Pain Medicine 2016; 17: 1821–1828
doi: 10.1093/pm/pnw020

OXFORD

Mathematical Validation and Credibility of Diagnostic Blocks for Spinal Pain

Andrew J. Engel, MD* and Nikolai Bogduk, MD, DSc†

*Affordable Pain Management, Chicago, Illinois, USA; †The University of Newcastle, Newcastle, New South Wales, Australia

Correspondence to: Andrew J. Engel, MD, Affordable Pain Management, 5600 N Sheridan Rd #104, Chicago, IL 60660, USA. Tel: 773-944-0365; Fax: 773-944-0470; E-mail: engel.andrew@gmail.com.

Disclosure and conflicts of interest: Neither of the authors have any disclosures related to this topic.

Abstract

Background. Diagnostic blocks are used in different ways for the diagnosis of spinal pain, but their validity has not been fully evaluated.

Methods. Four clinical protocols were analyzed mathematically to determine the probability of correct responses arising by chance. The complement of this probability was adopted as a measure of the credibility of correct responses.

Results. The credibility of responses varied from 50% to 95%, and was determined less by the agents used but more by what information was given to patients and if the agents were fully randomized for each block.

Conclusions. Randomized, comparative local anesthetic blocks offer a credibility of 75%, but randomized, placebo-controlled blocks provide a credibility of 95%, and are thereby suitable as a criterion standard for diagnostic blocks.

Key Words. Spine; Pain; Diagnostic blocks; Validity

Introduction

Like any diagnostic test, diagnostic blocks are liable to false-positive responses. Patients may report a positive response for reasons other than the pharmacological effect of the local anesthetic agent on their pain. Patients may suffer a placebo effect; or they may be confused, and think that a positive response is the expected or required response; or they think that a positive response is what will earn them a treatment for their pain. In medicolegal cases, other factors can bear on the response. A patient may believe that a positive response will vindicate their complaint of pain and injury, and will lead to compensation.

Empirical studies of single, diagnostic blocks have shown that false-positive rates can range from 25% to 45% [1–6]. These high rates compromise the diagnostic confidence that a physician can have, on the basis of a single block, that their diagnosis is correct. Depending on the prevalence of the condition being tested, positive responses may be false in one in three cases, or as many as two in three cases, or greater [7].

In order to improve diagnostic confidence, by reducing the likelihood of false-positive responses, authorities have proposed using various types of controls. A stringent form of control are placebo-controlled, triple blocks [8]. On the first occasion, a local anesthetic agent is used, that is either long acting or short acting. A local anesthetic is used in order to determine, prima facie, if the target structure is, indeed, the source of pain. A negative response concludes the investigation. A positive response invites verification. For the second block, either a placebo is administered or a local anesthetic, which may be long acting or short acting. The same agents are used for the third block. A correct, ostensibly genuine response would be long-lasting relief whenever a long-acting agent is used, short-lasting relief whenever a short-acting agent is used, and no relief when a placebo is used. This paradigm is based on the assumption that only a patient with a genuine source of pain can detect if an agent relieves their pain or not, and for how long the effect lasts.

Triple blocks are not attractive to physicians at large, for a variety of reasons. Foremost, they are consumptive of time and resources. Logistically they require additional personnel, such as a nurse, to maintain blinding. In some jurisdictions, physicians are not reimbursed for more than one, or perhaps two, diagnostic blocks.

Payers mistakenly believe that a diagnosis should be made, and can be made, using only one block.

An alternative to triple blocks, that is palatable for clinical practice, are comparative blocks [9]. On each of two occasions either a long-acting or a short-acting local anesthetic agent is used. A correct response is long-lasting relief whenever a long-acting agent is used, and short-lasting relief whenever a short-acting agent is used.

In routine practice, conventional patients are unlikely to choose to guess what response they should have. In these patients, false-positive responses might arise for natural, unpremeditated reasons, such as placebo effects. In contrast, patients with medicolegal claims, but no genuine source of pain, might be motivated to guess the correct responses, because without a genuine source of pain they are unable to detect if the agent injected relieved their pain or not. In such cases, the validity of the blocks is challenged not only by natural effects such as placebo, but also by premeditated mischief.

Fundamental to the diagnostic confidence that diagnostic blocks offer is the credibility of the response. Without access to other information, a physician cannot tell if a patient is responding honestly and accurately to a diagnostic block, and is not responding arbitrarily or guessing what the response should be. The present study was, therefore, undertaken in order to calculate the credibility of responses using different protocols for diagnostic blocks. The results serve to inform physicians about how confident they can be about the responses obtained by the protocols that they choose to use.

Methods

A set of protocols were developed that differed according to the number of diagnostic blocks performed, the agents used, and the information that would be disclosed in the informed consent to the patient. For each protocol, the probability was calculated of guessing the agents used and, thereupon, the probability of guessing the "correct" series of response that the patient should offer in order to achieve a positive result. The credibility of the overall response was then determined as the complement of this probability, i.e., credibility = [1 – probability of guessing correctly].

The four protocols were: 1) comparative blocks using alternative agents in random order; 2) comparative blocks using agents randomly on both occasions; 3) placebo-controlled, triple blocks using alternative agents in random order; and 4) triple blocks, with agents used randomly on each occasion.

Common to all protocols was the assumption that the physician who administers the blocks, the patient, and the person who assesses the responses, remain blinded to identity of the agents used. The physician needs to be blinded in order that they do not subconsciously try to be less accurate in placing their block when placebo agents are administered, or subconsciously to cue patients as to the duration of action when active agents are administered. Although other variants might be possible, a suitable protocol for blinding the physician could be as follows.

In preparation for a block, the physician acts with a nurse assistant. Each is dressed to handle syringes under sterile conditions. The physician draws up separate syringes containing each of the agents that might be used in a given protocol, being two or three agents according to the protocol used. Having drawn up all agents, the physician turns away while the nurse consults a randomization schedule; retains the syringe containing the selected agent; and disposes of the syringes containing the unwanted agents. Under this protocol the physician knows that only an agent that he or she has drawn up can be administered, and that no foreign agent with adverse effects is to be used; but the physician is unaware of the particular agent to be administered. The nurse maintains a record of which agent was used so that the response of the patient can later be correlated with the nature of the agent, and so that an appropriate agent can be selected on future occasions if the protocol requires using complementary agents. Under this protocol, the nurse who selects the agent is barred from assessing the response to blocks; but because the physician is blinded, he or she can assess the responses to the blocks. Otherwise a third party must assess the response.

Pertinent to the interpretation of responses to diagnostic blocks are the definitions of relief and the duration of relief. For the purposes of the present study relief was defined as complete relief of pain in the region targeted, and long-lasting relief was defined as relief lasting longer than short-lasting relief, provided that in both instances, relief ceased within a period measurable in minutes or hours. These considerations have been discussed in detail elsewhere [9–12] but for readers not familiar with that literature a summary of the data and arguments is provided in Appendix 1.

Results

Comparative Blocks; Alternative Agents in Random Order

In protocol 1, comparative blocks are administered, commencing randomly with either a short-acting local anesthetic or a long-acting local anesthetic. If the response is positive the complementary agent is used for the second block. The patient is informed that a short-acting agent and a long-acting agent will be used, but they will not know in which order they are used.

Under this protocol, the patient does not know which agent is to be used on the first occasion, but they do know that the opposite agent will be used for the

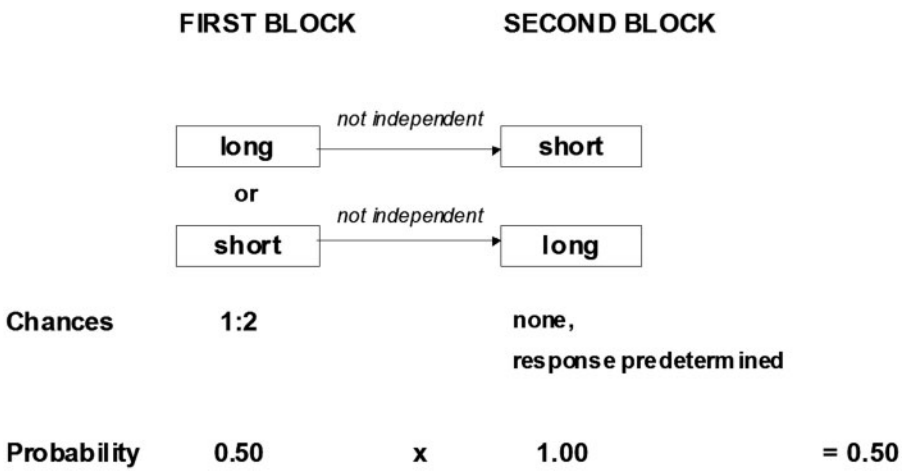

Figure 1 The sequence, possible responses, and chances of guessing responses correctly for comparative blocks administered in random but complementary order.

second block. For the first block, they have a 50:50 chance of guessing the agent used, but their decision for the first block locks in their choice for the second block: it must be the opposite from what they chose for the first block (Figure 1). The probability for guessing correctly is wholly determined by the first block, and amounts to 0.50. The resultant credibility is 1–0.50, which amounts to 0.50. This means that the physician can be 50% confident that the response is credible or genuine.

Comparative Blocks; Agents Randomly, Both Occasions

In protocol 2, comparative blocks are administered. For the first block, either a short-acting local anesthetic or a long-acting local anesthetic is randomly selected. If the response is positive, either agent is again randomly selected for the second block. The patient is informed that, on both occasions, either a short-acting agent or a long-acting agent will be used, but they will not know which is used.

Under this protocol, the patient may guess long-acting or short-acting for the first block, which amounts to a probability of 0.50 (Figure 2). However, that choice recurs for the second block, because the selection of agents remains random. Consequently, the probability of guessing both agents correctly is 0.50 * 0.50, which amounts to 0.25. The resultant credibility is 0.75 (Figure 2). The physician can be 75% confident that the response is genuine.

Placebo-Controlled, Triple Blocks; Alternative Agents in Random Order

In protocol 3, placebo-controlled, comparative blocks are administered. For the first block either a short-acting local anesthetic or a long-acting local anesthetic is used. If the response is positive, either the complementary local anesthetic agent or a placebo is randomly used for the second block. For the third block, the agent not used for the second block is used. The patient is informed that for the first block either a short-acting agent or a long-acting agent will be used, but for the second and third blocks either a placebo or the other local anesthetic will be used.

Under this protocol, on the occasion of the first block, the patient must chose if the response should be short-lasting or long-lasting, and they need to remember this choice, for later they must tender the opposite response (Figure 3). However, the use of a placebo introduces a new domain of response. The patient must choose between relief and no relief. However, given their response to the first block they do not have to choose if the relief is long or short for the second or third block. That choice is locked in as the opposite to their choice for the first block (Figure 3). The probabilities become 0.50 for the first block (long or short), and 0.50 for the second block (relief or no relief). Having made these two decisions, no choice remains for the third block; the response must be the opposite of each of the two earlier responses. The total probability of guessing the correct responses becomes 0.50 * 0.50, which amounts to 0.25, with a credibility of 75% (Figure 3).

Triple Blocks; Agents Used Randomly on Each Occasion

In protocol 4, placebo-controlled, comparative blocks are administered. For the first block either a short-acting local anesthetic or a long-acting local anesthetic is used. If the response is positive, a second block will be performed, randomly using either a short-acting local anesthetic or a long-acting local anesthetic, or a placebo. For the third block, again a short-acting local anesthetic or a long-acting local anesthetic, or a placebo is randomly used. The patient is informed that for the

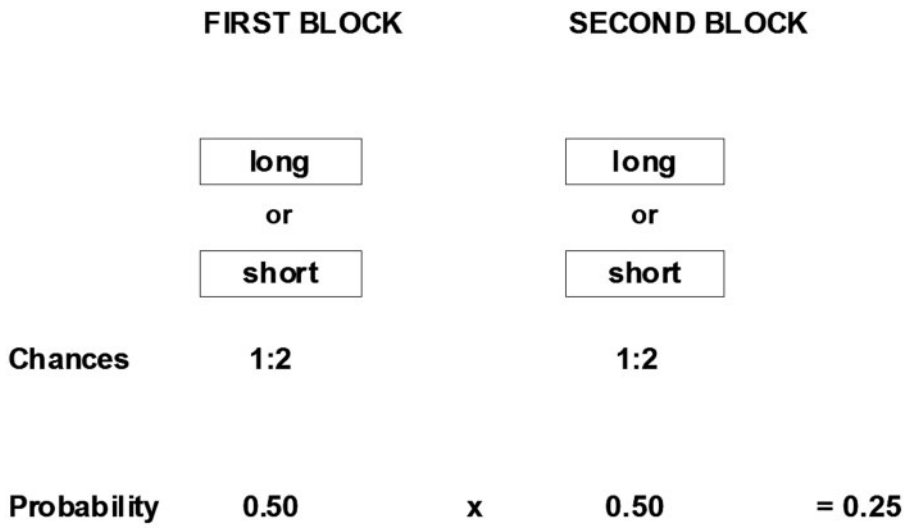

Figure 2 The sequence, possible responses, and chances of guessing responses correctly for comparative blocks administered randomly on each occasion.

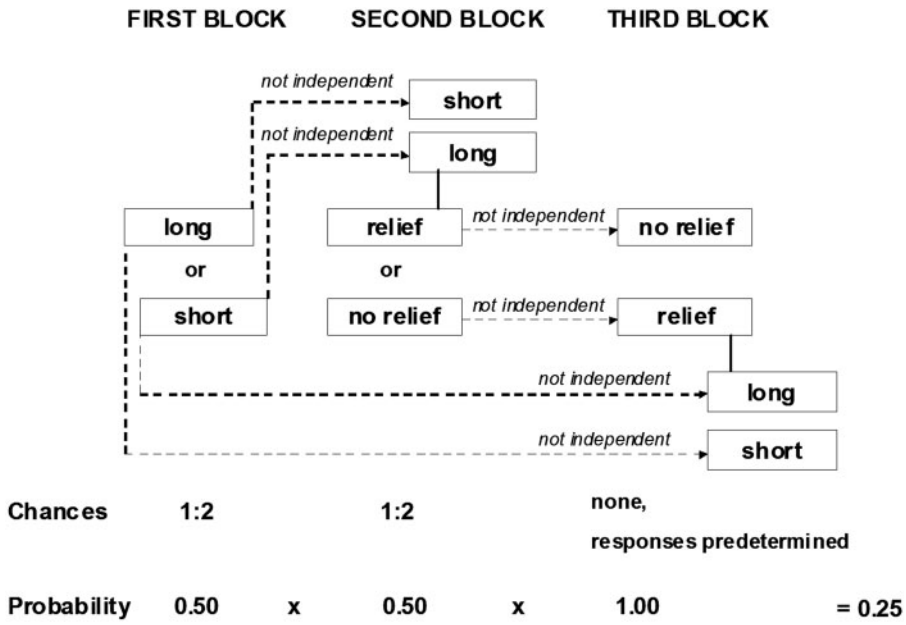

Figure 3 The sequence, possible responses, and chances of guessing responses correctly for placebo-controlled triple blocks administered randomly but in complementary order.

first block either a short-acting agent or a long-acting agent will be used, but for the second and third blocks a short-acting local anesthetic or a long-acting local anesthetic, or a placebo will be used.

Under this protocol, the probability of guessing correctly the response for the first block is 0.50 (long or short). For the second block, the choice is between relief or no relief, and long or short if relief is chosen (Figure 4). The options, therefore, number three, and the probability of guessing the correct one is 0.33. The same options apply for the third block, and the probability is again 0.33. The total probability of guessing the sequence correctly becomes 0.50 * 0.33 * 0.33, which amounts to 0.054 (Figure 4). The resultant credibility is 94.6%.

Discussion

Credibility is not the same as validity. Credibility is the measure of how unlikely the pattern of response to

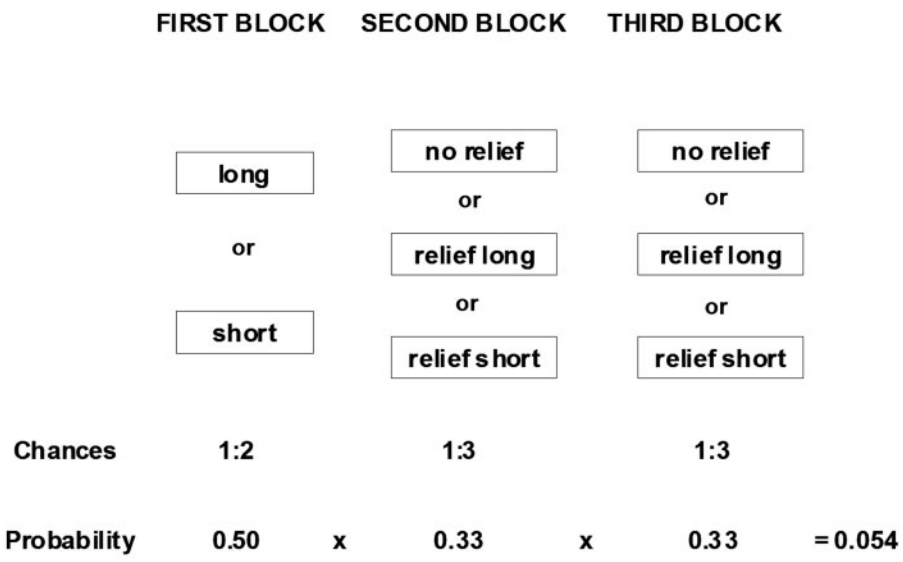

Figure 4 The sequence, possible responses, and chances of guessing responses correctly for placebo-controlled triple blocks administered randomly on each occasion.

diagnostic blocks is due to guessing or random behavior. In contrast, validity is the measure of how well the correct pattern of response establishes the presence of the condition being diagnosed. Nevertheless, credibility is pivotal to validity because theoretically it is possible for a block to appear valid if the pattern of responses is correct, but that validity of the implied diagnosis lapses if the responses themselves lack credibility, because they were guessed or random.

Validity is typically determined by matching the results of a particular test against the results of an independent test, known as the criterion standard. Most often, the criterion standard is a physical test, such as an operative finding or a pathology test, over whose results there is little or no dispute. However, such physical tests are rarely available for conditions typically diagnosed using diagnostic blocks.

In the absence of a physical criterion standard, the diagnostic block itself is promoted to become the only available criterion standard [12], and it is in this regard that credibility assumes a greater practical significance. Credibility becomes the surrogate for validity. When we cannot validate a diagnostic block against a physical criterion standard, we must apply philosophical techniques, such as declaring an axiom, which relies on the credibility of the responses to the block [12]. Blocks of lesser credibility (such as comparative blocks) could be validated against blocks of greater credibility (such as placebo-controlled blocks), as has been done [8], but eventually, the validity of that criterion standard depends on its credibility.

What the present study shows, by mathematical analysis, is that credibility is determined less by what agents are administered or the order in which they are administered, but more by if they are randomized and on what the patient is told. Credibility is low if blocks are alternated and if patients are told that long-acting and short-acting agents will be alternated, as in protocol 1. Chance applies only for the first block; but thereafter, the response for the second block is locked in because it must be opposite to that for the first block. The probability of a correct response arising—by guessing or by random behavior—is 50%.

Credibility is increased to 75% if a placebo control is introduced into a sequence of alternating comparative blocks (protocol 3), because chance occurs twice: once for long or short relief, and once for relief or no relief. However, the same credibility of 75% is provided by comparative blocks that are fully randomized on both occasions, i.e., short-acting or long-acting agent for the first block, and short-acting or long-acting agent again for the second block (protocol 2). As the two protocols are statistically equivalent for credibility, randomized comparative blocks become a legitimate and more practical substitute for placebo-controlled, alternating blocks. The same credibility is generated by two blocks instead of three blocks.

Of the four protocols, the highest credibility is generated by triple blocks in which placebo and a long-acting agent and a short-acting agent are randomized for the second and for the third block (protocol 4). A one in two chance is followed twice by a one in three chance, to produce a credibility of 94.6%.

Conspicuously, the value for this credibility is effectively equal to the *P* value of 0.05 that is conventionally

applied throughout statistics. That convention means that there is a less than 5% chance that the results arose by chance alone, and that the investigator can be 95% confident that the response was not due to chance.

This result serves to validate randomized, placebo-controlled, diagnostic blocks as a criterion standard. In the absence of a superior criterion standard, a 5% error margin renders randomized, placebo-controlled, diagnostic blocks an acceptable criterion standard, for practical purposes.

Reviewing the published literature, it is conspicuous that there are no reports of physicians using randomized blocks, with or without placebo controls. When reported, comparative blocks have been used in an alternating fashion, which carries a credibility of only 50%. Placebo-controlled blocks with complementary alternation of local anesthetic agents have a credibility of 75%, but not 95%. It is perhaps more than a coincidence that the empirical data resonate with this figure. A credibility of 75% is broadly compatible statistically with the observed success rate of 70% for treatment based on these blocks [13,14]. The possibility arises that applying the more rigorous protocol of randomized blocks in the future might improve the success rate.

The results of the present study provide several salient messages for physician who practice diagnostic blocks. Foremost, the results show that alternating agents, and telling patients that they will be alternated, compromises the credibility—and therefore the validity—of diagnostic blocks. Credibility is secured by randomization of all blocks, and informing patients that neither they nor the physician will know which agent will actually be used, until the entire series of blocks is completed.

For practical, pragmatic purposes physicians might chose to accept a credibility of 75% and, therefore, adopt randomized, comparative blocks as the standard of practice. Less stringent protocols offer a credibility no better than 50%. However, for cases in which a confidence greater than 75% is required—for example in medicolegal determinations—randomized, placebo-controlled, triple blocks offer 95% confidence, which is consonant with the burden of proof accepted for most statistical purposes.

References

1 Barnsley L, Lord S, Wallis B, Bogduk N. False-positive rates of cervical zygapophysial joint blocks. Clin J Pain 1993;9:124–30.

2 Schwarzer AC, Aprill CN, Derby R, et al. The false-positive rate of uncontrolled diagnostic blocks of the lumbar zygopophysial joints. Pain 1994;58:195–200.

3 Manchikanti L, Pampati V, Fellows B, Bakhit CE. Prevalence of lumbar facet joint pain in chronic low back pain. Pain Physician 1999;2:59–64.

4 Manchikanti L, Pampati V, Fellows B, Bakhit CE. The diagnostic validity and therapeutic value of lumbar facet joint nerve blocks with or without adjuvant agents. Curr Rev Pain 2000;4:337–44.

5 Manchikanti L, Boswell MV, Singh V, et al. Prevalence of facet joint pain in chronic spinal pain of cervical, thoracic, and lumbar regions. BMC Musculoskelet Disord 2004;5:15.

6 Manchukonda R, Manchikanri KN, Cash KA, Pampati V, Manchikanti L. Facet joint pain in chronic spinal pain: An evaluation of prevalence and false-positive rate of diagnostic blocks. J Spinal Disord Tech 2007;20:539–45.

7 Bogduk N. On the rational use of diagnostic blocks for spinal pain. Neurosurg Q 2009;19:88–100.

8 Lord SM, Barnsley L, Bogduk N. The utility of comparative local anesthetic blocks versus placebo-controlled blocks for the diagnosis of cervical zygopophysial joint pain. Clin J Pain 1995;11: 208–13.

9 Barnsley L, Lord S, Bogduk N. Comparative local anesthetic blocks in the diagnosis of cervical zygopophysial joints pain. Pain 1993;55:99–106.

10 Bogduk N. Diagnostic nerve blocks in chronic pain. In: Breivik H, Shipley M, eds. Pain. Best Practice & Research Compendium. Edinburgh: Elsevier; 2007:47–55.

11 Curatolo M, Bogduk N. Diagnostic and therapeutic nerve blocks. In: Fishman SM, Ballantyne JC, Rathmell JP, eds. Bonica's Management of Pain, 4th edition. Philadelphia: Wolters Kluwer; 2010:1401–23.

12 Engel A, MacVicar J, Bogduk N. A philosophical foundation for diagnostic blocks, with criteria for their validation. Pain Med 2014;15:998–1006.

13 Lord SM, Barnsley L, Wallis BJ, McDonald GJ, Bogduk N. Percutaneous radio-frequency neurotomy for chronic cervical zygopophysial-joint pain. N Engl J Med 1996;335:1721–6.

14 MacVicar J, Borowczyk JM, MacVicar AM, Loughnan BM, Bogduk N. Cervical medial branch radiofrequency neurotomy in New Zealand. Pain Med 2012;13:647–54.

15 McDonald GJ, Lord SM, Bogduk N. Long-term follow-up of patients treated with cervical

radiofrequency neurotomy for chronic neck pain. Neurosurgery 1999;45:61–7.

16 MacVicar J, Borowczyk JM, MacVicar AM, Loughnan BM, Bogduk N. Lumbar medial branch radiofrequency neurotomy in New Zealand. Pain Med 2013;14:639–45.

17 Bogduk N, ed. Practice Guidelines for Spinal Diagnostic and Treatment Procedures, 2nd edition. San Francisco: International Spine Intervention Society; 2013:137–8, 597–8.

18 Moore DC, Bridenbaugh LD, Bridenbaugh PO, Tucker GT. Bupivacaine for peripheral nerve block: A comparison with mepivacaine, lidocaine, and tetracaine. Anesthesiology 1970;32:460–3.

19 Moore DC, Bridenbaugh LD, Bridenbaugh PO, Tucker GT. Bupivacaine: A review of 2,007 cases. JAMA 1970;214:713–18.

20 Cousins MJ, Mather LE. Clinical pharmacology of local anaesthetics. Anaesth Intens Care 1980;8:257–77.

21 Rubin AP, Lawson DIF. A controlled of bupivacaine: a comparison with lignocaine. Anaesthesia 1968;23:327–31.

22 Watt MJ, Ross DM, Atkinson RS. A double blind trial of bupivacaine and lignocaine. Anaesthesia 1968;23:331–37.

23 Engel A, MacVicar J, Bogduk N. A philosophical foundation for diagnostic blocks, with criteria for their validation. Pain Med 2014;15:998–1006.

Appendix 1

Some physicians use operational criteria for relief of pain and for expected duration relief that are different from those used by others. Some readers may be unaware of the arguments for and against different criteria. This appendix summarizes the arguments and the available data, with reference to the pertinent literature.

Relief

The present authors recommend that positive responses to diagnostic blocks be defined as complete relief of pain. In turn, complete relief is defined numerically as a pain score of zero, complemented by opposite responses to two complementary questions: "Has your pain gone?" "Yes." and "Do you have any pain left?" "No."

The criteria are advocated because they are the ones applied in the seminal literature on comparative blocks and placebo-controlled blocks for spinal pain [7–11] and its successful treatment [13–16]. This is the only criterion consistent with the actual source of pain having been found.

Although some physicians accept partial responses, such as 80% relief or 50% relief as a positive response, these are ambiguous. Although one interpretation of partial responses is that there is some other, undisclosed source for the remnant pain, this is no more than a conjecture unless and until that other source is found. Competing conjectures are that the patient is uncertain about the effect of the block, or that the source of pain has not been fully anesthetized for technical or other reasons.

However, a pertinent consideration is that a patient is not required to report complete relief of all of their pain. In a patient with bilateral pain, blocks of the left side might provide complete relief of pain on the left but no relief on the right. Likewise, in a patient with neck-shoulder pain and headache, a block may completely relieve the headache but not the neck-shoulder pain, or vice versa. In patients who might have multiple sources of pain, the operation criterion becomes complete relief of pain in the region targeted by the block [17].

Duration

There is no singular or absolute number that defines the expected duration of action of a local anesthetic agent. For a given agent, clinical studies provide a mean duration of action accompanied by a considerable standard deviation around this mean [18–20]. Thus, although the mean duration may be regarded as "typical" (or modal in a statistical sense) it is not the only duration that legitimately applies for an agent. Moreover, in some patients the longest duration of action of short-acting agents is longer than the shortest duration of action of long-acting agents in other patients. This underscores that an absolute number cannot be applied for expected duration of action. However, it has been shown that—in a given patient—short-acting agents consistently have shorter duration of action than that of long-acting agents [21, 22].

Therefore, the operational criterion for defining long response and short response is a relative one, unique to the patient. The long-acting agent should relieve pain for longer than does the short-acting agent, irrespective of the absolute duration of action.

However, fundamental to the use of local anesthetic agents for diagnostic purposes is that the effect must cease at a time consistent with the known duration of action of the agent used. This may be up to 24 hours in the case of bupivacaine [9,18–20], but the duration is nevertheless sensibly measurable in hours. Relief that last for days or weeks may be merciful and welcomed, but it is not compatible with a simple pharmacologic

effect of the local anesthetic agent. Something else is going on; it may be physiological or it may be psychological; but in either event, the prolonged response defeats the purpose of the diagnostic block [23]. For a diagnostic block to be meaningful, the onset of relief must be complemented by the offset of relief, within a reasonable time frame [23].