Philosophy of science and the diagnostic process

Brian H Willis\textsuperscript{a,}\textsuperscript{*}, Helen Beebee\textsuperscript{b} and Daniel S Lasserson\textsuperscript{c}

\textsuperscript{a}Primary Care Clinical Sciences, University of Birmingham, Birmingham, UK, \textsuperscript{b}Department of Philosophy, University of Manchester, Manchester, UK and \textsuperscript{c}Primary Care Health Sciences, University of Oxford, Oxford, UK.

\*Correspondence to Brian H Willis, Primary Care Clinical Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK; E-mail: b.h.willis@bham.ac.uk

Received 9 April 2013; Revised 24 May 2013; Accepted 28 May 2013.

This is an overview of the principles that underpin philosophy of science and how they may provide a framework for the diagnostic process. Although philosophy dates back to antiquity, it is only more recently that philosophers have begun to enunciate the scientific method. Since Aristotle formulated deduction, other modes of reasoning including induction, inference to best explanation, falsificationism, theory-laden observations and Bayesian inference have emerged. Thus, rather than representing a single overriding dogma, the scientific method is a toolkit of ideas and principles of reasoning. Here we demonstrate that the diagnostic process is an example of science in action and is therefore subject to the principles encompassed by the scientific method. Although a number of the different forms of reasoning are used readily by clinicians in practice, without a clear understanding of their pitfalls and the assumptions on which they are based, it leaves doctors open to diagnostic error. We conclude by providing a case example from the medico-legal literature in which diagnostic errors were made, to illustrate how applying the scientific method may mitigate the chance for diagnostic error.

**Keywords.** Diagnostic tests (e.g. ultrasound), doctor–patient relationship, medical errors/patient safety, philosophy, primary care, risk assessment, science.

Introduction

Making a diagnosis is a fundamental component of the delivery of health care, and although the complexity of the primary care consultation requires many different skills in the practitioner, diagnostic decision-making is important. Diagnosis is a complex process beset with pitfalls—often representing the balancing of different uncertainties. Cases in the literature demonstrate how easy it is for clinicians to miss important diagnoses and the impact this can have on patients’ lives.\textsuperscript{1,3}

Although less controlled than the conditions experienced in a laboratory, the process of diagnosis is nonetheless an example of science in action. Whether it is in the choice of the test or the interpretation of results, the clinician, to a large degree, uses reasoning based on a scientific knowledge of diseases and how they manifest. Consequently, some have highlighted the role that philosophy could play in the diagnostic process.\textsuperscript{1}

The methods scientists use to acquire knowledge are the concern of philosophers of science. But are scientists justified in using such methods and can we truly believe the conclusions derived? What are their limitations? These are some of the questions we address in this discussion of the different forms of reasoning that take place in the context of the diagnostic process before applying the principles to a case from the medico-legal literature.

Deductive reasoning

Deduction dates back to the time of Aristotle\textsuperscript{4} and allows us to derive a new proposition from two other premises.\textsuperscript{5} An example would be all men are mortal; Aristotle is a man, therefore we deduce that Aristotle is mortal.\textsuperscript{5} In diagnosis, such a deductive argument could be applied to diseases with pathognomonic features, which are exhibited only by the disease in question.

For example, rice-watery stool is considered pathognomonic of cholera. Thus, if John has rice-watery stools, we may deduce John has cholera. Clearly, the argument rests on the truth that no other condition presents with rice-watery stools and that John actually has such stools. It must be questioned whether there exists any feature in medicine, which when present, unfailingly indicates the presence of a specific disease.

Furthermore, logical fallacies may result when the direction of reasoning is ignored.\textsuperscript{6} For example, consider the following: ‘all patients who are cancer-free have negative test results. John is cancer-free therefore he will have a negative test result.’ This is quite different
from reasoning that John is cancer-free because he had a negative test result, which may be falsely reassuring.

Inductive reasoning

In contrast, inductive reasoning does not require the conclusion to logically follow from the premises—just that it is supported by them. In particular, inductive reasoning relies upon using empirical evidence from observing a sample of existing cases to make assertions about the unobserved or a wider set of cases.7

This type of reasoning is not foreign to the diagnostic process where pattern recognition relies upon the reasoning that all previous cases that exhibited this pattern were X, therefore all future cases that exhibit this pattern will be X. Yet, such reasoning is fallible: the classic example is the inductive assertion that all swans must be white after observing only white swans. In particular, uncritical use of inductive reasoning can lead a clinician to ignore or, perhaps, not even consider other rare causes of the combination of features exhibited.

For instance, consider a 50-year-old man presenting to his primary care physician or GP with itchy, erythematous scaling patches on his arms. In primary care, where eczema is so common, usually such a presentation would be diagnosed as eczema, but it could also be, more rarely, mycosis fungoides. Importantly, taking such an uncritical inductive approach to diagnosis would miss the possibility of the rare but real occurrence of the occasional mycosis fungoides.

One of the chief proponents of induction was the English philosopher Sir Francis Bacon, who considered induction and theory confirmation as the final stages of his ‘scientific method’. The initial stage was a framework detailing how to generate a hypothesis in the first place. He believed that investigators should not have preconceived hypotheses and that observation should drive the process of forming the hypothesis.8

In practice, observing without at least some preconceived ideas is neither possible nor desirable.8 In making a diagnosis, clinicians have to be able to filter the ‘signal from the noise’ while listening to a clinical history that may well contain extraneous detail that has no bearing on the eventual diagnosis.

Falsification

Inductive reasoning attracted a number of critics but Sir Karl Popper was probably the first to suggest that we should dispense with it altogether.9 He recognized an important asymmetry between attempting to confirm a generalization and trying to falsify it.10 No matter how many cases we observe that allegedly confirm a generalization, we can never know whether it is true for all cases. In contrast, it takes only one case to demonstrate it is false—the black swan.

Popper believed that taking evidence to justify or confirm a theory is both irrational and unscientific.11 The only thing we can really be sure of is if we come across a counter-example, our theory is false. Investigators should therefore be relentless in their search for evidence that challenges a theory, which also means being clear on the conditions that would force them to discard it.11

While many philosophers have rejected Popper’s claim that no theory can ever be positively justified or confirmed, his method enshrines the important insight that we should not simply seek to confirm our theories; we should also seek out potential counter-examples.

The hypothetico-deductive approach starts by hypothesizing on the potential diagnosis in a patient. Diagnostic tests in the form of targeted questions, clinical examinations or more invasive tests are used to establish or refute the diagnosis. In the case of the 50-year-old, presenting with features of eczema, the hypothesis of eczema may be challenged if there was a poor response to a trial of steroid cream. Such an occurrence should prompt the hypothesis of eczema to be re-considered. So in the case of rare but important diagnoses, Popper’s method could be valuable. However, like inductive reasoning, it is not without flaws.

The Duhem–Quine thesis

Essentially, Popper’s method relies upon being able to test a single hypothesis. The problem is how to isolate the hypothesis of interest from all other assumptions. It seems there are always other assumptions that need to be considered. Thus we cannot be sure that our hypothesis has really been refuted when one of our auxiliary assumptions may be at fault.13

Consider the following case. A 65-year-old lady who has a history of migraines presents with a right-sided temporal headache. Her doctor would like to exclude temporal arteritis and decides to carry out an erythrocyte sedimentation rate (ESR) blood test believing that a normal result will do this.

The problem with this belief is that it does not separate the hypothesis of interest, namely that the patient has temporal arteritis, from the auxiliary assumption that temporal arteritis presents with an abnormal ESR. Hence, a normal ESR really falsifies the combined hypothesis that the patient has temporal arteritis and that temporal arteritis presents with a raised ESR.

Importantly, this combined hypothesis is false when either temporal arteritis is not present or the assumption that temporal arteritis presents with an abnormal ESR is false. In the latter instance, the patient could still have temporal arteritis and an important diagnosis would be missed.
Theory-laden observations

Consider the case of a child presenting with a headache. A doctor may start with a hypothesis of meningitis and seek to test this by looking for photophobia, a fever and a positive Kernig’s sign. In doing so, the clinician is making theory-laden observations based on their understanding of the disease process. But as a result, they may fail to observe or dismiss the relevance of features that do not fit their understanding of meningitis.14

Although Thomas Kuhn is probably most associated with the phrase ‘paradigm-shift’,15 he also brought to our attention the importance of theory-laden observations,16 and primary care especially suffers from this problem with the majority of medicine being taught from a secondary care perspective. Thus, our knowledge of target disorders is dependent upon a conceptual framework that arises predominantly from classical descriptions of disorders that are more usually seen in secondary care.

Until recently, the presence of leg pain or cold hands and feet in children would not usually prompt GPs to think of a potential meningococcal infection as our theory on the presenting features of meningococcal infection did not include such features.17 Indeed, GPs might not even have observed these features at all, even if they were present, any more than they might observe the colour of the patient’s eyes. Yet, it turns out that such features may be the only detectable clinical manifestations of sepsis in children at risk of death.17

Although Bacon’s advice would be to observe objectively first then form a hypothesis,8 the idea of theory-laden observation suggests that observation is always through the veil of our current understanding and is thus prejudicial. Consequently, as our theories of disease presentation evolve features that might have been once unobserved or dismissed become relevant and more likely to be noted by clinicians.

Inference to the best explanation

What happens when we have two competing hypotheses that seem equally likely? Bacon recognized this problem and suggested that we perform a ‘crucial experiment’93 that was designed so that the result would refute one of the hypotheses while confirming the other.18 For example, if a 65-year-old woman presented with breathlessness and bibasal crackles on auscultation of the chest, then left ventricular failure and pulmonary fibrosis are both potential diagnoses. Thus the ‘crucial experiment’ would be to choose a test that would help discriminate the two, such as a high-resolution computed tomography scan.

An alternative approach is to use ‘inference to the best explanation’.15 Broadly, this consists of deciding which theory best explains the evidence and is probably one of the more frequent modes of reasoning used in diagnosis.19 For example, consider a 25-year-old who presents after collapsing in the gym with a history of sudden death in the family. Such an event could be explained by a pulmonary embolus, aortic stenosis or even a myocardial infarction, but the best explanation would be to consider a hereditary disorder such as hypertrophic obstructive cardiomyopathy (HOCM).

Again this approach is not without shortcomings and can lead us to decide on ‘the best of a bad lot’—if the correct diagnosis is not considered a possibility, we will ultimately converge on the wrong diagnosis with this method.19 Even if the correct diagnosis is among the possibilities considered, the ‘best’ explanation may still lead to the wrong diagnosis owing to the fact that diseases do not always present as expected.

A further difficulty arises from the risk of cognitive biases influencing the diagnostic process.20 The clinician may diagnose HOCM by unwittingly accessing stereotypical descriptions of different diseases in their memory. Thus, the ‘best explanation’ becomes the stereotype which most closely matches the presentation—the so-called representativeness heuristic.21 Although such a process is rapid, automated and requires less effort than rational thought,22 it is also fallible.

Bayesian inference

So far, all the methods discussed have been qualitative in nature, but Bayesian inference allows us to quantify the likelihood of a diagnosis using probability theory. The methods have been described elsewhere23 but, briefly, comprise using estimates of both the probability of the diagnosis before testing (the ‘prior probability’) and the performance of the test.20 These two elements are then combined using Bayes’ rule to estimate how much the test result modifies the probability of the diagnosis—‘the posterior probability’.23 The source of ‘prior probability’ (disease prevalence) and test performance can be taken from the clinician’s experience and the research literature.

Thus, Bayesian inference lends itself to the predominant paradigm of evidence-based medicine. Although appealing, this too is not above criticism. Ultimately, the accuracy of the posterior probability depends not only on being able to identify the relevant studies from the literature but for those estimates derived from the studies being transferable into practice, which may not always be the case.24,25

Case example

We will now apply the principles to the following case in the medical-legal literature where diagnostic errors were made.26
A 47-year-old gentleman presented with heel pain to his GP. He had suffered with it for several months before it suddenly worsened during a squash game. The pain was confined to the heel and exercise seemed to exacerbate the symptoms. The examination was unremarkable and an X-ray showed only a bony spur on the calcaneus. Reassured, the GP decided to treat conservatively but following non-resolution the patient re-presented with a swollen calf and erythema. Over the following weeks, this led to investigation for a deep venous thrombosis and then treatment for suspected cellulitis. After 3 months, he was referred to an orthopaedic surgeon who diagnosed a ruptured Achilles tendon, which was confirmed on ultrasound.²⁶

So could applying some of the philosophical principles discussed have helped to diagnose a ruptured Achilles tendon sooner?

From the initial presentation there are a range of potential diagnoses and the risk is deciding upon the diagnosis too soon without adequately testing each hypothesis. For instance, from the initial presentation the GP may reasonably infer that the best explanation for the symptoms is not a ruptured Achilles tendon at all, as it does not match the classical description, and that plantar fasciitis is more likely. But such a diagnosis would have been made based on inadequate data and also demonstrates how susceptible ‘inference to the best explanation’ is to the representativeness heuristic.

In this case, it is not clear whether there was an initial diagnostic hypothesis and following an unremarkable examination and normal X-ray the GP was content to discount serious causes in favour of a more benign cause to the pain. Popper would argue that we need to be clear on the conditions by which we are prepared to discard a scientific hypothesis, but if we have not been explicit about the diagnostic hypothesis at the outset, then how can we know the conditions for discarding it?

Furthermore, the nature of the diagnostic process means that after confidently discarding a hypothesis, others have to be considered. But unless a ruptured Achilles tendon appears in the range of hypotheses, it will not be diagnosed, whichever approach we take.

Suppose the GP is an evidence-based practitioner and does consider Achilles tendon rupture as a possible diagnosis and decides to perform Simmonds’ calf squeeze test. They refer to the literature and find a study that reports the sensitivity and specificity of Simmonds’ test to be 96% and 93%, respectively, in orthopaedic outpatients.²⁷ In theory, if the GP estimated the prevalence of Achilles tendon rupture, using Bayesian inference they could calculate its probability once the test result was known.

Unfortunately, the Simmonds’ test like so many clinical tests was developed in a specialist setting where there was a highly selected group of patients and where the operators had a high degree of expertise. Many of these tests have not been evaluated in a primary care setting owing to the low prevalence of disease, and it is unlikely that such performance characteristics would transfer into general practice. In the study, the prevalence of rupture prior to testing was 86%,²⁷ which is significantly higher than that which would be experienced in general practice.

A positive Simmonds’ test would probably have prompted a referral, but the more difficult question is how to manage a negative result? Since both the accuracy of the test in primary care patients and the dependence of the test on operator expertise are uncertain, the real possibility of a false-negative result remains. Thus, to dismiss an Achilles tendon rupture as a diagnosis after a negative Simmonds’ test would be, according to the Duhem–Quine thesis, to make an unsubstantiated assumption on the accuracy of Simmonds’ test. In such an instance, further investigation would be warranted. In fact, the GP did not document whether the Simmonds’ squeeze test was performed at all.²⁸

Applying the scientific method in this case illustrates the importance of stipulating a range of potential diagnoses early in the diagnostic process. This provides a framework for the clinician to aid decision-making on testing and when to move to another hypothesis. In primary care, in particular, there is perhaps more uncertainty than in secondary care owing to definitive tests being not readily available. The uncertainty in test performance characteristics, prevalence of diseases and the potential diagnoses should serve to remind us that making strong assumptions in such an environment is a recipe for diagnostic error.

Conclusion

The diagnostic process is a form of science where different methods of enquiry based on our best scientific knowledge are used to explain the clinical course of a patient’s symptoms. One of the objectives of philosophy of science is to analyse these different forms of enquiry to establish whether we are justified in believing the conclusions drawn and the ways in which we are exposed to error.

Although clinicians probably use many of the forms of reasoning outlined above automatically, whether they are conscious of these methods or their associated shortcomings is another matter. Perhaps examples of missed diagnoses in the literature¹⁻³ suggest otherwise. Thus, increasing awareness of the different forms of reasoning and their pitfalls can only serve to improve understanding and help enable clinicians to take a rational approach to diagnosis.

The case example also illustrates the importance of formulating hypotheses at the beginning of the diagnostic process. Without clear hypotheses, it is difficult to know the conditions by which one should be discarded and another taken up.
Recent calls by some authors for clinicians to employ Popper\(^1\) or Bayesian\(^2\) methods ignore the more nuanced approaches to scientific reasoning that have emerged in the philosophy of science, many of which have direct relevance to medical diagnosis. As such, the scientific method applied in diagnosis encompasses not one single general method, but by necessity, a range of principles and modes of reasoning in response to the wide variation of cases experienced in practice.

The Duhem–Quine thesis demonstrates how falsificationism as a method is constrained by the need to make additional assumptions. Yet, this criticism is not just confined to Popper’s method. For instance, the method of ‘Inference to the best explanation’ assumes that the real diagnosis is contained in the list of possibilities being considered and Bacon’s crucial experiment assumes a test exists that is able to accurately discriminate between two competing diagnostic hypotheses. In fact, all the methods discussed are potentially undermined by uncertainty in the validity of the underlying assumptions.

The Bayesian response to this is to describe our uncertainties in terms of probabilities. But this too makes the assumption that these probabilities are measureable with accuracy and that, in particular, study estimates of test accuracy are applicable to the setting of interest. Of course, we have to be pragmatic: diagnosis is a decision-making process and by necessity involves making assumptions on the ‘usefulness’ of different pieces of evidence.

But if philosophy should tell us anything about diagnosis, it is that whichever method of reasoning we use, we must always be explicit at the outset on our diagnostic hypotheses and be prepared to test the underlying assumptions upon which the method is based, otherwise we leave ourselves open to diagnostic errors. Thus, philosophical reflection on the scientific method helps us to better understand, and hence better able to guard against, the problems inherent in patient diagnosis.

Acknowledgement

BW conceived the idea and all authors were involved in the writing and editing of the article.

Declaration

Funding: none.
Ethical approval: none.
Conflict of interest: none.

References

29. Lewis SL. Think it possible you may be mistaken (letter). BMJ 2011; 343: d7701.