Science for the Clinician

Jim Meadows

Contents

[Introduction 2](#_Toc227562245)

[Science 2](#_Toc227562246)

[The Scientific Method 2](#_Toc227562247)

[Levels of Evidence 2](#_Toc227562248)

[Clinical Use of Validity Numbers 2](#_Toc227562249)

[Clinical Reasoning in Pathoanatomical Diagnosis and Treatment 2](#_Toc227562250)

[The Pathoanatomical Diagnosis 2](#_Toc227562251)

[Clinical Reasoning 2](#_Toc227562252)

[Illness Scripts 2](#_Toc227562253)

[Making the Diagnosis 2](#_Toc227562254)

[Summary 2](#_Toc227562255)

# Introduction

Evidence comes in all shapes and sizes with some types better regarded than others and this (point) seems to be mainly based on taste. Of course, as with taste, if somebody doesn’t agree with yours then they are either vulgar, uneducated, simple-minded, naïve, ill-advised or unfortunate, but they are certainly without taste and unfortunate. There is generally this division between the thoughtful clinician and the knee-jerk researcher and much less of a division between the thoughtful clinician and the thoughtful researcher. Both the latter groups (and by the way I am not saying there are no thoughtless clinicians, but both these groups of people are really not worth discussing) are prepared to accept that all forms of evidence have their own usefulness and what we should be discussing is not whether research evidence is better than non-research evidence but how useful each is based on current levels of knowledge and even the individual case the evidence is going to be used for.

Research based evidence is primarily that of criterion validity where a gold standard is used to measure against the outcomes of a physical experiment that is where something is manipulated in a deliberate and determined manner to see what happens and that which happens is then compared to the outcome using a tool accepted as the best. There are variations of this such as concurrent validity, predictive validity, and prescriptive validity (some of these will be discussed in more detail later). For the clinician criterion validity is often of little use or are confusing for a number of reasons such as irrelevant topics, poor methodology, poor external validity, low power, conclusions that do not match the results, and vested interests swaying the research direction or conclusion (the Quebec Task Force report is an example of this). Often the best evidence for the clinician is face, content, and construct validity.

Research evidence is usually considered best evidence not because of its truthfulness or validity but because it can be objectively measured but this alone does not mean that it is more helpful to the conclusion. It should also be noted that the so-called lower forms of evidence listed above have their own problems possibly the biggest being that they are open to interpretation to a much greater extent than criterion forms of evidence. It should be noted that ironically for the researcher there is no evidence that research-based evidence is better than nonresearched based evidence but only a presumption that this is so and this presumption is being challenged by mounting evidence and opinion. What should be considered is not one form of evidence over the other but our profession’s body of knowledge, that is all forms of evidence that are based on our model; anatomy, pathology, physiology, biomechanics, observation, research of previous observations, reasoning intervention, analysis and conclusion. This list includes both thoughtful clinical practice and thoughtful academic research.

Each of the most impactful (impactful to the clinician that is) forms of evidence will be discussed as will how they may best be used in clinical practice. Clinical practice must be a meld of both research based and non-research based evidence. Before we do that it might be enlightening to discuss science in a broader sense than that typically be touted as what we do, that is practice using best evidence or doing random controlled trails because science is more than just those things.

# Science

According to Webster’s New Collegiate Dictionary science is “*knowledge attained through study or practice," or "knowledge covering general truths of the operation of general laws, esp. as obtained and tested through scientific method [and] concerned with the physical world*." A definition hugely richer than the impoverished definition that is usually implied when science is discussed in most PT circles when only a small part of the definition is generally considered, the part that mentions scientific method and even then scientific method is often taken to mean experiment.

According to “Science Made Simple” <http://www.sciencemadesimple.com/science-definition.html> *“The term science also refers to the organized body of knowledge people have gained using that system. Less formally, the word science often describes any systematic field of study or the knowledge gained from it.”*

Science can be thought of a systematic method of understanding how reality works. And while experiment plays a large role in coming to this understanding, it is not the sole method of doing so. Anybody who wants to argue this point needs to look at astronomy where any experiment is rare and extremely difficult, to cosmology where strict experiment with variable control is impossible and even physics where the early experiments into were Einstein’s thought experiments and even now the most popular answer to the underpinning of the cosmos is probably string theory which presently cannot even be tested experimentally. So science is larger than just experimenting but what does it encompass for the physical therapist. Well exactly the same as it does for the physicist or biologist or chemist and for a simplified and contained look at the definition of science it is worth checking out Wikipedia on the subject <http://en.wikipedia.org/wiki/Science>.

Even science is not a coherent body, Wikipedia (<http://en.wikipedia.org/wiki/Science#Philosophy_of_science>) reports the following;

*There are different schools of thought in the philosophy of scientific method.* [*Methodological naturalism*](http://en.wikipedia.org/wiki/Methodological_naturalism) *maintains that scientific investigation must adhere to* [*empirical*](http://en.wikipedia.org/wiki/Empirical) *study and independent verification as a process for properly developing and evaluating natural explanations for* [*observable*](http://en.wikipedia.org/wiki/Observation) *phenomena.*[*[19]*](http://en.wikipedia.org/wiki/Science#cite_note-18) *Methodological naturalism, therefore, rejects* [*supernatural*](http://en.wikipedia.org/wiki/Supernatural) *explanations,* [*arguments from authority*](http://en.wikipedia.org/wiki/Appeal_to_authority) *and biased* [*observational studies*](http://en.wikipedia.org/wiki/Observational_studies)*.* [*Critical rationalism*](http://en.wikipedia.org/wiki/Critical_rationalism) *instead holds that unbiased observation is not possible and a demarcation between natural and supernatural explanations is arbitrary; it instead proposes* [*falsifiability*](http://en.wikipedia.org/wiki/Falsifiability) *as the landmark of empirical theories and falsification as the universal empirical method. Critical rationalism argues for the ability of science to increase the scope of testable knowledge, but at the same time against its* [*authority*](http://en.wikipedia.org/wiki/Authority)*, by emphasizing its inherent* [*fallibility*](http://en.wikipedia.org/wiki/Fallibilism)*. It proposes that science should be content with the rational elimination of errors in its theories, not in seeking for their verification (such as claiming certain or probable proof or disproof; both the proposal and falsification of a theory are only of methodological, conjectural, and tentative character in critical rationalism).*[*[20]*](http://en.wikipedia.org/wiki/Science#cite_note-19)[*Instrumentalism*](http://en.wikipedia.org/wiki/Instrumentalism) *rejects the concept of truth and emphasizes merely the utility of theories as instruments for explaining and predicting phenomena.*[*[21]*](http://en.wikipedia.org/wiki/Science#cite_note-20)

It might thereforerequire us to be a little more discerning about our language and ask for which definition of science the person lecturing us is using. At the least if he/she cannot understand our question we can have a good idea that that person is, to be generous, on shaky ground and we should not pay too much attention to what they have to say on the subject.

Does the thoughtful clinician do science, of course. Does he/she do the same science as our researchers yes within the confines of the above discussion. Does he/she do it in the same manner as the researcher, of course not. The clinician’s objective is to help the patient get better while the main objective of the researcher should be the discovery of some new paradigm, principle, theory law, technique that will help the clinician better treat the patient. The main difference between the researcher and the clinician in the practice of science is that the researcher often wants to control most if not all variables not being manipulated and needs to make detailed notes on all aspects of the experiment, the clinician does not. There are differences in reporting the results of the treatment/experiment but these are minor and it cannot be argued that the methodological and thoughtful clinician is practicing science on an hourly basis.

The one thing that we can all agree on is that the researchers need to seek out the best evidence as to what works and the clinician needs to use the best available evidence. Most times this evidence will not be the result of experiment because the experiment has not yet been done and in a few instances it will be experimental evidence because not only has the experiment been done but it has been found to be useful to the clinician. The researcher must understand that the final consumer of their work is the clinician and so not only must the research be intrinsically sound but there must be good external validity and relevance.

On the other hand often the clinician is not the most discerning of consumers. Authority is all too frequently substituted for rationality and the critical acquisition of information. Things are accepted on the basis of who says it and what is fashionable (this criticism may also be leveled at researchers although not to the same degree) rather than what evidence is present for the practice and often a short search of the literature and sometimes even into the anatomy book will dispel the vomitus that is being spewed out.

To illustrate the dangers of authority, here are four quotes from Lord Kelvin (1824-1907) one of the great figures in physics.

* “There is nothing more to be discovered in physics, all that remains is more precise measurement.”
* “X-rays will prove to be a hoax”
* “I accept no theory of gravitation. Present science has no right to attempt explain gravity. We know nothing about it. We know simply nothing about it.”
* “Radio has no future.”

In addition to being a physicist this man was also a consultant to large corporations such as ATT and gave them such advice as not too worry about the wireless telegraph. Authority!

While authority, even when it is benign, is not the evidence that we should be using, there is nothing wrong with experience. Sometimes the statement “this is what I do and it seems to work” is better than “I have no idea what to tell you”, and it at least gives the seeker of knowledge a starting point in his/her search for clinical enlightenment. But even this should be based on a construct of the clinical sciences; otherwise, it is debatable whether this just wishful thinking, phantasy or the realm of the paranormal. Of course, there is always empiricism and trying things because you have run out of rational things to do is always an option as long as the risk to the patient is low.

While research evidence may not necessarily generate everything from assessment tools to treatments, some form of rational and scientific bedrock is required if we are going to practice in a sensible manner. We practice, or at least should practice, on this bedrock, which is our construct or matrix and, as already stated, consists of the clinical sciences. Almost everything we do and think must come from this construct and our practice and even our research must arise from it. Other constructs may be faith-healing, psychic surgery, aura therapy, healing touch, and others which while some may claim their efficacy have no place in our practice unless a reasonable scientific model based on our bedrock can be developed.

The clinical aspect of physical therapy must therefore incorporate the entire body of knowledge of our professsion including all forms of scientific evidence from opinion to criterion validity which when taken as a whole provides construct validity for what we do.

## The Scientific Method

Scientific method is not random controlled trials or single case study designs or any of the other research methodology; rather it is all of them including what you do with your patients every visit. It is the verification or falsification of an idea formed while interacting with your relevant environment, in your case the patient. For other scientists the environment may be the universe, a pond, a wolf-pack or in extreme cases the subjects of a scientific experiment. By making observations the scientist can generate a framework to explain these observations, this framework can then be formalized into a hypothesis which is then tested by making other observations or usually but not always by making an intervention and seeing if the results of the intervention strengthen or weaken a hypothesis. I think we are all familiar with this process as it applies to the researcher but maybe we do not really appreciate that it equally applies to the clinician. We look at the patient, ask questions and carry out an objective examination and as we go along we generate one, two or a series of hypotheses based on the incoming information from the observation, history and objective examination. The original hypothesis may be based on pattern recognition and prevalence of the condition, but further information may force the clinician to change the hypothesis for something more in keeping with this information. Then this is tested until, by the end of the examination, we have a diagnosis rather than a hypothesis although strictly speaking, the diagnosis is a hypothesis until it is successfully tested by the treatment deemed most suitable for this hypothesis.

A screenshot of a cell phone

Description automatically generated

The research essentially does no more than this other than use on or another methodologies to keep the hypothesis testing as objective as possible to prevent tester errors such as bias and mishearing or miss-seeing results and keep different records of the interaction as they will be used for different purposes. But the essential process is the same for both groups and each recognize the validity of the other.

It is worthwhile at this point to discuss some pivotal terminology and its place in physical therapy.

Theory

This has two major meanings, the lay-person’s and the scientist’s. The lay meaning of the term is an idea that is shaky at best and downright silly at worse. Statements such as “he has a theory that the earth is flat” or I have a theory that if I bet on the number 8 horse on the 8th day of the month, I’ll win eight times what I bet” are both of this type. It is most certainly not the scientific meaning of the word. This leads to opponents of the theory to say “well it’s only a theory” as if it had no basis. But this is exactly similar to say that Newton’s theory of gravity is only a good idea that is rationally debatable.

The scientific meaning of the term is that it is the over-arching concept that includes many other concepts and laws and that it has been successfully tested many, many times and has never once been shown to fail any test. Such theories include plate tectonics, cell theory, atomic theory, relativity, quantum mechanics, gravity and evolution. The scientific theory is the ultimate statement of how things are and how they work and are only accepted as theories after intensely hard work, long times and much testing and are usually based on mathematics or at least use mathematics in their formulation.

"...facts and theories are different things, not rungs in a hierarchy of increasing certainty. Facts are the world's data. Theories are structures of ideas that explain and interpret facts” Stephen Jay Gould.

In addition, a theory must be falsifiable, that is there must be a way of testing the theory to disprove it so all theories are conditional but the odds of something like relativity, quantum mechanics and evolution being disproved are extremely low. We really do not have theories in physical therapy or even in medicine even though many medical treatments have been established by experiment as sound they have not as yet reached a number or a broadness that an over-arching statement of how medicine works can be made. Germ theory is defined as The [theory](https://en.wiktionary.org/wiki/theory" \o "theory) (coherent scientific explanation of an observable fact) that some [diseases](https://en.wiktionary.org/wiki/disease" \o "disease) are caused by [microorganisms](https://en.wiktionary.org/wiki/microorganism" \o "microorganism). It’s not that the theory is on shaky grounds but the word ‘some’ doesn’t do it any favors as to rigor.

Theories subsume laws and other theories. Newton’s theory of gravity contained his Laws of Gravity and for about three and a half centuries was unchallenged then Einstein’s general theory of relativity swallowed Newton’s theory and made his Law more precise and more general to the extremely large and fast. It was not that Newton’s theory was wrong it just didn’t go far enough and incorporated less conditions than did Einstein’s.

Model

In lieu of a theory, we have a model or more correctly many models. In the hard sciences such as physics (even chemistry has its detractors as a hard science and has been described as mere stamp collecting by Kelvin) a model is a method of simplifying a complex system so as to work efficiently with it. A model’s function is to explain facts (observations) and to predict the result of an intervention. It is not intended to be the truth and, in fact, almost never is, but nonetheless, it is an essential scientific tool. To give an example, in physics relativity together with quantum mechanics is the most proven theories there is having been tested to 13 or 14 decimal places without ever once looking shaky. Yet, exactly how each works is unknown and may even be unknowable, but some working framework is needed to conceptualize them. What is really interesting is that there are at least two contradictory models for gravity, one involving the curvature of space by massive bodies and the other the interaction of particles to draw the bodies together, both work well at predicting behavior and each is used when to do so is optimal. For us, a model replaces the theory that we cannot generate and gives us such a framework to predict what our examinations will find and what our treatments will do. A good model will explain all of the observations and facts, a bad one will explain only a few and a fantasy will explain none. One strength of using formal models is that when new information comes in that tends to weaken the model, it can be modified to explain the new facts or if the model is so badly weakened, it can be rejected and replaced with one that does better without the profession losing credibility.

So in order to do science in physical therapy the following procedure must be followed.

1. The acquisition of the profession’s body of knowledge (or at least a reasonable chunk of it)
2. The observation of the facts of interest
3. The generation of a hypothesis
4. The testing of the hypothesis
5. The acceptance or rejection of the hypothesis
6. The integration of that hypothesis into the model

For the clinician this means:

1. Knowing enough about our profession that we can understand our way around the human body and how it functions and dysfunctions
2. Observing, talking to and objectively examining the patient appropriately
3. Making a diagnosis and treatment plan
4. Carrying out the treatment
5. Recognizing the correct hypothesis when the treatment is successful and changing it when not
6. Using this information the next time a similar patient attends and taking the opportunity for further testing of the hypothesis

## Levels of Evidence

The hierarchy of evidence is not really based on rational determinants but on pragmatic ones and as such the terms stronger and weaker forms of evidence should not really be used especially as there is no evidence for one form over another. But the convention has been set by the researchers and the academics and is based on their bias that measurability and objectivity is all important. While measurablility may be an attainable goal, it is doubtful that objectivity will ever be fully realized as there is reason to believe that even in the hard sciences bias is a factor. In medicine for there have been many studies that have looked at the publication of so-called scientific articles only to find that the conclusions did not match the results but more usually matched some other criterion probably the researcher’s preferred result. (THIS IS CONFUSING) But it will be instructive to look at the levels of evidence and how each side of the profession perceives them.

Face Validity

This is how much the instrument (we are going to confine ourselves for the moment to the researcher’s perspective) appears to measure the item of interest. Goniometers have face validity because it is obvious that they do what they do. It is considered as among the weaker forms of evidence because it cannot be measured as it is a cognitive and even abstract notion, in effect it is opinion and as such an instrument either has it or it hasn’t. It is challengeable by those with a different opinion and supposedly it is not possible to separate opinion. But of course you can, a panel of experts should be able to come to a group opinion on the instrument and this would presumably be stronger than a single person with less accomplishment, experience and qualification than the panel. But another thought is that experts cannot provide face validity and that it can only be provided by non-expers; I have to admit to not understanding this perspective and it is debated. In short face validity can be summed up as “does it appear to work”.

For the clinician consuming research, face validity may be considered to have another meaning such as how much does the result of this piece of research or even argument agree or disagree with my observations. Certainly when experienced clinicians criticize the results and conclusions of a piece of research as not matching their experience then surely it requires the researchers to have another look at their methodology, statistics, results and conclusions. The clinician’s perspective of face validity is therefore best summed up as “does this make sense to me or not”.

Content Validity

Each profession has a domain or mini-universe within which boundary lay all of the knowledge, practice, “theories”, information in fact everything there is that pertains to that universe. The variable of interest must therefore come from that domain; and therefore to have content validity, the instrument must test everything that pertains to that variable. For the most part content validity is most pertinent to exams, curricula, surveys and tests etc., but it can and does extend beyond that. In principle, to have content validity, the instrument must be free from factors that have nothing to do with the variable being measured. If a test is designed to determine the presence of stability of the shoulder it should not at the same time measure mobility. In addition, factors not associated with variable should not influence the test; for example, in the case of the stability tester, language difficulties between the tester and the subject should not be an issue.

For the clinician, content validity really becomes a case of is this piece of research of any interest to me or not. An example would be the clinical prediction rule concerning the criteria necessary to predict that a lumbar manipulation will be successful. If these criteria were not present, how much difference would it make to my decision to carry out such a manipulation once the decision to do so has been reached? The answer is none because I have no other treatment that is indicated by the diagnosis. It might have content validity if it could be shown to have some bearing on risk management but, currently, it does not and so is of no interest to the clinician.

Construct Validity

This is the ability of an instrument to measure a variable of a concept or even the concept itself, a very difficult thing to do as most if not all concepts are multidimensional as are their variables. In order to do this, the concept is redefined for the purposes of measurement. How well it is redefined is a measure of construct validity. For example pain is about as multidimensional as you can get so the instrument may measure pain behaviors such as activities of daily living, functional movements, physical activity levels, or it may use pain reports, physiological measures, concentration difficulties etc. Providing all of these are within the domain of the researcher, that is they have content validity, and that they have face validity then it can be said that pain is being measured to one extent or another.

Much of what we do as clinicians is based on construct rather than criterion validity. We know the systems we are working with from anatomy, physiology and biomechanics and we know how they become dysfunctional through pathology so we can design a treatment or test based on this knowledge.

Criterion Validity

This really means the same to both the researcher and the clinician; it is how an instrument performs when compared with a “gold standard” of known performance. To make the comparison fair the testing context must be the same. As far as variables that affect the outcome of the study are concerned, the patients must be from similar groups , the testing environment must be the same and the testers similar. The gold standard or criterion must obviously be valid and consistently reliable and not influenced by external phenomenon that have nothing to do with what it is measuring, it must therefore have content, face and construct validity. If done properly, therefore, this test may truly rank as the peak of validity, but it has its own set of pratfalls that can affect it to the point where the same instrument (O’Brien’s test in this example) can be tested against the same criterion using very comparable patient’s and conclude two very different results with one tester saying it is an almost perfect test and the other stating the exact opposite. Remember that both studies used comparable groups and the same criterion . How much faith can you then have in criterion validity regardless of magnificent the concept is in principle.

However, for the precise calculation of ow good a test is criterion validity must be used and the results of studies using this type of validity must not be thrown out without a good deal of thought and research. Nor should they be accepted simply because they have criterion validity and certainly if they disagree with your experience you have to question their face validity and read the entire paper with a critical eye before accepting or rejecting the study. The clinician obviously cannot use the methods of criterion validity in practice. He has not control and he/she cannot fix all variables but the patient’s treatment, but he/she can use information generated from good criterion valid information. The remainder of this section will be concerned using the numbers generated from validity studies in the practice of manual therapy.

# Clinical Use of Validity Numbers

For the clinician evaluating diagnostic tests only two numbers have any real relevance beyond those associated with methodology and determining the value of the research article, they are sensitivity and specificity. But the extension of these tests to likelihood ratios and the use of Bayesian calculations increases their usefulness exponentially.

Sensitivity

Defined as the number that tells the percentage or the proportion of people who are identified correctly as having the condition, which is the true positives.

A sensitivity of 1.0 or 100% means that all of the people with the condition are correctly identified. The number is arrived at by dividing the number of true positives (TP) by the number of true positives plus the number of false negatives (FN).

TP/TP+FN

The higher the sensitivity the more confidence you can have in a positive test and the more confidence you have in ruling out the people with the negative test. In the extreme case of a sensitivity of 100% or 1, all negative tests will be true negative and can be ruled out of having the disease; Hence the acronym SNOUT (SeNsitivity OUT).

Specificity

Defined as the number which tells the percentage or the proportion of people who are identified correctly as not having the condition, that is true negatives.

A specificity of 1 means that all people who do not have the condition are correctly identified by the test. The number is arrived at by dividing the number of true negatives (TN) by the number of true negatives plus the number of false positives (FP).

TN/TN+FP

The higher the sensitivity of test the more confidence you can have in a negative results and the more confidence you can have of ruling in those people with a positive test. In the extreme case of a specificity of 100% or 1.0 all positive tests will be true and so the condition can be included in. Hence the acronym SPIN (Specificity IN).

The problem with sensitivity and specificity numbers is that they are not very intuitively logical. That is simply looking at them does not give much of an indication to how good a test is unless both numbers are very high. For example, a sensitivity and specificity of .8 seems to be high on the scale of 0.1-10, but in fact, they are not very good at all. In addition, these numbers are not usable in the current form in changing the probability that a condition is present as a result of using a test whose sensitivity and specificity are known. For this a number incorporating both sensitivity and specificity is used, this is the likelihood ratio.

Baye’s Theorem

This is a mathematical calculation to change the odds of an event from that was estimated as more information comes in. It requires an estimate of the odds of the patient having a particular condition and this can come from the incidence of that disease, family history etc. For example you have a 37 year old male patient with severe back and leg pain whose onset followed a few hours after heavy lifting and progressed rapidly to the point where sitting was extremely painful. You may be thinking that this is a disc herniation and estimate the chances of this as say 60%, you then find out that walking has become increasingly difficult and that the patient has also started to experience parasthesia in the back of the leg and lateral foot. Now the chances of this being a disc herniation has increased considerably with evidence of radiculopathy and the difficulty in walking to say 90%. This is an example of changing probabilities as more information comes in and it should be noted that for the most part these estimates are subjective.

The new information may be in the form of data coming from the patient’s history or from an objective test say the straight leg raise. Baye’s theorem requires estimated or precise pre-test odds to be combined with precise post-test odds to change the likelihood of a condition existing. To do this pre-test probability must be converted to pre-test odds, this is then multiplied by the likelihood ratio and the resulting post-test odds must then be converted to a post-test probability.

Likelihood Ratio

This is an odds number that can be directly combined with the pre-test odds that the condition is present or absent to change the odds of the condition being present or absent after the test is carried out. That is, if we believed that the probability of a condition being present prior to a positive test is 67% (2/3), then by multiplying the likelihood ratio, which is say 7, by the pre-test odd, then a change in the probability of the condition being present occurs so that the post-test probability is now 93%.

A positive likelihood ratio is where the odds of a positive test being truly positive and a negative likelihood ratio where the odds of negative test being truly negative. The pre-test or anterior probability is gained by what you know about the condition and about the patient. Factors such as prevalence, family history, results of previous tests including the subjective examination as well as the objective, personal habits such as smoking drinking, activity levels etc. all go into forming the anterior probability. The probability must be converted to odds otherwise the calculation becomes more complicated and then multiplied by the likelihood ratio to get the post-test or posterior odds which then must be converted back to probability. The calculation to change confidence is derived from Bayes theorem and is used in many fields.

Likelihood Ratio Calculation

The positive likelihood ratio (+LR) is calculated by dividing the sensitivity (S) by one minus the specificity (Sp) and is really just a probability to odds conversion using both numbers.

S/(1 - Sp)

So that a positive test with S = .9 and a Sp of .75 would have a +LR = 3.6.

The negative likelihood ration (-LR) is calculated by dividing 1 minus the sensitivity by the specificity.

(1 – S)/Sp

So that a negative test with S = .9 and a Sp of .75 would have a –LR of 0.25.

As a general guideline +LR above 10 are considered excellent, between 5 and 10 good and below 5 poor. Similarly –LR with scores below 0.1 are considered excellent, between 0.1 and 0.2 good and above 0.2 poor. But the real value of likelihood ratio is its ability to change probabilities when combined with the probability of the condition being present prior to the test. (MAYBE AN EXAMPLE OF APPLICATION HERE?)

Probability to Odds Conversion

Probability is limited to 1.0 or 100% but odds run from 0 to infinity to 1. Consider betting, odds of 38:1 is what you get betting on a single number on the roulette table not a great bet unless you like huge risks. Odds of 2:1 is a reasonable bet, but you are not surprised when you do not win and a 40:1 winner in a horse race excites all sorts of investigations. So most of us are familiar with odds in the betting sense but not in a professional sense, at least not in our profession. We are more at home with probability. The terminal patient with a 50% chance of surviving the week, the cancer patient with a 30% chance of recurrence are examples of how we think but it would be alien to our way of thinking to say they have 1:1 and 0.42:1 odds even though they mean the same thing. So probability (p) needs to be converted to odds (o) and odds back to probability. It is a simple calculation.

Probability to Odds

Odds = p/(1 – p)

similar to calculating the positive likelihood ratio

Odds to Probability

Probability = (1 – p)/p

similar to calculating the negative likelihood ratio

So we can take the test and apply it to the patient, lets assume we are using the O’Brien test for a suspected labral tear and lets say that from the history and other tests we have a estimated that there is a 60% probability of a labral tear being the pathology. According to O’Brien’s study of the test it has a sensitivity of 1 and a specificity of 0.98. Assuming a positive test we calculate the positive likelihood ratio:

+LR = S/(1-sp) = 1/1 - 0.98 = 1 - 0.02 = 50

Then we convert the anterior (pre-test) probability into odds.

0.6/(1-0.6) = 1.5

The posterior (post-test) odds = the anterior odds x the +LR = 75:1

The posterior probability = posterior odds/(1+posterior odds) = 75/76 = 99%

Therefore, using O’Brien’s interpretation of the O’Brien test, a positive test would increase the chances of a labral tear from 60% to 99%, a virtual certainty for the diagnosis.

If O’Brien’s test was negative, then the –LR is required.

-LR = 1-S/Sp = 1 – 1/0.98 = 0/0.98 = 0

This means that O’Brien’s test cannot be used mathematically to determine how likely a negative test is to predict the absence of a disease, but in words, it means that a negative test means that there is no chance of the disease being present. But if we drop the sensitivity to .999 from 1.0, then the calculation can be done and we get.

1 – 0.999/.98 = 0.001

Now to calculate the posterior odds

.001x1.5 = 0.0015

And then to calculate the post probability

0.0015 or .15% or 1.5 tenth of one percent, essentially no chance of a labral tear being present.

This is a particularly useful example of how criterion validity can go wrong. Stetson and Templin studied the test using a very comparable, although substantially smaller subject group, and the same gold standard and arthroscopic examination; although, in this study, the test was carried out in sitting and in O’Brien’s in standing. The results were completely different as the table below shows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Researcher | Sensitivity | Specificity | +LR | -LR |
| O’Brien | 1 | 0.98 | 50 | .001 |
| Stetson | 0.54 | 0.31 | 0.78 | 1.48 |

So using 60% as an anterior probability again, Stetson’s figures would be:

Posterior probability with positive test = 0.53

Posterior probability with negative test = 69%

So in this case there is less chance of a labral tear being present when the test is positive than before the test and more chance of the labral tear being present when the test is negative. Go figure!

Obviously, somebody or both bodies messed up probably in their methodology, somewhere in their study. So in this case it is imperative that you read the studies and hope that you can gleam where the mess-up occurred. It may have been that the description of the technique given by O’Brien did not match his photograph of the technique, he talks of supinating and pronating the forearm, but in fact, he medially and laterally rotated the shoulder and if Stetson followed the written instructions, this may explain the differences in the results. Anyway, one thing is clear the fact that criterion validity is being used, there are major flaws in its use here and almost certainly in other cases, so be careful.

Short Cuts

For those not mathematically inclined, too busy or too lazy and most of those descriptions apply to me there are quicker ways of using Baye’s Theorem. You still have to estimate the anterior odds and to restate this is done from the individual patient’s data and the prevalence/incidence of the condition. To help with estimating these odds, keep the precision fairly low initially. The spectrum ranges from it absolutely cannot be the diagnosis to it absolutely is the diagnosis. In between there is 50/50, probably is and probably isn’t. If you feel you can be a bit more precise there is almost “certainly is” and “almost certainly isn’t”, but you probably will not need to get more precise than this. The figure below represents the probabilities and odds associated with these estimates and a translation into plain language.

A screenshot of a cell phone

Description automatically generated

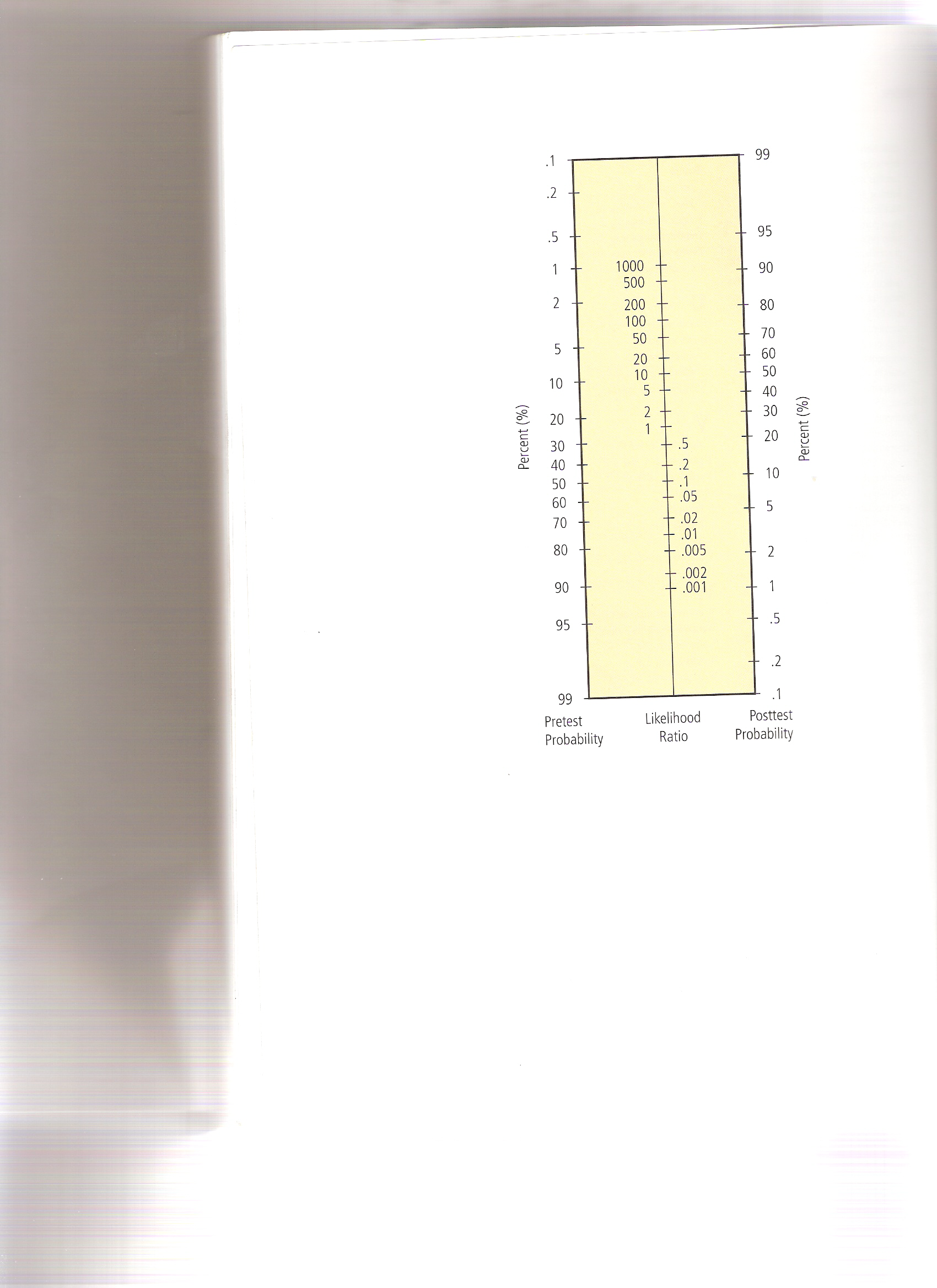
To save you calculating the likelihood ratios the two matrices below allow you to find it from the sensitivity and specificity. The likelihood ratio in red are extremely significant while those in blue moderately so.

**Positive Likelihood Ratios**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Spec** |  |  |  |  |  |  |  |  |  |  |
| **Sens** | 1 | 0.99 | 0.95 | 0.9 | 0.85 | 0.8 | 0.75 | 0.7 | 0.65 | 0.6 | 0.55 |
| 1 | **infinite** | ***100*** | ***20*** | ***10*** | **6.7** | **5** | 4 | 3.3 | 2.9 | 2.5 | 2.2 |
| 0.99 | **infinite** | ***99*** | ***19.8*** | ***9.9*** | **6.6** | **5** | 4 | 3.3 | 2.8 | 2.5 | 2.2 |
| 0.95 | **infinite** | ***95*** | ***19*** | **9.5** | **6.5** | 4 | 3 | 3.17 | 2.7 | 1.5 | 1.2 |
| 0.9 | **infinite** | ***90*** | ***18*** | **9** | **6** | 4.5 | 3.5 | 3 | 2.6 | 2.3 | 2 |
| 0.85 | **infinite** | ***85*** | ***17*** | **8.5** | **5.7** | 4.3 | 3.4 | 2.8 | 2.4 | 2 | 1.9 |
| 0.8 | **infinite** | ***80*** | ***16*** | **8** | **5.3** | 4 | 3.2 | 2.7 | 2.3 | 2 | 1.8 |
| 0.75 | **infinite** | ***75*** | ***15*** | **7.5** | **5** | 3.6 | 3 | 2.5 | 2.1 | 1.9 | 1.7 |
| 0.7 | **infinite** | ***70*** | ***14*** | **7** | 4.7 | 3.5 | 2.8 | 2.3 | 2 | 1.8 | 1.6 |
| 0.6 | **infinite** | ***60*** | ***12*** | **6** | 4 | 3 | 2.4 | 2 | 1.7 | 1.5 | 1.3 |
| 0.5 | **infinite** | ***50*** | ***10*** | **5** | 3.3 | 2.5 | 2 | 1.7 | 1.4 | 1.3 | 1.1 |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

**Negative Likelihood Ratios**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Spec** |  |  |  |  |  |  |  |  |  |  |  |
| **Sens** | 1 | 0.99 | 0.95 | 0.9 | 0.85 | 0.8 | 0.75 | 0.7 | 0.65 | 0.6 | 0.55 | 0.5 |
| 1 | 0 | **0** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0.99 |  | **0.01** | **0.01** | **0.01** | **0.01** | **0.0125** | **0.013** | **0.014** | **0.02** | **0.016** | **0.02** | **0.02** |
| 0.95 |  | **0.05** | **0.05** | **0.05** | **0.06** | **0.0625** | **0.067** | **0.07** | **0.08** | **0.08** | **0.09** | **0.1** |
| 0.9 |  | **0.1** | **0.11** | **0.11** | **0.12** | **0.125** | **0.13** | **0.14** | **0.15** | **0.17** | **0.18** | **0.2** |
| 0.85 |  | 0.15 | **0.16** | **0.17** | **0.18** | **0.19** | **0.2** | 0.21 | 0.23 | 0.25 | 0.27 | 0.3 |
| 0.8 |  | **0.2** | **0.2** | 0.22 | 0.24 | 0.25 | 0.27 | 0.29 | 0.31 | 0.33 | 0.36 | 0.4 |
| 0.75 |  | 0.25 | 0.26 | 0.28 | 0.29 | 0.31 | 0.33 | 0.36 | 0.38 | 0.42 | 0.45 | 0.5 |
| 0.7 |  | 0.3 | 0.32 | 0.33 | 0.35 | 0.375 | 0.4 | 0.43 | 0.46 | 0.5 | 0.55 | 0.6 |
| 0.6 |  | 0.4 | 0.42 | 0.44 | 0.47 | 0.5 | 0.53 | 0.57 | 0.62 | 0.67 | 0.73 | 0.8 |
| 0.5 |  | 0.5 | 0.53 | 0.56 | 0.59 | 0.625 | 0.67 | 0.71 | 0.77 | 0.83 | 0.91 | 1 |

For the truly lazy, below is a normogram that allows you to go from the anterior probability to the posterior probability by extrapolating a straight line linking the sensitivity and likelihood ratio to the posterior probability; but it is worthwhile at least understanding the math so that you understand the

concepts of changing confidence as more information comes in. By using Baye’s theorem either doing the math yourself or using the nomogram the sensitivity and specificity values can be converted into something really quite useful providing the research into the test is not rubbish.

# Clinical Reasoning in Pathoanatomical Diagnosis and Treatment

## The Pathoanatomical Diagnosis

Two make a pathoanatomical diagnosis there must be state of pathology and structure. For example, a second degree tear of the biceps brachi belly, an L5 disc herniation with an L5 radiculopathy or a traumatic arthritis of the knee. Not pathoanatomical diagnoses are whiplash, low back sprain, nerve irritation, cervicogenic headaches, metatarsalgia or fibromyalgia; these have neither a pathological statement nor a description of the affected structure. Sometimes names are substituted for a pathoanatomical diagnosis.

Conditions such as tennis elbow (but not lateral epicondylitis unless you really believe the epicondyl *is* inflamed), golfer’s elbow, Sever’s disease and the like are widely understood to contain the pathoanatomical diagnosis and may be used as the diagnostic statement. But there are many causes of metatarsalgia, cervicogenic headaches, low back pain and even fibromyalgia, and the term is really just a regurgitation of the patient’s symptoms in fancy language, with the exception of fibromyalgia which nobody has a good idea of what it is. In some cases, the pathology is not known and so the term syndrome, which is a collection of characteristic signs and symptoms that lead to a treatment and the delusion that we know what we are doing, is used. However, the term piriformis syndrome should not be used, as the exact pathology of this syndrome is known, and therefore should be called a piriformis entrapment neuropathy or something similar.

## Clinical Reasoning

Clinical reasoning can be defined as a process that at best integrates and analyzes the clinician’s knowledge of the individual patient’s information, disease prevalence, experience, intuition (?), scientific method, the value of the various tests used and other factors to come to a decision on the diagnosis and best treatment.

The expert is known to function differently to the novice in the reasoning process although most of the difference is in the formality used to make the decision. Experts tend to use short cuts (heuristics) based on previous experience while the novice tends to conform to a sequence set for them by authority. The problem with heuristics is that they can be full of bias and even prejudice and while a degree of bias is useful in that it sensitizes the clinicians to certain findings that may have otherwise been overlooked if it is taken to the extreme, it can bias the clinician so that things that are not present are found. On the other hand, the most extreme version of a novice sequence is an algorithm that has been set by authority and leaves little or no room for improvisation. As smothering as this must seem to anybody with any degree of experience or free thought it does have advantages in the early stages; for example, set algorithms do not have any bias and certainly they exclude prejudice. The questions must be first do set algorithms retard learning and the use of heuristics and second can using set algorithms really be considered reasoning.

Clinical reasoning uses three main methods usually in combination with each other:

1. Pathognomia (Pathognomonia)
2. Hypothetic-deductive reasoning
3. Pattern recognition
4. Pathognomia really isn’t clinical reasoning but recognition of a single sign or symptom that is found in only one condition, and therefore when present, can be used to definitively diagnose the condition. These are rare in medicine and practically non-existent in manual therapy. Ulnar deviation of the wrist in rheumatoid arthritis and Kayser-Fleisher rings of Wilson’s disease are two examples among few. Close to pathognomic is the sharp severe pain over the posterior SIJ on turning in bed which wakes the patient in SI arthritis.
5. Hypothetic-deductive reasoning is where a hypothesis is generated early in the examination and often just on the patient’s profile and the prevalence of the condition. As the history is taken and more information gathered the hypothesis is either accepted or rejected and changed. This continues through the entire examination until at the end, a single hypothesis is left which is considered the diagnosis. This is a process followed by almost everybody not using set algorithms at some point in most examinations. To use this type of reasoning efficiently and effectively the questions and tests must be chosen of the basis of how well they test the current hypothesis so that taking a history becomes focused and algorithmic (in this case an evolving algorithm) and special tests should be selected based on the hypothesis.
6. Pattern recognition is where the examiner is able to see a pattern emerge from the patient’s history that fits a given diagnosis. The pattern recognized in the history is confirmed by the objective examination that is also typical of the condition. This type of pattern is really of use only to the experienced or really well read clinician and even then is generally not usable unless the patient is allowed to disgorge all of the information he/she has in one go as the first question from the examiner moves it from pattern recognition to hypothetic-deductive reasoning. So pattern recognition plays a very prominent role in hypothetic-deductive reasoning in that the first hypothesis is based on recognizing to some extent at least the first things learned from the patient.

# Illness Scripts

The formal concept of illness scripts both as a clinical tool and as an education one is relatively new, but the general idea is not new and at first glance seems obvious. In effect they are the total story of the condition although with the plethora of people writing on the subject on the internet the definition is becoming a little fuzzy. But to define it for the purposes of this essay, the illness script is the history of the condition, which can be enlarged to include the objective findings to become the extended illness script.

Illness scripts can be thought of as being held by both the patient and the examiner with the patient’s illness script obviously the more, detailed, individualized and personal. The examiner’s illness script is one that is generated by the literature, training, personal experience with the condition and reports from fellow practitioners so it will necessarily be general and, in the first instance, typical or average for the condition. The aim is to match the patient’s description of his complaints with the illness script that the examiner holds of the condition, and when this is done, a diagnosis can be made. When this is meshed with hypothetic-deductive reasoning, it becomes a very powerful tool in clinical reasoning.

For educational purposes the illness script is also a very useful concept especially when it is reduced to the unique cluster identifier of the condition. We can call this an essential illness script (as far as I know my term for better or worse), which usually runs to about 5 characteristics from the history and maybe 8 characteristics when the objective examination identifying characteristics are added, the extended essential illness script. The more experienced the therapist is with the condition, the richer the illness script is so that atypical variations of the condition can be considered when the patient relates one or more characteristics that are at variance with the standard illness script. For example, a contained disc herniation occurring may behave as does an uncontained posterolateral herniation but may be central or even bilateral, and have negative straight leg raises but the rest of script, severe pain, difficulty with walking and severely limited trunk movements conform to the typical illness script. In this particular case the normal or near normal straight leg raise is wrong for the diagnosis but the remainder of the script remains stable. So a hypothesis of disc herniation seems likely, but the variation of the normal SLR has to be reconciled, and a contained herniation does just that. Also important to the illness script is the context that is, the patient’s age, gender, activity level, personal habits etc. For example, you would likely hypothesize that a person who had bilateral back and leg pain on walking and standing might have central stenosis but not if it was a 30 year old woman.

In the subject and objective components of the essential illness script one question and one test needs to be close to 100% sensitive the remainder need to be as specific as possible. If the questions and test are mostly sensitive then hundreds of diagnoses will fit the EIS but if most are specific then only one or two will. For example, in the case of a tendonosis, the question, where is the pain will be close to 100% sensitive and palpation for tenderness along the tendon will also be 100% sensitive providing the entire tendon can be palpated. One or both contractile tests (isometric contraction and stretching) will be close to 100 specific and when analysed in conjunction with the tenderness on palpation you’ve pretty much got a working diagnosis.

An example of an essential illness script is the following for a large prolasped lumbar disc herniation without radiculopathy.

1. Severe unilateral low back and leg pain
2. Severe pain with flexion and extension activities and postures
3. Pain on walking to the point where it causes limping
4. Obligate (not of the patient’s choosing) functional loss

The extended essential illness script

1. Severe limitation of flexion and extension movements
2. Severe limitation of SLR

While there are obviously many more signs and symptoms than this these combine to form a cluster which is, if not a unique identifier of disc herniation, very close to it and it is difficult to think of another diagnosis that would fit that is not so rare and exotic that it’s prevalence alone comes close to ruling it out.

# Making the Diagnosis

First the context or patient profile must be understood, age, gender, family history, trauma or no trauma etc. (you must be careful here to not let the patient profile and statistics cause framing bias). Then the area of the pain must carefully mapped and at that point an initial hypothesis and its illness script can be generated. Further questions based on the hypothesis’ illness script and algorithmic in nature (the next question is based on the patient’s last response) are asked and the answers will either strengthen or weaken the hypothesis. If the hypothesis is sufficiently weakened to the point where even outliers of the condition are less likely than another condition, the hypothesis is rejected and a more suitable one generated from the previous answers and with the questions now concentrate on the illness script for the new hypothesis. The process continues until at the end of the subjective examination, the examiner is sufficiently satisfied with the hypothesis that the objective examination can be undertaken. The differential diagnostic examination based on Cyriax’s selective tissue tension testing is a break with the idea of focused and algorithmically based testing, but it is a standard part of many therapist’s routines and if considered as a single test rather than as a series of tests the principle is not broken. For example, if the hypothesis based on the subjective examination is an uncontained posterolateral disc herniation then the lumbar scan is appropriate, but if the hypothesis is segmental dysfunction, then a segmental examination is what should be done. However, many therapists will complete the scan first, but even this may be considered as testing the illness script as one feature of segmental dysfunction is a negative scan (that is no diagnosis is made from it).

# Summary

Scientific method is not the exclusive domain of the researcher and does not necessarily entail strict control of variables, double-blinded trials nor is even formal note taking on methods and results. Rather it is an organized process of using observations and facts together with appropriate knowledge to generate a testable and reasonable hypothesis that is tested until accepted or rejected. Using our body of knowledge which includes empiricism, experience, research, theoretical models and clinical sciences, the clinician chooses the best available evidence for that case to determine through examination and clinical reasoning a diagnosis and treatment plan for each individual patient. Given this, it can be said that the clinician uses the scientific method far more routinely than does the researcher and the number of experiments he/she does is far more numerous.

For efficient and effective treatment of the patient the clinician must be able to integrate the context, the history and the objective examination into a cohesive whole that leads to a diagnosis and treatment plan. This integration is optimized if the clinician understands the processes used to do this and at least initially carries this process out in a formal manner, which later will become more casual as the ability to use heuristics takes over. We need to apply our knowledge of clinical sciences to the patient’s problem; we need to understand levels of evidence but not slavishly adhered to them as the researcher perspective would have us do; we need to be able to reconcile research evidence when it clashes with our experience or common sense; we need to use the various methods of clinical reasoning, pattern recognition, pathognomia and hypothetic-deductive reasoning in combination to generate intermediate hypotheses that will lead to a diagnosis and treatment; we need to organize our thought processes as much as our physical examination processes; and most difficult we need to know our bias’s, accept them, use them and control them but eliminate our prejudices. We should use tools such as illness scripts and essential illness scripts to make things easier and more efficient.

A screenshot of a cell phone

Description automatically generated