Diagnosis in medicine, as is well known, is the process of inferring a disease causing illness from data in a patient with symptoms. This inference is statistical for the most part as the relationship between data and disease is usually not unique. The goal in diagnosis is to infer this disease accurately so that it can be treated appropriately. It follows that the method employed for statistical inference (henceforth called 'inference' for brevity) in diagnosis needs to be such that it infers the disease accurately. We shall argue in this paper that the Bayesian method, which has been prescribed as the normatively correct method of inference in diagnosis (1) in theory is not such a method.

In the Bayesian method, the pretest probability of a disease suspected from a presentation (initially available information such as patient's age, sex, symptoms, risk factors etc.), which is usually derived from its prevalence, is combined with a likelihood ratio (LR) of a test result to generate a post-test probability. The post-test probability represents a degree of belief from which the disease is inferred (1).This inference is considered to be coherent, as it is equivalent to a bet made with odds based on the post-test probability, which prevents a Dutch book, that is a series of bets which ensures loss, from being created against the inferring agent. Due to this coherence, the Bayesian method is believed to be rational (2). Thus the Bayesian method is prescribed for inference during diagnosis due to its rationality based on its coherence. But what interests a practicing physician more than this coherence based rationality is if the Bayesian method leads to an accurate inference of a disease in a patient. This issue appears to be unsettled, for a well- known Bayesian statistician, Dennis Lindley, has written, " I am often asked if the (Bayesian) method gives the right answer: or, more particularly, how do you know if you have got the right prior. My reply is that I don't know what is meant by 'right' in this context. The Bayesian theory is about coherence, not about right or wrong"(3).

We shall now examine the performance of the Bayesian method for inference during diagnosis in a real patient who was presented and discussed in a clinical problem solving exercise (4). The patient is a 40 year old healthy woman without any cardiac risk factor who presents with highly uncharacteristic chest pain. The disease, acute myocardial infarction (MI) is suspected from this presentation and a test, an EKG, performed which reveals acute ST elevation EKG changes. In the Bayesian method, the pretest probability of acute MI, which is estimated to be 7 percent from its prevalence, is combined with the LR of 13 (5) for acute ST elevation EKG changes to generate a post-test probability of 50 percent (Appendix 1). This post-test probability represents an equivocal degree of belief, based on even odds of a bet placed on acute MI, from which it is inferred to be indeterminate in this patient. The discussing physician in this exercise however does not infer in this Bayesian manner. Instead, he infers acute MI conclusively (and accurately) from the test result, acute ST elevation EKG changes with LR of 13 alone which he considers to be strong evidence based on its known performance in inferring acute MI accurately in 86 percent patients with varying pretest probabilities (6). We note, the Bayesian method, while it may be coherent, is not accurate and is therefore not employed by the discussing physician for inference in this patient.

We propose the method employed for inference in this patient and in all patients in practice is the confidence frequentist method (7), which is the other major method of statistical inference (other than the Bayesian method). In this method, a pretest probability is not attached to a suspected disease (as it is in the Bayesian method), so that there is no pretest degree of belief in the suspected disease. This disease is inferred by a procedure which has a high probability of leading to an accurate inference. This procedure consists of performing a test and inferring the disease from a highly informative test result (LR greater than 10) (8) which has a high frequency, 85 percent or higher, of accurate inference of the disease in patients with varying pretest probabilities. In this method, the focus is on high inferential accuracy, which is why, we suggest, it is employed for inference during diagnosis in practice.

The application of the confidence method for inference during diagnosis in practice is made possible by the fact, we suggest, that different patients with varying pretest probabilities in whom we suspect a given disease, acute MI for example, over a period of time present to us purely by chance in no particular order. Thus we may encounter a patient with low pretest probability followed by two patients with high pretest probabilities or in any other sequence. Therefore this series of patients can be looked upon, we propose, as a random sample which is drawn from a population of patients with varying pretest probabilities, in whom we suspect acute MI, and which is known to have some patients with acute MI. Therefore, the observed frequency of acute MI of 86+/-2 percent (2 standard deviations, 95 percent probability) in patients with acute ST elevation EKG changes in the given series (6) is distributed normally in samples (series) drawn from this population with it being between 84 and 88 percent in 95 percent samples (series) by the Central Limit Theorem (9). We can thus say with a confidence level of 95 percent in a given patient with acute ST elevation EKG changes that this patient is drawn from a series in which the frequency of acute MI in patients with acute ST elevation EKG changes is between 84 and 88 percent which allows us to infer acute MI with diagnostic accuracy of 84 to 88 percent in this patient.

It is seen that acute MI is inferred from acute ST elevation EKG changes in practice with the same diagnostic accuracy of 84 to 88 percent by the confidence method in every patient regardless of its pretest probability. This is in sharp contrast to the Bayesian method in which this inference differs from patient to patient as it is made from a varying post-test probability that is generated from a varying pretest probability in different patients. In addition, the inferential accuracy of a confidence inference is known to be high at 84 to 88 percent, while the inferential accuracy of a Bayesian inference is unknown as it is made from a post-test probability which represents a subjective degree of belief.

It is for this reason, we believe, that a disease which has a test capable of generating a highly informative result (LR greater than 10) (8) is inferred by the confidence method in practice. For example, covid-19 disease is inferred from positive covid-19 test, LR 14 (10), pulmonary embolism from positive chest CT angiogram, LR 20 (11) and deep vein thrombosis from positive venous ultrasound, LR 16 (12) in any patient regardless of pretest probability of the disease by the confidence method in practice. The confidence method is employed for inference of a disease in all published diagnostic exercises in real patients such as in clinicopathologic conferences (CPCs) and clinical problem solving exercises

(13,14). We could not find the Bayesian method to be employed for inference in any of these exercises.

Apart from its highly accurate inference of a disease in patients with varying pretest probabilities, which is its main advantage, the confidence method has several other advantages over the Bayesian method as follows:

(1) As the inferential accuracy of a disease is independent of pretest probability, it appears to be the same all over the world in the confidence method despite varying prevalence of a given disease in different countries which is bound to influence pretest probability. For example, acute MI is inferred from acute ST elevation EKG changes and a patient sent to a cardiac catheterization laboratory for treatment in a similar manner in every country where such a facility is available (15). In the Bayesian method, on the other hand, we expect the inference of acute MI from acute ST elevation EKG changes to differ from country to country depending on its prevalence which has not been observed indicating this method is not employed in practice.

(2) The goal in diagnosis is to infer if a disease is present (or absent) in a given patient. In the confidence method, this inference is easily made by employing a rule such as 'Infer a disease (acute MI), if a test result with LR greater than 10 (acute ST elevation EKG changes) is observed in any patient regardless of pretest probability of disease'. In the Bayesian method, it is not clear how high the post-test probability (how strong degree of belief) needs to be before a disease is inferred to be present in a patient which introduces ambiguity in inference.

(3) The confidence method introduces a clear division between roles of a presentation and a highly informative test result (LR greater than 10) in inference of a disease. A presentation only makes us suspect a disease which is formulated as a diagnostic hypothesis while a highly informative test result alone provides evidence from which a disease is inferred in any patient regardless of pretest probability. Specifically, a presentation is not a source of pretest evidence in this method.

This division places all things that happen by chance about which we do not know anything and which differ from patient to patient, in a presentation, and a law like, more or less invariant action in a highly informative test result. Thus we do not know why acute MI occurs in a 40 year old healthy woman in whom its pretest probability is very low or in a 65 year old man in whom its pretest probability is very high. It is the law like action of a highly informative test result such as acute ST elevation EKG changes which makes accurate inference of acute MI possible in every patient regardless of pretest probability. This division is similar, we believe, to the division in physics between initial conditions such as position, velocity of a body which vary from situation to situation and a law such as Newton's second law of motion which is invariant in its action in every situation regardless of initial conditions. And just as this division is recognized as a key factor in successful application of physics to natural phenomena (16), we suggest, this division is an important factor in successful inference by the confidence method during diagnosis.

In the Bayesian method, there is no such division between a presentation and a test result as both are considered to be sources of evidence. The absence of this division creates confusion in inference of a disease in different patients. Let us assume that a post-test probability of 86 percent represents strong degree of belief from which a disease, acute MI, for example, is inferred in a patient in this method. We find that a test result with LR of 75 is required to generate this post-test probability in the 40 year old woman mentioned above in whom the pretest probability is 7 percent (Appendix 2). And a test result with LR of only 1 (such as non-specific T wave EKG changes) is required in a 65 year old man with multiple cardiac risk factors and highly characteristic pain in whom the pretest probability, let us say, is 86 percent to generate a post-test probability of 86 percent (Appendix 3). In other patients in whom the pretest probability varies from 7 to 86 percent, a test result with LR varying from 75 to 1 is required to generate a post-test probability of 86 percent. We find that test results with different LRs are required to infer acute MI with the same high degree of belief in different patients depending upon pretest probabilities in these patients. We note a distinction is not made between a highly informative test result such as ST elevation EKG changes, LR 13 and a non-informative, worthless test result such as non-specific T wave EKG changes, LR 1 in the Bayesian method. The Bayesian approach is thus impractical and likely to lead to inferential errors, which is why, we believe, it is not employed in practice.

(4) The confidence method is open-ended in the sense that any number of diseases with whatever pretest probabilities can be suspected from a presentation initially and formulated as diagnostic hypotheses in it. And if one of these diseases is not found to be present on testing, some other disease can be suspected, tested and accurately inferred at a later date. This is seen in the case of the well-known country singer, Kris Kristofferson, who developed memory loss. (17). A number of diseases were initially suspected and tested but none was accurately inferred. Many years later, another physician suspected Lyme disease, tested it and correctly inferred it to be present. This disease was successfully treated and Kris Kristofferson was able to resume his singing career after 3 weeks.

In the Bayesian method, on the other hand, the pretest probabilities of all initially suspected diseases are required to add up to 1 and the disease which a patient has, is assumed to be among these diseases (18). If none of the initially suspected diseases is found after testing, there is no provision in this method for suspecting some other disease and testing it later. Thus the correct inference (diagnosis) of Lyme disease in Kris Kristofferson could not have been made by the Bayesian method.

(5) The confidence method of inference bears a close resemblance to the scientific method (19), which is universally recognized as the most reliable method of investigation and inference in any field. For example, in his investigation of the cause of explosion of space shuttle Challenger, which occurred in 1986, the eminent physicist, Richard Feynman (20) suspected malfunction of rubber O- ring valve due to cold weather at time of shuttle

launch as the cause of explosion from study of available information. He formulated this suspicion as a hypothesis and tested it by performing the experiment of dipping a replica of rubber 0-ring in a glass of ice cold water (on television) which revealed the O-ring to lose its resilience. From this experimental result, he correctly inferred O-ring malfunction to be the cause of shuttle explosion.

We note the method of Feynman's investigation and inference to be closely similar to inference of acute MI in the 40 year old woman by the discussing physician in the diagnostic exercise mentioned above. The disease, acute MI is suspected from the presentation, and formulated as a diagnostic hypothesis for which a test, an EKG, is performed. The highly informative test result, acute ST elevation EKG changes, is considered strong evidence based on its performance of having the high frequency of 86+/-2 percent of accurate inference of acute MI in other patients, from which acute MI is accurately inferred in this patient. The Bayesian inference of acute MI being indeterminate in this patient cannot be considered scientific by any stretch of imagination.

The situation that is faced in reality in practice in diagnosis is that any given disease is known to occur in different patients with varying presentations and thus with varying pretest probabilities, with our goal being to infer a disease accurately in every patient. In this situation, it does not appear to be very helpful, we believe, to look upon a pretest probability as degree of pretest belief in presence or absence of a disease, as is done in the Bayesian method. In this method, a very low pretest probability of a disease is interpreted as strong degree of pretest belief against this disease which may lead to it not being suspected or tested causing a diagnostic (inferential) error. The Bayesian method thus appears to encourage diagnostic errors due to failure to suspect a disease with an atypical presentation (low pretest probability) which have been reported in several studies (21,22)). The accurate diagnosis (inference) of diseases with low pretest probabilities (diseases which are rare or have highly atypical presentations) in published diagnostic exercises (13,14) in real patients is made possible to a great extent, we believe, by not interpreting the low pretest probabilities of these diseases as strong pretest degree of belief against these diseases in the confidence method which is employed in these exercises. In these exercises and in practice in general, a pretest probability is interpreted, we suggest, as chance of a disease in a patient and not as a degree of belief. Its only role in diagnosis, we suggest, is in prioritizing testing of various suspected diseases in a differential diagnosis in a non-urgent diagnostic situation. The disease with the highest pretest probability is tested first as it has the greatest chance of being present.

We find the availability of a test capable of generating a result with LR greater than 10 plays a key role in increasing inferential (diagnostic) accuracy in patients with varying pretest probabilities in the confidence method. This is seen most clearly in the case of pulmonary embolism whose accurate inference in patients with varying pretest probabilities increased dramatically with availability of perfusion lung scan (23) and chest CT angiogram (11) which can generate such results. In the case of diseases, which do not have such tests, they are inferred, we believe, from combination of two or three test results whose combined LR is greater than 10, but this matter needs to be investigated further.

We cannot think of a single feature of the Bayesian method which promotes inferential accuracy during diagnosis while it has several features, discussed above, which encourage inferential errors. We have not found any published case report or series of patients in which the Bayesian method has been employed for inference of a disease. The non-role of the Bayesian method in diagnosis in practice has been well summarized, with which we agree completely, by the eminent clinical investigator, Alvan Feinstein in 1977, who wrote: “I know of no published work in which the initial claims of a Bayesian enthusiast have been confirmed by the results found in clinical reality. I know of no clinical setting or institution in which the Bayesian diagnostic methods are being regularly used for practical diagnostic purposes in a routine or specialized manner. I know of no specific constructive, practical diagnostic decisions-involving real patients, data and doctors-in which the Bayesian methods have made a prominent contribution that could not have been achieved as easily without Bayes’ formula. (If readers know of any, I hope they will tell me)” (24).

By contrast, the confidence method. in which a suspected disease is formulated as a diagnostic hypothesis without any pretest probability attached to it and inferred with high inferential accuracy from a highly informative test result (LR greater than 10), is found to be employed in all published diagnostic exercises in real patients and in practice in general. The overall inferential accuracy of the confidence method is very high at 85 to 90 percent in practice in general (25) and 98 percent in CPCs (13). The inferential accuracy of the Bayesian method, on the other hand, is unknown as it is not employed for inference during diagnosis in practice. The prescription of the Bayesian method for inference in diagnosis without knowing its inferential accuracy appears to us to be similar to prescription of a treatment whose therapeutic efficacy is unknown.

In conclusion, the Bayesian method has been prescribed in diagnosis due to its rationality based on its coherence, which however is not our goal in diagnosis in practice. Our goal in diagnosis, as is well known, is to infer (diagnose) a disease accurately in a patient regardless of its pretest probability. This goal is not achieved, as we have discussed in this paper, by the Bayesian method which is not employed therefore in practice. The method which is employed for inference in practice, as we have discussed above, is the confidence frequentist method which is designed to achieve high inferential accuracy in any patient regardless of pretest probability of a disease. What we find impressive is that the confidence method, unlike the Bayesian method, has not been prescribed, but has been discovered, so to speak, by practicing physicians on their own. It is important to recognize, we believe, that the correct method of inference during diagnosis is the confidence method which is employed in practice and not the prescribed Bayesian method. This recognition has important implications, we suggest, in minimizing diagnostic errors and in teaching diagnosis to medical students and novice physicians.

Appendix 1

Pretest prob. of 7 percent = Pretest odds of 7/93

In odds form of Bayes' theorem,

Pretest odds x Likelihood ratio = Post-test odds

Thus 7/93 x 13 = 1/1 = Post-test prob. of 50 percent

Appendix 2

Pretest prob. Of 7 percent = Pretest odds of 7/93

Post-test prob. of 85 percent = Post-test odds of 85/15

In odds form of Bayes' theorem,

Likelihood ratio = Post-test odds/Pretest odds = 85/15 / 7/93 = 75

Appendix 3

Pretest prob. of 85 percent = Pretest odds of 85/15

Post-test prob. of 85 percent = Post-test odds of 85/15

In odds form of Bayes' theorem,

Likelihood ratio = Post-test odds/Pretest odds = 85/15 / 85/15 = 1

References

1. Weinstein MC, Fineberg HV. Clinical Decision Analysis. Philadelphia: WB Saunders Company 1980.
2. Edwards W, Lindman H, Savage LJ. Bayesian Statistical inference for psychological research. Psych Review 1963; 70: 193-242.
3. Lindley D. Bayesian statistics In Harper W. and Hooker C (eds) Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science, Vol 2,Boston MA. D. Reidel 1976, pp 353-62.
4. Pauker SG, Kopelman RI. How sure is sure enough? N Engl J Med 1992; 326: 688-91.

5. Rude RE, Poole WK, Muller JE, Turi Z, Rutherford J, Parker et al. Electrocardiographic and clinical criteria for recognition of acute myocardial infarction based on analysis of 3,697 patients. Am J Card 1983; 52: 936-42.
6. Larson DM, Menssen KM, Sharkey SW, Duval S, Schwartz RS, Harris J et al. "False positive" cardiac catheterization laboratory activation among patients with suspected ST-segment elevation myocardial infarction. JAMA 2007; 298: 32754-59.
7. Mayo DG. Statistical inference as Severe Testing: How to get beyond the Statistics Wars. Cambridge University Press 2018, pp 190-201.
8. Guyatt G, Rennie D, Meade MO, Cook DJ. Users' guide to the medical literature: A manual for evidence-based clinical practice. New York: The McGraw-Hill Companies 2008, p 428.
9. Adams WJ. The life and times of the Central Limit Theorem. American Mathematical Society, Providence RI, 2009.
10. Watson J, Whiting PF, Brush JE. Interpreting a covid-19 result. BMJ 2020.369m1808 doi 10.1136/bmj.m1808 (Published online on 12 May 2020)
11. Stein PD, Fowler SE, Goodman LR, Gottschalk A, Hales C, Hull RD et al. Multi-detector computed tomography for pulmonary embolism. N Engl J Med 2006; 353: 2317-27.
12. Zierler BK. Ultrasonography and diagnosis of venous thromboembolism. Circulation 2004; 353: 1-9-1-14.
13. Jain BP. An investigation into method of diagnosis in clinicopathologic conferences (CPCs). Diagnosis 2016; 3: 61-64.
14. Jain BP. Why is diagnosis not probabilistic in clinicopathologic conferences (CPCs): Point. Diagnosis 2016; 3: 95-97.
15. Heart of England, NHS Trust. Management of ST elevation myocardial infarction guidelines, Version 2, Update 2016.
16. Wigner EP. Symmetries and Reflections: Scientific essays. Woodbridge, Connecticut, Oxbow Press 1979 p 3.
17. Strauss N. Kris Kristofferson: an outlaw at 80. Rolling Stone Magazine, June 6, 2016.

18. Romeijn Jan-Willem "Philosophy of Statistics" The Stanford Encyclopedia of Philosophy (Spring 2017 Edition) Edward N Zalta (ed), URL=https://plato.stanford.edu/archives/spr2017/entries/statistics/
19. Jain BP. The scientific nature of diagnosis. Diagnosis 2017; 4: 17-19.
20. Feynman RP. Mr. Feynman goes to Washington: Investigating the space shuttle Challenger disaster, In "What do you care what other people think" ed. Leighton R. New York: WW Norton & Co. 2001 pp 113-238.
21. Singh H, Giardina TD, Meyer AN, Forjuoh SN, Reis MD,Thomas EJ. Types and origins of diagnostic errors in primary care settings. JAMA Intern Med 2013; 173: 418-25.
22. Ely JW,Kaldjian LC, D'Alessandro DM. Diagnostic errors in primary care: Lessons learnt. J Am Board Fam Med 2012; 25: 85-97.
23. The PIOPED Investigators. Value of the ventilation/perfusion scan in acute pulmonary embolism diagnosis (PIOPED). JAMA 1990; 263: 2753-9.
24. Feinstein AR. The haze of Bayes', the aerial palaces of decision analysis and the computerized Ouija board. Clin Pharmacol Ther 1977; 21: 482-95.
25. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. Am J Med 2008; 121: S2-S23.