

[ORIGINAL ARTICLE]

Availability Bias Causes Misdiagnoses by Physicians: Direct Evidence from a Randomized Controlled Trial

Ping Li, Zi yan Cheng and Gui lin Liu

Abstract:

Objective Empirical evidence on the availability bias associated with diagnostic errors is still insufficient. We investigated whether or not recent experience with clinical problems can lead physicians to make diagnostic errors due to availability bias and whether or not reflection counteracts this bias.

Methods Forty-six internal medicine residents were randomly divided into a control group (CG) and experimental group (EG). Among the eight clinical cases used in this study, three experimental cases were similar to the disease of dengue fever (DF) but exhibited different diagnoses, one was actually DF, and the other four filler cases were not associated with DF. First, only the EG received information on DF, while the CG knew nothing about this study. Then, six hours later, all participants were asked to diagnose eight clinical cases via nonanalytic reasoning. Finally, four cases were diagnosed again via reflective reasoning.

Results In stage 2, the average score of the CG in the diagnosis of experimental cases was significantly higher than that of the filler cases (0.80 vs. 0.59, $p < 0.01$), but the EG's average score in the two types of cases was not significantly different (0.66 vs. 0.64, $p = 0.756$). The EG and CG had significantly different scores for each experimental case, while no difference was observed in the filler cases. The proportion of diseases incorrectly diagnosed as DF among experimental cases ranged from 71% to 100% in the EG. There were no significant differences between the mean diagnostic accuracy scores obtained by nonanalytic reasoning and those obtained by the reflective reasoning in any cases.

Conclusion Availability bias led to diagnostic errors. Misdiagnoses cannot always be repaired solely by adopting a reflective approach.

Key words: availability bias, recent experience, physicians' knowledge, diagnostic error repair

(Intern Med 59: 3141-3146, 2020)

(DOI: 10.2169/internalmedicine.4664-20)

Introduction

People, including physicians, often do not use calculated or rational decision strategies, instead relying on heuristics strategies to make decisions (1, 2). Heuristics can reduce the time and effort in decision making but may also lead to systematic cognitive biases in medicine settings (3).

Many studies have reported that heuristics biases play an important role in medical diagnostic errors (4-7), which are most prevalent in internal, family and emergency medicine (8-10). There are over 50 heuristics related to the cognitive biases (11), and physicians' use of the availability heuristic is one of the most common cognitive biases related

to diagnostic errors (12). The availability bias is the tendency to overestimate the likelihood of events when they readily come to mind (13). Thus, physicians who are influenced by a recent experience with a certain kind of disease may be more likely to diagnose other similar diseases as that particular disease.

There have been many descriptive studies on availability bias, but to our knowledge, only two experimental studies have been conducted: one by Mamede and one by Schmidt (14, 15). More direct evidence is therefore necessary to understand what roles availability biases play in physicians' diagnostic errors. In this study, we used the definition of diagnostic errors established by the National Academy of Medicine: the failure to establish an accurate and

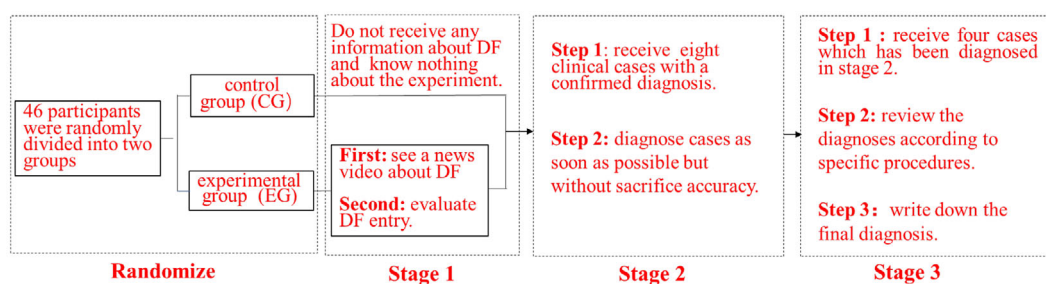


Figure. Research flowchart.

timely explanation of the patient's health problem or communicate that explanation to the patient.

Experimental studies provide more direct evidence than descriptive studies. However, only the diagnostic accuracy from filler cases and experimental cases on their own is compared in these studies (14, 15). In Mamede's study, it was argued that participants were affected by availability bias (14). However, that conclusion was based on the assumption that there was no difference in participants' knowledge between the experimental cases and the filler cases. If participants' knowledge of filler cases and experimental cases differs due, for example, to differences in participants' clinical experiences or for other reasons, the participants' diagnostic accuracy may differ between experimental cases and filler cases, giving a false-positive result. Unfortunately, previous studies (14, 15) have simply assumed that there were no differences in participants' knowledge between filler and experimental cases, and no strict tests to confirm this were conducted.

Once availability bias has been documented, it is particularly important to find ways to counteract a physicians' availability bias. One way to counteract such biases, as suggested by previous studies (7, 14-16), is to encourage physicians to adopt more reflective reasoning, which is analytical in nature and requires careful and effort-driven consideration of the findings of a case. However, while some studies have found support for the claim that reflection can lead to a more accurate diagnosis, this finding is not completely universal (17, 18). Mamede et al. also argued that diagnostic errors are actually more likely to be provoked by reasoning processes, as the errors made by participants due to nonanalytic reasoning can be repaired by reflective reasoning (14).

Except for recent experience with clinical problems, the diagnostic reasoning mode or the physicians' knowledge about experimental cases and filler cases also contribute to significant differences in the diagnostic accuracy. However, due to limitations in the study design, the previous studies cannot distinguish one reason from the other two in a specific case.

Therefore, evidence concerning the availability bias in medical diagnosis is still unclear, and more rigorous experiments, such as randomized controlled trials, should be conducted.

In this study, several blank control trials and strict tests

were performed in order to exclude the possibility of false-positive results caused by difference in participants' knowledge between filler and experimental cases. The present study thus investigated whether or not availability bias occurs after physicians were given specific disease information. This study also explored whether or not reflection can counteract the availability bias and produce significant improvements in the diagnostic accuracy.

Materials and Methods

Participants

The 46 volunteers for this study were internal medicine residents from the Second Affiliated Hospital & Yuying Children's Hospital of Wenzhou Medical University in Zhejiang Province, China. The participants were invited for the study through their residency program directors. They were randomly divided into two groups, each with 23 residents: a blank control group (CG) and an experimental group (EG).

This study was approved by the ethics committee of Wenzhou Medical University.

Study design and procedure

This study consisted of three stages and was conducted sequentially in one day. The procedure is shown in Figure.

In stage 1, the EG was asked to view a news video about dengue fever (DF), which had recently been reported in the mass media. They then received a copy of a DF entry from Baidu encyclopedia, a popular online encyclopedia in China, and were requested to evaluate the accuracy and completeness of statements about the epidemiology, transmission, symptoms and laboratory tests encountered in the text. Subsequently, they wrote down the inaccurate and incomplete statements and assigned a score to the entry on a five-point scale. After completing that test, the EG were thanked for their contributions, and then they returned to their daily clinical work. During stage 1, the CG did not receive any information about DF and knew nothing about the experiment.

In stage 2, which was conducted about six hours later, the EG and CG were asked to diagnose eight clinical cases. Great care was taken to ensure that stage 2 appeared to be an unrelated study.

In stage 2, the participants were required to diagnose the cases described in the reading materials under the supervision of an experimenter who had not appeared in stage 1.

The eight clinical cases were based on actual patients with a confirmed diagnosis that had been prepared and judged by two experts of internal medicine. These 2 experts, each with over 10 years' experience in clinical practice and teaching internal medicine, were not a part of this study and were not aware of the study hypothesis or the DF entry. One expert prepared the case, and the other checked it and confirmed the diagnosis.

Four of the eight cases were "filler" cases, in which the descriptions of clinical manifestations were unrelated to those frequently encountered in patients with DF. The remaining cases were "experimental" cases, among which three had clinical manifestations similar to DF but had different diagnoses, while the other was consistent with DF. The diseases of bronchiectasis, alcoholic hepatitis, lung cancer and acute left heart failure were designated as filler cases, and measles, scarlatina, DF and typhoid were treated as experimental cases.

Each case included a brief description of the patient's medical history, signs and symptoms, test results and four possible diagnoses (see an example of case in Box 1).

Box 1. Example of a Case (Diagnosis: DF fever): A 42-year-old man visited his relatives in Indonesia for 3 months before returning to Fujian Province (located in southeast China) in March. Five days ago, the patient presented with sudden-onset chills with no obvious cause. An hour later, his temperature rose to 39°C, accompanied by arthralgia, headache (located behind the eyes), myalgia and asthenia. Symptoms of cough, pharyngalgia, nausea, vomiting, abdominal pain, diarrhea and melena were not observed. Three days ago, a rash described as "islands of white on a sea of red" appeared on the face and bilateral upper limbs, and the small red spots disappeared when the skin was pressed. Red maculopapules the size of needle tips were also found over both lower extremities in last three days. The patient lost his appetite during the course of the disease. His sleep, feces and urine were normal.

Physical Examination

No abnormalities other than mental fatigue, an acute febrile appearance and a maculopapular rash were observed. BP: 110/75 mmHg; pulse: 88/min; temperature: 38.4°C.

Laboratory Tests

White blood cells: $2.5 \times 10^9/L$ (normal, $3.5-9.5 \times 10^9/L$), normal; red blood cells: $5.3 \times 10^{12}/L$ (normal, $4.0-5.5 \times 10^{12}/L$); hemoglobin: 157 g/L (normal, 120-160 g/L); Blood platelet: $44 \times 10^9/L$ (normal, $100-300 \times 10^9/L$).

The diagnosis of this case is most likely to be:

A. Measles B. Scarlet Fever C. Dengue fever D. Typhoid

The cases were presented on a test paper in random sequence. The CG and EG diagnosed the same eight cases in stage 2. The instructions for the test paper were as follows:

"Please review the cases and then select the most likely diagnosis from the listed diseases. Make a decision as quickly as possible without sacrificing accuracy." At this stage, a rapid and largely unconscious diagnostic approach was adopted by the participants because of the limited amount of time available to make the diagnosis (14).

In stage 3, each resident received another test paper that had four cases. It included one filler cases (Bronchiectasis) and three experimental cases (DF, measles and typhoid) that the participants had diagnosed in stage 2. To induce reflective reasoning, the participants were told that they should review the four cases and comply with the instructions as follows: "1) Peruse these cases carefully again; 2) write down your initial diagnosis for the cases in stage 2; 3) list the evidence in the case statement that supports the initial diagnosis; 4) list the items that speak against the initial diagnosis; 5) list the evidence that should be present if the initial diagnosis was accurate but was not mentioned in the case." Subsequently, the participants were asked to list alternative diagnoses if they felt that their initial diagnosis was incorrect and to follow the same procedure (steps 3-5) for each alternative diagnosis considered for the case. Finally, they were asked to draw a conclusion by writing down their final diagnosis for the case.

Data analyses

Information on the sex, age, years of clinical experience and educational background of all 46 participants was collected. The mean age of the participants was 27.0 years old [standard deviation (SD)=2.16 years], and they had an average of 2.1 years of clinical experience (SD=0.95 years). The ratio of men to women was 18:28, and there were 17 participants with bachelor's degrees, 26 with master's degrees and 3 with PhDs.

Because the data characteristics did not meet the requirements of parameter tests, the Mann-Whitney U test, which is the most commonly used nonparametric test for two-independent-sample tests, was used to compare the differences in the baseline situation between the CG and EG.

All eight cases had a specific diagnosis that could be used as a standard to evaluate the accuracy rate of the participants' diagnoses. The diagnoses were considered correct or incorrect. A score of 1 was given if the participant selected or wrote down the correct diagnosis; otherwise, 0 points were given. For each group, the mean scores (or accuracy rate) obtained in stages 2 and 3 for the filler and experimental cases were computed.

A second Mann-Whitney U test was performed to clarify whether or not differences in physicians' knowledge between the filler cases and experimental cases led to a difference in diagnostic accuracy and whether or not a false-positive result had occurred in those studies (14, 15).

A third Mann-Whitney U test was used to compare the diagnostic performances between the EG and CG in stage 2. We tested the hypothesis that recent exposure to a similar disease would induce an availability bias and that this bias

Table 1. The Difference of Diagnostic Accuracy Scores between the Experimental Cases and Filler Cases.

Group	Mean scores (SD) of all experimental cases	Mean scores (SD) of all filler cases	p value ^a
CG	0.80 (0.212)	0.59 (0.207)	0.001*
EG	0.66 (0.234)	0.64 (0.211)	0.756

^aMann-Whitney U Tests; *p<0.05. SD: standard deviation, CG: control group, EG: experimental group

Table 2. The Mean Scores and Number of Incorrectly Diagnosed as DF in Stage 2.

Cases	Mean score (SD)			Number of incorrectly diagnosed as DF in gross diagnostic errors			
	CG	EG	p value ^a	CG ^b	EG ^b	EG vs. CG ^c	
Experimental cases	Measles	0.74 (0.45)	0.43 (0.51)	0.038*	4/6	11/13	+7 (100%)
	Scarlatina	0.96 (0.21)	0.74 (0.45)	0.042*	1/1	5/6	+4 (80%)
	DF	0.70 (0.47)	0.96 (0.21)	0.021*	0/6	0/1	-5 (100%)
	Typhoid	0.83 (0.39)	0.52 (0.51)	0.029*	2/4	7/11	+5 (71%)
Filler cases	Bronchiectasis	0.65 (0.49)	0.52 (0.51)	0.374	-	-	-
	Alcoholic hepatitis	0.17 (0.39)	0.30 (0.47)	0.305	-	-	-
	Lung cancer	0.87 (0.34)	0.91 (0.29)	0.639	-	-	-
	Acute left heart failure	0.65 (0.49)	0.83 (0.39)	0.184	-	-	-

^aMann-Whitney U tests. *for all p<0.05. ^bnumber of incorrectly diagnosed as DF / all incorrect diagnostics. ^cthe difference of incorrectly diagnosed as DF between CG and EG and the percentage of incorrectly diagnosed as DF in gross wrong diagnostics. DF: dengue fever, SD: standard deviation, CG: control group, EG: experimental group

would not occur in filler cases. The proportion of incorrect diagnoses in stage 2 was calculated. Then the number of incorrect diagnoses of DF between the EG and CE was compared to understand the role of availability bias in physicians' diagnostic errors.

A fourth Mann-Whitney U test was used to compare the diagnostic accuracy between participants in stage 2 (no analytical reasoning) and stage 3 (reflective reasoning). Whether or not reflective reasoning promoted diagnostic accuracy in the EG and CG was then assessed.

In this study, significance was set at p<0.05. The SPSS software program, version 19 (SPSS, Chicago, USA), for Windows was used for the analyses.

Results

All 46 residents finished the test. The Mann-Whitney U test revealed no significant difference in the sex, age, years of clinical practice or educational background between the EG and CG.

The mean diagnostic accuracy scores obtained from the experimental cases and filler cases by CG and EG in stage 2 are shown in Table 1. The Mann-Whitney U tests revealed a significant difference in the mean diagnostic accuracy scores between the experimental and filler cases in the CG but not in the EG. The significant finding in the CG suggested that the participants' knowledge of experimental cases might have been better than that of the filler cases. Further tests should be performed to clarify whether or not the availability bias caused the non-significant result in the EG.

Table 2 shows the mean scores obtained by the CG and EG when cases were diagnosed through non-analytical reasoning. The Mann-Whitney U tests revealed a significant difference in the mean diagnostic accuracy scores between the CG and EG in experimental cases but not filler cases. The EG score was significantly lower than the CG for the diagnostic test except in the DF cases. This can be explained by the fact that the EG showed higher scores in the DF cases due to the interference of availability bias. The right-most column of Table 2 shows that the diagnosis of other diseases as DF by the EG primarily contributed (contribution from 71% to 100%) to the low scores for experimental cases. Furthermore, the diagnostic accuracy of experimental cases decreased when the EG was exposed to information on DF in advance, as a good consistency of diagnostic accuracy was noted between the CG and EG in the filler cases. Because they had not received any information on DF, the CG was uninfluenced by such information when all cases were diagnosed.

We may therefore draw the following inferences: 1) the participants' diagnostic ability for all cases did not differ markedly between the CG and EG, 2) the information in DF received by the EG six hours earlier did not affect the EG's diagnostic accuracy for filler cases, 3) the lower mean scores for the EG than the CG in stage 2 were consistent with an availability bias; and 4) if differences in participants' knowledge between experimental and filler cases are not considered, a false-positive results may be obtained.

The mean diagnostic accuracy scores obtained through nonanalytic reasoning (stage 2) were compared with those

Table 3. The Mean Scores and Mann-Whitney U Tests between Stage 2 and Stage 3 for 4 Cases.

Cases	Mean score (SD) of CG			Mean score (SD) of EG		
	Stage 2	Stage 3	p value	Stage 2	Stage 3	p value
Bronchiectasis	0.65 (0.49)	0.65 (0.49)	1.000	0.52 (0.51)	0.57 (0.51)	0.770
Measles	0.74 (0.45)	0.87 (0.34)	0.270	0.44 (0.57)	0.65 (0.49)	0.143
DF	0.70 (0.47)	0.91 (0.29)	0.066	0.96 (0.21)	0.96 (0.21)	1.000
Typhoid	0.83 (0.39)	0.70 (0.47)	0.305	0.52 (0.51)	0.78 (0.42)	0.066

SD: standard deviation, CG: control group, EG: experimental group, DF: dengue fever

obtained through reflective reasoning (stage 3) by the CG and EG for four cases (one filler cases and three experimental cases) as presented in Table 3. Our findings showed that the changes in the mean diagnostic accuracy scores of stage 2 and stage 3 in the EG and CG were complicated, and Mann-Whitney U tests revealed no significant difference between the scores at the two stages. Reflection therefore does not significantly improve all participants' diagnostic accuracy compared with nonanalytic reasoning. Of note, however: reflection reduced the diagnostic accuracy of the CG in the typhoid case.

Discussion

Previous studies that have investigated availability bias in clinical settings through self-control experiments have failed to consider differences in participants' knowledge between filler cases and experimental cases, potentially resulting in false-positive results (14, 15). In the present study, a randomized control trial was performed, and strict tests were carried out to check for difference in the participants' knowledge between the filler cases and the experimental cases. In this way, an availability bias can be considered as an independent variable, and more accurate results can be acquired.

The results of this study are consistent with those of previous studies and demonstrated that diagnostic errors can occur as a result of availability bias. Several reasons for this are proposed. First, there was no marked difference in the characteristics of the participants between the CG and EG, and they used the same reasoning process (nonanalytic reasoning) in stage 2. If the diagnostic errors had been mainly caused by the decision process, no marked differences in the diagnostic accuracy scores would have been noted between the CG and EG in all cases. Obviously, the evidence obtained in this study did not support this hypothesis. The significant differences in diagnostic accuracy scores between the CG and EG in experimental cases can be interpreted to mean that because the EG had recently been exposed to information on DF, placing thoughts of DF readily in their mind, they were more likely than the CG to diagnose experimental cases as DF, and this phenomenon did not appear in the CG because they did not receive information in advance on DF. Second, the incorrect diagnosis of other diseases as DF by the EG markedly contributed (from 71% to

100%) to their lower scores for the diagnoses of the experimental cases.

In previous studies on the availability bias, reflective reasoning has been considered an effective method for improving the diagnostic accuracy (5, 14, 15, 19). However, the present study found that participants' diagnostic accuracy was not significantly improved by reflective reasoning. In fact, surprisingly, the diagnostic accuracy of typhoid by the CG was even lower than that with nonanalytic reasoning. Several reasons for why the diagnostic accuracy with reflective reasoning failed to achieve significant improvement are proposed. First, if logical and analytical processes (reflective reasoning) fail to identify and correct the errors which stem from nonanalytic processes, then such errors will continue (18). Furthermore, when a wrong initial hypothesis is triggered by availability bias, other biases, such as the anchoring effect, confirmation bias and premature closure, may activate, thereby hindering the correction of the incorrect diagnoses (9). For those reasons, in some cases, the diagnostic accuracy scores did not change at all when reflective reasoning was introduced or actually decreased compared with nonanalytic reasoning in the present study. The findings of our study also agree with those of other studies (20-22), providing some evidence to support this hypothesis (9). Second, the diagnoses in previous studies were usually considered correct, partially correct or incorrect and scored as 1, 0.5 or 0 points (14, 15). In the present study, however, the experts suggested that it was more appropriate to classify the diagnosis as correct or incorrect, as a simply partially correct diagnosis may lead to inappropriate treatment and cause serious adverse consequences. Differences in the scoring method may thus have affected the score gap between nonanalytic and reflective reasoning. This change may also have led to the differences in statistical results. Improving the diagnostic accuracy by reflective reasoning remains an issue that needs further investigation.

Nonanalytic reasoning, such as that using the availability heuristic, is a rapid and largely unconscious diagnostic approach that can work well in many situations (23). However, reliance on nonanalytic reasoning may be more easily affected by bias than reflective reasoning (24, 25), although diagnostic errors are not simply a consequence of over-reliance on one way of thinking (21). Our study indicated that availability bias is mainly prompted by recent experience with a similar disease rather than nonanalytic reasoning

itself. This finding may provide some insight into the mechanisms underlying availability bias.

Several limitations associated with the present study warrant mention. First, some degree of both nonanalytic and analytic reasoning may take place when physicians make diagnostic decisions, even if one reasoning process may dominate a specific scene. It is difficult to discriminate strictly between nonanalytic and analytic reasoning through our experiment design. Some degree of reflective reasoning may therefore have occurred even though participants were asked to respond promptly (17). Second, although the cases were based on real patients and the tasks simulated medical decision-making procedures, there may be some limitations in extending our findings to real-world situations. This is because the factors that affect diagnostic decisions are always more complex in real problem solving situations. Third, the role of the reasoning approach in availability bias is still unclear, and further investigations should be performed.

Conclusion

In summary, the present study conducted several blank control trials and strict tests to exclude potential false-positive results caused by differences in participants' knowledge between filler cases and experimental cases, finding that the recent experiences of participants with similar cases induced an availability bias in medical situations. Furthermore, the availability bias seemed to account for the bulk of diagnostic errors and was not well repaired by reflective reasoning.

The authors state that they have no Conflict of Interest (COI).

Financial Support

This material is based upon work funded by the Zhejiang Provincial Philosophy and Social Science Foundation of China (17NDJC159YB)

Acknowledgement

The authors thank the residents of the 2nd Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University for participating in the study and Dr. Yixing Huang and Dr. Jing Sun from the 2nd Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University for collaborating with the case development and the analysis of participants' responses. The authors also thank Dr. Jiaming Liu, Dr. Yong Cheng and Qiaoyan Zhu from Wenzhou Medical University for their valuable suggestions concerning the revision of the manuscript.

References

1. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science* **185**: 1124-1131, 1974.
2. Cain DM, Detsky AS. Everyone's a little bit biased (even physicians). *JAMA* **299**: 2893-2895, 2008.
3. Marewski JN, Gigerenzer G. Heuristic decision making in medicine. *Dialogues Clin Neurosci* **14**: 77-89, 2012.
4. McDonald CJ. Medical heuristics: the silent adjudicators of clinical practice. *Ann Intern Med* **124**: 56-62, 1996.
5. Redelmeier DA. The cognitive psychology of missed diagnoses. *Ann Intern Med* **142**: 115-120, 2005.
6. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* **165**: 1493-1499, 2005.
7. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med* **78**: 775-780, 2003.
8. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study I. *Qual Saf Health Care* **13**: 145-151, 2004.
9. Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The quality in Australian health care study. *Med J Aust* **163**: 458-471, 1995.
10. Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care* **38**: 261-271, 2000.
11. Baron J. *Thinking and Deciding*. Cambridge University Press, New York, 2000.
12. Lucchiari C, Pravettoni G. Cognitive balanced model: a conceptual scheme of diagnostic decision making. *J Eval Clin Pract* **18**: 82-88, 2012.
13. Tversky A, Kahneman D. Availability: a heuristic for judging frequency and probability. *Cogn Psychol* **5**: 207-232, 1973.
14. Mamede S, van Gog T, van den Berge K, et al. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA* **304**: 1198-1203, 2010.
15. Schmidt HG, Mamede S, Van Den, Berge K, et al. Exposure to media information about a disease can cause doctors to misdiagnose similar-looking clinical cases. *Acad Med* **89**: 285-291, 2014.
16. Evans JS. Dual-processing accounts of reasoning, judgment, and social cognition. *Annu Rev Psychol* **59**: 255-278, 2008.
17. Sherbino J, Dore KL, Wood TJ, et al. The relationship between response time and diagnostic accuracy. *Acad Med* **87**: 785-791, 2012.
18. Kahneman D, Tversky A. On the study of statistical intuitions. *Cognition* **11**: 123-141, 1982.
19. Mamede S, Schmidt HG, Rikers RM, et al. Influence of perceived difficulty of cases on physicians' diagnostic reasoning. *Acad Med* **83**: 1210-1216, 2008.
20. Croskerry P. Achieving quality in clinical decision making: cognitive strategies and detection of bias. *Acad Emerg Med* **9**: 1184-1204, 2002.
21. Norman GR, Rosenthal D, Brooks LR, et al. The development of expertise in dermatology. *Arch Dermatol* **125**: 1063-1068, 1989.
22. Lehr JL, Lodwick GS, Farrell C, et al. Direct measurement of the effect of film miniaturization on diagnostic accuracy. *Radiology* **118**: 257-263, 1976.
23. Norman GR, Eva KW. Diagnostic error and clinical reasoning. *Med Educ* **44**: 94-100, 2010.
24. Schmidt HG, Boshuizen HP. On acquiring expertise in medicine. *Educ Psychol Rev* **5**: 205-221, 1993.
25. Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implication. *Acad Med* **65**: 611-621, 1990.

The Internal Medicine is an Open Access journal distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view the details of this license, please visit (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).