Interventional Pain Medicine 1 (2022) 100127

Contents lists available at ScienceDirect

Interventional Pain Medicine

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

On understanding the validity of diagnostic tests

Nikolai Bogduk

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

ARTICLE INFO

Keywords:
Diagnosis
Validity
Sensitivity
Specificity
Likelihood ratio

ABSTRACT

For clinical practice to be professionally responsible, any diagnostic tests used need to be valid because, if a test lacks validity the information that it provides is wrong. Of the several subtypes of validity, the crucial one is construct validity, which determines how well a diagnostic test discriminates simultaneously between the presence and the absence of the condition being diagnosed. Its key parameters are the sensitivity and specificity of the test, and its (positive) likelihood ratio. The likelihood ratio serves mathematically as a coefficient in an equation that measures the confidence one can have that a positive result of the test is true-positive, given the prevalence of the condition being diagnosed. There is no ideal value for likelihood ratios that make a diagnostic test worthwhile. The value required depends on how much confidence a physician needs to have in a diagnosis before undertaking treatment, which must be calculated using the likelihood ratio and prevalence of the condition being diagnosed.

1. Introduction

In general terms, validity means that a diagnostic test correctly detects the condition that it is designed to detect. If a diagnostic test lacks validity it serves no useful purpose in clinical practice, because the information that it produces will be wrong.

A problem in the past was that when physicians were taught a diagnostic test, they – and their teachers – assumed that the test was valid, and that whenever the test was positive it was correctly positive. Research into the validity of diagnostic tests has repeatedly dispelled this fallacy. Diagnostic tests that were purported to work have been shown not to work.

2. Terminology

With respect to the terminology used, validity is a vexatious topic to address. Validity comes in various forms, and different disciplines apart from medicine, such as psychology and sociology, use different types of validity for different applications and with different definitions; but there is no single, authoritative source that provides universally accepted definitions that cover all concepts.

This essay reviews the types of validity that are pertinent to the practice of interventional pain medicine. Some people might dispute some of the names used, but that is a matter of semantics. The associated concepts are what matters.

The types of validity pertinent to interventional pain medicine are concept validity, content validity, face validity, and construct validity. Perhaps curiously, perhaps artificially, these different types of validity can be stratified against the evolution of diagnostic tests, from conception, through investigation, and finally to consolidation. Different types of validity assume different levels of relevance in the course of this evolution.

2.1. Concept validity

Concept validity asks if the test is plausible in principle, such as having a rational, biological basis. This is particularly relevant when a new test is first conceived, and before it is investigated. Having concept validity invites further development. Lack of concept validity invites reservations, which is unusual but not without precedent.

An extreme example is the proposition that fixation of the temporoparietal suture causes back pain, and can be diagnosed by palpating that fixation. This example lacks concept validity because there is no known link between cranial sutures and back pain, and it is seriously questionable if fixation could be detected in a joint that is normally strongly locked.

A converse example is provocation discography for lumbar discogenic pain. In the past, discography was held to lack concept validity because the disc was not innervated and, therefore, could not be a source of pain [1]. This objection was subsequently overcome by demonstrating that the lumbar discs were, indeed, innervated and, therefore, notionally could be a source of pain.

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

https://doi.org/10.1016/j.inpm.2022.100127
Received 12 June 2022; Accepted 16 June 2022

Otherwise, diagnostic tests might be proposed for a particular structure being the source of pain but denied *ex cathedra* on the grounds that this structure is not a known source of pain. In such cases, content validity can be initiated by showing that the structure can be source of pain when it is noxiously stimulated experimentally. This occurred in the early days of cervical zygapophysial joint blocks. The establishment resisted the concept of cervical zygapophysial joint pain because no-one had ever heard of these joints being a source of neck pain. This objection was eliminated by showing that neck pain could be evoked in normal volunteers by stimulating these joints experimentally [2,3].

2.2. Content validity

Content validity resembles an administrative requirement. It asks that the test be comprehensively and accurately defined, with no ambiguities, as opposed to "you know what I mean". The purpose is to ensure that whenever the test is performed it is performed in the same manner.

If all aspects of a diagnostic test are not strictly defined, one version of the test might resemble another version but could differ in some critical aspect that was not covered in the definition. Such differences may create significant differences in the outcome of the test. Having content validity is akin to defining words accurately, so that when a word is used its meaning is neither misunderstood nor misrepresented.

Pertinent to the content validity of diagnostic blocks is if the operational criteria for diagnostic blocks proposed by Engel et al. [4,5[are satisfied (Table 1). If the answers to these questions are all "yes", the block has been performed with strong validity. If any of the questions are answered "no", the block has been performed differently and with less validity, even if the block has been given the same name.

2.3. Face validity

Face validity asks if the diagnostic test has actually been shown to operate by the anatomical and physiological mechanisms by which it is purported to operate. In this regard, concrete evidence is required to replace assumption or assertion.

Some physical examination tests may have purported mechanisms by which they operate, but if these mechanisms have not been demonstrated by experiment, those who use the test cannot know if the test is doing what it is supposed to be doing. An example would be the assertion that moving a joint in a particular way stretches a particular ligament and only that ligament. Face validity would require demonstration that performing that movement actually does stretch the ligament but no other ligament. Conversely, showing that the movement does not stretch the ligament immediately invalidates the claimed mechanism of the test, and calls into question the validity of the test. An example of the latter pertains to the use of Ober's test for tightness of the iliotibial tract over the knee. Anatomical studies have shown that the iliotibial tract is too strongly attached to the greater trochanter and the linea aspera for it to be able to transmit tension from the pelvis to the knee [6,7].

For diagnostic blocks, it is well established that local anesthetics will block nerves; so, there is no concern about the face validity of that aspect of diagnostic blocks. Face validity pertains instead to whether or not anaesthetising a structure, or the nerves that supply it, successfully relieves pain stemming from it. Establishing this face validity relies on studies in normal volunteers.

Face validity is demonstrated by showing that normal volunteers are protected from experimentally-evoked pain after performing the block of that structure. In the field of interventional pain medicine, face validity has been established in this way for medial branch blocks for lumbar zygapophysial joint pain [8], and for sacral lateral branch blocks for pain stemming from the posterior sacroiliac ligaments [9]. For other target structures, studies have shown that they can be a source pain in normal volunteers [10–12], but no studies have formally shown that normal volunteers are protected from such pain by blocking the index structure.

Table 1
The operational criteria for valid diagnostic blocks, proposed by Engel et al. [4,5].

Was the block target-specific?
Was the effect temporary?
Was relief partial or complete?
Did repeat blocks reproduce the response?
Were control blocks used?
Were comparative blocks used?
Was placebo used?
Were agents fully randomised?
Have these blocks been replicated?

2.4. Construct validity

Construct validity is the most crucial type of validity for any diagnostic test. It directly assesses the idea (construct) that the test is able to make a diagnosis. Construct validity supersedes all other types of validity. It does not rely on concept validity and face validity having been established. If a test has clearly been shown to work empirically, it does not matter if we do not know how or why it works.

Construct validity pertains to the extent to which a diagnostic test simultaneously detects the condition of interest when it is present, and excludes the condition when it is absent. Simultaneity is an essential requirement for construct validity, because it underpins the discriminative ability of the test. A test is of little use if it cannot distinguish the presence from the absence of the condition being diagnosed. Although it might offer positive results, it does not indicate if these results are true-positive or false-positive.

In the past, physicians were taught that whenever a diagnostic test was positive it was always correctly so. Under those conditions, obtaining a positive result is satisfying to physicians because it allows them to feel that they used their skills to detect something. To be told that what they found is wrong is shattering and an anathema to them.

Showing that diagnostic tests can have false-positive results, and getting physicians to admit this and recognise it is one of the greatest, but unpublished, achievements of Evidence-Based Medicine. A false-positive result is a wrong result. Adjusting clinical practice to this possibility is critical to practice with intellectual integrity. The requirement is not to abandon diagnostic tests that have false-positive results, but to know how often this occurs and, thereby, to know the chances of the test being wrong. This is done by measuring the validity of the test.

3. Measurement

For testing a test, a study requires a sample of patients with and without the condition being diagnosed. Each patient is tested by the test in question and by a criterion standard. A criterion standard is another test, whose diagnostic accuracy either is not in doubt or is more trusted than that of the test being tested. For example, for testing the validity of medical imaging, post-mortem findings or biopsies might be the criterion standard. If palpation of masses is being tested, scans of the target organ can be the criterion standard.

Once patients have completed the diagnostic test and the criterion standard, the results of those tests are entered into a contingency table, according to if the condition to be diagnosed is present or absent by either test (Table 2).

In such a table, the "a" cell indicates the number of patients who have

Table 2
A contingency table for determining the construct validity of a diagnostic test.

		Criterion Standard		
		Present	Absent	
Diagnostic Test	**Present**	a	b	a + b
	Absent	c	d	c + d
		a + c	b + d	

the condition according to the criterion standard, and in whom the diagnostic test in question was positive. These patients are the ones in whom the test was true-positive. Cell "b" shows the number of patients who did not have the condition, but in whom the diagnostic tests was, incorrectly, positive. These results were false-positive.

Cell "c" indicates the number of patients who had the condition, according to the criterion standard, but in whom the diagnostic test was, wrongly, negative. These results of the test were false-negative. Cell "d" represents the patients who did not have the condition and in whom the diagnostic test was correctly negative. Those results were true-negative.

The ratio [a/(a + c)] indicates how many of the cases, in which the condition was present, the diagnostic test correctly detected. This ratio is known as the sensitivity of the test, but can also be interpreted as the true-positive rate. There were [a+c] positive cases that the test should have detected, but the test was correctly positive in only [a] of these cases. Note how sensitivity (and the true-positive rate) are read "down" the first column, from "a" to "a+c".

The ratio [d/(b + d)] indicates in how many cases, in which the condition was absent, the diagnostic test correctly detected absence. This ratio is known as the specificity of the test, and can be understood as how often the test correctly rules out the condition to be diagnosed. Note how specificity is read "up" the second column, from "b + d" to "b".

A third ratio is [b/(b + d)]. It is read "down" column 2, and mathematically it is the complement of the specificity of the test, i.e.

$$specificity = \frac{d}{[b+d]}$$

$$\frac{[b+d]}{[b+d]} = 1$$

$$\frac{b}{[b+d]} + \frac{d}{[b+d]} = 1$$

$$\frac{b}{[b+d]} = 1 - \frac{d}{[b+d]}$$

$$\frac{b}{[b+d]} = 1 - specificity$$

The value [b] is the number of cases in whom the condition was absent but the diagnostic test wrongly rated them as positive. The ratio [b/(b + d)] can, therefore, be interpreted as the false-positive rate, i.e. the rate at which the test should have been negative but was wrongly positive.

These ratios can be used to derive statistics that indicate how well the diagnostic test discriminates between the presence and absence of the condition being diagnosed. Subsequently these statistics can be used to determine how informative and useful the diagnostic is.

4. Statistics

Inspecting Table 2 should indicate to the reader that high values in cell "a" and in cell "d" are good. High values in these cells indicates that the test is correctly detecting large proportions of the cases in which the condition of interest is either present or absent. Reciprocally, the values in cells "b" and "c" should be low, for they indicate how often the diagnostic test is wrong. High values in these latter cells indicate that the diagnostic test is performing badly. In essence, the performance of the test is reflected by the imbalance of the diagonals of Tables 2 and i.e. the degree to which the values [a] and [d] are large compared with the values [b] and [c].

The statistic of interest is known the positive likelihood ratio (+LR), and is defined as:

$$+LR = \frac{sensitivity}{[1 - specificity]}$$

When expressed in these terms the positive likelihood ratio might seem rather anonymous: just a collection of new terms. Closer inspection shows that it can also be expressed as:

$$+LR = \frac{true\ \ positive\ \ rate}{false\ \ positive\ \ rate}$$

This makes the definition more meaningful. It is the ratio between true-positive results and false-positive results. Colloquially it can be regarded as the true-positive rate discounted by false-positive rate. This interpretation conveys the flavour that the positive likelihood ratio reflects the extent to which positive results are contaminated by false results.

Before examining the application of the positive likelihood ratio to assess the utility of a diagnostic test, readers troubled by the question" "what does all this mean?", or who are interested in understanding the derivation of the positive likelihood ratio, should digress temporarily to Appendix 1.

As explained in Appendix 1, the positive likelihood ratio is a fortuitous coefficient that emerges when we explore the algebra of validity. Appendix 1 also explains that the algebra does not work if we think in terms of chances. The algebra works only if think in terms of odds. Therefore, we need to stop thinking in terms of chances, and instead think in terms of odds.

Appendix 1 shows us that:

[the odds that a positive test is true-positive] = [prevalence odds]. [likelihood ratio],where.

- [prevalence odds] is the prevalence of the condition being diagnosed expressed as odds; and
- [likelihood ratio] is the measure of the validity of the diagnostic test used.

Once the odds of a positive test being true-positive are known, they can be converted back to chances using simple arithmetic. If the odds are m: n, the chances that the test is correct become [m/(m + n) x 100%]. These chances constitute the diagnostic confidence that a physician can have in the result of the test being correct.

For example, let there be a good test, with a sensitivity of 0.90 and a specificity of 0.80, looking for a condition that has a prevalence of 60%.

Prevalence odds = 60:40

Likelihood ratio = 0.90 / (1–0.80) = 4.5

The odds that a positive test is true-positive = [60 : 40] x 4.5 = 270 : 40

The diagnostic confidence is 270 / (270 + 40) = 270 / 310 = 87%

Let there be a weaker test for the same condition with the same prevalence.

Prevalence odds = 60:40.
Sensitivity = 0.60.
Specificity = 0.50

Likelihood ratio = 0.60 / (1–0.60) = 1.5

The odds that a positive test is true-positive = [60 : 40] x 1.5 = 90 : 40

The diagnostic confidence is 90 / (90 + 40) = 900 / 130 = 69%

In summary.

- the positive likelihood ratio is derived from the sensitivity and the specificity of the test;
- the positive likelihood ratio indicates how well the diagnostic test distinguishes true-positive results from false-positive results;
- to know how well a diagnostic test works we need to know the prevalence of the condition being diagnosed;
- all calculations need to be expressed as odds, not chances;
- the product of the positive likelihood ratio and the prevalence odds generates the odds that a positive result is true-positive;

- the odds of a true-positive result can be converted to the chances that it is positive;
- those chances indicate how confident the physician can be that their positive diagnosis is correct, and can be referred to as the diagnostic confidence that the physician can have.

5. Utility

There is no single, magic value for positive likelihood ratio that makes a diagnostic test a good test. The likelihood ratio measures the discriminative power of the diagnostic test, but the feature that defines if it is a good test is the diagnostic confidence that likelihood ratio generates in any particular context. The utility of any given test requires not only calculating the diagnostic confidence but also then considering how useful that confidence is for any given situation.

Table 3 illustrates some examples. Five different tests have been applied to diagnose a particular condition whose prevalence is 40%.

Table 3 shows that Test 1 has high sensitivity, high specificity, and a likelihood ratio of 4.0. Against a prevalence of 40%, it generates a diagnostic confidence of 73%. Whether or not this is good enough is a matter of judgement for the physician who uses it. Against the legal standard of "more likely than not" Test 1 provides a handsome surplus of confidence. For selecting a patient for low-risk conservative therapy, the test provides enough confidence that the diagnosis – and the associated treatment – is probably correct. For an irreversible, surgical treatment the physician might prefer a diagnostic test that provides a greater degree of diagnostic confidence. In that event the physician will need to find one.

For a physician seeking a particular level of diagnostic confidence, the odds equation can be used in reverse to calculate how powerful this desired test needs to be. In the equation

odds that a positive test is true-positive = [prevalence odds]. [likelihood ratio]

we can set [likelihood ratio] as the unknown value, but we know that the prevalence is 40%, and the desired confidence is, say, 90%.

First we convert prevalence and desired confidence into odds.

Desired confidence as chances = 90%

Desired confidence as odds = 90 : 10 = 90 / 10

Prevalence as chances = 40%

Prevalence as odds = 40 : 60 = 40 / 60

Substituting these odds into the equation provides

[90 / 10] = [40 / 60]. [LR]

[LR] = [90 / 10] / [40 / 60] = 13.5

So the physician, who wants 90% diagnostic confidence rather than the 73% provided by Test 1, would need a test with a likelihood ratio of 13.5. Although such likelihood ratios can be found in other realms of Medicine, they are virtually unheard of in Musculoskeletal Medicine or Pain Medicine.

Test 2 has the same specificity as does Test 1 but has a lower sensitivity, and a lower likelihood ratio. Against a prevalence of 40% it provides a diagnostic confidence of 67%. This is good enough for legal purposes and for selecting conservative therapy, but is probably too low for undertaking surgery.

Test 3 has the same sensitivity as Test 1 but a lower specificity, and a lower likelihood ratio of 2.0. It generates a diagnostic confidence of only 57%, which might not be appealing for any clinical application, because chances of 57% are little better than guessing.

Test 4 has a likelihood ratio of 1.0. It generates a diagnostic confidence of 40%; but this value is the same as the prevalence of the condition being diagnosed. These figures show that tests with likelihood ratios of 1.0, or near to 1.0, are clinically useless. They produce no gain in diagnostic confidence. The physician would be just as confident in their diagnosis if they guessed it based on the prevalence of the condition, and would be no wiser for having applied the test.

Test 5 has a likelihood ratio less than 1.0. It generates a diagnostic confidence of 34%, which is less than the prevalence of the condition being diagnosed. If a physician used such a test they would be less informed after using it than they were before using it. To underscore this point, using the test makes the physician more ignorant. Just knowing the prevalence of the condition is more informative than the result of using that diagnostic test.

The summary messages from these examples are:

- tests with likelihood ratios less than 1.0 are misinformative, and have no place in clinical practice.
- tests with a likelihood ratio of 1.0, or near to 1.0, are not informative. Guessing the diagnosis based on its prevalence rate is faster, and just as accurate.
- tests with likelihood ratios up to 3.0 are of questionable or limited value.
- tests with likelihood ratios greater than 3.0 are worth considering, but
- the final grounds for assessing a test are the prevalence of the condition being diagnosed and the diagnostic confidence that the test produces. The physician needs to judge if that diagnostic confidence is enough for the decision that they are about to make.

6. Examples

Looking at some examples from the literature can show us just how good (useful) commonly used tests are.

Table 4 shows the pooled data from the literature on the validity of straight leg raise for the diagnosis of lumbar disc herniation. The data indicate that straight leg raise is very sensitive for detecting disc hernations but has a very low specificity, which means that it is next to useless for distinguishing disc herniations from other conditions that cause radiculopathy, such as stenosis. This lack of discrimination is reflected in the likelihood ratio of 1.1, which tells us that straight leg raise has virtually no discriminative power. It might confirm radicular pain, but it does not determine the cause.

Table 5 shows the data on the validity of degenerative changes being the cause of back pain. The sensitivity is low, and the likelihood ratio is 1.3. This tells us that looking for degenerative chances in the diagnosis of back pain is uninformative. This arises because degenerative changes are almost as common in asymptomatic subjects as they were in subjects with

Table 3
The properties of five diagnostic tests, with different sensitivities and specificities all used to test for the same condition that has a prevalence of 40%.

Properties	Test 1	Test 2	Test 3	Test 4	Test 5
Sensitivity	0.80	0.60	0.80	0.60	0.40
Specificity	0.80	0.80	0.60	0.40	0.50
+Likelihood Ratio	4.0	3.0	2.0	1.0	0.8
Prevalence	40%	40%	40%	40%	40%
Diagnostic Confidence	73%	67%	57%	40%	34%

Table 4
A contingency table for determining the construct validity of a straight leg raise for the diagnosis of lumbar disc hernation. The data shown are pooled data from five studies in the literature [13].

		Disc Herniation	
		Present	Absent
Straight Leg Raise	**Positive**	2324	55
	Negative	93	63
Sensitivity		0.96	
Specificity			0.15
Likelihood Ratio		1.1	

Table 5
A contingency table for determining the construct validity of finding degenerative changes on plain radiographs as the cause of back pain. The data shown are pooled data from six studies in the literature [14].

		Back Pain	
		Present	Absent
Degenerative Changes	**Present**	1422	943
	Absent	1113	1285
Sensitivity		0.56	
Specificity			0.58
Likelihood ratio		1.3	

back pain. So the magnitude of the likelihood ratio should be of no surprise.

Table 6 summarises the data from studies that investigated if high-intensity zones (HIZ) were diagnostic of the affected disc being painful. The studies reported a range of sensitivities, specificities, and likelihood ratios, probably because of differences in imaging protocols and the criteria used for distinguishing HIZ from latent fissures in the disc. The pooled data accommodate these differences and provide a representative likelihood ratio of 3.8. For a condition that has a prevalence of about 40%, a likelihood ratio of 3.8 provides a diagnostic confidence of 72%, which is quite reasonable and useful.

7. Confidence intervals

Because the positive likelihood ratio involves proportions it is subject to the potential errors inherent in calculating proportions. Different raw data may generate ratios larger or smaller than the true, or generalizable, value of the ratio. The range of this variation can be determined by calculating the 95% confidence intervals of the ratio. The formula for these intervals is more complex than the formula for estimating the confidence intervals of a proportion, because the likelihood ratio is calculated from a collection of proportions and ratios [16].

$$+LR = \exp\left[\ln\left[\frac{SEN}{[1-SPEC]}\right] \pm 1.96\sqrt{\frac{[1-SEN]}{a} + \frac{SPEC}{b}}\right]$$

Calculating these confidence intervals provides two sets of information for physicians. Firstly, it tells them the range of possible values that they might encounter if they adopted a particular diagnostic test, instead of the value reported by a study. Secondly, if the confidence intervals are too wide, the physician can conclude that the study was underpowered. Its sample size was too small to provide an informative estimate of what the true value of the likelihood ratio is. Table 7 illustrates an example.

Assessment for centralisation has good sensitivity and good specificity that generate a likelihood ratio of 2.5. Given that lumbar discogenic pain has a prevalence of about 40%, such a likelihood ratio generates a diagnostic confidence of 63%, which is possibly acceptable for pursuing conservative therapies.

However, in this study, the confidence interval of this likelihood ratio is 1.4–4.4. Consequently, the true likelihood ratio could be as small as half the reported value or nearly twice that value, and the diagnostic confidence ranges from 48% to 74%. This variance arises because the study sample was small. A physician could be forgiven for not being impressed by this study. A larger study would be required to determine if the true likelihood ratio is substantially smaller or larger than one reported, or approximately of the same magnitude.

Table 6
The results of 12 twelve studies of the validity of high-intensity zone on magnetic resonance imaging being diagnostic of the affected disc being painful. Data based on Bogduk et al. [15].

Sample Size	Sensitivity	Specificity	Likelihood Ratio	95% CI
142	0.37	1.00	∞	
120	0.82	0.89	7.5	4.0–14.1
256	0.45	0.94	7.5	3.7–15.1
152	0.27	0.95	5.4	1.7–17.1
101	0.52	0.90	5.2	2.4–11.2
155	0.81	0.79	3.9	2.5–6.0
178	0.57	0.84	3.6	2.2–5.7
109	0.45	0.84	2.8	1.4–5.5
152	0.26	0.90	2.6	1.2–5.8
97	0.56	0.70	1.9	1.2–3.0
116	0.27	0.85	1.8	0.9–3.8
80	0.09	0.93	1.3	0.3–5.4
Combined =1658	0.45	0.88	3.8	3.1–4.5

Table 7
A contingency table for determining the construct validity of McKenzie assessment for diagnosing internal disc disruption, the criterion standard being provocation discography. The data shown are from Donleson et al. [17].

		Discogenic Pain	
		Yes	No
Centralisation	**Yes**	21	10
	No	8	24
Sensitivity		0.72	
Specificity			0.71
Likelihood ratio		2.5	
95% confidence interval		1.4–4.4	

8. Relevance

The principles of construct validity are pertinent to all diagnostic tests in medical practice. If the sensitivity and specificity of a test are not known, physicians are operating in the dark; they do not know if the test works well enough or not.

In the field of interventional pain medicine, validity is particularly pertinent to understanding the diagnostic power and utility of diagnostic blocks. The physician should know what the chances are that the positive response that they encounter is correct and not a false-positive. There is no universal answer to this dilemma, because the utility of a diagnostic block changes with the prevalence of the condition being diagnosed.

A placebo-controlled study has shown that the sensitivity and specificity of cervical medial branch blocks are different for different definitions of a positive response [18]. For concordant responses, meaning complete relief of pain lasting longer when bupivacaine was used than when lidocaine was used, the sensitivity was 0.54 and the specificity was 0.88 generating a positive likelihood ratio of 4.5. For discordant responses, meaning lignocaine lasting longer than bupivacaine, the respective figures were 1.00, 0.65, and 2.9. Table 8 shows how diagnostic confidence differs according to the likelihood ratio and according to prevalence.

Diagnostic confidence is less for discordant responses because of their lesser likelihood ratio. However, although the likelihood ratio for discordant responses is nearly half the size of that for concordant

Table 8
The diagnostic confidence obtained from concordant and discordant responses to diagnostic blocks for different prevalences of the condition diagnosed.

Prevalence	Positive Response	
	Concordant	Discordant
	LR = 4.5	LR = 2.9
80%	95%	92%
70%	91%	87%
60%	87%	81%
50%	82%	74%
40%	75%	66%
30%	66%	55%
20%	53%	42%
10%	33%	24%
	Diagnostic Confidence	

responses, the diagnostic confidences are only about 10% less. For clinical purposes, this could be tolerated, particularly when discordant responses have a greater sensitivity, and will detect a greater number of positive cases.

The effect of prevalence is less tolerable. When the condition being diagnosed is common, for example with a prevalence of 60%, discordant responses and concordant responses both produce diagnostic confidence levels above 80%. As the prevalence drops, however, diagnostic confidence plummets. With a prevalence of 20% diagnostic confidence is 50% or less, meaning that a positive diagnosis will be wrong in half the cases. For a prevalence of 10%, diagnostic confidence will be 33% or 24%, meaning that there will be three or four incorrect diagnoses for every correct one.

Cervical zygapophysial joint pain is common, with a prevalence of about 60% [19–21]. Under those conditions, cervical medial branch blocks will have a validity of greater than 80%, irrespective of concordant or discordant responses.

Lumbar zygapophysial joint pain is uncommon, with a prevalence of about 15% [22]. Under those conditions the validity of comparative diagnostic blocks is low, with two or more diagnoses being wrong for every one that is correct.

Informed of these data, physicians can choose what to do in order to improve the validity of their lumbar medial branch blocks. They could adopt fully randomised, placebo-controlled blocks [5]. They could compromise and adopt fully randomised comparative blocks [5]. Performing blocks without improving their validity amounts to polluting the practice of interventional pain medicine with diagnostic noise and consequent therapeutic noise.

Disclosures

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

Appendix 1

Derivation of the positive likelihood ratio

In most textbooks of statistics or clinical epidemiology, and on most websites, the positive likelihood ratio is introduced as some mystical entity that can be learned and used. Explanations of what it actually is, where it came from, and what it means, are hard to come by. This Appendix takes readers through a journey of discovery.

Readers who use a diagnostic test should be concerned about how often a positive result, in their hands, is true-positive, or how often it is contaminated by false-positive results. Seeking to find out, they might consult the results of a hypothetical study as shown in Table Appendix 1.

Table Appendix 1
A contingency table for determining the construct validity of a diagnostic test.

		Criterion Standard	
		Present	Absent
Diagnostic Test	**Present**	a	b
	Absent	c	d

Readers might be tempted to use the first row of such a table to address their concerns. That row shows that, in the study, the diagnostic test was positive in [a+b] cases, and was correct in [a] of these cases. Surely, the ratio [a/(a + b)] reflects the accuracy of the test. Unfortunately, this is not the case. The ratio [a/(a + b)] is known as the positive predictive value of the test, but it applies only to the particular sample that was reported in the study. It cannot be generalised to all situations.

The sensitivity and specificity of a test are defined by the columns of the contingency table. For practical purposes, the sensitivity and specificity can be regarded as fixed, intrinsic properties of the test. They will be the same in any and all samples whenever the test is applied.

Although not immediately apparent, the same does not apply to the rows of the contingency table. It can be shown that the numbers in the rows, and the ratios between them, change as the prevalence of the condition being diagnosed changes. Figure Appendix 1 illustrates this phenomenon.

In each of the contingency tables in Figure Appendix 1 the diagnostic test has the same sensitivity and specificity, but the tables differ in the prevalence of the condition being diagnosed. Across the three tables, prevalence increases from 50% to 80%. At the same time, the positive predictive values increase, from 73% to 91%.

	Table 1			Table 2			Table 3		
		Criterion Standard			Criterion Standard			Criterion Standard	
		Pos	Neg		Pos	Neg		Pos	Neg
Diagnostic	**Pos**	40	15	Pos	48	12	Pos	64	6
Test	**Neg**	10	35	Neg	12	28	Neg	16	14
Sensitivity		0.80			0.80			0.80	
Specificity		0.70			0.70			0.70	
Prevalence		50%			60%			80%	
PPV		73%			80%			91%	
+Likelihood Ratio		2.7			2.7			2.7	

Fig. Appendix 1. Three tables with the same sensitivity and specificity but increasing prevalence of the condition being diagnosed. The positive predictive values (PPV) increase as the prevalence increases, but the positive likelihood ratio remains constant. **Pos:** positive. **Neg:** negative.

These data show that the positive predictive value is not constant. It increases as prevalence increases. The explanation for this phenomenon is simply that a diagnostic test is more likely to be correctly positive in samples dominated by the condition being present.

Perhaps intriguingly, across all three tables the positive likelihood ratio remains constant. We can now explore why this is so, and if it is of any use to us.

The intuition that perhaps the ratio [a/(a + b)] could be useful is not entirely wrong. Disheartened by our experience with Figure Appendix 1, we can ask a more specific question. Is there a positive predictive value of the test that applies regardless of the prevalence of the condition, or perhaps might be corrected for prevalence? There is such an entity, but to discover it requires a colourful exercise in algebra.

To cut a long story short, if we were to pursue this algebra we would find there is an expression for [a/(a + b)] but it is somewhat complicated in appearance, and difficult to relate directly to clinically useful concepts. However, an elegant and meaningful expression can be derived if temporarily we stop thinking in terms of chances and, instead, think in terms of odds. This requires that, as a prelude, we reprise the definition and properties of odds.

Remember that chances are by convention expressed as proportions of 1, e.g. 0.60. These chances can be converted to the colloquially more familiar idiom of "percentage chances" by multiplying the chances by 100, e.g. 0.60 × 100 = 60%.

If m = the chances that an event occurs, and.

If n = the chances that an event does not occur

$$the\ Odds\ that\ an\ event\ will\ occur = \frac{the\ chances\ that\ an\ event\ will\ occur}{the\ chances\ that\ an\ event\ will\ not\ occur}$$

$$the\ Odds\ that\ an\ event\ will\ occur = \frac{m}{n}$$

Some additional properties and useful expressions of odds are:

$$[the\ chances\ that\ an\ event\ will\ occur] + [chances\ that\ an\ event\ will\ not\ occur] = 1$$

So,

$$m + n = 1$$

and

$$1 = m + n \qquad \text{Eq. 1}$$

If the chances that an event will occur = m. $= \frac{m}{1}$

Using Equation (1) creates

$$the\ chances\ that\ an\ event\ will\ occur = \frac{m}{(m+n)}$$

Likewise,

$$the\ chances\ that\ an\ event\ will\ not\ occur = \frac{n}{1}$$

$$the\ chances\ that\ an\ event\ will\ not\ occur = \frac{n}{(m+n)}$$

$$Since \quad \frac{(m+n)}{(m+n)} = 1$$

$$\frac{m}{(m+n)} + \frac{n}{(m+n)} = 1$$

$$\frac{n}{(m+n)} = 1 - \frac{m}{(m+n)}$$

$$[chances\ that\ an\ event\ will\ not\ occur] = 1 - [chances\ that\ an\ event\ will\ occur]$$

If

$$Odds\ that\ an\ event\ will\ occur = \frac{chances\ that\ an\ event\ will\ occur}{chances\ that\ an\ event\ will\ not\ occur} \qquad \text{Eq. 2}$$

$$Odds\ that\ an\ event\ will\ occur = \frac{[chances\ that\ an\ event\ will\ occur]}{1 - [chances\ that\ an\ event\ will\ occur]}$$

When required, Odds can be converted back to chances using the following relationships.

$$Odds\ that\ an\ event\ will\ occur = \frac{m}{n}$$

$$Chances\ that\ an\ event\ will\ occur = \frac{m}{(m+n)} \quad or \quad 100\left[\frac{m}{(m+n)}\right]\%$$

So, if the numerical values for "m" and "n" are known from the Odds, they can be substituted into the expression for chances.
More formally,

$$Chances = \frac{[Odds]}{[Odds]+1}$$

We can now resume our pursuit of an expression for an expression of [a/(a + b)] corrected for prevalence.
For the data in Table Appendix 1 we know that the sensitivity of the test (SEN) is:

$$SEN = \frac{a}{[a+c]}$$

From which we can calculate

$$a = (a+c)\ .\ SEN \qquad \text{Eq. 3}$$

This gives us a handle on "a" but we need to eliminate "c".
By definition, the prevalence (PREV) of the condition being diagnosed is:

$$PREV = \frac{[a+c]}{N}$$

From which we can state

$$[a+c] = N\ .\ [PREV] \qquad \text{Eq. 4}$$

By combining equations (3) and (4), we find:

$$a = N.[PREV].[SEN]$$

Now that we have a handle on [a] we can look for a handle on [b].
For the data in Table Appendix 1 we know that the specificity of the test (SPEC) is:

$$SPEC = \frac{d}{[b+d]}$$

$$\frac{b+d}{b+d} = 1$$

$$\frac{b}{b+d} = 1 - \frac{d}{b+d} \qquad \text{Eq. 5}$$

$$\frac{b}{b+d} = 1 - SPEC$$

$$b = [b+d]\ .\ [1 - SPEC]$$

Equation (4) gives us a handle on "b" but is contaminated by "d". We can deal with this.

$$N = [a + b + c + d]$$
$$[b + d] = N - [a + c]$$
$$From\ \ Equation\ \ 7$$
$$[a + c] = [N].[PREV] \qquad \text{Eq.6}$$
$$So,$$
$$[b + d] = N - [N].[PREV]$$
$$[b + d] = [N].[1 - PREV]$$

By substituting Equation (6) into Equation 5
b = N.[1-PREV].[1-SPEC].
We now have expressions for [a] and for [b], which we can use to calculate [a/b].

$$\frac{a}{b} = \frac{N.[PREV][SEN]}{N.[1 - PREV][1 - [SPEC]}$$

Cancelling "N" gives us.

$$\left[\frac{a}{b}\right] = \left[\frac{PREV}{1 - PREV}\right] \cdot \left[\frac{SEN}{1 - SPEC}\right] \qquad \text{Eq. 7}$$

This expression gives us a neat relationship between [a] and [b] and the known values of prevalence, sensitivity, and specificity. The expression contains three entities, each of which bears interpretation.

If we realise that

$$\frac{a}{(a + b)} = chances\ \ that\ \ a\ \ positive\ \ result\ \ is\ \ true\ \ positive$$

$$\left[\frac{a}{b}\right] = Odds\ \ that\ \ a\ \ positive\ \ result\ \ is\ \ true\ \ positive$$

From equation (2) we can recognise that

$$\left[\frac{PREV}{1 - PREV}\right] = \text{Oddsthattheconditionbeingdiagnosedispresent}$$

The third entity requires closer inspection. It contains the sensitivity and the specificity of the diagnostic test and, therefore, is a measure of the properties of the test. More specifically,

$$Since$$
$$SEN = \frac{a}{(a + c)} = True\ \ Positive\ \ Rate$$
$$[1 - SPEC] = \frac{b}{(b + d)} = False\ \ Positive\ \ Rate$$
$$\left[\frac{SEN}{1 - SPEC}\right] = \frac{True\ \ Positive\ \ Rate}{False\ \ Positive\ \ Rate}$$

So, the third entity is the ratio between rate at which the diagnostic test generates true-positive results and the rate at which it generates false-positive results. This ratio tells us the likelihood with which the test generates true-positive results for every false-positive result. In qualitative terms, this reflects the power of the test to discriminate between true-positive and false-positive results.

Equation (7) can now be translated into words as meaning:

ODDS that positive result is true	=	ODDS condition is present	X	Positive LIKELIHOOD RATIO

So,

- if we calculate the likelihood ratio of the test from its sensitivity and specificity, and
- if we express the prevalence as odds, not chances,
- we can calculate the odds that a given positive results is true-positive.

Once we have the odds of a positive test being true, we can convert those to the chances of it being true positive.

$$Odds = m\ \ :\ \ n$$
$$Chances = 100\ \ .\ \ \left[\frac{m}{(m + n)}\right]\ \ \%$$

These chances constitute the diagnostic confidence that the physician can have that a given positive result is true positive.

For example,

if sensitivity = 0.80, specificity = 0.60, and prevalence is 40%,
the true-positive test odds = [40/60]. [0.80/(1–0.60)] = 80/60.
and
the chances of the test being true-positive = 80/(80 + 60) = 57%,
and diagnostic confidence = 57%.

In summary, the positive likelihood ratio is an entity that fortuitously arises when we perform the algebra to find the odds of a positive result being true-positive. The algebra shows that likelihood ratio acts as a coefficient that adjusts the positive predictive value of the test for the prevalence of the condition being diagnosed.

Footnote

[This Appendix has explicitly focussed on the derivation of the positive likelihood ratio. A similar process can be used to derive what is known as the negative likelihood ratio, which indicates the power of a diagnostic test to rule out the condition being diagnosed; but we do not need to complicate the present message by going through that process. Negative likelihood ratios are of relevance to Radiologists, who perform scans in order to rule out conditions, such as breast cancer, but negative likelihood ratios are not of immediate relevance to the practice of interventional pain medicine.]

References

[1] Bogduk N, Aprill C, Derby R. Lumbar discogenic pain: state-of-the-art review. Pain Med 2013;14:813–36.
[2] Dwyer A, Aprill C, Bogduk N. Cervical zygapophyseal joint pain patterns I: a study in normal volunteers. Spine 1990;15:453–7.
[3] Fukui S, Ohseto K, Shiotani M, Ohno K, Karasawa H, Nagaauma Y, Yuda Y. Referred pain distribution of the cervical zygapophyseal joints and cervical dorsal rami. Pain 1996;68:79–83.
[4] Engel A, MacVicar J, Bogduk N. A philosophical foundation for diagnostic blocks, with criteria for their validation. Pain Med 2014;15:998–1006.
[5] Engel AJ, Bogduk N. Mathematical validation and credibility of diagnostic blocks for spinal pain. Pain Med 2016;17:1821–8.
[6] Nesham-West R, Mercer S. Ober's test – is it a valid test of iliotibial band tightness? N Z J Sports Med 1999;27:35–6.
[7] Mercer SR, Rivett DA, Nelson RA. Stretching the iliotibial band: an anatomical perspective. N Z J Physiother 1998;26:5–7.
[8] Kaplan M, Dreyfuss P, Halbrook B, et al. The ability of lumbar medial branch blocks to anesthetize the zygapophysial joint. Spine 1998;23:1847–52.
[9] Dreyfuss P, Henning T, Malladi N, et al. The ability of multi-site, multi-depth sacral lateral branch blocks to anesthetize the sacroiliac joint complex. Pain Med 2009;10:679–88.
[10] Dwyer A, Aprill C, Bogduk N. Cervical zygapophyseal joint pain patterns I: a study in normal volunteers. Spine 1990;15:453–7.
[11] Dreyfuss P, Michaelsen M, Fletcher D. Atlanto-occipital and lateral atlanto-axial joint pain patterns. Spine 1994;19:1125–31.
[12] Fortin JD, Dwyer AP, West S, et al. Sacroiliac joint: pain referral maps upon applying a new injection/arthrography technique. Part I. Asymptomatic volunteers. Spine 1994;19:1475–82.
[13] Bogduk N, Govind J. Medical management of acute lumbar radicular pain. Newcastle, Australia: Newcaste Bone and Joint Institute; 1999.
[14] Bogduk N. Degenerative joint disease of the spine. Radiol Clin 2012;50:613–28.
[15] Bogduk N, Aprill C, Derby R. Lumbar discogenic pain: state-of-the-art review. Pain Med 2013;14:813–36.
[16] Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. J Clin Epidemiol 1991;44:763–70.
[17] Donelson R, Aprill C, Medcalf R, Grant W. A prospective study of centralization of lumbar and referred pain. Spine 1997;33:1115–22.
[18] Lord SM, Barnsley L, Bogduk N. The utility of comparative local anaesthetic blocks versus placebo-controlled blocks for the diagnosis of cervical zygapophysial joint pain. Clin J Pain 1995;11:208–13.
[19] Barnsley L, Lord SM, Wallis BJ, Bogduk N. The prevalence of chronic cervical zygapophysial joint pain after whiplash. Spine 1995;20:20–6.
[20] Lord S, Barnsley L, Wallis BJ, Bogduk N. Chronic cervical zygapophysial joint pain after whiplash: a placebo-controlled prevalence study. Spine 1996;21:1737–45.
[21] Yin W, Bogduk N. The nature of neck pain in a private pain clinic in the United States. Pain Med 2008;9:196–203.
[22] MacVicar J, MacVicar AM, Bogduk N. The prevalence of “pure” lumbar zygapophysial joint pain in patients with chronic low back pain. Pain Med 2021;22:41–8.