

The origins of the treatment of uncertainty in clinical medicine. Part I: ancient roots, familiar disputes

Robert AJ Matthews

Department of Mathematics, Aston University, Birmingham B4 7ET, UK

Corresponding author: Robert AJ Matthews. Email: rajm@physics.org

Introduction

[T]he research of truth requires much labour, and is beset with difficulty. (Louis,¹ p 28)

These words appear in *An Essay on Clinical Instruction* by the 19th-century French physician Pierre Charles Alexandre Louis, but they could have been written at any time in the long history of medicine. They reflect an aspiration of physicians since at least the time of Hippocrates, but which is now rarely stated so bluntly: the use of evidence to establish *certainty* in diagnosis, treatment and outcome. In the first part of this brief history of this aspiration, I examine the origins of attempts to turn it into a reality, beginning in the 5th century BCE with simple reasoning and intuition, followed by the use of logic and quantification. The second part will describe the emergence of attempts to use probability theory to deal with the challenges of uncertainty around 1700.

I also describe the controversies engendered by these approaches. Perhaps their most striking feature is their durability. Concerns raised millennia ago remain at the forefront of current debate in clinical medicine: how to infer reliable conclusions from observations, the meaning of ‘convincing’ evidence and the recognition of sources of uncertainty beyond the play of chance. No less striking is the vociferous nature of the arguments engendered by attempts to reduce uncertainty, especially through the application of mathematics.

Early approaches: simple reasoning

In contemporary clinical medicine, the concept of uncertainty is inextricably linked to probability. Error bars and confidence intervals are used to quantify the impact of the play of chance on a finding; sample size calculations are performed with the aim of controlling this source of uncertainty; significance tests are performed in the belief they can extract truth

from uncertain findings. The entire armamentarium rests on the mathematical theory of probability, which began to take its modern form in the mid-17th century.^{2,3} Initially, its focus was so-called *aleatory* probability (from the Latin for ‘dice-player’) and analysis of games of chance. The idea that it may be applicable to uncertainties of events beyond the gambling house began to be explored in quantitative terms by the late 17th century, while its systematic use in clinical medicine has become established only since the second half of the 20th century. Thus, if the history of medicine is condensed into a year starting with the work of Hippocrates (c. 460 to c. 375 BCE) on 1 January, the quantitative treatment of uncertainty only became routine around the last week of December.

In contrast, recognition of both the existence of uncertainty in medical practice and the need to address it dates back at least as far as the Ancient Greeks. The Hippocratic text *On Ancient Medicine* (c. 475 BCE) argued that the uncertainty involved in ascribing a patient’s response to a specific treatment implied that as a discipline medicine lacks the precision (‘ἀκριβεία’) and reliability of an exact science (‘τέχνη’), and that this affects the ability of physicians to prescribe appropriate treatments.⁴ Nevertheless, the author(s) of that text saw no reason why the lack of certainty should preclude the use of medicine, with treatments being based on associations and rules of thumb gained from experience with individual patients. A celebrated example is Hippocrates’ observation that ‘Persons who are naturally very fat are more apt to die quickly than those who are thin’ (*Aphorisms* 2.44, cited in Tsiompanou and Marketos).⁵ This points to an approach to uncertainty that is essentially probabilistic, in the sense of being based on comparisons of frequencies. This becomes more explicit in the commentaries that follow case histories, which often end with statements of the form ‘It is *probable* that, by means of . . . , this patient was cured/the death is to be attributed to . . .’ (*Of the Epidemics* cited in Sheynin,⁶ emphasis added).

Such an approach to extracting insight from uncertain information is far from infallible, of course. Lacking a quantitative theory of probability, Hippocrates could not have known that even with random sampling, using simple majorities to assess the value of treatments is potentially misleading, especially with small samples. Even so, the belief that clinically useful insights can be obtained even in the presence of uncertainty shows a level of sophistication far from universal even in contemporary clinical research.

Aristotle and the use of logic

Nevertheless, the belief that precision and certainty are the defining characteristics of an exact science persisted. It received considerable kudos through its advocacy by Aristotle (c. 384 to 322 BCE), who argued that there can be no science of chance, and so, by implication, phenomena made uncertain by chance events cannot be the subject of scientific study. Indeed, this has been cited in explanation for one of the most vexed questions in the history of science: why did the theory of probability take so long to emerge? (Franklin,⁷ pp.335–337). Nevertheless, Aristotle's writings do include an attempt to bridge the gap between the 'more often than not' argument of Hippocrates and his acolytes and the certainty presumed necessary for an exact science. The *Prior Analytics*⁸ is best known for its introduction of deductive logic based on syllogisms such as 'All men are mortal/Socrates is a man/Therefore Socrates is mortal'. Less well-known is its treatment of cases where formally valid arguments cannot be based on syllogisms because of the presence of potentially unreliable observations. Aristotle accepted that a reasonable, if not formally valid, argument could be based on observations that suggest a conclusion is true more often than not. Elsewhere he gives the example of lactating women, pointing out that it is generally true that such women have given birth, but that this is not always the case.⁹

Aristotle went further, however, combining this approach with his conception of the nature of chance. In *On the Heavens*, he argues that the stars must all be fixed to some celestial sphere because if one assumes their orderly procession across the night sky is the result of mere chance, such movement would be extremely improbable (Franklin,⁷ pp.133–134). Such reasoning has obvious parallels with the modern practice of significance testing, according to which a finding is deemed 'statistically significant' if the probability of obtaining at least as impressive an effect by chance alone is deemed too low. Lacking the necessary mathematics, Aristotle made no attempt to

quantify this form of argument. But more surprisingly, he appears not to have noticed a basic logical error in his reasoning. He blithely assumes that because chance appears to have been ruled out as an explanation of the observations, his preferred (and, we now know, entirely incorrect) hypothesis about the existence of a celestial sphere must therefore be correct. The same logical error can be found throughout today's research literature, not least in clinical medicine where evidence against the null hypothesis is routinely taken as support for the specific hypothesis under investigation.

A second, more subtle, fallacy lies in Aristotle's misuse of a rule in propositional logic, first stated explicitly by Theophrastus (c. 371–287 BCE), Aristotle's successor as head of the Lyceum. Known as *modus tollens* ('denying the consequent'), it leads to valid inferences when used with indisputable premises or statements such as axioms, but becomes unreliable when applied to statements subject to uncertainty such as observations.^a The means of handling such cases emerged only 2000 years later in the form of Bayes' Theorem; even so, the unreliability of this form of argument can be shown without mathematics. In the case of his deduction of the reality of a celestial sphere, Aristotle argues that since independence of motion of the stars cannot give rise to the regularity of their procession across the night sky, they cannot be moving independently, and thus must be affixed to a celestial sphere. However, this use of *modus tollens* is fallacious because it involves an observational statement that is *not* indisputable: that independently-moving stars cannot give rise to regularity of motion. Indeed, we now know they *do* move independently, but lie at distances so great as to make this challenging to detect. In contrast, the apparent regularity is due to the Earth's rotation, the existence of which Aristotle denied. Again, the use of flawed inferential reasoning to extract insight from uncertain evidence remains common in the current research literature and is now regarded as a major threat to scientific enterprise.¹⁰

Rationalism v. empiricism

Another strikingly familiar and directly relevant debate broke out following the death of Aristotle: how best to reach reliable conclusions in medicine. The first known account of the debate was given by the Roman encyclopaedist Aulus Cornelius Celsus (c. 25 BCE–c. 50 CE) in *De Medicina*, the first complete textbook on medicine to be printed.¹¹ Celsus identified various schools of thought concerning the acquisition of reliable medical knowledge. The so-called Rationalists (sometimes also called

Dogmatists) held that medical practice is best guided by an understanding of the fundamental mechanisms of diseases and treatment. In contrast, the Empiricist school argued that reliable knowledge comes from simply observing large numbers of cases. Both approaches had obvious flaws. Rationalists were vulnerable to basic misconceptions about how the body worked, leading to misguided principles of treatment such as the balancing of humours. Empiricists, in contrast, eschewed speculation about mechanisms, which led them to use prevalences and correlations without reference to aetiological insight, making them vulnerable to bias and confounding. Yet, the rivalry between the schools provoked a dispute that remains of critical importance: how much evidence is sufficient to make a compelling case? The dispute is described in detail in the text *On Medical Experience* by the Greek physician Galen (130–210 CE), and centres on an argument used against the purely data-driven approach of the Empiricists. Given the latter school's insistence that truth is revealed through observing a phenomenon very many times, Galen reports the Rationalists' question: 'Can you tell us, Empiricists, how many times 'very many' times is?'.¹² The challenge was pressed home using a philosophical conundrum dating from the time of Aristotle known as the Sorites Paradox. Usually attributed to Eubulides of Miletus, its name comes from its original formulation: how many grains of sand suffice to form a heap ('σωρός')? Clearly one grain does not, nor does adding a few more. However, a million grains certainly does make a heap, so at some point the transformation takes place – but where? The Rationalists argued that the same applies to observations: a single observation is clearly not enough, while thousands may be – but where is the tipping point? Clinical medicine faces the same issue almost two millennia on. The requirement that evidence be 'compelling' or 'convincing' before being acted upon introduces the vague predicates that lead to the Sorites Paradox. The use of hard thresholds in the analysis of data, such as the $p < 0.05$ criterion for statistical significance, have helped conceal this difficulty. Nevertheless, its presence can be seen in the absurdity of results with $p = 0.049$ being regarded as real effects, while those with $p = 0.051$ are dismissed as 'null findings'. Despite this, the treatment of uncertainty in contemporary clinical research – in the guise of evidence-based medicine – remains closer to Empiricism than Rationalism, prompting concern about over-reliance on data divorced from insights about plausible mechanisms.¹³ The use of thresholds to create the illusion of definitive inferences from uncertain data has long been criticised by the statistical community and is

now under sustained attack.¹⁰ This in turn has prompted increased interest in Bayesian methods, which are seen as better at dealing with clinical uncertainty – not least because they allow both 'rationalist' and 'empiricist' sources of insight to be combined. As such, a principled resolution of this ancient epistemological dispute may now be in sight.

The shift towards quantification

The Sorites dispute suggests that by the start of the first millennium CE, the treatment of uncertainty in medicine was starting to shift away from purely qualitative considerations. It would take another two millennia before it became substantively quantitative, however. The transition received significant assistance from the establishment of the Abbasid Caliphate in 786 CE, which encouraged the translation and transmission of classical knowledge by Arab and Persian scholars writing in Arabic. Arguably, the two most important for medical science were Al-Kindi (c. 801–873 CE) and Ibn Sina (often known as Avicenna, c. 980–1037 CE).

Educated in Baghdad, Al-Kindi was tasked with supervising the translation of Greek and Greco-Roman texts in many fields and led him to build on classical knowledge. The outcome was over 270 texts, of which around 30 concerned medicine.¹⁴ Of these, the most important is *De gradibus* ('On degrees'), in which Al-Kindi attempts to express quantitatively Galen's vague eponymous method of describing the quality or impact of a remedy. While medieval scholars found the result difficult to understand, Pioreschi credits Al-Kindi with the first serious attempt at quantification in medicine.

The second key figure from the so-called Islamic Golden Age is Ibn Sina, whose five-volume *Canon of Medicine* (c 1012) draws on the works of Hippocrates, Aristotle and Galen.¹⁵ Like Al-Kindi, Ibn Sina also made important contributions of his own, most famously by stating seven rules for the testing of remedies. These explicitly address the problem of uncertainty arising from variability of response. The lingering influence of Aristotelian reasoning can be seen in the sixth rule, which states that a remedy must be effective either always or at least in many cases, as otherwise its action must be accidental since remedies with genuine, innate effectiveness 'act according to their natures either always or for the most part' (quoted in Franklin,⁷ p 178). However, Ibn Sina generally puts greater emphasis than Aristotle on the use of observational evidence rather than logical deduction, perhaps reflecting the influence of Quranic teachings on the value of such evidence. By doing so, Ibn Sina is regarded by some modern authors as the earliest exponent of evidence-based medicine.^{16,17}

It took another half-millennium for the next major development in the treatment of uncertainty in medicine to emerge: recognition of the role of *quantitative* measurement. The second part of this brief history will examine the impact of this transition and the controversies sparked by its combination with the emerging theory of probability.

Declarations

Competing Interests: None declared.

Funding: None declared.

Ethics approval: Not applicable.

Guarantor: RAJM.

Contributorship: Sole authorship.

Provenance: Invited article from the James Lind Library.

Acknowledgements: I am most grateful to Sir Iain Chalmers and his dogged persistence in asking me to address the issues covered in this paper, and to Prof Ulrich Tröhler for helpful discussions.

Note

- a. Modus tollens ('denying the consequent') is a form of valid argument in propositional logic. It states that if we have two statements A and B, both of which are either axioms or logical consequences of them, and A necessarily entails B, then the fact that B is false implies A is also false. So, for example, it is by definition axiomatically true that all even numbers are exactly divisible by 2. If some number X is not divisible by 2, it then follows from modus tollens that X is not even. But this fails with statements subject to uncertainty. For example, all patients infected with typhus have symptoms like fever. But it is clearly clinically absurd to argue that a patient with no symptoms logically cannot be infected: there is uncertainty surrounding factors such as plausibility of exposure, timing of infection and incubation period. Bayes' Theorem allows explicit incorporation of these external factors.

References

1. Louis PCA. *An Essay on Clinical Instruction*. London: S. Highley. See <https://archive.org/details/b22384285> (1834, last checked 6 April 2020).
2. Garber D and Zabell S. On the emergence of probability. *Arch History Exact Sci* 1979; 21: 33–53.
3. Hacking I. *The Emergence of Probability*. Cambridge: University Press, 2006.
4. Schiefsky MJ. *Hippocrates "On Ancient Medicine"*. Leiden: Brill, 2005, p.34.
5. Tsiompanou E and Marketos SG. *Hippocrates: Timeless Still*. *JLL Bulletin: Commentaries on the history of treatment evaluation*. See <https://www.jameslindlibrary.org/articles/hippocrates-timeless-still/> (2012, last checked 6 April 2020) (reprinted in *J Roy Soc Med* 2013 106: 288–292).
6. Sheynin OB. On the prehistory of the theory of probability. *Arch History Exact Sci* 1974; 12: 97–141.
7. Franklin J. *The Science of Conjecture: Evidence and Probability before Pascal*. Baltimore: Johns Hopkins University Press, 2001.
8. Aristotle *Prior Analytics* (trans A J Jenkinson). See <http://classics.mit.edu/Aristotle/prior.html> (350 BCE, last checked 6 April 2020).
9. Burnyeat MF. *Explorations in Ancient and Modern Philosophy*. (Vol. 1). Cambridge: Cambridge University Press, 2012: Ch 6.
10. Wasserstein RL and Lazar NA. The ASA's statement on p-values: context, process, and purpose. *Am Stat* 2016; 70: 129–133.
11. Howick J. Aulus Cornelius Celsus and 'empirical' and 'dogmatic' medicine. *JLL Bulletin: Commentaries on the history of treatment evaluation*. See <https://www.jameslindlibrary.org/articles/aulus-cornelius-celsus-and-empirical-and-dogmatic-medicine/> (2016, last checked 6 April 2020).
12. Tuominen M. Heaps, experience, and method: on the Sorites argument in ancient medicine. *History Philos Q* 2007; 24: 109–125.
13. Webb W. Rationalism, empiricism, and evidence-based medicine: a call for a new Galenic synthesis. *Medicines* 2018; 5: 40.
14. Pioreschi P. Al-Kindi, a precursor of the Scientific Revolution. *J Int Soc Hist Islamic Med* 2002; 2: 17–19.
15. Nasser M, Tibi A and Savage-Smith E. Ibn Sina's Canon of Medicine: 11th century rules for assessing the effects of drugs. *JLL Bulletin: Commentaries on the history of treatment evaluation*. See <https://www.jameslindlibrary.org/articles/ibn-sinas-canon-of-medicine-11th-century-rules-for-assessing-the-effects-of-drugs/> (2007, last checked 6 April 2020) (reprinted in *J Roy Soc Med* 2009 102: 78–80).
16. Shoja MM, Rashidi MR, Tubbs RS, Etemadi J, Abbasnejad F and Agutter PS. Legacy of Avicenna and evidence-based medicine. *Int J Cardiol* 2011; 150: 243–246.
17. Akhondzadeh S. Avicenna and evidence based medicine. *Avicenna J Med Biotechnol* 2014; 6: 1.