The goal in diagnosis, as is well known, is accurate determination of a disease causing illness in a patient with symptoms. This goal is achieved mostly by statistical inference of a disease from data (symptoms, tests etc.) as the relationship between a disease and data is usually not unique. In this paper, we shall investigate the method employed for statistical inference (henceforth) called inference for brevity) during diagnosis in practice.

We start our investigation by examining the method employed for inferring acute myocardial infarction (MI) during its diagnosis in different patients in practice.

It is well known from experience that acute MI, like any other disease, occurs in different patients with varying presentations and therefore with varying prior probabilities that range from being very low to being very high. For example, it occurs in a healthy 40 year old woman without any cardiac risk factor presenting with highly uncharacteristic chest pain in whom its prior probability is very low at 7 percent (1). It occurs as well in a 65 year old man with multiple cardiac risk factors presenting with highly characteristic chest pain in whom its prior probability is very high, say, 85 percent.

When we see a patient with chest pain, we usually suspect acute MI regardless of whether its prior probability is low or high and perform an EKG to evaluate it.

If acute ST elevation EKG changes with likelihood ratio (LR) OF 13 (2) for acute MI are observed, this test result is interpreted as strong evidence based on the high frequency of 85 percent (3) with which it leads to an accurate inference of acute MI in patients with varying prior probabilities.

Based on this strong evidence, acute MI is inferred to be present in any patient regardless of its prior probability.

The accuracy of inferring acute MI in this manner is 85 percent (3) in patients with varying prior probabilities seen by us, for example, in an emergency room (ER) , from where these patients with STEMI are transferred to a cardiac catherization laboratory for further management.

We note that prior probability does not play any role as prior evidence in inference of acute MI in this manner.

We find this method of inference, which consists of suspecting a disease and inferring it from a test result with high frequency of accurate inferences is employed in practice for inference of any disease which has a test capable of generating a result with LR greater than 10 (4). Thus, pulmonary embolism is inferred from chest CT angiogram, LR 20 (5) and deep vein thrombosis from positive venous ultrasound study, LR 16 (6) in any patient regardless of prior probability in practice.

This method is clearly not Bayesian, as it lacks two key features of the Bayesian method (7) as pointed out below:

(a) Prior probability of a disease does not play any role as prior evidence for it in a patient
(b) A disease is not inferred from its posterior probability which is generated by combining prior probability and likelihood ratio.

We suggest this method is closely similar or identical to the frequentist method of statistical inference introduced by Ronald Fisher and Jerzy Neyman (8) as we discuss below.

In the frequentist method, a cause or explanation is formulated from given phenomena and formulated as a hypothesis which does not have a prior probability attached to it so that it does not have any prior evidence for or against it.

The hypothesis is evaluated by performing a test and inferred to be correct if the test result has a high frequency of accurate inference of the cause as determined experimentally by a test of significance or by a confidence argument.

It is implicitly assumed in the frequentist method that the cause being evaluated is distributed normally in the sample being studied.

In the method employed for inference in practice, we suggest, a presentation functions as phenomena from which a disease is suspected and formulated as a diagnostic hypothesis which does not have any prior probability attached to it so that it does not have any prior evidence for or against it just as is done in the frequentist method.

This hypothesis is evaluated by performing a test and inferred to be correct if the test result has a high frequency of accurate inference of the disease in a series of patients with varying prior probabilities, which is very similar to what is done in the frequentist method.

In addition, a disease tends to be distributed normally in an unselected series of patients in whom we suspect it in practice. For example, in the well known PIOPED study about diagnosis of pulmonary embolism (9), 68 percent of 252 cases of pulmonary embolism occurred in patients with intermediate (20-79 percent) prior probability, 9 percent in patients with low (0-19 percent) and 23 percent in patients with high (80-100 percent) prior probability.

The finding of 68 percent cases of pulmonary embolism in patients with intermediate prior probability clearly indicates, we believe, a trend for its distribution to be normal.

We suggest, the reason for a trend towards normal distribution of a disease is the large number of independent factors (10) such as patient's age, sex, various risk factors, various symptoms etc. which make up a presentation from which its prior probability is derived.

Due to this striking resemblance, we conclude the method employed for inference in practice is frequentist with one difference from the standard method.

The difference is that in practice, the high frequency of accurate inference of a disease which functions as strong evidence is obtained naturally from an unselected series of patients seen over a period of time, while in the standard method, this high frequency is generated experimentally by a test of significance or by a confidence argument.

The key difference between the Bayesian and the frequentist methods is that evidence in the former is represented by probability of a disease while in the latter it is represented by probability, in the form of a frequency, of a procedure in leading to an accurate inference (8).

The other important difference to which we have already alluded to, is that a presentation from which a hypothesis is formulated plays no further role in inference which is done purely from the test result. In the Bayesian method, on the other hand, a prior probability derived from a presentation represents prior evidence.

We shall now examine why the frequentist method is employed for inference during diagnosis and not the Bayesian method which has actually been prescribed as the normatively correct method.

1. The most important reason, we suggest, is that the frequentist method leads to an accurate inference in practically every patient regardless of prior probability of a disease. This is made possible by the fact that prior probability does not represent prior evidence which allows a disease with atypical presentation (low prior probability) to be suspected, tested and inferred if it is present.
   It is due to this feature of the frequentist method that diseases which are rare or have highly atypical presentations are inferred correctly in published diagnostic exercises in real patients such as clinicopathologic conferences (CPCs) and clinical problem solving exercises (11, 12).
   In the Bayesian method, on the other hand, a disease with an atypical presentation may not be suspected or tested due to its low prior probability being interpreted as prior evidence against it leading to a serious diagnostic error. This is likely to occur, for example, in the 40 year old woman mentioned above in whom the prior probability of acute MI is only 7 percent.
   The Bayesian method thus encourages failure to suspect a disease with an atypical presentation which has been reported as an important cause of diagnostic errors in several studies (13,14).
2. A frequentist inference in practice is experience based as it is made from a frequency as evidence which is derived from a heterogenous series of

patients with varying prior probabilities which we have actually observed. For example, the inference of acute MI from acute ST elevation EKG changes is validated by our experience of finding this inference to be correct in 85 percent (8 to 9 out of 10) patients with varying prior probabilities.

This experience derived from a heterogenous series is legitimate, we believe, because acute MI is the same disease with the same pathophysiology in all patients, as far as we know, regardless of prior probability.

A Bayesian inference made from a posterior probability would not be experience based as it refers to a frequency of a disease in a homogenous series of patients who are similar. For example, the Bayesian inference of acute MI being indeterminate which is made from a posterior probability of 50 percent generated by combining the prior probability of 7 percent and LR of 13 in the 40 year old woman (Appendix 1) is not likely to be validated by our experience as we are unlikely to have observed a frequency of acute MI of 50 percent in a series of patients similar to her.

3. A frequentist inference is made from a test result with high LR which indicates a key feature of the disease being inferred. For example, acute MI is inferred from acute ST elevation EKG changes, LR 13 which indicates acute myocardial injury that is a key feature of acute MI. As anyone can observe this test result, this inference is objective.

   On the other hand, there may be a difference of opinion about a Bayesian inference as it is made from a posterior probability which is generated from a prior probability that is often derived subjectively.

4. The frequentist method is consistent with the clinically useful practice of creating a comprehensive differential diagnosis which includes diseases with high as well as those with low prior probabilities (15) as prior probability does not represent prior evidence in this method.

   In the Bayesian method, on the other hand, a disease with a low prior probability may not be included in a differential diagnosis due to prior evidence against it.

Also the clinically useful practice of asking 'What else could it be' in a difficult diagnostic situation (16) which encourages us to think of a disease with a very low prior probability may not be possible with the Bayesian method.

5. The frequentist method is simple to execute in practice as it makes a clear distinction between a presentation and a test result with a high frequency of accurate inference. In this method, estimation of an accurate prior probability is not required as it plays no role as prior evidence .
   We suspect a disease from a presentation and infer it from a test result with a high frequency. There is no mathematical calculation required to combine evidence from two sources such as is needed in the Bayesian method.
   Due to its simplicity, the frequentist method is employed in practice by a wide range of healthcare providers which include physicians, nurse practitioners, nurses.
   The Bayesian method is comparatively difficult as it requires a mathematical calculation which combines a prior probability and a likelihood ratio to generate a posterior probability. It is our impression most healthcare providers are not familiar with this calculation.
6. In the frequentist method, only a test result with high LR (usually greater than 10), which leads to a high frequency of accurate inference counts as strong or conclusive evidence.
   In the Bayesian method, even a test result with LR of 1, which is known to be clinically worthless, may appear to provide strong evidence. For example, in the 65 year old man mentioned earlier, in whom the prior probability of acute MI is 85 percent, the test result, non-specific T wave EKG changes with LR of 1, when combined with the prior probability leads to a posterior probability of 85 percent, from which acute MI would be inferred with near certainty (Appendix 2).
7. In the frequentist method, we may include any number of suspected diseases which have high prior probabilities as diagnostic hypotheses in a differential diagnosis.

In the Bayesian method, we are limited by the rule that the prior probabilities of all suspected diseases should add up to 1 (8). Therefore, strictly speaking, we cannot include a disease with prior probability of 0.7 (70 percent) as well as a disease with prior probability of 0.6 (60 percent) in a differential diagnosis in the same patient in this method.

8. The diagnostic accuracy of the frequentist method which is employed for inference in practice is known to be high at 85 to 90 percent (17) which is why we employ it.
   The diagnostic accuracy of the Bayesian method, on the other hand, is unknown as it is not employed in practice.

Thus we do not find the Bayesian method to have any advantage over the frequentist method as far as diagnostic accuracy is concerned which makes us wonder why it has been prescribed as the normatively correct method of inference during diagnosis as diagnostic accuracy in every patient is our goal in diagnosis.

It turns out on examining its prescription in the early 1960s that it was not prescribed due to its diagnostic accuracy but because it is rational ,as a Bayesian inference is (like) a bet placed on it with odds based on its probability (18). It is believed to be rational as it ensures we shall not be wrong in the long run (that is not lose a bet in the long run). This notion of rationality, we believe, is not applicable to diagnosis, because our goal in diagnosis is not long run accuracy but accuracy in every individual patient. As this individual accuracy cannot be provided by a probability based Bayesian method, it is not employed, we believe, for inference in practice as we have discussed above.

We believe if the goal and process of inference in diagnosis had been carefully studied prior to prescription of the Bayesian method , it would not have been prescribed and some other more suitable method such as the frequentist method would have been considered for prescription. The frequentist method is as much a well established method of statistical inference as the Bayesian method is.

The manner in which the Bayesian method was prescribed appears to us to be similar to performing diagnosis in a patient without a careful study of the diagnostic situation and without creation of a differential diagnosis. It was done in this manner perhaps because in the early 1960s, when it was done, the cognitive revolution overthrew the behavioral theory of thinking and Bayesian reasoning was held to be normatively correct in all fields due to its rationality (19).

There have been misgivings earlier about Bayesian reasoning in diagnosis which were expressed most forcefully by the eminent clinical investigator, Alvan Feinstein, in 1977 (20) as follows:

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).

We find Feinstein's words to be as true today, more than 40 years later, as they were when he wrote them in 1977.

We too have not found the Bayesian method to be employed for inference in any diagnostic exercises in real patients such as clinicopathologic conferences (CPCs) and clinical problem solving exercises which are published regularly in the New England Journal of Medicine (12,13). We found the frequentist method to be employed for inference in all these exercises.

The disparity between theory, in which the Bayesian method is prescribed, and practice, in which the frequentist method is used for inference during diagnosis, creates ambiguity, we believe, about the correct method of inference. We believe it is important to align theory and practice for the sake of reducing diagnostic errors by discontinuing prescription of the Bayesian method based on the discussion in this paper. This alignment would also be of value, we believe, in teaching diagnosis to medical students and in developing more accurate computer programs for diagnosis.

Appendix 1

Prior probability of 7 percent = Prior odds of 1/13

In odds form of Bayes' theorem,

Posterior odds = Prior odds x Likelihood ratio

Therefore,

Posterior odds = 1/13 x 13 = 1/1 = Posterior probability of 50 percent.

Appendix 2

Prior probability of 85 percent = Prior odds of 85/15

In odds form of Bayes' theorem,

Posterior odds = Prior odds x Likelihood ratio

Therefore,

Posterior odds = 85/15 x 1 = Posterior probability of 85 percent.

References

1. Pauker SG, Kopelman RI. How sure is sure enough? N Engl J Med 1992; 326: 688-91.
2. 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; 52: 936-42.
3. Larson DM, Menssen KM, Sharkey SW, Duval S, Schwartz RS, Harris J et al. "False positive" cardiac catherization laboratory activation among patients with suspected ST-segment elevation myocardial infarction. JAMA 2007; 298: 2754-2759.

4. 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,2008,428.
5. Stein PD, Fowler SE, Goodman LR, Gottschalk A, Hales C, Hull et al. Multidetector computed tomography for pulmonary embolism. N Engl J Med 2006; 354: 2317-27.
6. Zierler BK. Ultrasonography and diagnosis of venous thromboembolism. Circulation 2004; 109: 1-9-1-14.
7. Weinstein MC, Fineberg HV. Clinical Decision Analysis. Philadelphia: WB Saunders Company, 1980.
8. Mayo DG. Statistical Inference as Severe Testing: How to get beyond the Statistics Wars. Cambridge: Cambridge University Press 2018.
9. The PIOPED Investigators. Value of the ventilation/perfusion scan in acute pulmonary embolism. Results of the prospective investigation investigation of pulmonary embolism diagnosis (PIOPED). JAMA 1990; 263: 2753-9.
10. Tao T. E pluribus unum. From complexity, universality. P 32-46 in The Best Writing in Mathematics 2013. Ed. M. Pitici. Princeton: Princeton University Press 2014.
11. Jain BP. An investigation into method of diagnosis in clinicopathologic conferences (CPCs). Diagnosis 2016; 3: 61-64.
12. Jain BP. Why is diagnosis not probabilistic in clinicopathologic conferences (CPCs): Point. Diagnosis 2016: 3: 95-97.
13. 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.
14. Ely JW, Kaldjian LC, D'Alessandro DM. Diagnostic errors in primary care: Lessons learnt. J Am Board Fam Med 2012; 25: 85-97.
15. Jain BP. The key role of differential diagnosis in diagnosis. Diagnosis 2017; 4: 239-40.
16. Bordage G. Why did I miss the diagnosis? Some cognitive explanations and educational implications. Acad Med 1999; 74: S138-43.
17. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. Am J Med 2008; 121: S2-S23.

18. Edwards W, Lindman H, Savage LJ. Bayesian statistical inference for psychological research. Psych Review 1963; 70: 193-242
19. Gigerenzer G, Swijtink Z, Porter T, Daston L, Beatty , Kruger L. The Empire of Chance: How Probability changed Science and Everyday Life. Cambridge: Cambridge University Press 1989.
20. Feinstein AR. The haze of Bayes, the aerial palaces of decision analysis and the computerized Ouija board. Clin Pharmacol Ther 1977; 21: 482-95.