71)	4.19(2.92)	3.78(2.93)	0.076(0.928)	0.269(0.874)
15–75	5	4.70(2.75)	4.36(2.93)	3.90(2.93)	0.199(0.822)	0.917(0.632)
15–90	6	4.92(2.77)	4.45(3.02)	3.97(2.97)	0.403(0.675)	1.905(0.386)
15–105	7	5.04(2.79)	4.53(2.98)	3.97(2.97)	0.698(0.511)	3.125(0.210)
15–120	8	5.08(2.83)	4.53(2.98)	3.97(2.97)	1.079(0.358)	4.594(0.101)
15–135	9	5.1(2.82)	4.56(2.99)	3.97(2.97)	1.539(0.235)	**6.183(0.045)**
15–150	10	5.1(2.82)	4.58(3.00)	3.97(2.97)	2.075(0.145)	**7.936(0.019)**

SD= Standard Deviation. Significant findings are highlighted in bold typeface. Single ratings at individual time points would have failed to show any differences; averaged and averged maximal pain ratings were sensitive only at late time intervals; individual patient slopes showed significant differences when slope calculations encompassed at least 4 data time points and only with use of nonparametric testing by Kruskal Wallis Test.

Table 3

Breast Biopsy Trial.

Time Point (min)	Number of Ratings	Average (SD) of Measures			F(p) Value for Group Differences	χ²(p) Kruskal-Wallis Test
		Standard	Empathy	Hypnosis		
		Average of Original (Single) Ratings at Each Time Point n= number of patients giving ratings at each time point				
10	1	1.16(1.95) n=76	0.81(1.57) n=81	1.16(2.08) =n-78	0.902(0.407)	1.339(0.512)
20	1	1.97(2.43) n=76	1.14(1.87) n=81	1.21(2.10) n=77	3.562(0.030)	**7.409(0.025)**
30	1	2.33(2.62) n=76	1.64(2.20) n=81	1.30(2.13) n=75	3.459(0.033)	**6.281(0.043)**
40	1	2.87(2.88) n=70	2.23(2.37) n=71	1.77(2.51) n-65	2.535(0.082)	5.788(0.055)
50	1	3.68(3.18) n=59	2.60(2.70) n=57	2.03(2.77) n=52	3.425(0.036)	**7.010(0.030)**
60	1	2.97(2.85) n=41	3.16(3.30) n=45	2.81(3.38) n=39	0.089(0.915)	0.436(0.804)
70	1	3.91(2.91) n=32	2.52(2.52) n=31	2.80(3.25) n=26	1.476(0.236)	3.316(0.191)
80	1	4.67(2.94) n=23	3.33(2.53) n=23	3.38(3.81) n=20	0.809(0.454)	2.063(0.356)
90	1	5.50(3.02) n=15	2.86(2.73) n=12	3.00(4.24) n=8	1.530(0.247)	3.302(0.192)
100	1	7.57(2.70) n=8	2.60(1.95) n=7	2.80(3.11) n=4	**7.007(0.008)**	**7.347(0.025)**
		Average (SD) of Averaged Ratings				
10–20	2	1.56(2.23)	0.98(1.73)	1.19(2.09)	**3.346(0.036)**	**6.906(0.032)**
10–30	3	1.80(2.38)	1.18(1.9)	1.22(2.09)	**5.958(0.003)**	**11.756(0.003)**
10–40	4	2.02(2.53)	1.38(2.04)	1.33(2.18)	**8.281(<0.001)**	**16.645(<0.001)**
10–50	5	2.24(2.67)	1.55(2.18)	1.41(2.27)	**11.012(<0.001)**	**21.954(<0.001)**
10–60	6	2.30(2.69)	1.68(2.33)	1.52(2.40)	**9.690(<0.001)**	**22.679(<0.001)**
10–70	7	2.40(2.73)	1.73(2.35)	1.59(2.47)	**10.957(<0.001)**	**25.653(<0.001)**
10–80	8	2.49(2.77)	1.78(2.36)	1.63(2.51)	**12.363(<0.001)**	**28.589(<0.001)**
10–90	9	2.55(2.80)	1.80(2.37)	1.65(2.53)	**13.779(<0.001)**	**31.156(<0.001)**

Time Point (min)	Number of Ratings	Average (SD) of Measures			F(p) Value for Group Differences	χ^2(p) Kruskal-Wallis Test
		Standard	Empathy	Hypnosis		
10–100	10	2.63(2.87)	1.81(2.37)	1.66(2.54)	**16.197(<0.001)**	**34.155(<0.001)**
Average (SD) of Individual Patient Slopes						
10–20	2	0.08(0.20)	0.03(0.18)	0.01(0.15)	**3.604(0.029)**	5.359(0.069)
10–30	3	0.06(0.12)	0.04(0.09)	0.01(0.10)	**4.230(0.016)**	**7.434(0.024)**
10–40	4	0.05(0.09)	0.04(0.07)	0.02(0.10)	1.814(0.165)	**6.554(0.038)**
10–50	5	0.05(0.09)	0.04(0.07)	0.02(0.09)	2.054(0.131)	5.761(0.056)
10–60	6	0.04(0.08)	0.04(0.07)	0.03(0.09)	0.834(0.436)	3.194(0.203)
10–70	7	0.04(0.08)	0.04(0.07)	0.03(0.09)	0.618(0.540)	2.128(0.345)
10–80	8	0.04(0.08)	0.04(0.07)	0.03(0.09)	0.508(0.602)	2.034(0.362)
10–90	9	0.04(0.08)	0.04(0.07)	0.03(0.09)	0.454(0.636)	1.729(0.421)
10–100	10	0.04(0.08)	0.04(0.07)	0.03(0.09)	0.502(0.606)	2.047(0.359)
Average Maximal Pain(SD)						
10	1	1.16(1.95)	0.81(1.57)	1.16(2.08)	-	-
10–20	2	2.06(2.45)	1.35(2.01)	1.53(2.29)	0.680(0.571)	0.956(0.620)
10–30	3	2.60(2.77)	1.73(2.19)	1.72(2.36)	1.210(0.362)	1.412(0.494)
10–40	4	2.98(2.95)	2.07(2.39)	2.01(2.48)	1.672(0.241)	2.094(0.351)
10–50	5	3.56(3.2)	2.45(2.60)	2.27(2.59)	2.047(0.172)	2.750(0.253)
10–60	6	3.68(3.19)	2.88(2.92)	2.44(2.65)	2.284(0.136)	3.428(0.180)
10–70	7	3.73(3.21)	2.89(2.92)	2.58(2.72)	2.706(0.094)	4.379(0.112)
10–80	8	3.75(3.19)	2.90(2.91)	2.63(2.71)	3.246(0.059)	5.261(0.072)
10–90	9	3.80(3.18)	2.93(2.90)	2.65(2.74)	**3.888(0.034)**	**6.378(0.041)**
10–100	10	3.86(3.24)	2.94(2.89)	2.65(2.74)	**4.660(0.018)**	**7.688(0.021)**

SD = Standard Deviation. Significant findings are highlighted in bold typeface. Averaged ratings were the most sensitive in detecting differences. Use of the nonparametric Kruskal Wallis Test should be favored, since the prerequisite of a normal distribution for performance of the ANOVA F test was not given. Use of single time points or individual patient slopes was relatively insentive with this data set. Averaged maximal pain ratings showed significant differences only at very late time intervals when few patients remained in the procedure room.

Tumor Embolization Trial.

Table 4

Time Point (min)	Number of Ratings	Average (SD) of Measures			F(p) Value for Group Differences	χ^2(p) Kruskal-Wallis Test
		Standard	Empathy	Hypnosis		
		Average of Original (Single) Ratings at Each Time Point n= number of patients giving ratings at each time point				
15	1	2.00(2.55) n=70	1.41(2.00) n=65	1.20(1.80) n=66	2.536(0.082)	3.764(0.152)
30	1	2.11(2.5) n=69	1.85(2.37) n=65	1.00(1.88) n=66	4.346(0.014)	10.308(0.006)
45	1	2.21(2.38) n=69	2.2(2.58) n=64	1.21(2.28) n=66	3.741(0.025)	11.473(0.003)
60	1	2.36(2.78) n=69	2.17(2.4) n=62	1.57(2.43) n=61	1.624(0.200)	4.452(0.108)
75	1	2.72(2.72) n=66	2.68(2.65) n=58	1.84(2.19) n=55	2.148(0.120)	3.945(0.139)
90	1	3.17(2.88) n=58	3.43(2.75) n=49	2.02(2.24) n=41	3.431(0.035)	6.603(0.037)
105	1	3.15(3.13) n=45	3.31(3.13) n=37	2.26(2.56) n=34	1.256(0.289)	2.409(0.300)
120	1	3.6(3.81) n=34	3.71(3.07) n=29	2.05(2.36) n=21	1.958(0.149)	3.539(0.170)
135	1	3.5(3.84) n=25	4.96(3.06) n=14	1.94(2.33) n=17	3.748(0.032)	7.593(0.022)
150	1	3.78(3.8) n=13	4.35(3.25) n=10	1.22(1.92) n=10	2.675(0.089)	5.439(0.066)
		Average (SD) of Averaged Ratings				
15–30	2	2.05(2.52)	1.63(2.20)	1.10(1.83)	6.31(0.002)	12.784(0.002)
15–45	3	2.11(2.47)	1.82(2.34)	1.14(1.99)	9.56(<0.001)	23.714(<0.001)
15–60	4	2.16(2.54)	1.9(2.36)	1.24(2.11)	10.87(<0.001)	27.400(<0.001)
15–75	5	2.26(2.58)	2.04(2.42)	1.34(2.13)	12.8(<0.001)	30.132(<0.001)
15–90	6	2.37(2.63)	2.23(2.51)	1.42(2.15)	15.76(<0.001)	35.293(<0.001)
15–105	7	2.44(2.68)	2.33(2.59)	1.50(2.2)	16.64(<0.001)	36.160(<0.001)
15–120	8	2.50(2.76)	2.42(2.65)	1.52(2.21)	18.81(<0.001)	39.382(<0.001)
15–135	9	2.53(2.8)	2.5(2.69)	1.54(2.21)	20.56(<0.001)	42.422(<0.001)

Time Point (min)	Number of Ratings	Average (SD) of Measures			F(p) Value for Group Differences	χ²(p) Kruskal-Wallis Test
		Standard	Empathy	Hypnosis		
15–150	10	2.56(2.82)	2.54(2.72)	1.53(2.20)	**22.45(<0.001)**	**46.037(<0.001)**
Average (SD) of Individual Patient Slopes						
15–30	2	0.003(0.14)	0.029(0.15)	−0.012(0.13)	1.419(0.245)	2.085(0.353)
15–45	3	0.006(0.08)	0.025(0.09)	0.001(0.07)	1.694(0.187)	2.005(0.367)
15–60	4	0.005(0.07)	0.017(0.05)	0.007(0.06)	0.754(0.472)	0.847(0.655)
15–75	5	0.012(0.05)	0.016(0.04)	0.016(0.05)	0.215(0.807)	1.352(0.509)
15–90	6	0.017(0.05)	0.019(0.04)	0.018(0.05)	0.151(0.860)	0.631(0.729)
15–105	7	0.019(0.05)	0.02(0.03)	0.020(0.05)	0.047(0.954)	0.360(0.835)
15–120	8	0.019(0.04)	0.019(0.03)	0.020(0.05)	0.047(0.954)	0.174(0.917)
15–135	9	0.019(0.04)	0.020(0.03)	0.019(0.05)	0.116(0.890)	0.653(0.722)
15–150	10	0.018(0.04)	0.021(0.03)	0.019(0.05)	0.161(0.851)	0.928(0.629)
Average Maximal Pain(SD)						
15	1	2.00(2.55)	1.41(2.00)	1.20(1.80)	-	-
15–30	2	2.69(2.72)	2.39(2.42)	1.62(2.13)	1.627(0.332)	2.571(0.276)
15–45	3	3.15(2.75)	3.12(2.58)	2.02(2.44)	1.906(0.229)	2.756(0.252)
15–60	4	3.72(3.01)	3.56(2.64)	2.45(2.78)	2.174(0.170)	3.231(0.199)
15–75	5	4.08(2.91)	4.11(2.74)	3.05(2.74)	2.051(0.171)	3.500(0.174)
15–90	6	4.47(2.91)	4.56(2.73)	3.33(2.79)	2.112(0.156)	3.556(0.169)
15–105	7	4.69(2.93)	4.78(2.78)	3.70(2.81)	2.151(0.145)	3.792(0.150)
15–120	8	4.76(2.93)	4.95(2.76)	3.81(2.82)	2.310(0.124)	4.085(0.130)
15–135	9	4.85(2.89)	5.06(2.80)	3.81(2.82)	2.603(0.095)	4.672(0.097)
15–150	10	4.86(2.91)	5.08(2.81)	3.81(2.82)	2.980(0.068)	5.473(0.065)

SD = Standard Deviation. Significant findings are highlighted in bold typeface. Averaged ratings were the most sensitive in detecting differences with the caveat that the nonparametric Kruskal Wallis Test should be favored, since the prerequisitie of a normal distribution for performance of the ANOVA F test was not given. Use of individual patient slopes and averaged maximal pain measures were completely insensitive. Use of single time points showed significant differences at 4 time points.