**Constructing diagnostic likelihood: clinical decisions using subjective versus statistical probability**

John Kinnear,1,2 Ruth Jackson1

1. Postgraduate Medical Institute, Anglia Ruskin University
2. Department of Anaesthetics, Southend University Hospital NHS Foundation Trust

*Corresponding author:*

John Kinnear

Postgraduate Medical Institute

Anglia Ruskin University

Bishop Hall Lane

Chelmsford

CM1 1SQ

Email: [John.Kinnear@southend.nhs.uk](mailto:John.Kinnear@southend.nhs.uk)

**Abstract**

**Background**

Although physicians are highly trained in the application of evidence-based medicine, and are assumed to make rational decisions, there is evidence that their decision-making is prone to biases. One of the biases that has been shown to affect accuracy of judgments is that of representativeness and base rate neglect, where the saliency of a person’s features lead to overestimation of their likelihood of belonging to a group. This results in the substitution of ‘subjective’ probability for statistical probability.

**Methods**

This study examines clinicians’ propensity to make estimations of subjective probability when presented with clinical information that is considered typical of a medical condition. The strength of the representativeness bias is tested by presenting choices in textual and graphic form. Understanding of statistical probability is also tested by omitting all clinical information.

**Results**

For the questions that included clinical information, 46.7% and 45.5% of clinicians made judgments of statistical probability respectively. Where the question omitted clinical information, 79.9% of clinicians made a judgment consistent with statistical probability. There was a statistically significant difference in responses to the questions with and the question without representativeness information (P < 0.0001).

**Conclusions**

Physicians are strongly influenced by a representativeness bias, leading to base rate neglect, even though they understand the application of statistical probability. One of the causes for this representativeness bias may be the way clinical medicine is taught where stereotypic presentations are emphasized in diagnostic decision-making.

**Introduction**

It is said that ‘life is a school of probability’,1 because all aspects of existence are subject to statistical principles. A good understanding of these principles is particularly relevant to the science-based professions. Effective healthcare depends on the application of evidence-based medicine, at the heart of which is Bayesian probability. Unsurprisingly then, instruction in evidence-based medicine and statistics is integral to most medical curricula. It is therefore natural to expect that clinical decision-making, being steeped in this scientific tradition, should be rational, predictable, and rules-based. However, evidence does not support this expectation.

The theory of clinical decision-making derives largely from cognitive psychology and describes reasoning as a dual systems process, where system 1 is intuitive, automatic, effortless and instant (sometimes described as pattern recognition), while system 2 is analytical, effortful, and deliberative (also described as hypothetico-deductive).2 These systems do not function independently, and are moderated by each other in a more or less effective way depending on the influences of various biases.3

Making a judgment of the likelihood of the presence of a disease or stateis important at several levels. Since no test or treatment is without hazard a physician must constantly weigh up risks in the face of uncertainty, i.e. assess her degree of belief that the patient has a condition, knowing when to test or treat, and providing information so that the patient can make rational choices. Various probabilistic approaches have been described to optimise making effective diagnoses, i.e. improving the physician’s judgment of the likelihood of the presence of a disease or state.4-7 One such approach is the ‘threshold approach’, which describes the thresholds at which a test is ordered or a treatment commenced. However, application of the threshold approach to decision-making relies on a common understanding of what probability is as a concept, and then to apply this approach requires the estimation of the statistical probability in an accurate and unbiased way.

There are several non-clinical influences which impact on clinical decision-making, sometimes consciously when clinicians use their expert judgment based on personal experience to overrule evidence-based practice, and sometimes unconsciously when decisions are shaped by bias.8 One such unconscious process is ‘subadditivity’, where physicians tend to overestimate probability when asked to assess the likelihood of different differential diagnoses, their summated estimations adding up to greater than 100%, when logically the maximum should be 100%.9

Several studies have explicitly questioned clinicians’ understanding of the practical application of statistical principles, especially statistical probability, to their practice, and they consistently find that clinicians show a lack of basic understanding of how to apply statistics to patient care.8,10-17 Their prediction of the presence of a disease improves with clinical experience,18 but it is not clear whether this is matched by an improved understanding of probability?

A characteristic of system 1 processes is that they require minimal information to make confident assumptions, using a heuristics-based approach to achieve rapid decisions. A heuristic is a mental shortcut that bypasses step-by-step deliberation, to achieve an adequate answer in a complex environment. Under many clinical conditions, these heuristics allow fast and efficient thinking by allowing effective decisions to be made with limited information. This is the ‘mastery’ endpoint of deliberate practice. It is not about optimising diagnostic accuracy, but about being effective, and heuristics achieve this by ignoring information in excess of what is needed.10 Physicians have taken advantage of this fact by developing formal heuristics as algorithms, which have the added advantage of applying rules consistently, without the fallability of variable human judgment.19 For individual physicians to achieve accuracy using heuristics requires a regular (high validity) environment and the opportunity for repeated practice.20 Part of this environmental validity is its statistical structure, i.e. how commonly the event is likely to occur. In a low validity environment heuristics are prone to fail, when they become cognitive biases that bypass the rules of logic and result in inaccurate assumptions.21,22

One such bias, where physicians follow a logic of representativeness rather than a Bayesian logic, statistical probability is confused with subjective probability.23 Representativeness is the bias where the subjective likelihood of an event is judged by its similarity to a parent population, or by the saliency of its features. This leads to systematic errors of judgment based on stereotypic representation in contrast to normative likelihood.24

The representativeness heuristic was first described by Kahneman and Tversky,24 who describe it as also giving rise to the base rate fallacy (or base rate neglect). When presented with a general state (the base rate), and a particular state, the representativeness bias will result in overestimation of the likelihood of the latter. The particularity of the state commonly includes irrelevant information that is mistakenly believed to have relevance because it is associated with that state. However, logically it is a necessary condition that the particular state is included within the group that exists in the more general state.25-27 For example, a tall, dark, handsome stranger is more likely to belong to a group of people with a) hair; than b) dark hair, since the former group includes the latter. This bias, applied in a clinical context, will lead to an estimation of how stereotypical a patient is for a diagnosis, rather than to how statistically likely it is. There are several clinical implications, including over-diagnosis and over-treatment of disease resulting from over-estimation of disease frequency.28

A clinical example of suboptimal probabilistic thinking would be the diagnosis of Ebola virus infection in a tourist who has developed a pyrexial illness and recently returned from West Africa. Such a patient is far more likely to be suffering an influenza-like illness (approximate 5% chance) than Ebola, which has a vanishingly small incidence in this population group. However, the salience of a highly publicised epidemic and the associated limited (but typical) information may trigger representativeness and undermine probabilistic reasoning.

The representativeness bias has been described in the clinical psychology literature as being dominant over the rules of logic, but it is not clear if this dominance is displayed in professionals who are explicitly trained to apply evidence and the rules of logic to their decision-making. Doctors are one of the few professions where the importance of the base-rate is emphasised in their training. Undergraduates are exhorted to remember that ‘common things are common’, and not to diagnose exotic diseases in favour of more common ailments. However, their training equally emphasises stereotypical presentations of diseases in order to instil the ‘pattern recognition’ that makes rapid diagnosis possible. These distinct goals of learning seem at odds and this study aims to test the presence and the strength of the representativeness bias in practicing clinicians, as well as an applied understanding of base-rate.

**Ethics**

Full ethics approval for the study was granted by the Faculty of Medical Science Research Ethics Panel for Anglia Ruskin University (FREP no. 15/16 085) in June 2016. The survey questionnaire included a summary description of the study aims. It also included a statement that guaranteed full anonymity for all participants and confirmed that data would only be presented in aggregated form. No clinical information related to real patients. A contact email was provided in the event of any queries related to the study. In view of the assessed low risk and voluntary participation no explicit consent was obtained from participants.

**Methodology**

Postgraduate doctors were sampled using a regional database of physicians in the UK. A total of 1050 questionnaires were distributed electronically. Three questions were administered sequentially using a proprietary electronic platform. The separation and sequencing of the questions ensured that later questions did not ‘contaminate’ judgments on the earlier questions. The three different questions were designed with the same logical construction, where respondents were asked to rank the likelihood of a patient belonging to a given group, where probabilistically one group wholly included another. The addition of items of clinical information had the effect of making a choice statistically less likely, but at the same time added saliency for that particular condition.

The information presented in the diagnostic groups was intentionally incomplete, since the application of heuristics is insensitive to either the quantity or the quality of information on which to construct a decision. Although all the response choices had the same logical relationship, the different questions were framed such that they tested different end points. Questions 1 and 2 included clinical information characteristic of the conditions they asked about. Their function was to trigger the representativeness heuristic, while question 3 contained no clinical information.

The objective of question 1 (figure 1) was to test for the presence and strength of a representativeness bias. Respondents were asked to assign a patient with a specified diagnosis (myocardial infarction) to categories of clinical findings that, in their opinion, had the highest likelihood of representing that diagnosis. Limited clinical information is purposely given to trigger type 1 processing, which is susceptible to representativeness. The ‘critical’ items in the question were those that overlapped such that one set of clinical findings entirely included the other, which in this case were items B and D.

Question 2 (figure 2) also gives a clinical choice using the same representativeness construction, but presents the choice such that the statistical relationship is more obvious. The two possibilities are graphically represented as a Venn diagram where one group is wholly included in the other. The objective was to test whether this graphic illustration would influence the representativeness bias, if any was present, with the hypothesis being that although the added clinical features should trigger a representativeness heuristic, the graphical relationship would dominate the decision-making.

For question 3 (figure 3) all representativeness information is removed, and the decision only requires application of the rules of statistical probability. For the first two questions an assumption is made that doctors understand the principle of statistical probability, but by removing all clinical information, question 3 tests this assumption.

**Statistical Analysis**

Statistical calculations were performed using SPSS 20.0. The level of significance was set at P ≤ 0.05 and the X2 test was used to compare categorical variables.

**Results**

Of the 1050 distributed questionnaires, there were 24 automated responses indicating failed delivery. This left a total of 1026 questionnaires administered for each question. The number of respondents for questions 1, 2 and 3 were 259, 255 and 254 respectively, giving response rates of 25.2%, 24.9% and 24.7%.

Figure 4 illustrates the distribution of respondents who chose the correct answer based on statistical probability (P choice), and the incorrect answer based on representativeness (non-P choice). Of the 259 respondents for the first question, 121 (46.7%) ranked the answer based on probability higher; of 255 respondents for question 2, 116 (45.5%) chose the answer based on statistical probability; while In question 3, where all representativeness information was removed, of the 254 respondents 203 (79.9%) chose the statistically more likely answer.

When the chi-square test was performed to examine the relationship between P and non-P choices for questions 1 to 3, the relationship was non-significant in Question 1, X2 (1, *N* = 259) = 0.107, *p* = 0.744; and Question 2, X2 (1, *N* = 255) = 0.237, *p* = 0.626; but was highly significant for Question 3, X2 (1, *N* = 254) = 54.45, *p* < 0.0001.

**Discussion**

There is often a mistaken impression that physicians operate within a model of *unbounded rationality*, that is, strive for maximisation of expected utility by collecting all information at their disposal, dispassionately evaluating it, and applying it to a well-defined and objective goal to achieve an optimal outcome.10 This model suffers from being information greedy for decision-makers (humans) that have a notable limitation of cognitive capacity. Even recognition of imperfect knowledge in the decision-maker tends to only constrain irrationality rather than block it. In reality, humans display systematic errors of judgment owing to numerous cognitive biases, and physicians are no less prone to these deviations from rational optimisation.

One of these systematic errors is estimating probabilities by stereotyping, or representativeness. This is an irrational decision-making strategy since it serves no useful purpose, which should be distinguished from functional heuristics, which serve the purpose of fast and frugal decision making, particularly in pressured, complex environments – a type of *ecological rationality*.10

A high proportion of the doctors studied were strongly influenced by the representativeness bias when making a diagnosis based on clinical features. When offered a range of responses based on limited clinical information, the majority ranked a stereotypical presentation as more likely despite it being statistically less probable based on a lower base-rate occurrence. Framing the question in the form of a Venn diagram where the logical relationship was presented graphically had no effect on judgment, where representativeness still dominated and led to a decision of subjective likelihood in a similar majority of respondents. Instead of constructing their decision-making around a question of statistical likelihood, which is a rules-based question of base-rate occurrence, they were more likely to ask how closely the clinical feature complex matched the diagnosis based on representativeness. These are different questions, and may relate to the way clinical decision-making is learned.

Thinking is believed to be the result of two distinctive processing types - type 1 processes are rapid and autonomous, and characterized by offering default responses; and type 2 processes are capable of reflecting on the system 1 suggestions and either modifying or rejecting them in favor of a more or less rationalised response. The former results in minimal cognitive load, whereas the later is load intensive.29 Stanovich makes a further useful separation of type 2 processes into two distinct *modes* of thinking, which describe *thinking dispositions*.30 The modes share the characteristic of always requiring working memory, but the styles of thinking occupy a continuum of slow and careful to quick and casual, and they constantly vary.29 Conceptually therefore, there are a trio of ‘minds’ – the autonomous mind (type 1 processes), the algorithmic mind (responsible for complex computation and traditionally thought of as responsible for innate intelligence), and the reflective mind (responsible for weighing choices). These minds display a hierarchical structure of control. The autonomous mind can be overridden by the algorithmic mind, which in turn is subordinate to the reflective mind, the highest level of regulation. This tripartite model goes some way to explain ‘irrationality’ in otherwise intelligent people, with the reflective mind as its locus. Intelligence and rationality are separate constructs, and whereas the former is the domain of the algorithmic mind, the latter is firmly the province of ‘reasoning’.31

The representativeness heuristic is the domain of system 1 thinking. It is up to the reflective mind of system 2 to apply logical rationalisation to the suggested solution, but it fails for several reasons. These may include ‘laziness’ of system 2 (an energy minimising strategy), fatigue, ‘ego depletion’, priming through prior exposure, or an incorrect learned response. The last may be amenable to educative intervention. Type 1 processes tend to be associative and are responsible for implicit learning. They therefore give rise to the automatic responses of highly learned associations, where the associations that result from repeated experiences are dominant to the point of automaticity.20

Although there are advantages to the ‘fast and frugal’ approach to clinical decision-making that heuristics allow, there are also potential dangers that clinicians need to be aware of. Fitting patients into stereotypic clinical categories runs the risk of misdiagnosing those who present atypically, and ‘closing’ the diagnosis prematurely without considering alternative possibilities. An example in anaesthetic practice is to not consider accidental awareness under anaesthesia when the typical signs of tachycardia, sweating and lachrymation are absent. In this example, patients who are aware are far more likely to belong to the group of patients having general anaesthetic rather than those having general anaesthetic and displaying these ‘typical’ signs. Representativeness constrains clinical decisions to prototypical presentations. From a teaching and learning perspective it is important to make a clear distinction between the application of base-rate logic, which is a probability question, and recognising distinctive clinical presentations.

Limitations of this study are that the survey questions had a relatively low response rate, which carries the risk that the samples are not representative of the population of UK doctors. We did not use unique user identifiers for the surveys, and although we have made an assumption that the respondents for the three questions were the same participants, we cannot verify this.

**Conclusions**

Decision-making is not always a rational process. Despite physicians being trained in the scientific tradition, their clinical judgments are strongly influenced by a representativeness bias, which makes them vulnerable to making systematic errors of judgment. Base rate neglect is well described in lay public as well as academically astute groups, but its relevance to medical decision-making is only emerging. This study indicates that statistical probability is one of those elements of medical curricula that is frequently talked about but rarely ‘done’. Although we have shown evidence for representativeness bias and base rate neglect in physicians at a cognitive level, the next step is to determine if these impact on dynamic decision-making in a simulated clinical environment, which will be a more proximate evaluation of authentic clinical decision-making.

**Main messages**

* Doctors are prone to a representativeness bias in clinical decision-making.
* Clinicians understand the application of the base rate principle where representativeness is not triggered.
* Cognitive ‘intelligence’ is often a dominant selection criterion to study medicine, but does not protect against ‘irrationality’.

**Current research questions**

* Does representativeness affect dynamic decision-making in authentic clinical environments?
* Can modes of thinking be identified and used as a basis for selection to study medicine?
* Does the teaching method contribute to the representativeness heuristic?

**Figure legends**

**Figure 1.** The question and response choices, with the ‘critical’ elements being items B and D.

**Figure 2.** A Venn diagram illustrating the logical relationship between the choices, where Group A wholly includes Group B.

**Figure 3.** Question with no representativeness (clinical) information.

Figure 4. Responses to all questions where the P choice denotes the correct probability choice, and non-P choice the incorrect choice. In questions 1 and 2 the non-P choice is the representativeness choice.

**References**

1. AC Grayling. (Proverbs 99:7) The Good Book. A Secular Bible. London: Bloomsbury.
2. Djulbegovic B, Hozo I, Beckstead J, *et al*. Dual processing model of medical decision-making. *BMC Medical Informatics and Decision Making* 2012;12:94.
3. Kahneman D. Thinking, Fast and Slow. London: Penguin Books, 2011.
4. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med* 1980;302:1109–17.
5. Sox HC Jr., Blatt MA, Higgins MC, *et al*. Medical decision making. Stoneham MA: Butterworth-Heinmann, 1988.
6. Diamond GA, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. *N Engl J Med* 1979;300:1350–8.
7. Diamond GA, Forrester JS, Hirsch M. Application of conditional probability analysis to the clinical diagnosis of coronary artery disease. *J Clin Invest* 1980;65:1210–21.
8. Hajjaj F, Salek M, Basra M, *et al*. Non-clinical influences on clinical decision-making: a major challenge to evidence-based practice. *J R Soc Med* 2010;103:178–187.
9. Cahan A, Gilon D, Manor O, *et al*. Probabilistic reasoning and clinical decision-making: do doctors overestimate diagnostic probabilities? *Q J Med* 2003;96:763–769.
10. Marewski J, Gigerenzer G. Heuristic decision making in medicine. *Dialogues Clin Neurosci* 2012;14:77-89.
11. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, *et al*. Helping doctors and patients make sense of health statistics. *Psychol Sci Public Interest* 2007;8:53-96.
12. Gigerenzer G, Hoffrage U, Ebert A. AIDS counselling for low-risk clients. *AIDS Care* 1998;10:197- 211.
13. Casscells W, Schoenberger A, Grayboys T. Interpretation by physicians of clinical laboratory results. *N Engl J Med* 1978;299:999-1001.
14. Eddy DM. Probabilistic reasoning in clinical medicine: problems and opportunities. In: Kahneman D, Slovic P, Tversky A (eds). Judgment Under Uncertainty: Heuristics and Biases. Cambridge: Cambridge University Press, 1982;249-267.
15. Ghosh AK, Ghosh K. Translating evidence-based information into effective risk communication: current challenges and opportunities. *J Lab Clin Med* 2005;145:171-180.
16. Hoffrage U, Lindsey S, Hertwig R, *et al*. Communicating statistical information. *Science*. 2000;290:2261-2262.
17. Young JM, Glasziou P, Ward JE. General practitioners’ self rating of skills in evidence based medicine: a validation study. *BMJ.* 2002;324:950- 951.
18. Cipoli D, Martinez E, de Castro M, *et al*. Clinical judgment to estimate pretest probability in the diagnosis of Cushing’s syndrome under a Bayesian perspective. *Arq Bras Endocrinol Metab*. 2012;56:633-637.
19. Green L, Mehr DR. What alters physicians’ decisions to admit to the coronary care unit? *J Fam Pract.* 1997;45:219–226.
20. Kahneman D, Klein G. Conditions for intuitive expertise: A failure to disagree. *American Psychologist* 2009;80:237–251.
21. Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. *Science* 1974;185:1124–31.
22. Redelmeier DA, Koehler DJ, Liberman V, *et al*. Probability judgment in medicine: Discounting unspecified possibilities. *Med Decis Making* 1995;15:227–30.
23. Bar-Hillel M. Representativeness and fallacies of probability judgment. *Acta Psychologica* 1984;55: 91-107.
24. Kahneman D, Tversky A. Subjective probability: A judgment of representativeness. *Cognitive Psychology* 1972;3:430-454.
25. Kahneman D; Tversky A. Evidential impact of base rates. In Daniel Kahneman, Paul Slovic & Amos Tversky (eds.). Judgment under uncertainty: Heuristics and biases. 1985, 153–160.
26. Bar-Hillel M. The base-rate fallacy in probability judgment. *Acta Psychologica* 1980;44:211–233.
27. Kahneman D, Tversky A. On the psychology of prediction. *Psychological Review* 1973;80:237–251.
28. Payne V, Crowley R. Assessing Use of Cognitive Heuristic Representativeness in Clinical Reasoning. *AMIA Annu Symp Proc*. 2008;2008:571–575.
29. Evans J, Stanovich K. Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science* 2013;8:223–241.
30. Stanovich KE, West RF, Toplak ME. The complexity of developmental predictions from dual process models. *Developmental Review* 2011;31:103–108.
31. Stanovich KE. Rationality and the reflective mind. New York, NY: Oxford University Press. 2011.