Interventional Pain Medicine 1 (2022) 100125

Contents lists available at ScienceDirect

Interventional Pain Medicine

journal homepage: www.journals.elsevier.com/interventional-pain-medicine

Criteria for determining if a treatment for pain works

Nikolai Bogduk

The University of Newcastle, Newcastle, Australia, PO Box 431, East Maitland, NSW, 2323, Australia

ARTICLE INFO

Keywords:
Outcomes
Categorical data
Success rate
Pain
Function
Distress

ABSTRACT

Claims that a treatment works are hollow unless qualified in terms of: in what respects, by how much, how often, and for how long. Essential co-requisites for improvements in pain are improvements in function, psychological distress, and use of health care. Validated instruments are available for these outcome measures. Mean scores and p-values are not informative. Categorical data are required to reveal by how much a treatment works and how often. In order to provide a full picture, outcomes need to be followed until they plateau. Readers of studies should not rely on what authors claim. Instead, readers should demand comprehensive, transparent data on outcomes so that they can decide for themselves if a treatment works to their satisfaction.

1. Introduction

Using treatments that do not work is an indictment of professional practice. In doing so a physician fools themselves, fools their patients, and fools those who pay for the treatment. Knowing that a treatment works is, therefore, crucial for intellectually honest and professionally responsible practice. However, it can be difficult to know if a treatment actually does work.

Hearsay, assertions, and reputation do not constitute evidence. Evidence comes in the form of empirical data on the outcomes of treatment. In that regard, two words apply, which can be confused as if they are synonymous. Efficacy refers to how well a treatment works under ideal conditions, in which the patients selected for treatment are ones who are most likely to show benefit from the treatment. Effectiveness refers to how well the treatment works in the general population, under real-world conditions, in which patients may have comorbidities or other features that might prevent expression of an optimal response to treatment. When new treatments are introduced, efficacy is typically measured first, because if the treatment does not work well in ideal patients it is unlikely to work well in the general population.

Whether the objective is to determine efficacy or effectiveness, crucial to the appraisal of any treatment is what authors and their readers mean by "it works". Discerning readers would understand that simple definitions cannot apply; qualifications are required in order to provide an unambiguous definition. Those qualifications amount to answers to four subordinate questions:

- in what respects,
- by how much,
- how often, and
- for how long?

These questions serve to distinguish treatments that relieve pain only slightly, in some patients, for only a short time, from treatments that improve function as well as pain, each to a large extent, in a large proportion of patients, and for appreciable periods measured in months if not years.

2. In what respects?

Patients with pain typically present with more than just pain. They may have loss of function (reciprocally referred to as disability). They may suffer psychological distress. They may be relying on continuing health care, such as therapeutic drugs or physical therapy; or they may be desperately seeking a successful treatment.

A belief commonly held, but rarely expressed in the literature, is that these additional features are all a consequence of the pain and, therefore, would all be relieved if the pain is relieved. However, there is no guarantee that this is the case, and especially not if the pain is only partially relieved. If it is the case, it needs to be shown.

A treatment cannot be held as having worked completely if, despite pain having been relieved, the patient remains disabled, distressed, and continuing to rely on other health care, or continuing to seek further care. Either the treatment has not worked, or the treatment is incomplete, and the persisting features need separate, additional attention.

For these reasons, in order to gauge how well a treatment works, outcome measures are required not just for pain alone but also for function, psychological distress, and use of other health care. These are

E-mail address: nbogduk@bigpond.net.au.

https://doi.org/10.1016/j.inpm.2022.100125
Received 12 June 2022; Accepted 16 June 2022

not secondary outcomes, meaning that they are subordinate or of lesser significance; they are critical co-requisites to the relief of pain.

Return to work is another outcome measure but a capricious one. If return to work is achieved, it strongly corroborates success of treatment; but failure to return to work does not necessarily reflect failure of treatment, because socio-economic factors may prevent return to work. For example, employers may be reluctant to hire a worker with a history of back pain, even if that pain has been successfully relieved.

Several validated instruments are available for measuring pain and its change after treatment. These include the visual analog scale (VAS) and the numerical pain rating scale (NRS), or various adaptations of these [1–5].

For other outcome variables, a variety of composite and specific instruments are available. Composite instruments are ones that simultaneously measure several variables, such as general health status, quality of life, physical function, social function, and psychological distress. Specific instruments measure just one variable.

Examples of composite instruments include, but are not limited to: the SF-36 [6] and SF-12 [7], various quality of life measures [8–11], and Patient Reported Outcomes Measurement Information System (PROMIS) [12,13]. These instruments provide general measures of how an individual has been affected by their primary complaint, but to various degrees they provide scales for particular variables such as physical function, social function, and elements of psychological distress. The scales are based on values drawn from the normal population, and can show the extent to which a given patient with pain is less than normal, and the extent to which if they return to normal after treatment.

Specific measures of disability are available for specific primary complaints. These include the Oswestry Disability Index [14], the Roland-Morris Disability Questionnaire [15,16], and others [17] for back pain, and the Neck Disability Index [18–20] for neck pain. Like the composite instruments for measuring health outcomes, these specific measures of disability are calibrated against population norms. The patient's scores are compared with those of the normal population for activites considered relevant or clinically important by those who developed the instruments and, implicitly by those who use the instrument. Others use a different approach.

The patient-specific functional scale (PSFS) ignores what others consider important, and instead evaluates changes in what individual patients consider important for themselves [21–23]. At baseline, this instrument asks patients to nominate four or five activities of daily living that are affected by their pain, and which most dearly they would want restored by successful treatment. Outcomes are then measured by the number of activities restored after treatment, or the degree to which each is restored. This approach tailors outcome assessment to what the affected patient considers important, not what an investigator considers important. It deals with ironic situations such as asking an elderly lady how many stairs she can climb, when she lives in a single-storey house and most dearly would want to be able knit again.

The various composite instruments for measuring health outcomes can provide some index of psychological distress, but more explicit measures can be obtained by instruments such as the Symptom Checklist (SCL-90) [24]. For specific psychological symptoms, such as anxiety and depression, a variety of instruments are available, such the Beck Depression Inventory, the Zung Self-Rating Depression Scale, the Depression, Anxiety and Stress Scale, The Pain anxiety symptom Scale, and the Depression, Anxiety, and Positive Outlook Scale [25]. Other specific scales can measure behaviours such as coping strategies [26] and catastrophizing [27].

For use of other health care there are no sophisticated instruments. Improvements in usage can be rated in a binary fashion, such as “still used” and “no longer used”, or quantified by ordinal scales, such number of doses used or morphine-equivalents used.

It is not the role of this essay to discuss the relative merits of the many instruments available. The critical message is that a good study will report not only on pain but also on other variables that impact on the quality of the patient's life. If studies do not do so, a good reader will nonetheless want to see such data. The relevance of doing so is perhaps better expressed in a negative sense. Can a treatment for pain be considered to have worked if the patient is still disabled, is still distressed, and still requiring treatment? A specific example is: can surgery for back pain be considered to have been successful if 80% of patients treated are still taking opioids? Unless the reader is explicitly told about outcomes other than pain, they are prevented from being able to evaluate fully if the treatment works.

Table 1
The pain scores in a hypothetical study, and their differences in each of 10 patients. A paired *t*-test shows a significant difference, with p = 0.001.

Pain Score		
Before	After	Difference
4	3.9	- 0.1
4	3.8	- 0.2
5	4.9	- 0.1
6	5.9	- 0.1
6	5.9	- 0.1
7	6.9	- 0.1
8	7.6	- 0.4
8	9.9	- 0.1
9	8.8	- 0.2
10	9.9	- 0.1

Table 2
The pain scores from Table 1 together with their mean values. A two-sample *t*-test shows no statistically significant difference, with p = 0.875.

Pain Score		
Before	After	Difference
4	3.9	- 0.1
4	3.8	- 0.2
5	4.9	- 0.1
6	5.9	- 0.1
6	5.9	- 0.1
7	6.9	- 0.1
8	7.6	- 0.4
8	9.9	- 0.1
9	8.8	- 0.2
10	9.9	- 0.1
6.70	6.55	- 0.15
	Mean Values	

3. By how much?

This question applies to each and every outcome measure. The gold standard answers would be: pain completely relieved; function restored to normal; no psychological distress; and no need for continuing health care. Lesser outcomes would be partial improvements in outcome measures; and these can be quantified for any outcome, using numerical scales.

Absolute reductions in pain can be easily determined by subtracting the pain score at follow-up from the pain score before treatment. Relative reductions can calculated as the percentage reduction in pain, which amounts to the absolute reduction divided by the original pain score.

A similar approach can be applied to improvements in physical and social function functions. Absolute changes are calculated as the difference between original scores and scores at follow-up. Additionally or alternatively, improvements can be quantified as the extent to which the patient's scores at follow-up have reverted towards normal levels.

What becomes vexatious or contentious in this regard is how much change indicates that the treatment has worked. Various approaches can be used.

The least informative change is one that is found to “statistically significant”. Statistics looks at the behaviour of collections of numbers,

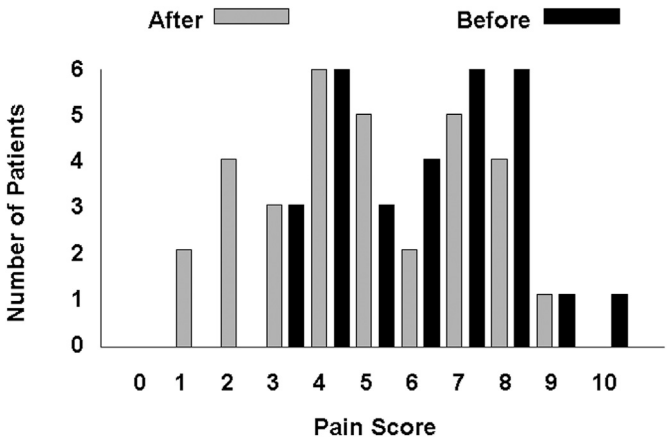

Fig. 1. The pain scores of a group of patients before and after treatment.

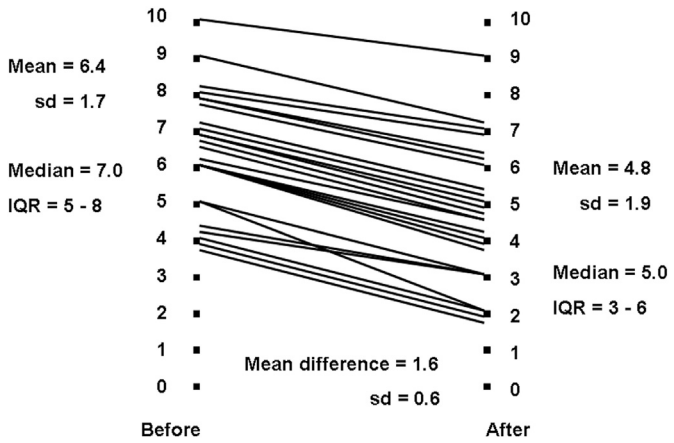

Fig. 3. A graph of the pain scores of 25 individual patients before and after treatment.

and is only remotely related to how well a treatment works. This is particularly so when investigators report the results of paired *t*-test, i.e. comparing the scores, before and after, for each individual patient.

Consider a hypothetical study. The authors claim that the treatment works, with a p-value of $p = 0.001$, which sounds compelling. However, Table 1 reveals the raw data, and shows how meaningless and false the claim is. Despite the statistical significance the changes in pain scores are trivial. The statistical significance arises not because of the magnitude of change but because of the consistency of the changes across all patients.

Table 2 shows the same data analysed in a different manner. The mean scores before and after treatment are barely different numerically, and a two-sample *t*-test shows no statistically significant difference, with $p = 0.875$. The lack of difference on a two-sample *t*-test indicates that, although the scores have improved, the patients as a group have not changed state: their state after treatment is indistinguishable statistically from their state before treatment.

The first message from this example is that statistically significant p-values do not necessarily mean that the treatment works. Reserve your own judgement until you see the raw data and their distribution and magnitudes. The second message is to beware of paired t-tests. They are generous for generating low p-values, because they reward consistency of change more than magnitude of change.

Fig. 1 introduces another consideration of statistics. It shows the pain scores of a group of patients before and after treatment. On inspection there seems to have been an improvement in scores which might be statistically significant. A two-sample *t*-test confirms this impression with a p-value of 0.040.

The *t*-test is classed as a parametric test because the formula for calculating significant difference uses parameters of the data such as the mean values, their standard deviations, and the sample sizes. However, that formula supposes that the data have a reasonably normal distribution.

Fig. 2 shows that this is not the case for the data in Fig. 1. Before treatment the pain scores show two distinct peaks, and after treatment there are three, distinct peaks. This lack of a normal distribution indicates that parametric tests are not appropriate.

Under such conditions non-parametric tests, such as the Mann-Whitney test, become more appropriate for testing for significant difference. Non-parametric tests do not rely on variables such as mean values. In essence, they compare individual scores, one-by-one, from lowest to highest, looking for the frequency of consistent differences.

When a Mann-Whitney test is applied to the data of Fig. 1, it emerges that the p=value is 0.063, which falls short of satisfying the threshold for statistically significant difference. So, although there has been a change in pain scores, there has not been a statistically significant effect.

The take-home message is beware of inappropriate statistical tests. Reserve your judgement until you have seen the distribution of the data, and determine if they are normally distributed or not. Non-parametric tests may be more appropriate to use.

A third issue related to statistics arises when considering the data in Fig. 3. The graphic shows a distinct and universal decrease in pain scores that might be statistically significant. A two-sample *t*-test provides a p-value of 0.002, making the difference clearly significant. Likewise, a Mann-Whitney test provides a p-value of 0.004.

The question that arises is does this statistically significant difference mean that the treatment works? The answer lies in magnitude of change in the pain scores. The group mean scores changed from 6.4 to 4.8, which amounts to 1.6. The mean value of all of the changes per patient also amounts to 1.6. So, although the group changed state statistically, the

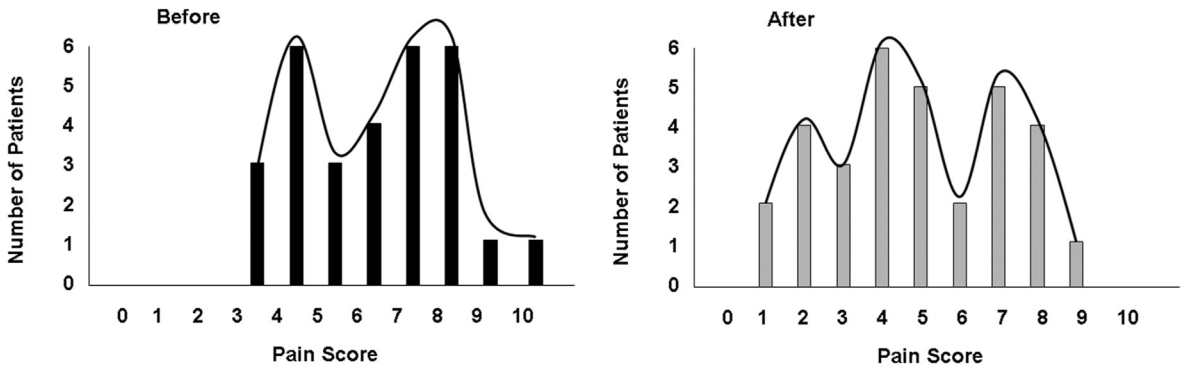

Fig. 2. The data shown in Fig. 1 but separated, and with curves fitted to highlight the distribution of the data. Both before and after treatment the data do not conform to a typical, symmetrical, normal distribution.

Table 3
A list of the minimal clinically important changes for various outcomes measure in pain treatment.

Outcome Variable	MCIC	Reference
Back pain	2.5–4.5	[28]
	2.5	[29]
	2.5–4.3	[30]
Neck pain	4.1	[31]
	4.3	[32]
	3	[33]
	2.6	[34]
Lumbar Radicular Pain	3	[35]
Cervical Radicular Pain	3	[33]
	4.1	[34]
Oswestry Disability	17.8	[36]
	10	[37]
	24	[38]
Roland Morris	5.5–13.8	[30]
Neck Disability Index	10.5	[32]
	10	[33]
	13.4	[39]
	17	[34]
	16	[38]
PROMIS physical function	8	[38]
SF-36 physical function	10–15	[40]

question becomes: is this a clinically significant change?

One approach to answering this question has been to define Minimal Clinically Important Change (MCIC). The MCIC is the magnitude of change in scores that, on average, patients require to experience before they consider that they have noticeably improved. This is perhaps better expressed in the opposite sense: patients consider changes less than the MCIC to have not made a significant difference to their condition.

Many studies have been conducted to establish values for MCIC for various common outcome measures. A selection of these values is summarised in Table 3.

Returning to Fig. 3, it should now be clear that, despite being statistically significant, a mean improvement by 1.6 does not constitute evidence that the treatment has worked. That change is less than the MCIC for back pain, neck pain, and cervical or lumbar radicular pain. It means that patients achieving a change of 1.6 could not tell if they are different after the treatment.

In many modern publications the MCIC has been adopted as a surrogate measure for success. This is misuse. The MCIC constitutes the least improvement that patients consider constitutes a change from their present state. It does not constitute the change that patients consider satisfying or desirable.

When this issue has been studied, patients report that they want complete relief of pain, and that a 50% reduction in pain might be acceptable or tolerable [41]. Such improvements are well in excess of typical values for MCIC for spinal pain.

In that regard, there is no prescribed or universally accepted value that constitutes the threshold for worthwhile, clinically significant relief, and which might be used as the threshold for establishing that a treatment "works". Different consumers may apply different thresholds. In palliative care even a small improvement of pain might be considered merciful. At the other extreme, insurers might prefer treatments that eliminate pain completely (just as the patients would). For most clinical situations, accepted values lie somewhere in between these two extremes.

Consequently, the answer to the question: by how much, is open-ended. Therefore, a responsible author would provide transparent data about all degrees of improvement, i.e. how many patients achieved 50% relief, 60%, 70%, 80%, and 90% relief, and complete relief. Similarly, for other outcome measures they would provide the data for a range of levels

Table 4
Categorical outcome data showing the number, proportions, and cumulative proportions of patients who obtained the listed degrees of improvement.

Improvement (%)	Number	Proportion	Cumulative Proportion
100	4	0.13	0.13
90	0	0.00	0.13
80	4	0.13	0.26
70	5	0.17	0.43
60	1	0.03	0.46
50	4	0.13	0.59
40	0	0.00	0.59
30	1	0.03	0.62
20	2	0.07	0.69
10	2	0.07	0.76
0	5	0.17	0.93
Worse	2	0.07	1.00

of outcome. Given such data, any reader could find how many patients achieved the level of improvement that the reader considers indicates that the treatment has worked, or has worked well enough.

4. How often?

This question is perhaps the most pivotal of the question about outcomes. It asks how often the treatment achieves particular grades of outcome, for each outcome measure. That information constitutes the success rate of the treatment.

Elsewhere it has been explained how group data do not provide success rates [42]. Group data are outcomes described in terms of means and standard deviations, or medians and interquartile ranges. Such data might indicate that, on average, the treatment has some degree of effect, but they do not reveal if all patients benefit to the same degree, or if some patients benefit but others do not. Those features are provided only by categorical data [42]. Accordingly, readers should look for, and demand categorical data in order to determine if a treatment works.

Categorical data are easily displayed – and disclosed – in tables such as Table 4. Such a table shows various degrees of improvement and the numbers and proportions who achieved those degrees of improvement. The cumulative proportions show the proportion of patients who achieved at least the improvement given the corresponding row.

Once provided with such categorical data, any reader can perform their own sensitivity analysis. They can select the degree of improvement in which they are interested, and then read the success rate for achieving that improvement or at least that improvement. They can also see if success rates are impressive for less stringent definitions of success. For example, from the data in Fig. 4 the success rate for achieving complete improvement is only 13%, but it rises to 43% for achieving at least 70% improvement, and 59% for at least 50% improvement.

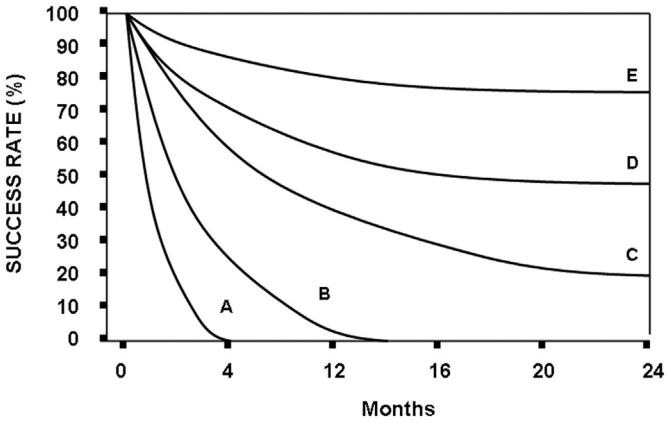

Fig. 4. A set of survival curves showing the changes over time of the success rates of five hypothetical treatments.

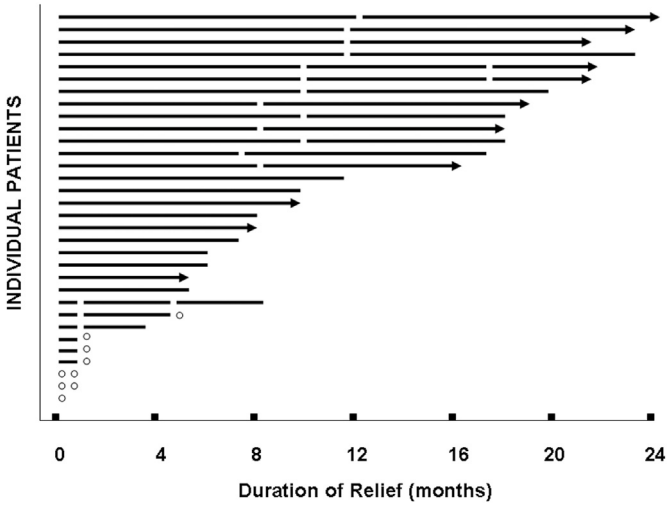

Fig. 5. A bar graph of the response to treatment of 32 individual patients. Each line is the history of a given patient. A circle indicates no response to treatment. A solid bar indicates a successful treatment, and its length indicates the duration of successful outcome. An arrow at the end of a bar indicates successful outcome persisting. A second bar or circle indicates the response to repeat treatment.

Such transparency allows any reader to apply their own criterion for success, and read the success rate for that criterion. It provides readers with the freedom to choose any criterion for success, and not rely on what the authors selectively choose to reveal or to emphasise.

5. How long?

The reason for this question is that some treatments may have a short duration of effect, while other treatments have a long duration of effect; and the effect of a treatment may attenuate (decay) over time. Consequently, readers should expect and demand full information about the duration of effect of a treatment, so that they can decide how useful the treatment may be for their patients.

There is no single figure that defines how long patients should be followed in studies of treatments for pain. The required duration of follow-up will differ according to the nature and objective of treatment. However, a useful guideline is developed in Fig. 4.

This Figure shows the evolution over time of the success rates of five, different, hypothetical treatments. All of the treatments are initially very successful, but subsequently they differ in their evolution. The results of treatment A attenuate rapidly, such that the success rate falls to zero by 4 months. The results of treatment B persist longer but also attenuate, until they become zero at 14 months. Treatments C, D, and E each attenuate but successful outcomes persist in 20%, 50%, and 80% of patients at 24 months, respectively.

If the outcomes of each of these treatments were assessed at, say, two weeks, all the treatments would all appear to be very effective. At two months, it would become evident that the success rate of treatment A had plummeted to 20%, and that of treatment B had dropped to 50%; but the success rates of treatments C, D, and E remain high. At 12 months, the success rate of treatment B will have all but evaporated; the success rates of treatments C, D, and E will have fallen, although to different degrees.

If outcomes are assessed at prescribed or arbitrary periods, the full picture of effectiveness will not be disclosed. At particular times, the success rate may still be attenuating. If the follow-up is too early, treatments that rapidly attenuate will not be identified. However, a guideline can be applied that provides a reasonable assessment of the evolution of success.

Follow-up should be conducted for long enough to see if the outcomes plateau, or approach an asymptote. Doing so allows the calculation of the half-life of the success rate, and the long-term success rate. Provided with such data a reader can judge if a particular treatment is worthwhile for their patients.

Table 5
The number of patients who achieved various degrees of relief of pain at various times after receiving a novel treatment.

Relief (%)	Follow-up (months)							
	1	2	3	6	9	12	15	18
100	3	2	2	2	1	1	1	1
50–90	4	3	3	2	2	2	1	
10–50	5	4	1	1				
0	1							
Worse	1	1	1	1				
Total	14	10	7	6	3	3	2	1

Treatments with a short duration of effect may be worthwhile if they can be readily repeated. This is the case for drug therapy, because although a drug may have a half-life of, say, 6 h, it can readily be prescribed twice or three times daily. However, this may be impractical and not cost-effective for other treatments with short half-lives.

Treatments with durations of effect amounting to several months might be repeatable. Surgery, however, is not repeatable; for which reason follow-up over two years or more is required, in order to determine what the long-term success rate is.

Another form of data display is the survival bar graph (Fig. 5). This display shows the history of every patient treated, and whether or not they responded to treatment (for a given definition of success), and their response to any repeat treatment.

The example in Fig. 5 tells the reader that three patient did not respond to treatment, with two of them not responding to repeat treatment. Three patients had a successful outcome that did not last, and repeat treatment was not successful. Three patients had successful outcomes initially, which did not last; but repeat treatment improved the duration of response, albeit for short periods. Nine patients had successful outcomes for up to 12 months, with three of them having persisting relief at the time of follow-up. Fourteen patients had successful outcomes lasting more or less 12 months, and when that relief lapsed, repeat treatment reinstated relief for periods totalling between 16 months and 24 months.

This display portrays an imperfect treatment. Sometimes if fails. Sometimes it is successful for only short periods. But in some 50% of cases, it is successful for several months; and when relief wanes, relief can be reinstated by repeat treatment. The particular virtue of this type of display is that it can show the success rate of initial treatment and the success rate of repeat treatment, as well the duration of relief after each treatment, and the cumulative duration of relief achieved by repeat treatment.

6. Discussion

Stating that a treatment "works" lacks meaning unless and until it is qualified. For readers of studies to make a fully informed decision, authors need to provide comprehensive, transparent data on the effects of treatment not only on pain but also on function, distress, and use of other health care, each expressed in terms of the magnitude of benefit, how often it occurs, and how long it lasts. Readers are entitled to expect and demand such information, so that they can decide if the treatment is suitable and worthwhile for their own patients. Some examples from the literature serve to illustrate applying these principles.

In a study of a (then) new treatment, the authors reported that, with an average follow-up of 23.5 weeks, 14 patients reported a reduction in mean pain scores from 7.6/10 to 3.6/10, which was statistically significant and amounted to a 52% reduction in pain, with 6 of the 11 patients who were taking narcotics reducing or eliminating their use. This might appear to show that the treatment "works". To their credit, the authors provided transparent data that allow for a more incisive analysis.

Table 6
Summary statistics of the key outcomes of a contentious treatment. **ODI**: Oswestry Disability Index. **p**: p-value, two-sample *t*-test for ODI and Pain, and chi-squared for Change Pain.

Outcome	Treatment		Sham		P
ODI (0–100)	mean	sd	mean	sd	
Before	31	10	33	11	
6 months	20	12	28	15	0.02
Pain (0-10)					
Before	6.6	1.4	6.5	1.9	
6 months	4.2	2.6	5.4	2.7	0.09
Change Pain (%)	**N**	**%**	**n**	**%**	
< 0	2	6%	8	33%	
0-24	11	34%	6	23%	
25-49	7	22%	2	8%	
50-74	5	16%	7	29%	
75-99	4	13%	0	0%	0.03
100	3	9%	1	4%	

Table 7
Summary statistics for outcomes at 3 years or longer after open lumbar interbody fusion for patients with lumbar radicular pain.

	Before	>3 years	Change Pain				
Leg pain (0-10)	8	0		Leg Pain		Back Pain	
Back pain (0-10)	8	0		**n**		**n**	
SF-36 (0–100)			**100%**	14	64%	7	32%
Physical Functioning	20	75	**>50%**	4	18%	8	36%
Social Functioning	38	88	**<50%**	4	18%	8	36%
Mental Health	64	84					

Table 5 plots the number of patients who obtained various grades of relief at various periods of follow-up. Those data show that fewer than half of the patients were followed beyond 3 months. So, the study provides no valid information about long-term effects of treatment, but it does imply that there is substantial deterioration over time. Pain scores for all patients were available only at one month follow-up. So, the group scores reported apply only at one month. Readers might be concerned that this is too short a follow-up, and the data uninformative, for an invasive treatment such as MILD [43].

Table 6 shows the key, summary statistics of a placebo-controlled trial of a contentious treatment. The data show that Oswestry Disability Scores were significantly improved, by at least the MCIC in the treatment group, and a significantly greater proportion of patients achieved greater than 50% relief of pain after active treatment. However, the success rates for achieving 50% and complete relief of pain were only 13% and 9% respectively, which explains why the group pain scores were not significantly better in the treatment group. Despite its statistical significance, this low success rate makes the treatment unappealing, particularly for a technically demanding, invasive procedure such as intradiscal electrothermal therapy [44].

Table 7 summarises the outcomes of a study of open lumbar interbody fusion for lumbar radicular pain [45]. The group data show large improvements in leg pain, back pain, physical functioning, social functioning, and mental health, all being statistically significant, with p-values less than 0.01. The categorical data show a 64% success rate at achieving complete relief of leg pain for 3 years or more, with a further 18% of patients achieving at least 50% relief. Outcomes for the relief of back pain were more modest. Such data provide a comprehensive picture of what a physician and their patients can expect from treatment. There is a 1 in 5 chance of surgery not helping, but a 64% chance of obtaining complete relief of leg pain along with clinically significant improvements in function and mental health. Armed with such information the physician can discuss with their patient whether these chances are worth the risk of major surgery or if worthwhile alternatives are a better option.

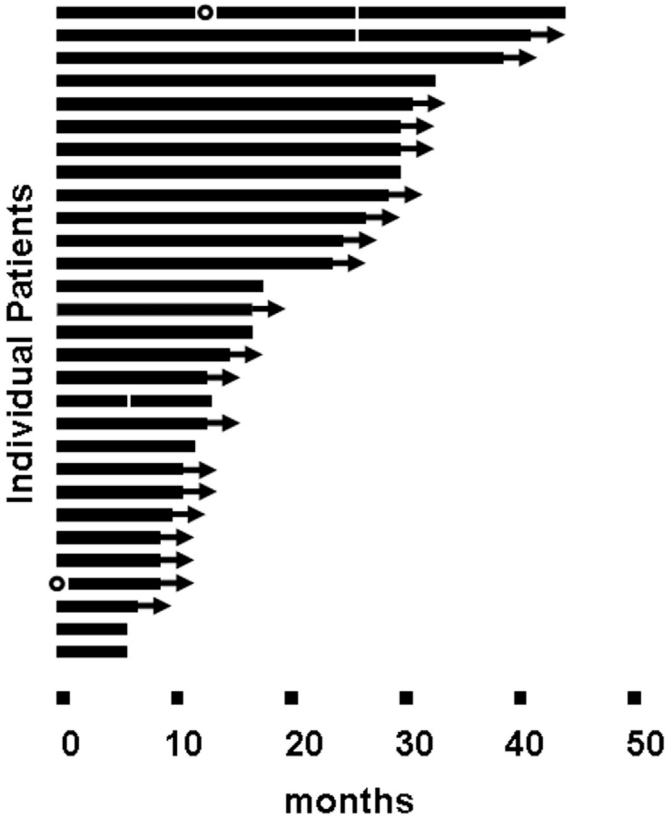

Fig. 6. A bar graph showing the duration of successful outcomes after treatment and repeat treatment.

Fig. 6 shows the long-term history of patients with successful outcomes after treatment with radiofrequency cervical medial branch coagulation [46], with success defined as complete relief of pain, restoration of all activities of daily living, and no need for other health care for neck pain. The data show that patients obtain complete relief of pain for different durations, but with a median duration of about 13 months; and complete relief can very often be reinstated by repeat treatment, to maintain relief beyond 10 months and up to 50 months. A physician can decide if these data are sufficiently convincing and attractive to adopt this treatment, and the need for repeat treatment, for their patients.

There is a dark side to these principles. If authors do not provide transparent, comprehensive data bout their treatment, are they just being irresponsible and disrespectful of their readers, or are they hiding unflattering elements of their results? Or are they trying to "sell" a cheap definition of "it works". Faced with insufficient data, educated readers are entitled to ignore or discount any claims made by the authors, or at least not be swayed by authors' rhetoric. Read the data not the prose.

Disclosures

The author was not funded to produce this article, and has no conflicts of interest with its contents.

References

[1] Huskisson EC. Measurement of pain. Lancet 1974;2:1127–31.
[2] Strong J, Ashton R, Chant D. Pain intensity measurement in chronic low back pain. Clin J Pain 1991;7:209–18.
[3] Chien CW, Bagraith KS, Khan A, Deen M, Strong J. Comparative responsiveness of verbal and numerical rating scales to measure pain intensity in patients with chronic pain. J Pain 2013;14:1653–62.
[4] Hjermstad MJ, Fayers PM, Haugen DF, Caraceni A, Hanks GW, Loge JH, Fainsinger R, Aass N, Kaasa S. European palliative care research collaborative (EPCRC). Studies comparing numerical rating scales, verbal rating scales, and visual

analogue scales for assessment of pain intensity in adults: a systematic literature review. J Pain Symptom Manag 2011;41:1073–93.
[5] Thong ISK, Jensen MP, Miró J, Tan G. The validity of pain intensity measures: what do the NRS, VAS, VRS, and FPS-R measure? Scand J Pain 2018;18:99–107.
[6] Stewart AL, Hays RD, Ware JE. The MOS short-form general health survey: reliability and validity in a patient population. Med Care 1988;26:724–35.
[7] Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34(3): 220–33.
[8] EuroQol Group. EuroQol–a new facility for the measurement of health-related quality of life. Health Pol 1990;16:199–208.
[9] Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: past, present and future. Appl Health Econ Health Pol 2017;15:127–37.
[10] Hawthorne G, Richardson J, Osborne R. The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of health-related quality of life. Qual Life Res 1999;8:209–24.
[11] Busija L, Pausenberger E, Haines TP, Haymes S, Buchbinder R, Osborne RH. Adult measures of general health and health-related quality of life: medical outcomes study short form 36-item (SF-36) and short form 12-item (SF-12) health surveys, Nottingham health profile (NHP), sickness impact profile (SIP), medical outcomes study short form 6D (SF-6D), health utilities index mark 3 (HUI3), quality of well-being scale (QWB), and assessment of quality of life (AQoL). Arthritis Care Res 2011;63(Suppl 11):S383–412.
[12] Hays RD, Spritzer KL, Schalet BD, Cella D. PROMIS®-29 v2.0 profile physical and mental health summary scores. Qual Life Res 2018;27:1885–91.
[13] Promis. Patient-reported outcomes measurement information System. https://www.healthmeasures.net/explore-measurement-systems/promis. https://commonfund.nih.gov/promis/index.
[14] Fairbank JC, Pynsent PB. The Oswestry disability index. Spine 2000;25:2940–52.
[15] Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. Spine 1983;8: 141–4.
[16] Roland M, Fairbank J. The roland-morris disability questionnaire and the Oswestry disability questionnaire. Spine 2000;25:3115–24.
[17] Smeets R, Köke A, Lin CW, Ferreira M, Demoulin C. Measures of function in low back pain/disorders: low back pain rating scale (LBPRS), Oswestry disability index (ODI), progressive isoinertial lifting evaluation (PILE), Quebec back pain disability scale (QBPDS), and roland-morris disability questionnaire (RDQ). Arthritis Care Res 2011;63(Suppl 11):S158–73.
[18] Vernon H, Mior S, The Neck Disability Index: a study of reliability and validity. L Manipulative Physiol Ther 1991l 14:409-415.
[19] Vernon HJ. The neck disability index: state-of-the-art, 1991-2008. J Manip Physiol Ther 2008;31:491–502.
[20] McCarthy MJ, Grevitt MP, Silcocks P, Hobbs G. The reliability of the Vernon and Mior neck disability index, and its validity compared with the short form-36 health survey questionnaire. Eur Spine J 2007;16:2111–7.
[21] Stratford P, Gill C, Westaway M, Binkley J. Assessing disability and change on individual patients: a report of a patient specific measure. Physiother Can 1995;47: 258–63.
[22] Westaway MD, Stratford PW, Binkley JM. The patient-specific functional scale: validation of its use in persons with neck dysfunction. J Orthop Sports Phys Ther 1998;27:331–8.
[23] Hefford C, Abbott JH, Arnold R, Baxter GD. The patient-specific functional scale: validity, reliability, and responsiveness in patients with upper extremity musculoskeletal problems. J Orthop Sports Phys Ther 2012;42:56–65.
[24] Derogatis LR, Melisaratos N. The brief symptom inventory: an introductory report. Psychol Med 1983;13:595–605.
[25] Woo AK. Depression and anxiety in pain. Rev Pain 2010;4:8–12.
[26] Rosenstiel AK, Keefe FJ. The use of coping strategies in chronic low back pain patients: relationship to patient characteristics and current adjustment. Pain 1983; 17:33–44.
[27] Sullivan MJL, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. Psychol Assess 1995;7:524–32.
[28] van der Roer N, Ostelo RW, Bekkering GE. Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. Spine 2006;31:578–82.
[29] Ostelo RW, de Vet HC. Clinically important outcomes in low back pain. Best Pract Res Clin Rheumatol 2005;19:593–607.
[30] Kovacs FM, Abraira V, Royuela A, et al. Minimal clinically important change for pain intensity and disability in patients with nonspecific low back pain. Spine 2007; 32:2915–20.
[31] Kovacs FM, Abraira V, Royuela A, et al. Minimum detectable and minimal clinically important changes for pain in patients with nonspecific neck pain. BMC Muscoskel Disord 2008;9:43.
[32] Pool JJ, Ostelo RW, Hoving JL. Minimal clinically important change of the neck disability index and the numerical rating scale for patients with neck pain. Spine 2007;32:3047–51.
[33] Carreon LY, Glassman SD, Campbell MJ, Anderson PA. Neck Disability Index, short form-36 physical component summary, and pain scales for neck and arm pain: the minimum clinically important difference and substantial clinical benefit after cervical spine fusion. Spine J 2010;10:469–74.
[34] Parker SL, Godil SS, Shau DN, Mendenhall SK, McGirt MJ. Assessment of the minimum clinically important difference in pain, disability, and quality of life after anterior cervical discectomy and fusion: clinical article. J Neurosurg Spine 2013;18: 154–60.
[35] Giraudeau B, Rozenberg S, Valat JP. Assessment of the clinically relevant change in pain for patients with sciatica. Ann Rheum Dis 2004;63:1180–1.
[36] Nagata K, Nakamoto H, Kato S. Minimum clinically important change for outcome scores among patients aged 75 or over undergoing lumbar spine surgery. Eur Spine J 2021;30:1226–34.
[37] Hägg O, Fritzell P, Nordwall A, et al. The clinical importance of changes in outcome scores after treatment for chronic low back pain. Eur Spine J 2003;12:12–20.
[38] Hung M, Saltzman CL, Kendall R, et al. What are the MCIDs for PROMIS, NDI, and ODI instruments among patients with spinal conditions? Clin Orthop Relat Res 2018;476:2027–36.
[39] Auffinger BM, Lall RR, Dahdaleh NS, et al. Measuring surgical outcomes in cervical spondylotic myelopathy patients undergoing anterior cervical discectomy and fusion: assessment of minimum clinically important difference. PLoS One 2013;8:6.
[40] Spratt KF. Minimal clinically important difference based on clinical judgment and minimally detectable measurement difference: a rationale for the SF-36 physical function scale in the SPORT intervertebral disc herniation cohort. Spine 2009;34: 1722–31.
[41] Yelland MJ, Schluter PJ. Defining worthwhile and desired responses to treatment of chronic low back pain. Pain Med 2006;7:38–45.
[42] Bogduk N, Stojanovic M. Group data or categorical data for outcomes of pain treatment? Pain Med 2020;21:2046–52.
[43] Chopko BW. A novel method for treatment of lumbar spinal stenosis in high-risk surgical candidates: pilot study experience with percutaneous remodeling of ligamentum flavum and lamina. J Neurosurg Spine 2011;14:46–50.
[44] Pauza KJ, Howell S, Dreyfuss P, et al. A randomised, placebo-controlled trial of intradiscal electrothermal therapy for the treatment of discogenic low back pain. Spine J 2004;4:27–35.
[45] Cheung NK, Ferch RD, Ghahreman A, et al. Long-term follow-up of minimal-access and open posterior lumbar interbody fusion for spondylolisthesis. Neurosurgery 2013;72:443–50.
[46] MacVicar J, Borowczyk JM, MacVicar AM, et al. Cervical medial branch radiofrequency neurotomy in New Zealand. Pain Med 2012;13:647–54.