

ACADEMIC MEDICINE

Journal of the Association of American Medical Colleges

Uncomposed, edited manuscript published online ahead of print.

This published ahead-of-print manuscript is not the final version of this article, but it may be cited and shared publicly.

Author: Mamede Sílvia MD, PhD; Hautz Wolf E. MD, MME; Berendonk Christoph MD, MME; Hautz Stefanie C. PhD; Sauter Thomas C. MD, MME; Rotgans Jerome PhD; Zwaan Laura PhD; Schmidt Henk G. PhD

Title: Think Twice: Effects on Diagnostic Accuracy of Returning to the Case to Reflect Upon the Initial Diagnosis

DOI: 10.1097/ACM.00000000000003153

ACCEPTED

Academic Medicine

DOI: 10.1097/ACM.0000000000003153

Think Twice: Effects on Diagnostic Accuracy of Returning to the Case to Reflect Upon the Initial Diagnosis

Sílvia Mamede, MD, PhD, Wolf E. Hautz, MD, MME, Christoph Berendonk, MD, MME, Stefanie C. Hautz, PhD, Thomas C. Sauter, MD, MME, Jerome Rotgans, PhD, Laura Zwaan, PhD, and Henk G. Schmidt, PhD

S. Mamede is associate professor, Institute of Medical Education Research Rotterdam, Erasmus MC, and Department of Psychology, Education & Child Studies, Erasmus University Rotterdam, Rotterdam, the Netherlands; ORCID: <https://orcid.org/0000-0003-1187-2392>.

W.E. Hautz is assistant professor of medical education and chief of service, Department of Emergency Medicine, Inselspital University Hospital Bern, Bern, Switzerland; ORCID: <https://orcid.org/0000-0002-2445-984X>.

C. Berendonk is deputy head, Department for Assessment and Evaluation, Institute for Medical Education, University of Bern, Bern, Switzerland; ORCID: <https://orcid.org/0000-0002-3740-9358>.

S.C. Hautz is educational consultant, Department of Emergency Medicine, Inselspital University Hospital Bern, University of Bern, Bern, Switzerland; ORCID: <https://orcid.org/0000-0003-4715-8465>.

T.C. Sauter is senior consultant, Department of Emergency Medicine, Inselspital University Hospital Bern, Bern, Switzerland; ORCID: <https://orcid.org/0000-0002-6646-5789>.

J. Rotgans is assistant professor, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore.

L. Zwaan is assistant professor, Institute of Medical Education Research Rotterdam, Erasmus MC, Rotterdam, the Netherlands.

H.G. Schmidt is professor, Institute of Medical Education Research Rotterdam, Erasmus MC, and Department of Psychology, Education & Child Studies, Erasmus University Rotterdam, the Netherlands.

Correspondence should be addressed to Sílvia Mamede, Institute of Medical Education Research Rotterdam, Erasmus MC, University Medical Center Rotterdam, Erasmus Medical Center – Ae-242, Wytemaweg 80, 3015 CN Rotterdam, the Netherlands; telephone: + 31-10-703-8270; email: s.mamede@erasmusmc.nl.

Supplemental digital content for this article is available at <http://links.lww.com/ACADMED/A791>.

Acknowledgments: Contributors: The authors would like to thank the assessment committee of the Swiss society for internal medicine, in particular Dr. Ullrich Stoller, Thun and Ursula Käser, Bern; and acknowledge Dr. Simone Ehrhard and Dr. Karin Ernst, Department of Emergency Medicine at Inselspital Bern, for their support in data acquisition.

Funding/Support: The study costs were covered by a guest professorship grant for Dr. Mamede awarded to Dr. Hautz by Inselspital Bern.

Other disclosure: None reported.

Ethical approval: The ethics committee of the Cantone Bern, Switzerland deemed the study exempt from full ethical review (Req-2017-00967) on December 14, 2017.

Abstract

Purpose

Diagnostic errors have been attributed to failure to sufficiently reflect on initial diagnoses. However, evidence of the benefits of reflection is conflicting. This study examined whether reflection upon initial diagnoses on difficult cases improved diagnostic accuracy and whether reflection triggered by confrontation with case evidence was more beneficial than simply revising initial diagnoses.

Method

Participants were physicians in Bern, Switzerland, registered for the 2018 Swiss internal medicine certification exam. They diagnosed written clinical cases, providing an initial diagnosis by following the same instructions and returning to the case to provide a final diagnosis. The latter required different types of reflection depending on the physician's experimental condition: return without instructions; identify confirmatory evidence; identify contradictory evidence; or identify both confirmatory and contradictory evidence. The authors examined diagnostic accuracy scores (range 0–1) as a function of diagnostic phase and reflection type.

Results

One hundred and sixty-seven physicians participated. Diagnostic accuracy scores did not significantly differ between the 4 groups of physicians in the initial (I) or the final (F) diagnostic phase [mean (95% CI): return without instructions, I: 0.21 (0.17, 0.26), F: 0.23 (0.18, 0.28); confirmatory evidence, I: 0.24 (0.19, 0.29), F: 0.31 (0.25, 0.37); contradictory evidence, I: 0.22 (0.17, 0.26), F: 0.26 (0.22, 0.30); confirmatory and contradictory evidence, I: 0.19 (0.15, 0.23), F: 0.25 (0.20, 0.31)]. Regardless of type of reflection employed while revising the case, accuracy increased significantly between initial and final diagnosis: I, 0.22 (0.19, 0.24) vs. F, 0.26 (0.24, 0.29); $P < .001$.

Conclusions

Physicians' diagnostic accuracy improved after reflecting upon initial diagnoses provided for difficult cases, independently of the evidence searched for while reflecting. The findings support the importance attributed to reflection in clinical teaching. Future research should investigate whether revising the case can become more beneficial by triggering additional reflection.

ACCEPTED

Diagnostic error has attracted the attention of the public and researchers since the 1999 Institute of Medicine report *To Err is Human* demonstrated the large societal cost of medical errors.¹ A recent follow-up report points to diagnostic error as one of the most common and most harmful patient safety problems.² Based on research showing diagnostic errors to affect around 12 million adults each year in U.S. outpatient settings alone, the report estimates that most people are likely to experience at least 1 diagnostic error in their lifetime. Many of these errors have minor consequences, but patients can also be severely harmed,³⁻⁶ and diagnostic error remains the most common and most costly reason internationally for malpractice claims in every large health system.⁷

Retrospective studies of malpractice claims⁸ and patients files^{9,10} have suggested that physicians' cognitive processes are implicated in most cases of diagnostic errors. What can go wrong in physicians' reasoning and how mistakes can be minimized have been a subject of much debate.¹¹⁻¹³ Research on clinical reasoning over the last decades has shown that physicians tend to generate diagnostic hypotheses early in a clinical encounter, subsequently verifying them by gathering additional information.^{14,15} While hypotheses are generated through an intuitive, largely unconscious process of pattern-recognition, their verification takes place under conscious control. The diagnostic process tends to involve, therefore, both intuitive and reflective reasoning modes, but the extent to which clinicians adopt one or the other mode while diagnosing a particular case seems to vary substantially depending on several factors.^{16,17}

Diagnostic errors have been frequently associated with failure to engage in reflection with a consequent excessive reliance on first impressions.^{11,13,18,19} Several authors have argued that returning to the case to verify the grounds of initial diagnosis would repair eventual errors made by rapid, intuitive judgments, thereby reducing diagnostic mistakes. Such errors can happen when physicians' attention is caught by findings in the case that, though salient, are

actually irrelevant (or not so relevant), and a wrong initial diagnosis is generated. Repairing this wrong diagnosis is not always easy, because we all have a natural tendency to look for (and value) evidence that supports rather than refutes our impressions.^{20,21} Only when physicians engage in critically scrutinizing the case evidence can they counteract this tendency, opening the door for recognition of actually relevant findings to occur, which eventually brings the right diagnosis to mind.

This explanation for diagnostic error and the role of reflection builds upon research in psychology of reasoning.^{22,23} There is also some empirical evidence that the findings from this research in fact apply to medical diagnosis. An approach to foster deliberate reflection upon initial diagnosis increased diagnostic accuracy in several studies.²⁴⁻²⁷ Deliberate reflection corrected initial mistakes, at least when cases were not straightforward,^{24,25} and also counteracted the adverse effect of cognitive bias induced, for example, by recent experiences with a similar-looking (but in fact different) disease.^{26,27} The approach employed in these experiments requires physicians to return to the case to search for evidence that speaks in favor of and against their initial diagnosis, then consider which other diagnoses would be plausible and submit each diagnosis to similar analysis before making a final decision. Studies using checklists to guide reflection upon the problem during verification of initial diagnostic hypothesis have also found substantial increase in accuracy after reflection.^{28,29} However, other studies found reflection to have no added value. An experiment that requested physicians to diagnose clinical cases by following instructions either to be as quick as possible or to be careful and reflective found no differences in diagnostic accuracy.³⁰ The negative relationship between accuracy and time to diagnosis observed in a study with medical residents was interpreted as a sign that there would be no advantage of spending more time to reflect further on the case.³¹ A study by Ilgen and colleagues³² requested physicians to diagnose cases either by trusting the sense of familiarity

and giving the first seemingly plausible diagnosis or by first summarizing the case information, then listing alternative diagnoses and, only after that, deciding which one was the most likely in light of the case features. The more reflective approach did not lead to higher diagnostic accuracy relative to the first diagnostic impression.

It therefore remains unclear whether physicians and their patients would actually benefit from further reflection upon initial diagnoses to increase diagnostic accuracy, and, consequently, whether medical teachers should teach the value of reflection. Case difficulty and participants' expertise apparently influence what can be gained from reflection,^{24,25} but none of these factors differed substantially in the studies that arrived at discrepant findings. They cannot therefore explain the discrepancies, and other factors might play a role. Specifically, the different methodological approaches that the studies have employed suggest that what results from further reflection depends on what reflection entails, that is, on the *type* of reflection. The deliberate reflection approach that has been shown to improve initial diagnosis²⁴⁻²⁷ confronted physicians with confirmatory and contradictory evidence from the case. Because this confrontation directs attention to findings that may have remained initially unnoticed, it fosters retrieval of appropriate knowledge and reorganization of diagnostically relevant information. This restructuring of initial reasoning may be required for reflection to help, and it is likely to take place particularly when the evidence that physicians are requested to search for speaks *against* the initial diagnosis. This claim seems reasonable, and is supported by psychological research,^{33,34} but to our knowledge has not been empirically investigated. Indeed, it is not clear whether a minimal search for any type of evidence would already suffice to correct initial mistakes.

This study aimed to examine, first, whether reflection upon an initial diagnosis improved diagnostic accuracy and, second, whether reflection triggered by confrontation with different types of case evidence was more beneficial than simply revising the initial diagnosis.

Physicians diagnosed clinical cases, first providing an initial diagnosis and then returning to the case to reflect further before making a final diagnosis. The final diagnosis was preceded by 1 of 4 “types” of reflection that differed in the extent to which physicians were confronted with evidence from the case. We expected reflection to improve diagnostic accuracy relative to initial diagnosis, with the improvement possibly increasing with the amount of reflection involved. As a secondary research question, we examined whether accuracy of the first diagnosis was associated with time spent on diagnosis.

Method

Overview

The study was a randomized experiment with a mixed design in which all participants diagnosed a set of clinical cases one by one, first providing a diagnosis and then returning to the case to give a final diagnosis. All participants followed the same instructions to give the first diagnosis, but the final diagnosis was preceded by one of 4 different “types” of reflection, depending on the experimental condition to which they had been randomly allocated: return without instructions; identification of confirmatory evidence; identification of contradictory evidence; or identification of both confirmatory and contradictory evidence. Figure 1 presents the study design.

Setting and participants

All physicians registered for the 2018 Swiss board exam for general internal medicine in Bern, Switzerland, were considered eligible for the study. Senior residents and practicing physicians are allowed to take the licensing exam to be certified as specialists in internal medicine. We sent a letter to registered physicians with an invitation to participate in the

study, which would take place immediately after the exam. Those who accepted the invitation were recruited as participants, randomly allocated to 1 of the 4 experimental conditions, and tested after the exam.

A priori power analysis, assuming to-be-detected effects of medium size (Cohen's $f = 0.25$)³⁵ and the standard alpha level of .05, provided the estimation that a sample of 136 residents and practicing physicians would be sufficient to achieve a power of 0.80.

We used a coding scheme to ensure that responses would be anonymous and could be linked to the scores obtained by the participants in the board exam. All participants provided written informed consent to participate in the study, including matching their responses to their exam score, and received \$40.00 for their participation. The ethics committee of the Cantone Bern, Switzerland, deemed the study exempt from full ethical review (Req-2017-00967).

Material and procedure

The study used 8 written clinical cases. Each case consisted of a brief description of a patient's history, complaints, symptoms, and findings from physical examination and tests. All cases had a confirmed diagnosis and had been used in previous studies with internal medicine residents.²⁴⁻²⁷ We chose cases to which a mean diagnostic accuracy score around 0.3–0.4 (max 1) was observed in these previous studies. We aimed at difficult cases as they provide a basis for mistakes and repair by reflection to occur. The diagnoses of the cases were: small cell lung cancer, inflammatory bowel disease, bacterial pneumonia with sepsis, acute bacterial endocarditis, pseudomembranous colitis, Vitamin B12 deficiency, celiac disease, and peripheral arterial occlusion disease. The cases were presented in a booklet, 1 per page. To control for order effects, we prepared 2 versions of the booklets by alternating sequence of presentation.

For each case, we requested first that participants read the case and write down the most likely diagnosis for the case as quickly as possible but without jeopardizing accuracy. This instruction has been used to induce a more intuitive reasoning mode.^{24,25,30} To be clear, we understand that any diagnostic reasoning will always involve some degree of reflection, particularly with difficult cases. However, this instruction was intended to result in an initial processing as fast as possible to make review of the initial diagnosis meaningful. After that, the same case was presented again, and we requested that the participants follow different instructions, depending on the condition to which they had been assigned. In the return without instructions condition, they were presented the case again and requested to write down their final decision on the most likely diagnosis. In the confirmatory condition, we asked them to write down findings in the case that spoke in *favor* of the initial diagnosis, and the final most likely diagnosis. In the contradictory condition, they had to write down findings in the case that spoke *against* the initial diagnosis, and the final most likely diagnosis. Finally, in the confirmatory and contradictory condition, we asked that they write down findings in the case that spoke in *favor* of the initial diagnosis; findings in the case that spoke *against* the initial diagnosis, and the final most likely diagnosis for the case. The participants registered the time before and after each page by looking at a large digital clock visible in the room. So, in brief, for each case, participants read the case, provided an initial diagnosis, reflected upon it in a manner determined by the experimental condition, and then provided a final diagnosis. They then moved to the next case and repeated this procedure. An example case is available in Supplemental Digital Appendix 1, at <http://links.lww.com/ACADMED/A791>.

Before diagnosing the cases, the participants provided information age, gender, and number of years in professional practice and, after completing the study, were asked “how often did you encounter the following diseases in the past?” followed by a list of the correct diagnoses for the 8 cases. Responses were collected on a 5-point Likert scale (from 1 = never to 5 = very frequently).

The licensing exam itself consists of 120 single best answer multiple choice questions.

Data analysis

The accuracy of participants’ diagnoses was evaluated by considering the confirmed diagnosis of each case as a standard. Two board-certified internists (C.B., T.C.S.) blinded toward the experimental condition independently evaluated each diagnosis as correct, partially correct, or incorrect (scored as 1, 0.5, or 0 points, respectively). We considered a response correct when it mentioned the core diagnosis, and partially correct when the core was not cited but a constituent element of the diagnosis was mentioned. The interrater agreement was high [ICC(3,2) = 0.96], and disagreements were decided upon by a third rater (W.E.H.).

To verify whether the 4 groups were similar in variables that could eventually influence the results, we performed separate analysis of variance (ANOVA) with experimental condition as between-subjects factors on age, number of years of clinical practice, experience with the diseases included in the study, and score obtained in the certification exam. We performed a repeated-measures ANOVA with experimental condition as between-subjects factor (return without instructions; confirmatory; contradictory; confirmatory and contradictory) and diagnostic phase (initial diagnosis; revised diagnosis) as within-subjects factor on the mean diagnostic accuracy scores. To examine the relationship between time to diagnosis and diagnostic accuracy, we computed Pearson’s correlation coefficient for the first phase, when

the diagnostic process was not under the influence of our treatment. The data were analyzed using SPSS statistical software, version 25 for Mac (IBM Corp, Armonk, New York).

Results

One-hundred and sixty-seven physicians enrolled in the study and were randomized to 1 of the 4 groups. Table 1 presents the participants' background information. The groups did not significantly differ in age [$F(3,162) = 0.28; P = .84$], gender [$\chi^2(6) = 7.23; P = .30$], number of years of clinical practice [$F(3,161) = 0.49; P = .69$], reported experience with the diseases included in the study [$F(3,162) = 0.45; P = .71$], or the score obtained in the board certification exam [$F(3,152) = 2.25; P = .09$].

Table 2 presents the mean diagnostic accuracy scores for the initial diagnosis and the final diagnosis as a function of experimental condition. The mixed ANOVA showed a significant main effect of diagnosis phase, [$F(1,163) = 31.29; P < .001; \eta_p^2 = 0.16$], with all groups showing higher diagnostic accuracy in the second phase (revised diagnosis) relative to the first phase (initial diagnosis). The main effect of the experimental condition under which the physicians performed was not significant [$F(3,163) = 1.19; P = .31; \eta_p^2 = 0.02$], and there was no significant interaction effect [$F(3,163) = 1.80; P = .15; \eta_p^2 = 0.03$].

Time spent to diagnose the cases is presented in Table 3. There was a positive correlation between accuracy of the initial diagnosis and time spent to make this diagnosis, $r = .23, P = .004$.

Discussion

In this study, we investigated whether reflection upon initial diagnoses improved the diagnostic accuracy of a group of physicians and whether improvement was influenced by the type of confrontation with evidence from the case. The findings are in line with our hypothesis that reflection would increase diagnostic accuracy. Overall, accuracy scores increased significantly between initial and revised diagnoses, though the effect size was

small. Contrary to our expectation, this increase did not depend on whether physicians were exposed to confrontation with evidence from the case. Simply returning to the case and having the chance to revise the initial diagnosis before making a final decision was enough to improve accuracy. Time to diagnose was associated with accuracy of initial diagnoses. These findings are in line with studies showing that reflection on initial diagnoses helps physicians repair errors and improves diagnostic performance.²⁴⁻²⁷ Other studies, however, have found no benefit of reflection to diagnostic performance.³⁰⁻³² What could explain these discrepant findings? Evidence of the positive effect of reflection emerged from studies with difficult cases, but the diagnostic accuracy scores observed in some studies with negative results suggest that their cases were not straightforward either.^{30,31} Other factors besides differences in case difficulty might therefore explain why reflection led to improved accuracy in some studies but not in others. There may be a key conceptual and methodological difference that affects what can be gained from reflection. In this study, reflection is conceived as a deliberate consideration of initial judgments aimed at verifying its grounds. This conceptualization is shared by studies that found reflection to improve initial diagnosis.³⁶ Conversely, studies that found no advantage of reflection to diagnostic performance encouraged physicians to reflect throughout the whole diagnostic process, including both the generation of diagnostic hypotheses and their verification. The primary mechanism through which reflection helps is by leading physicians to recognize relevant case findings that were initially overlooked or misinterpreted, which seems a main source of diagnostic error among experienced physicians.^{37,38} This is why reflection only helped when there was enough knowledge to recognize actually relevant features while revising the case.²⁵ It also explains why reflection improved on initial diagnoses when physicians were allowed to review the case features but was not beneficial when physicians could not go back to the case.²⁸ It may be that the request to return to the case to revise initial judgments allows

reflection to act because it triggers a search for possible mistakes, with a more critical check of the grounds of initial judgments that is not present (or not so much) when physicians are simply asked to be careful from the start. One can expect a request to review to induce such scrutinization of initial hypotheses even when participants spend more time to give the initial diagnosis than to reflect upon it, as it happened in our study (possibly because the highly difficult cases demanded time to make sense of and integrate all the information).

Regarding the *type* of reflection, confronting physicians with evidence from the case during reflection did not make it more beneficial. Even when physicians received no instruction on what to search for, returning to the case increased accuracy. The effect of reflection was small, with a gain of 22% in accuracy. In other studies with similar participants, diagnostic accuracy improved 40% or even more after physicians deliberately reflected upon their initial diagnosis for complex cases.^{24,25} In these previous studies, participants were not only requested to search for both confirmatory and contradictory evidence, but they were also required to generate alternative diagnoses and submit them to a similar analysis before making a final diagnostic decision. In the current study, participants were not asked to consider alternatives and, though they may have eventually done so, they certainly did not engage in scrutinizing the grounds for each of these alternative diagnoses. The amount of reflection triggered by the confrontation with the case evidence in the present study is therefore much less extensive than in previous studies, therefore reducing its potential to restructure initial reasoning. Indeed, while deliberate reflection upon a case took 5–7 minutes in those studies,^{24,25} our participants invested around 2 minutes in reflection. This may explain why confrontation with evidence as operationalized in this study did not lead to more substantial improvement than simply revising the case. It may also be the reason why gains after reflection were smaller than in previous studies.

To the best of our knowledge, this is the first study to empirically investigate the effect the type of reflection has on diagnostic accuracy. Our findings add to what we know about it and show issues requiring further investigation. Future research should explore whether an approach that makes the confrontation with case evidence more “reflection-triggering,” for example, by requesting physicians to generate alternative diagnoses, would increase the potential of reflection to improve diagnostic accuracy. While the deliberate reflection procedure used in several studies has proven very powerful to repair diagnostic errors, it is too time consuming and effortful to be applicable in real settings. In our study, taking a second look at an initial diagnosis required physicians around 2 minutes, and this short time was sufficient to allow for initial mistakes to be corrected. The increase in accuracy was small, but our findings suggest that suspending a decision, returning to the case, and revising an initial diagnosis is worthwhile, at least when cases are difficult. Returning to reflect is likely to be feasible in most situations in clinical practice. Whether it is possible to increase the potential benefit of this approach by triggering more reflection while keeping it within the boundaries of what is feasible in practice requires further investigation.

This study has limitations. First, our participants had on average around 5 years of clinical practice, and it is not clear whether the findings would apply to more experienced physicians. Because difficulties in restructuring initial diagnostic reasoning seem to increase with experience,³⁹ it may be that more experienced physicians would benefit more from a more extensive approach to reflection. This idea requires further investigation. Second, the study was conducted immediately after a high-stakes exam that may have tired participants, potentially hindering their performance both while giving an initial diagnosis and during reflection. However, the thoroughness of the participants’ responses indicate how seriously the physicians took their task. Finally, we used written clinical cases, which do not provide physicians with all the cues that would be available in real settings. To what extent these

findings generalize to actual practice is to be determined. Nevertheless, empirical research has shown that written cases allow for reliably detecting group-level differences and are a good proxy for the investigation of physicians' performance in real settings.^{40,41}

Summing up, this study has found that returning to the case to reflect upon initial diagnoses increased diagnostic accuracy on difficult clinical cases, reinforcing the value of further reflection to reduce diagnostic errors. The improvement in accuracy was small and not dependent on what physicians were required to search for during reflection. Future research should investigate whether revising the case can be made more beneficial by triggering additional reflection while maintaining feasibility for real settings.

ACCEPTED

References

1. Kohn LT, Corrigan J, Donaldson MS. *To Err Is Human : Building a Safer Health System*. Washington, DC: National Academy Press; 2000.
2. Balogh E, Miller BT, Ball J, Institute of Medicine (U.S.). *Committee on Diagnostic Error in Health Care. Improving Diagnosis in Health Care*. Washington, DC: The National Academies Press; 2015.
3. Shojania KG, Burton EC, McDonald KM, Goldman L. Changes in rates of autopsy-detected diagnostic errors over time: A systematic review. *JAMA*. 2003;289:2849–2856.
4. Hautz WE, Kammer JE, Hautz SC, et al. Diagnostic error increases mortality and length of hospital stay in patients presenting through the emergency room. *Scand J Trauma Resusc Emerg Med*. 2019;27:54.
5. Sauter TC, Capaldo G, Hoffmann M, et al. Non-specific complaints at emergency department presentation result in unclear diagnoses and lengthened hospitalization: A prospective observational study. *Scand J Trauma Resusc Emerg Med*. 2018;26:60.
6. Zwaan L, de Bruijne M, Wagner C, et al. Patient record review of the incidence, consequences, and causes of diagnostic adverse events. *Arch Intern Med*. 2010;170:1015–1021.
7. Graber ML. The incidence of diagnostic error in medicine. *BMJ Qual Saf*. 2013;22(Suppl 2):ii21–ii27.
8. Poon EG, Kachalia A, Puopolo AL, Gandhi TK, Studdert DM. Cognitive errors and logistical breakdowns contributing to missed and delayed diagnoses of breast and colorectal cancers: A process analysis of closed malpractice claims. *J Gen Intern Med*. 2012;27:1416–1423.

9. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med.* 2005;165:1493–1499.
10. 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;173:418–425.
11. Croskerry P. From mindless to mindful practice—Cognitive bias and clinical decision making. *N Engl J Med.* 2013;368:2445–2448.
12. Norman GR, Monteiro SD, Sherbino J, Ilgen JS, Schmidt HG, Mamede S. The causes of errors in clinical reasoning: Cognitive biases, knowledge deficits, and dual process thinking. *Acad Med.* 2017;92(1):23-30.
13. Redelmeier DA. Improving patient care: The cognitive psychology of missed diagnoses. *Ann Intern Med.* 2005;142:115–120.
14. Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise: Theory and implications. *Acad Med.* 1990;65:611–621.
15. Schmidt HG, Rikers RM. How expertise develops in medicine: Knowledge encapsulation and illness script formation. *Med Educ.* 2007;41:1133–1139.
16. Mamede S, Schmidt HG, Rikers RM, Penaforte JC, Coelho-Filho JM. Breaking down automaticity: Case ambiguity and the shift to reflective approaches in clinical reasoning. *Med Educ.* 2007;41:1185–1192.
17. Mamede S, Schmidt HG, Rikers RMJP, Penaforte JC, Coelho-Filho JM. Influence of perceived difficulty of cases on physicians' diagnostic reasoning. *Acad Med.* 2008;83:1210–1216.
18. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med.* 2003;78:775–780.
19. Klein JG. Five pitfalls in decisions about diagnosis and prescribing. *BMJ.* 2005;330:781–783.

20. Kostopoulou O, Russo JE, Keenan G, Delaney BC, Douiri A. Information distortion in physicians' diagnostic judgments. *Med Decis Making*. 2012;32:831–839.
21. Wallsten TS. Physician and medical student bias in evaluating diagnostic information. *Med Decis Making*. 1981;1:145–164.
22. Evans JST. The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*. 2006;13:378–395.
23. Kahneman D. A perspective on judgment and choice: Mapping bounded rationality. *Am Psychol*. 2003;58:697–720.
24. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Med Educ*. 2008;42:468–475.
25. Mamede S, Schmidt HG, Rikers RMJP, Custers EJFM, Splinter TAW, van Saase JLCM. Conscious thought beats deliberation without attention in diagnostic decision-making: At least when you are an expert. *Psychological Research-Psychologische Forschung*. 2010;74:586–592.
26. 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*. 2010;304:1198–1203.
27. Schmidt HG, Mamede S, van den Berge K, van Gog T, van Saase JL, Rikers RM. Exposure to media information about a disease can cause doctors to misdiagnose similar-looking clinical cases. *Acad Med*. 2014;89:285–291.
28. Sibbald M, de Bruin AB, Cavalcanti RB, van Merriënboer JJ. Do you have to re-examine to reconsider your diagnosis? Checklists and cardiac exam. *BMJ Qual Saf*. 2013;22:333–338.
29. Sibbald M, de Bruin AB, van Merriënboer JJ. Checklists improve experts' diagnostic decisions. *Med Educ*. 2013;47:301–308.

30. Norman G, Sherbino J, Dore K, et al. The etiology of diagnostic errors: A controlled trial of system 1 versus system 2 reasoning. *Acad Med.* 2014;89:277–284.
31. Sherbino J, Dore KL, Wood TJ, et al. The relationship between response time and diagnostic accuracy. *Acad Med.* 2012;87:785–791.
32. Ilgen JS, Bowen JL, McIntyre LA, et al. Comparing diagnostic performance and the utility of clinical vignette-based assessment under testing conditions designed to encourage either automatic or analytic thought. *Acad Med.* 2013;88:1545–1551.
33. de Neys W, Glumicic T. Conflict monitoring in dual process theories of thinking. *Cognition.* 2008;106:1248–1299.
34. De Neys W, Vartanian O, Goel V. Smarter than we think: When our brains detect that we are biased. *Psychol Sci.* 2008;19:483–489.
35. Cohen J. A power primer. *Psychol Bull.* 1992;112:155–159.
36. Mamede S, Schmidt HG. Reflection in medical diagnosis: A literature review. *Health Professions Education.* 2017;3:17–25.
37. Groves M, O'Rourke P, Alexander H. Clinical reasoning: The relative contribution of identification, interpretation and hypothesis errors to misdiagnosis. *Med Teach.* 2003;25:621–625.
38. Groves M, O'Rourke P, Alexander H. The clinical reasoning characteristics of diagnostic experts. *Med Teach.* 2003;25:308–313.
39. Eva KW. The aging physician: Changes in cognitive processing and their impact on medical practice. *Acad Med.* 2002;77(10 Suppl):S1–S6.
40. Mohan D, Fischhoff B, Farris C, et al. Validating a vignette-based instrument to study physician decision making in trauma triage. *Med Decis Making.* 2014;34:242–252.

41. Peabody JW, Luck J, Glassman P, et al. Measuring the quality of physician practice by using clinical vignettes: A prospective validation study. *Ann Intern Med.* 2004;141:771–780.

ACCEPTED

Figure 1

Diagram of the study design and participants flow for 167 physicians enrolled for the Swiss internal medicine certification exam, from a study of reflection on initial diagnoses and final diagnostic accuracy, Bern, Switzerland, 2018.

ACCEPTED

Table 1**Characteristics of 167 Physician Participants Enrolled for the the Swiss Internal Medicine Certification Exam, From a Study of Reflection On Initial Diagnoses and Final Diagnostic Accuracy, Bern, Switzerland, 2018^a**

Characteristic	Return without instruction	Confirmatory	Contradictory	Confirmatory and contradictory	Total
Age, mean (95% CI)	31.34 (30.09, 32.59)	31.24 (29.71, 32.77)	31.73 (30.62, 32.85)	30.87 (29.53, 32.21)	31.30 (30.66, 31.94)
Female, no. (%)	20 (44.4)	28 (66.7)	20 (48.8)	20 (51.3)	88 (52.7)
Number of years in clinical practice, mean (95% CI)	5.29 (4.36, 6.23)	5.07 (4.23, 5.91)	5.73 (4.66, 6.80)	4.95 (3.87, 6.02)	5.27 (3.4.79, 5.74)
Experience with the diseases (range 1–5), mean (95% CI)	2.78 (2.60, 2.97)	2.70 (2.55, 2.86)	2.70 (2.56, 2.86)	2.81 (2.64, 2.98)	2.75 (2.67, 2.83)
Exam score, mean (95% CI)	90.10 (86.70, 93.49)	91.29 (88.61, 93.97)	86.21 (82.45, 89.97)	91.75 (88.19, 95.31)	89.85 (88.19, 91.50)

Abbreviation: CI, confidence interval.

^aAll participants were recruited from the pool of physicians registered for the Swiss board exam for general internal medicine were considered eligible for the study. Senior residents and practicing physicians are allowed to take the licensing exam to be certified as specialists in internal medicine.

Table 2

Mean Diagnostic Accuracy Scores Obtained by 167 Physician Participants at the First and Second Diagnostic Phases, From a Study of Reflection On Initial Diagnoses and Final Diagnostic Accuracy, Bern, Switzerland, 2018^a

Diagnosis	Return without instruction	Confirmatory	Contradictory	Confirmatory and contradictory	Overall
Initial diagnosis, mean (95% CI)	0.21 (0.17, 0.26)	0.24 (0.19, 0.29)	0.22 (0.17, 0.26)	0.19 (0.15, 0.23)	0.22 (0.19, 0.24)
Final diagnosis, mean (95% CI)	0.23 (0.18, 0.28)	0.31 (0.25, 0.37)	0.26 (0.22, 0.30)	0.25 (0.20, 0.31)	0.26 (0.24, 0.29)

Abbreviation: CI, confidence interval.

^aRange of accuracy scores was 0–1, where 0 = incorrect and 1 = correct. All participants were recruited from the pool of physicians registered for the Swiss board exam for general internal medicine. Senior residents and practicing physicians are allowed to take the licensing exam to be certified as specialists in internal medicine.

Table 3

Mean Time Spent by 167 Physician Participants in Diagnosis (Seconds) at the First and Second Diagnostic Phases, From a Study of Reflection On Initial Diagnoses and Final Diagnostic Accuracy, Bern, Switzerland, 2018^a

Diagnosis	Return without-instruction	Confirmatory	Contradictory	Confirmatory and contradictory	Overall
Initial diagnosis, mean (95% CI)	140.05 (123.85, 156.24)	151.76 (137.35, 166.17)	155.91 (138.21, 173.61)	129.17 (114.92, 143.43)	144.59 (136.81, 152.37)
Final diagnosis, mean (95% CI)	79.00 (65.70, 92.29)	108.05 (93.95, 122.15)	116.57 (101.46, 131.68)	129.37 (108.05, 150.69)	107.57 (99.31, 115.84)

Abbreviation: CI, confidence interval.

^aAll participants were recruited from the pool of physicians registered for the Swiss board exam for general internal medicine. Senior residents and practicing physicians are allowed to take the licensing exam to be certified as specialists in internal medicine.

Figure 1

