The Effect of Information Presentation Order on Residents' Diagnostic Accuracy of Online Simulated Patients With Chest Pain

René A. Tio, MD, PhD
Marco A. Carvalho Filho, MD, PhD
Marcos F. de Menezes Mota, MS
André Santanchè, PhD
Sílvia Mamede, MD, PhD

ABSTRACT

Background Physicians may receive diagnostic information in different orders, and there is a lack of empirical evidence that the order of presentation may influence clinical reasoning.

Objective We investigated whether diagnostic accuracy of chest pain cases is influenced by the order of presentation of the history and electrocardiogram (EKG) to cardiology residents.

Methods We conducted an experimental study during a resident training in 2019. Twelve clinical cases were presented in 2 diagnostic rounds. Residents were randomly allocated to seeing the EKG first (EKGF) or the history first (HF). The mean diagnostic accuracy scores (range 0-1) and confidence level (0-100) in each diagnostic round and time needed to make the diagnosis were evaluated.

Results The final diagnostic accuracy was higher than the initial in both groups. After the first round, diagnostic accuracy was higher in HF (n=24) than in EKGF (n=28). Time taken to judge the history was comparable in both groups. Time taken to judge the EKG was shorter in HF (40±11 vs 64±13 seconds; $P<.01$). Time invested in the second round was significantly correlated with changing the initial diagnosis. A significant difference was observed in confidence ratings after the initial diagnosis, with EKGF reporting less confidence relative to HF.

Conclusions The order in which history and EKG are presented influences the clinical reasoning process.

Introduction

It is unclear whether the order in which information is presented influences the clinical reasoning process. In the evaluation of patients with chest pain, clinicians have 3 cornerstones to lean on: the history, electrocardiogram (EKG), and laboratory findings.[1] EKG and history are the first at clinicians' disposal and will prime the clinical reasoning process.

Knowledge of the patient history influences EKG interpretation. Studies by Hatala et al showed that EKG interpretation was more often correct when preceded by a history suggestive of the correct diagnosis, compared to an alternative diagnosis or no scenario at all.[2,3] Empirical research has also shown that a diagnostic hypothesis affects recognition and interpretation of findings encountered subsequently in a clinical case, leading for instance to overvaluing features consistent with the initial diagnosis and the other way around.[4,5] Diagnostic reasoning involves a hypothetico-deductive method. An initial diagnostic hypothesis is followed by gathering additional information to verify this hypothesis.[6,7] This generation is largely unconscious, based on activation of illness scripts.[8,9] The activated script guides the search for additional findings expected for that diagnosis. In case of an incorrect diagnostic hypothesis, its influence on the subsequent search for additional findings hinders the verification phase of the diagnosis. Verifying the initial interpretation of EKGs has been shown to improve diagnostic decisions.[10,11] Deliberate reflection upon the initial diagnosis has been shown to reduce diagnostic error, especially when cases are complex or contextually irrelevant information tends to mislead reasoning.[12-14] Even simply returning to the case to verify the initial diagnosis improved diagnostic accuracy.[15] However, physicians do not often recognize the need for further verification.[16] Whether the success of verification, and of the whole diagnostic process, depends on time invested remains an open question. While some studies have found no association between diagnostic accuracy and time spent in diagnosis,[17,18] other experimental studies have shown time restrictions can reduce accuracy,[19,20] as well as

DOI: http://dx.doi.org/10.4300/JGME-D-21-01053.1

correct diagnoses being made faster than incorrect ones.[21]

The order in which information is available may influence diagnostic performance. To address this issue, we conducted a randomized controlled trial. We investigated the effect of seeing the EKG first (EKGF) or obtaining the history first (HF) on diagnostic performance. Our hypotheses were: (1) diagnostic accuracy would depend on order of information presented, with EKGF having lower diagnostic accuracy relative to HF; (2) time dedicated to each component of the problem (EKG and history) would depend on the order of presentation, with more time dedicated to the component when it comes first relative to when it comes second; (3) changing the initial diagnosis would depend on the amount of time dedicated to the second component of the problem; and (4) confidence in the initial diagnosis would be inversely related to time invested in the second component of the problem and to changing the initial diagnosis.

Objectives
To investigate whether diagnostic accuracy of chest pain cases is influenced by the order of presentation of the history and electrocardiogram.

Findings
The final diagnostic accuracy was higher than the initial in both groups. Time invested in the second diagnostic round was significantly correlated with changing the initial diagnosis.

Limitations
As this is an experimental study, it is unclear if our findings could be extrapolated to other residents and fellows or other specialties as well to real-life situations under time pressure.

Bottom Line
Be aware that the order in which information is presented may influence the reasoning process in your educational setting as well as in the clinical setting.

Methods

Design

Participants were third-year cardiology residents from all 15 cardiology training programs in the Netherlands who were attending a course on acute cardiac care (course director R.A.T.), an obligatory part of the residency program.

The face-to-face course took place in November 2019. The cases were presented online toward the beginning of the course, and participants were told that the cases would be discussed by the group at the end of the course for teaching purposes. They were blinded to the experimental nature of the intervention. Twenty written clinical cases were prepared based on cases selected from a database of chest pain patients visiting the emergency department of the Catharina Ziekenhuis hospital in Eindhoven, Netherlands. Cases were selected based on their ambiguity by an experienced cardiologist (R.A.T.). The written cases were subsequently loaded de-identified in our Jacinto online platform (https://jacinto.harena.org). A specific module was developed in the platform to randomly assign customized cases to the participants. All cases were piloted by 2 other experienced cardiologists using the platform. Cases were excluded if one of them deemed it not sufficiently ambiguous on history or EKG. Finally, 12 cases remained to be used in the experiment.

All cases were categorized based on level of EKG abnormalities (completely normal, minor abnormalities, or apparent ischemic abnormalities) and the level of how typical the complaints were according to the Diamond-Forrester classification.[22] The final diagnosis was myocardial ischemia (acute coronary syndrome) in 7 of the cases and non-anginal chest pain in the other 5, ranging from pericarditis to gastroesophageal reflux and muscle pain.

All participants diagnosed the same 12 cases, which were presented in 2 diagnostic rounds in the online platform. Participants were given a unique login to be used on their own devices.

Intervention

Participants were randomly assigned to EKGF or HF. At the end of each round, the participant gave the most likely diagnosis as well as their level of confidence by placing a digital ruler on a scale from 0 to 100. Time needed to come to a diagnosis was automatically registered. All data were stored in a log file anonymously.

Outcome

Diagnostic accuracy was evaluated by 2 cardiologists independently and blinded for participants allocation to EKGF or HF. All 491 diagnoses given by the residents were listed for each case and judged by the cardiologists to be correct (score=1), incorrect (score=0), or in between (score=0.5). For example, the answer "possible angina/unstable angina" in an unstable angina case was rated 0.5. In case of disagreement (n=152) consensus was reached after discussing the given diagnosis. A mean diagnostic accuracy score for all cases was obtained first for each participant and then for each experimental condition.

A mixed analysis of variance (ANOVA) with diagnostic round as within-subjects factor (initial diagnosis and final diagnosis) and experimental condition (EKGF vs HF) as between-subjects factor

TABLE
Differences in Diagnostic Accuracy, Confidence in the Diagnosis, and Time Spent in the Diagnosis Between EKG First and History First Groups

Accuracy, Confidence, and Time	EKG First	History First	*P* Value
Accuracy after first round, mean (SD)	0.49±0.13	0.73±0.10	<.01
Accuracy after second round, mean (SD)	0.74±0.14	0.75±0.10	NS
Confidence after first round, mean (SD)	54±15	64±13	NS
Confidence after second round, mean (SD)	65±13	67±14	NS
Time spent on EKG, mean (SD), seconds	64.01±12.86	40.23±10.86	<.01
Time spent on history, mean (SD), seconds	64.67±18.63	60.80±12.86	NS
Diagnostic accuracy variation, mean (SD)	0.25±0.16	0.019±0.10	<.01

Abbreviations: EKG, electrocardiogram; NS, nonsignificant.

was performed on mean diagnostic accuracy scores. This analysis tested our first hypothesis that diagnostic accuracy would depend on the order of presentation of the components of the problem and vary across the diagnostic rounds. Post hoc analysis with independent and paired *t* test examined the significant interaction effect.

To test our second hypothesis regarding time spent in each component of the problem, we performed a mixed ANOVA with type of the component as within-subjects factor (EKG and history) and experimental condition as between-subjects factor (EKGF vs HF) with mean time dedicated to the problem component as dependent variable. A significant interaction effect was further explored by independent and paired *t* tests.

To test the third and fourth hypotheses, we first computed, for each participant, the "diagnostic accuracy variation" by subtracting the initial diagnostic accuracy score from the final diagnostic accuracy score. We first computed this variable for each case and subsequently the mean for all cases. This variable indicates the extent of change in the initial diagnosis, with zero indicating no change, a positive value showing that diagnostic accuracy improved between the initial and the second diagnostic rounds, and a negative value pointing to a decrease. Subsequently, we computed correlation coefficients between diagnostic accuracy variation and time spent in the second component of the problem (because the latter was not normally distributed, we used Spearman's correlation coefficient) and between mean confidence in the initial diagnosis and time spent in the second component (Spearman's correlation coefficient) and diagnostic accuracy variation (Pearson's correlation coefficient). For all correlations, the coefficient of determination, R^2, was also computed as a measure of the amount of variability in one variable that is explained by the other. The statistical analysis was performed on SPSS Statistics 25 (IBM Corp, Armonk, NY), and the significance level was set at $P<.05$ (2-tailed) for all analyses.

The board of the organizing national body gave a waiver for ethical approval since the quiz was part of the course and data were collected anonymously.

Results

A total of 52 residents (out of 55) participated in this study: 28 in EKGF and 24 in HF. The TABLE presents means and SDs for all outcome measurements. Results of the statistical tests performed to test the hypotheses are described below.

Diagnostic Accuracy

Overall, the final diagnostic accuracy was higher than the initial accuracy as shown by a significant main effect of diagnostic round, $F(1, 50)=52.37$; $P<.001$; $\eta_p^2=0.51$. After the first round diagnostic accuracy was higher in HF than in EKGF, $F(1, 50)=21.49$; $P<.001$; $\eta_p^2=0.30$. A significant interaction effect, $F(1, 50)=38.72$; $P<.001$; $\eta_p^2=0.44$, was present. While the final diagnostic accuracy was not significantly different between the 2 groups, $t(50)=0.28$; $P=.78$, the initial diagnostic accuracy was lower in EKGF, $t(50)=7.69$; $P\leq.001$, relative to HF. The accuracy of EKGF significantly increased after also knowing the history, $t(27)=8.39$; $P<.001$. This gain in diagnostic accuracy observed in EKGF did not happen in HF, whose performance was already high in the first diagnostic round, $t(23)=0.94$; $P=.36$ (see FIGURE 1 and TABLE).

Time Spent in Each Component of the Problem

FIGURE 2 presents the results relative to our second hypothesis. There was a significant main effect of the type of component, with more time spent overall in the history than in the EKG, $F(1, 50)=23.95$; $P<.001$; $\eta_p^2=0.32$. The main effect of experimental condition was also significant, with EKGF spending longer total

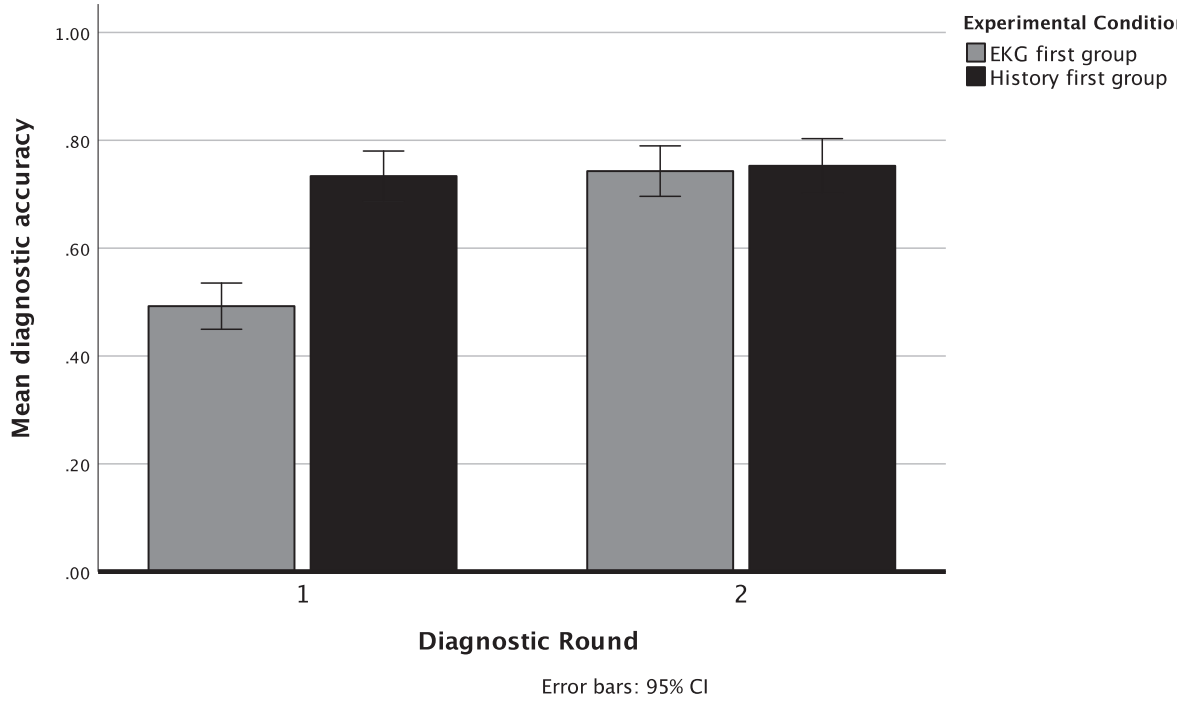

FIGURE 1
Diagnostic Accuracy as a Function of Experimental Condition and Diagnostic Round

time to diagnose a case than HF, $F(1, 50)=18.24$; $P<.001$; $\eta_p^2=0.27$. A significant interaction effect emerged, $F(1, 50)=21.04$; $P<.001$; $\eta_p^2=0.30$, with time needed to interpret the EKG depending on whether history was known—in case of EKGF significantly more time (approximately 24 seconds or 60% more) was needed to read the EKG than HF, $t(50)=7.53$; $P<.001$. In contrast in both groups a

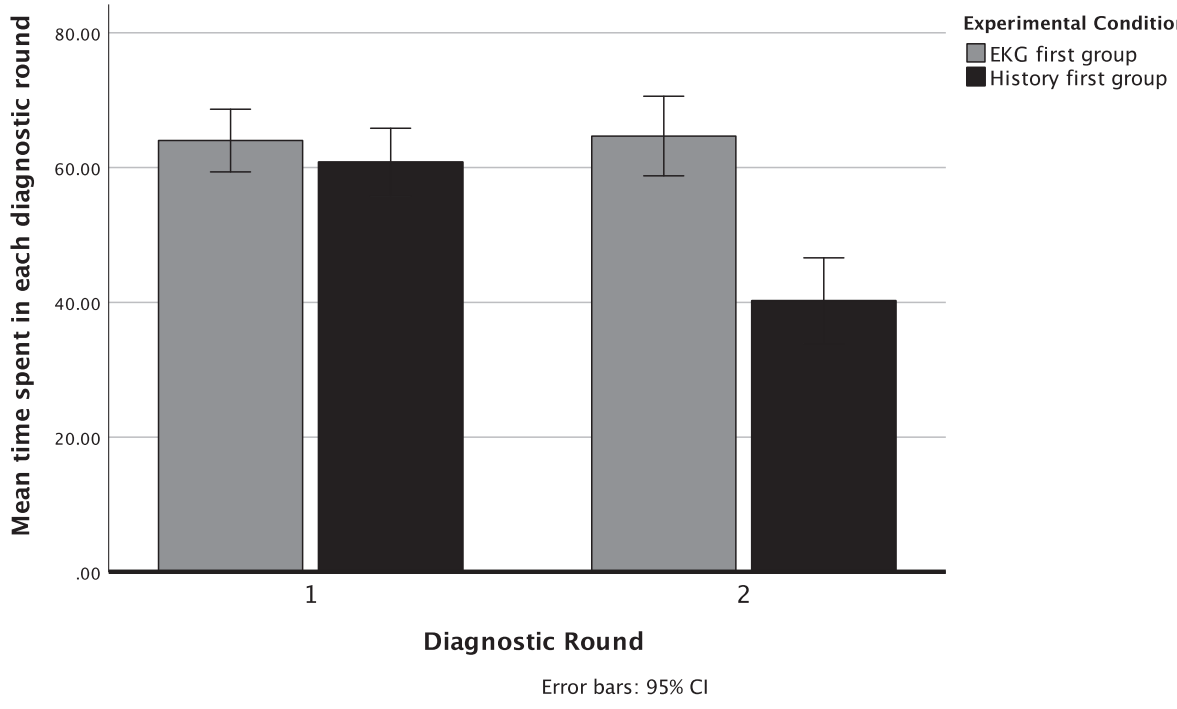

FIGURE 2
Time Spent in Each Component of the Problem as a Function of Experimental Condition

similar amount of time was spend on history, $t(50)=0.88$; $P=.40$. EKGF spent as much time in the EKG as in the history, $t(27)=0.20$; $P=.84$. HF dedicated more time to the history than to the EKG, $t(23)=7.49$; $P<.001$ (TABLE).

Changing the Initial Diagnosis and Time Dedicated to the Second Component of the Problem

Diagnostic accuracy variation was significantly correlated with time invested in the second component of the problem, $r_s=0.60$; $P<.001$; $R^2=0.36$, and changing the initial diagnosis happened more often when participants used more time on processing the subsequent information.

Confidence in the Initial Diagnosis and Changing it After the Second Round

Both groups indicated a similar amount of confidence about their diagnosis after the second diagnostic round, but a significant difference was observed in confidence ratings after the initial diagnosis, with EKGF reporting less confidence (TABLE). Confidence levels in the first diagnostic round were not significantly related to time spend in the second round, $r_s=-0.19$; $P=.17$; $R^2=0.04$. A weak but significant negative correlation was found between confidence and diagnosis change, $r=-0.28$; $P=.04$; $R^2=0.08$.

Discussion

Contrary to our expectations, we found that the order in which the EKG and history were presented did not influence the final diagnostic accuracy. The initial diagnostic accuracy however was higher for HF. Interestingly, HF and EKGF spent equal time on history, but EKGF spent more time on EKG. Consequently, EKGF spent more time in total. As hypothesized, changing the initial diagnosis was strongly associated with the amount of time spent in the second component of the problem. Finally, the level of confidence in the initial diagnosis had a negative relation with changing the diagnosis.

Equal time spent on history with a comparable diagnostic accuracy in both groups suggests that EKGF does not influence the speed of reasoning or lead to missing essential elements in the history. We expected that EKGF would reduce the amount of time dedicated to the history, thereby hindering the verification of initial impressions. However, both groups spent similar time in the history. HF needed less time in total to diagnose the case while reaching equal accuracy, mainly due to less time spent on the EKG. This finding suggests that judging the EKG in HF is easier and quicker without leading to more mistakes. These findings are in line with previous experiments by Hatala et al where EKGs having the right clinical context significantly increased EKG interpreting accuracy.[2] Comparable findings interprets cardiac auscultation with or without prior knowledge of the clinical context.[23] One may argue that the clinical context may have activated few (and more relevant) illness scripts to be confirmed by the EKG, whereas seeing only the EKG causes generation of a large number of hypotheses in a broad differential diagnosis that can only be narrowed down by going through the clinical context.

The improvement in diagnostic accuracy in EKGF after also knowing the history was influenced by the time taken for interpreting the history: the more time taken the higher the improvement in accuracy. This finding seems in line with previous research showing that final diagnostic accuracy benefits from efforts to scrutinize initial diagnostic impressions.[10-13] It is also important that the confidence in the diagnosis increased after the second round. Apparently, being less confident about the diagnosis may have facilitated changing it. These findings are only correlational, and it remains unclear whether awareness of a possible wrong diagnosis and enough time spent on the history may have helped to prevent closing the case with a wrong diagnosis.

It seems reasonable to expect that lower confidence in the diagnosis would tend to increase the willingness to more thorough processing of subsequent information. However, our findings did not show a significant relation between confidence in the initial diagnosis and time spent in the second component of the problem. Nevertheless, we observed a significant though small negative correlation between confidence and diagnostic accuracy variation. Moreover, EKGF reported lower confidence in their (actual lower) initial diagnostic accuracy than HF. This may be due to the scarcity of information available when only the EKG was presented. Previous studies that did not show an alignment between accuracy and confidence presented information in the standard order of first history and then physical examination followed by additional test results.[24]

There are limitations to this study. It is unclear if our findings in ambiguous cases would apply to easier cases or more typical cases; it also is unclear how this experimental study could be extrapolated to other residents and fellows or other specialties, as well as to real-life situations under time pressure. Our findings do have implications for how we teach clinical reasoning regarding patients with chest pain: residents should check their confidence and should take

their time interpreting the history. By doing so diagnostic errors may be reduced.

Future research would be needed to elucidate whether these observations can be extrapolated to real-life situations. In addition, thinking experiments out aloud could be helpful in giving more insight on the thinking process.

Conclusions

Initial diagnostic accuracy was lower in EKGF. Subsequently, the more time spent on history the higher the correction rate as well as diagnostic accuracy. EKGF does not lead to a lower diagnostic accuracy in the end. However, knowledge of the history makes judgment of the EKG quicker and easier.

References

1. Collet JP, Thiele H, Barbato E, et al. 2020 ESC guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. *Eur Heart J*. 2021;42(14):1289-1367. doi:10.1093/eurheartj/ehaa575
2. Hatala RA, Norman GR, Brooks LR. The effect of clinical history on physicians' ECG interpretation skills. *Acad Med*. 1996;71(suppl 10):68-70. doi:10.1097/00001888-199610000-00047
3. Hatala RA, Norman GR, Brooks LR. Impact of a clinical scenario on accuracy of electrocardiogram interpretation. *J Gen Intern Med*. 1999;14(2):126-129. doi:10.1046/j.1525-1497.1999.00298.x
4. Kostopoulou O, Russo JE, Keenan G, Delaney BC, Douiri A. Information distortion in physicians' diagnostic judgments. *Med Decis Making*. 2012;32(6):831-839. doi:10.1177/0272989X12447241
5. Leblanc VR, Brooks LR, Norman GR. Believing is seeing: the influence of a diagnostic hypothesis on the interpretation of clinical features. *Acad Med*. 2002;77(suppl 10):67-69. doi:10.1097/00001888-200210001-00022
6. Elstein AS, Shulman LS, Sprafka SA. *Medical Problem Solving: An Analysis of Clinical Reasoning*. Cambridge, MA: Harvard University Press; 1978.
7. Norman G. Research in clinical history: past history and current trend. *Med Educ*. 2005;39(4):418-427. doi:10.1111/j.1365-2929.2005.02127.x
8. Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implication. *Acad Med*. 1990;65(10):611-621. doi:10.1097/00001888-199010000-00001
9. Schmidt HG, Rikers RM. How expertise develops in medicine: knowledge encapsulation and illness script formation. *Med Educ*. 2007;41(12):1133-1139. doi:10.1111/j.1365-2923.2007.02915.x
10. Sibbald M, de Bruin AB, van Merrienboer JJ. Checklists improve experts' diagnostic decisions. *Med Educ*. 2013;47(3):301-308. doi:10.1111/medu.12080
11. Sibbald M, De Bruin AB, van Merrienboer JJ. Finding and fixing mistakes: do checklists work for clinicians with different levels of experience? *Adv Health Sci Educ Theory Pract*. 2014;19(1):43-51. doi:10.1007/s10459-013-9459-3
12. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on accuracy of medical diagnoses. *Med Educ*. 2008:42(5):468-475. doi:10.1111/j.1365-2923.2008.03030.x
13. 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(11):1198-1203. doi:10.1001/jama.2010.1276
14. Schmidt HG, Mamede S, Van den Berg 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(2):285-291. doi:10.1097/ACM.0000000000000107
15. Mamede S, Hautz WE, Berendonk C, et al. Think twice: effects on diagnostic accuracy of returning to the case to reflect upon the initial diagnosis. *Acad Med*. 2020;95(8):1223-1229. doi:10.1097/ACM.0000000000003153
16. Monteiro SD, Sherbino J, Patel A, Mazzetti I, Norman GR, Howey E. Reflecting on diagnostic errors: taking a second look is not enough. *J Gen Intern Med*. 2015;30(9):1270-1274. doi:10.1007/s11606-015-3369-4
17. Sherbino J, Dore KL, Wood TJ, et al. The relationship between response time and diagnostic accuracy. *Acad Med*. 2012;87(6):785-791. doi:10.1097/ACM.0b013e318253acbd
18. 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(2):277-284. doi:10.1097/ACM.0000000000000105
19. AlQahtani DA, Rotgans JI, Mamede S, et al. Does time pressure have a negative effect on diagnostic accuracy? *Acad Med*. 2016;91(5):710-716. doi:10.1097/ACM.0000000000001098
20. AlQahtani DA, Rotgans JI, Mamede S, Mahzari MM, Al-Ghamdi GA, Schmidt HG. Factors underlying suboptimal diagnostic performance in physicians under time pressure. *Med Educ*. 2018;52(12):1288-1298. doi:10.1111/medu.13686
21. Diamond GA, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary-artery

disease. *N Engl J Med*. 1979;300(24):1350-1358. doi:10.1056/NEJM197906143002402

22. Sibbald M, Panisko D, Cavalcanti RB. Role of clinical context in residents' physical examination diagnostic accuracy. *Med Educ*. 2011;45(4):415-421. doi:10.1111/j.1365-2923.2010.03896.x
23. Meyer AND, Payne VL, Meeks DW, Rao R, Singh H. Physicians' diagnostic accuracy, confidence, and resource requests: a vignette study. *JAMA Intern Med*. 2013;173(21):1952-1959. doi:10.1001/jamainternmed.2013.10081
24. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med*. 2005;165(13):1493-1499. doi:10.1001/archinte.165.13.1493

René A. Tio, MD, PhD, is a Cardiologist, Department of Cardiology, Catharina Hospital Eindhoven, Netherlands, and Medical Education Researcher, Department of Educational Development and Research, Faculty of Health, Medicine, and Life Sciences, University of Maastricht, Netherlands; **Marco A. Carvalho Filho, MD, PhD,** is a Medical Education Researcher, Department of Clinical Sciences, Faculty of Veterinary Medicine, Utrecht University, Utrecht, Netherlands, and Lifelong Learning, Education & Assessment Research Network (LEARN), University Medical Center Groningen, Groningen, Netherlands; **Marcos F. de Menezes Mota, MS,** is a Computer Scientist and PhD Candidate, Institute of Computing, University of Campinas, São Paulo, Brazil; **André Santanchè, PhD,** is a Computer Scientist, Institute of Computing, University of Campinas, São Paulo, Brazil; and **Sílvia Mamede, MD, PhD,** is a Medical Education Researcher, Institute of Medical Education Research Rotterdam, Erasmus Medical Center, Rotterdam, Netherlands, and Associate Professor, Department of Psychology, Education, and Child Studies, Erasmus University, Rotterdam, Netherlands.

Funding: This study was funded in part by CNPq grant number 428459/2018-8.

Conflict of interest: The authors declare they have no competing interests.

This work was previously presented at AMEE 2020: The Virtual Conference, September 7-9, 2020.

Corresponding author: René A. Tio, MD, PhD, Catharina Hospital Eindhoven, The Netherlands, rene.tio@catharinaziekenhuis.nl

Received October 28, 2021; revisions received February 18, 2022, and April 20, 2022; accepted May 4, 2022.