OPEN ACCESS

‘Immunising’ physicians against availability bias in diagnostic reasoning: a randomised controlled experiment

Sílvia Mamede ,[1,2] Marco Antonio de Carvalho-Filho,[3,4] Rosa Malena Delbone de Faria,[5,6] Daniel Franci,[3] Maria do Patrocinio Tenorio Nunes,[7] Ligia Maria Cayres Ribeiro,[8] Julia Biegelmeyer,[7] Laura Zwaan,[1] Henk G Schmidt[1,2]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/bmjqs-2019-010079).

For numbered affiliations see end of article.

Correspondence to
Dr Sílvia Mamede, Institute of Medical Education Research Rotterdam, Erasmus MC, Rotterdam 3015CN, The Netherlands; s.mamede@erasmusmc.nl

Received 15 July 2019
Revised 24 December 2019
Accepted 16 January 2020
Published Online First 27 January 2020

Linked

► http://dx.doi.org/10.1136/bmjqs-2020-010831

Check for updates

To cite: Mamede S, de Carvalho-Filho MA, de Faria RMD, *et al*. *BMJ Qual Saf* 2020;**29**:550–559.

ABSTRACT

Background Diagnostic errors have often been attributed to biases in physicians’ reasoning. Interventions to ‘immunise’ physicians against bias have focused on improving reasoning processes and have largely failed.

Objective To investigate the effect of increasing physicians’ relevant knowledge on their susceptibility to availability bias.

Design, settings and participants Three-phase multicentre randomised experiment with second-year internal medicine residents from eight teaching hospitals in Brazil.

Interventions Immunisation: Physicians diagnosed one of two sets of vignettes (either diseases associated with chronic diarrhoea or with jaundice) and compared/contrasted alternative diagnoses with feedback. Biasing phase (1 week later): Physicians were biased towards either inflammatory bowel disease or viral hepatitis. Diagnostic performance test: All physicians diagnosed three vignettes resembling inflammatory bowel disease, three resembling hepatitis (however, all with different diagnoses). Physicians who increased their knowledge of either chronic diarrhoea or jaundice 1 week earlier were expected to resist the bias attempt.

Main outcome measurements Diagnostic accuracy, measured by test score (range 0–1), computed for subjected-to-bias and not-subjected-to-bias vignettes diagnosed by immunised and not-immunised physicians.

Results Ninety-one residents participated in the experiment. Diagnostic accuracy differed on subjected-to-bias vignettes, with immunised physicians performing better than non-immunised physicians (0.40 vs 0.24; difference in accuracy 0.16 (95% CI 0.05 to 0.27); p=0.004), but not on not-subjected-to-bias vignettes (0.36 vs 0.41; difference −0.05 (95% CI −0.17 to 0.08); p=0.45). Bias only hampered non-immunised physicians, who performed worse on subjected-to-bias than not-subjected-to-bias vignettes (difference −0.17 (95% CI −0.28 to −0.05); p=0.005); immunised physicians’ accuracy did not differ (p=0.56).

Conclusions An intervention directed at increasing knowledge of clinical findings that discriminate between similar-looking diseases decreased physicians’ susceptibility to availability bias, reducing diagnostic errors, in a simulated setting. Future research needs to examine the degree to which the intervention benefits other disease clusters and performance in clinical practice.

Trial registration number 68745917.1.1001.0068.

BACKGROUND

Diagnostic errors pose an important threat to patient safety.[1] The diagnosis is estimated to be wrong 10%–15% of the time.[2] While many errors have minor consequences, harm inflicted to patients is often serious,[3] and diagnostic error remains the most common and most costly reason for malpractice claims in every large system.[2 4 5] For example, a large study of claims in UK[6] found failure or delay in diagnosis to account for 50% of the cases originated in primary care, with the death of the patient recorded in 21% of the cases.

Diagnostic errors are usually multifactorial, but errors in physicians’ *reasoning* have been detected in around 75% of the mistakes investigated in studies of malpractice claims[5] and patients’ files.[7 8] Such reasoning errors are frequently attributed to the use of heuristics, ‘rules of thumbs’ often employed by physicians, largely unconsciously, to make routine judgements.[9–11] Usually efficient, heuristics may sometimes induce biases. For example, we often decide on the likelihood of an event (for instance a diagnosis) based on how easily examples of it come to mind.[12] This usually helps but may induce *availability bias* when an *inappropriate* diagnosis comes more easily to mind. Availability bias caused errors when recent experiences with a

particular disease[13 14] made physicians confuse a subsequent case that looked like this disease (but had in fact another diagnosis) with the disease seen before. When irrelevant cues bring a wrong diagnosis to mind,[13–16] if findings that are actually relevant remain unnoticed, an error will occur.[17 18]

There have been many interventions to 'immunise' physicians against bias. (We use the word 'immunisation' here as an apt metaphor for the characteristics of the intervention investigated in the study: (1) immunisation efficacy is always partial, which probably also applies to the intervention, and multiple doses are usually required to restore immunity; (2) immunisation is always disease specific, which also happens with a knowledge-based intervention; (3) immunisation increases resistance against a threat faced in future situations, an important point to highlight because our study is not concerned with interventions that support physicians at the moment of problem solving). These interventions have focused on improving the *process* of reasoning by increasing physicians' ability to recognise circumstances that tend to induce bias and apply reasoning strategies to counteract bias. Courses on metacognitive skills, the basics of diagnostic reasoning and its possible cognitive pitfalls exemplify these interventions.[19–21] Although such interventions eventually succeeded in increasing physicians' awareness about biases,[19 20] they have largely failed to change actual performance.[22 23] Rates of diagnostic errors, whenever measured, remained unchanged.[24–27]

The present study deviated from these previous attempts by focusing on the *content* knowledge involved in diagnosis rather than the *process* of reasoning. We designed and tested an intervention directed at refining physicians' knowledge of diseases, particularly knowledge of 'discriminating features'. These features are findings that help distinguish between alternative diagnoses for a particular clinical presentation, because their presence substantially increases the likelihood of one of the diagnoses to be correct. Our assumption is that when this knowledge is robust, these features, when encountered in a case, will not be overlooked.[28] This would tend to counteract the influence of irrelevant, bias-inducing cues. If this assumption is correct, immunising physicians against bias would require increasing the amount and organisation of physicians' knowledge about these discriminating features.

To test this idea, we conducted an experiment in which an immunisation intervention was administered 1 week before a test that required physicians to diagnose clinical vignettes under conditions that were known to induce bias.[13] We hypothesised that physicians who had gone through the immunisation phase would be less vulnerable to bias and demonstrate better diagnostic performance than 'non-immunised' physicians.

METHOD

Study design and setting

A multicentre randomised controlled experiment was conducted in eight teaching hospitals in five cities in Brazil from August 2017 to August 2018. Online supplementary 1 presents the study protocol.

The experiment consisted of three phases: an immunisation intervention, a biasing phase and a diagnostic performance test. In the immunisation phase, physicians diagnosed one of two sets of vignettes (either diseases associated with chronic diarrhoea or with jaundice) and compared and contrasted their diagnoses of these diseases, receiving feedback. The biasing and the test phases replicated a procedure that had been shown to induce availability bias in a previous study.[13] In the biasing phase, physicians were exposed to a vignette of either inflammatory bowel disease (IBD) or acute viral hepatitis. Subsequently, in the test, all physicians diagnosed the same set of vignettes, half of them displaying diarrhoea-related diseases similar to IBD, the other half jaundice-related diseases similar to hepatitis, but all with different diagnoses. In the previous experiment, availability bias caused more mistakes to happen when the vignette was diagnosed after exposure to a similar-looking case in the biasing phase than when it was not (eg, physicians who encountered IBD in the biasing phase misdiagnosed the diarrhoea-related test vignettes as IBD more frequently than physicians who encountered hepatitis).[13] In the present study, it was assumed that the intervention would 'immunise' physicians against bias either on the diarrhoea-related diseases or on the jaundice-related diseases (figure 1).

The study involved therefore *two* different treatments. Each physician diagnosed the same vignettes in the test, but three test vignettes would look like the disease encountered in the biasing phase (hereafter 'subjected-to-bias' vignettes) and three would not (hereafter 'not-subjected-to-bias' vignettes) depending on the disease that the physician encountered in the biasing phase, and the physician would be either immunised or not immunised, depending on the diseases that the physician diagnosed in the immunisation intervention. (Notice that if bias depends on possessing specific knowledge, immunisation would also be specific to sets of related diseases.) For instance, for physicians who encountered hepatitis in the biasing phase, the jaundice test vignettes would be subjected to bias, while the diarrhoea test vignettes not subjected to bias. Among these physicians, those who diagnosed the jaundice vignettes in the immunisation intervention would be immunised against bias for the disease presented in the biasing phase, but not those who diagnosed the diarrhoea vignettes in the immunisation. The reverse would apply for the physicians who encountered IBD in the biasing phase. The combination of the two treatments would lead, therefore, to four 'types' of vignettes—subjected to bias with immunisation; subjected to bias without immunisation; not

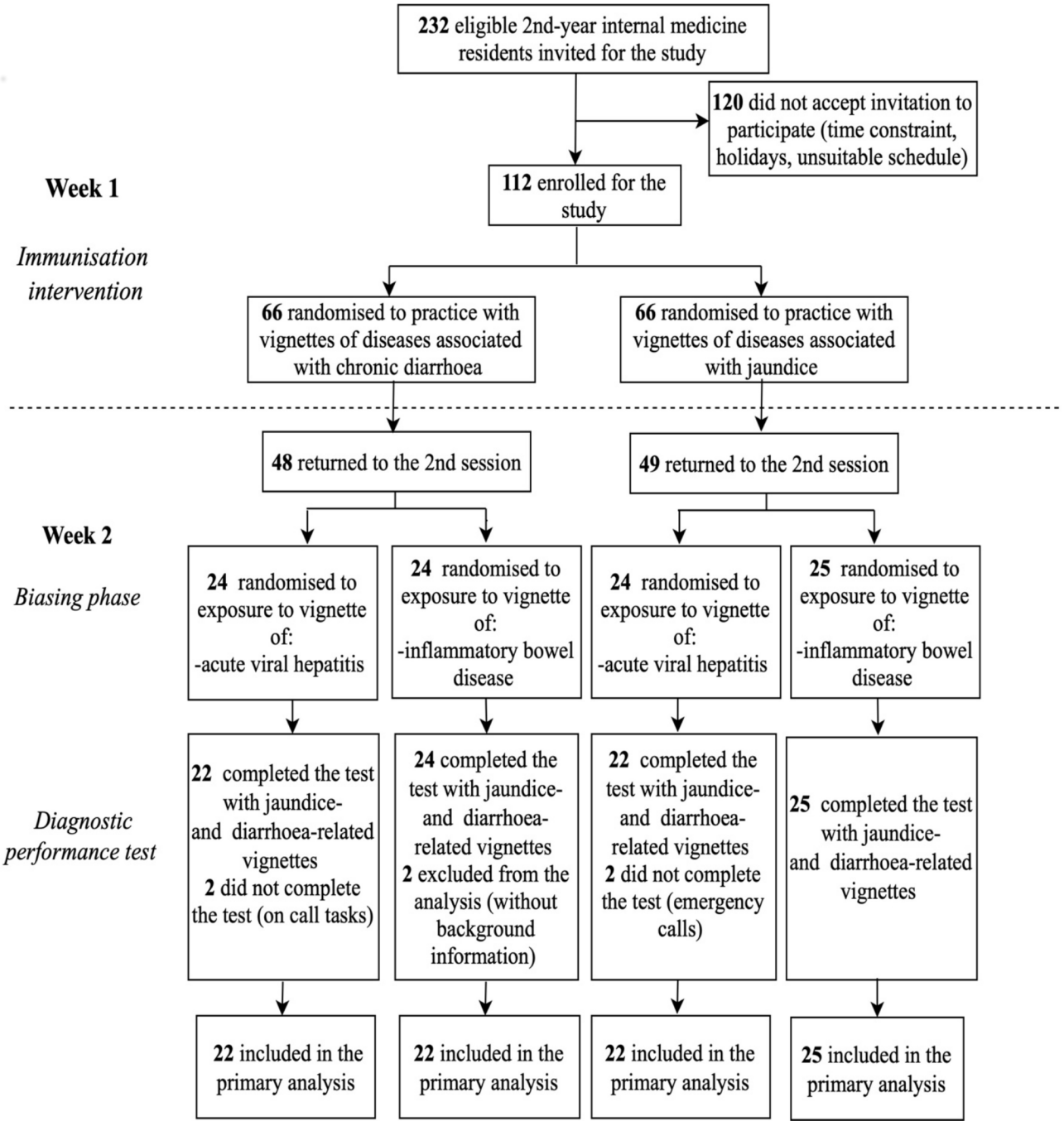


Figure 1 Diagram of the study and flow of participants.

subjected to bias with immunisation; and not subjected to bias without immunisation—with each participant acting as each other's control (figure 2).

Participants

We recruited participants from the pool of internal medicine residents in the teaching hospitals. Residents in Brazil have an MD degree, obtained on completion of 6-year undergraduate education, after which they are allowed to engage in clinical practice. All residents enrolled in the second year of the training programme were considered eligible and invited by the programme director to voluntarily participate in the study (see online supplementary 1 section 2.2 for additional information). Written consent was obtained from participants.

Sample size determination

A priori power analysis using to-be-detected effect of medium size (Cohen's f=0.25) and the standard alpha level of 0.05 indicated that a sample size of 98 participants would be sufficient to achieve a power of 0.80.[29] Enrolment rate was lower than expected, and data analysis was performed after completion of the planned sessions (see online supplementary 1 section 2.2 for additional information).

		Biased towards	
		Hepatitis	**IBD**
Immunised against bias for	**Jaundice diseases**	Jaundice vignettes: • subject-to-bias with immunisation Diarrhoea vignettes: • not-subject-to-bias without immunisation	Jaundice vignettes: • not-subject-to-bias with immunisation Diarrhoea vignettes: • subject-to-bias without immunisation
	Chronic diarrhoea diseases	Jaundice vignettes: • subject-to-bias without immunisation Diarrhoea vignettes: • not-subject-to-bias with immunisation	Jaundice vignettes: • not-subject-to-bias without immunisation Diarrhoea vignettes: • subject-to-bias with immunisation

Figure 2 Types of test vignettes as a function of the diseases that the participant diagnosed in the immunisation intervention and the disease encountered in the biasing phase. IBD, inflammatory bowel disease.

Materials

The study used 25 written clinical vignettes prepared by board-certified internists (MACF, DF, MPTN, JB) based on real patients or by adjusting cases of previous studies.[13 14] We aimed at using difficult cases to leave room for errors to occur. Two internists worked together to prepare each vignette, which was subsequently validated by the other internists. All vignettes contained sufficient information to arrive at the most likely diagnosis. Nine vignettes were 'fillers', used only to disguise the combination of diseases. Sixteen vignettes were relevant and actually considered for the analysis (we refer to the relevant vignettes hereafter). Half of the vignettes displayed diseases associated with jaundice and the other half diseases associated with chronic diarrhoea (online supplementary appendix 1). These diseases were chosen because, besides clinically important, they allowed us to use mostly vignettes validated in previous studies. In all phases, the vignettes were presented in booklets, each one prepared in two versions to counterbalance the presentation sequence.

Intervention

The immunisation intervention consisted of two exercises carried out sequentially, combining deliberate reflection on clinical cases[30] and feedback. Exercise 1 required physicians to diagnose a set of clinical vignettes, one by one, by following a procedure intended to increase knowledge of the clinical features that distinguish between diseases that share a similar clinical presentation. First, physicians read the vignette and gave the most likely diagnosis. Turning the page, they compared/contrasted alternative diagnoses presented in a table. They were requested to (1) list findings that speak in favour of their initial diagnosis, findings that speak against it and findings expected to be present if the initial diagnosis were correct but were absent in the vignette; (2) do the same for each alternative diagnosis; (3) rate the likelihood of each diagnosis under consideration; (4) underline findings shared by more than one diagnosis and circle those associated with only one of the diagnoses; and (5) list 'discriminating features', findings that help decide between the alternative diagnoses, because their presence is strongly associated with only one of them (see online supplementary 2 for an example).

In exercise 2, physicians received the same booklet but with the tables filled in, through a consensus model, by four expert internists (MACF, DF, MPTN, JB). For each vignette, participants compared their responses with the experts' tables, underlying which discriminating features they had overlooked in exercise 1.

Two different sets of vignettes were used in the immunisation phase, one containing diarrhoea-related diseases and the other jaundice-related diseases (figure 1). Participants were randomly allocated to work either with the diarrhoea vignettes or with the jaundice vignettes (see online supplementary 1 for additional information). The intervention lasted 2 hours, with physicians proceeding through it in their own pace.

Biasing phase and diagnostic performance test

The biasing phase and the test were conducted in a single session, purportedly as two independent studies.

In the biasing phase, the physicians received a set of clinical vignettes, each one with a diagnosis, and indicated (in percentage) the likelihood that the diagnosis was correct. Two different sets of vignettes were used, each set containing the same four fillers (intended to hide the purpose of the biasing event) and one bias-inducing vignette, either IBD or acute viral hepatitis. Participants were randomly allocated to receive either one or the other set.

Subsequently, in the test, all participants received the same new set of vignettes. They were requested to read the vignette and write down the most likely diagnosis. Three vignettes displayed diseases that resemble IBD; three others resembled acute viral hepatitis, all with different diagnoses however.

Finally, the physicians provided demographic information and indicated how frequently they saw patients with the diseases included in the study by using a 5-point Likert scale (1=none; 5=very frequently).

Outcomes

The primary outcome was diagnostic accuracy, measured by the score obtained in the test. Using a procedure proved reliable in previous studies,[13 14 31] two board-certified internists (MACF, DF) independently and blindly classified all diagnoses provided for each vignette as correct, partially correct or incorrect (scored, respectively, as 1, 0.5 or 0). The inter-rater agreement was high (ICC=0.98). Discordant classifications were solved by discussion.

Additionally, we measured the occurrence of availability bias by counting the number of times that the disease of the bias-inducing vignette was mentioned as the diagnosis of the similar-looking test vignettes (IBD on the diarrhoea-related vignettes; hepatitis on the jaundice-related vignettes). This measurement was necessary to check if errors were actually caused by availability bias, because even physicians who had not encountered the similar-looking vignette in the biasing phase could incorrectly give its diagnosis to a test vignette that shares similar findings (eg, the test vignette of coeliac disease could be misdiagnosed as IBD even by physicians who saw hepatitis in the biasing phase).

Data analysis

For each participant, we separately summed the diagnostic scores obtained in the test on the three subjected-to-bias vignettes and on the three not-subjected-to-bias vignettes. Mean diagnostic accuracy scores (0–1) were computed for each type of vignette. Similarly, the mean frequency (range 0–3) with which the diagnosis of the bias-inducing vignette was mentioned on the similar-looking test vignettes was computed for subjected-to-bias and not-subjected-to-bias vignettes. A mixed analysis of variance with immunisation against bias for the disease of the biasing phase (immunised vs non-immunised) as between-subjects factor and exposure to bias (subjected to bias and not subjected to bias) as within-subjects factor was performed on the mean diagnostic accuracy scores. This analysis assessed whether diagnostic accuracy decreases as a result of exposure to a similar-looking disease but is counteracted by the immunisation. Post hoc independent t-tests compared diagnostic accuracy of immunised and non-immunised physicians on the two types of vignettes (subjected to bias and not subjected to bias). Paired t-tests compared performance on each type of vignette within the same group of physicians. To verify whether availability bias actually occurred and was counteracted by immunisation, similar analyses were performed on the frequency with which the diagnosis of the bias-inducing vignette was given to the similar-looking test vignettes. Mean ratings of experience (range 0–5) with the diseases of the study were compared by performing independent t-test. All analyses were performed in SPSS V.25. The level of significance was set at two-sided $p<0.05$.

RESULTS

Ninety-one residents participated in the study (online supplementary 3, table 1). They reported moderate clinical experience with the diseases of the study, and there were no significant differences in participants' characteristics at baseline (table 1).

Figure 3 presents the diagnostic accuracy scores obtained on subjected-to-bias and not-subjected-to-bias vignettes by immunised and non-immunised physicians. As expected, overall, diagnostic accuracy did not differ between not-subjected-to-bias and

Table 1 Baseline characteristics of physicians immunised and non-immunised against bias for the disease encountered in the biasing phase

	Immunised	Non-immunised	Overall
Age (years)			
Mean (95% CI)	27.39 (26.52 to 28.26)	27.91 (27.18 to 28.64)	27.67 (27.11 to 28.22)
Sex			
Male	23 (52%)	21 (45%)	44 (48%)
Female	21 (48%)	26 (55%)	47 (52%)
Experience with the diseases of the study (range 0–5)			
Mean (95% CI)	2.77 (2.65 to 2.90)	2.66 (2.48 to 2.85)	2.72 (2.61 to 2.83)

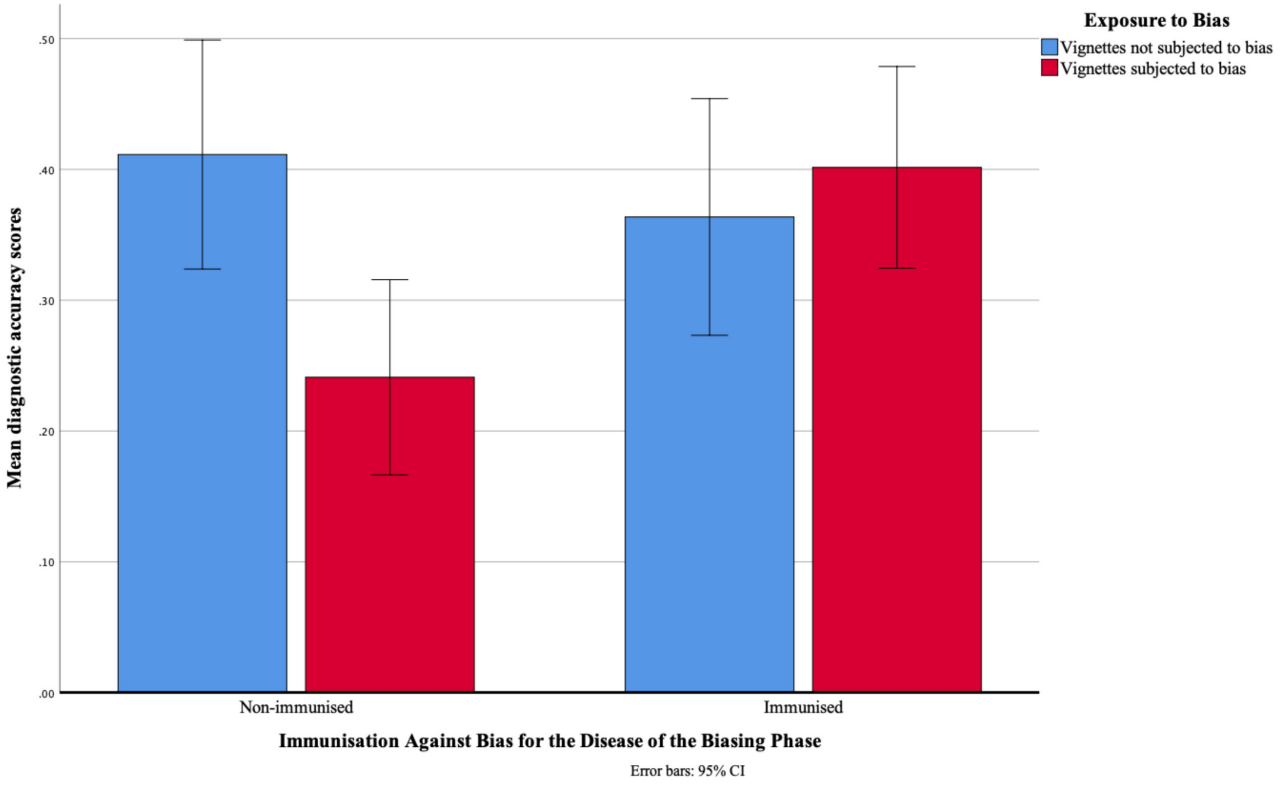


Figure 3 Diagnostic accuracy scores (range 0–1) as a function of previous exposure to a similar-looking disease in the biasing phase and immunisation against bias for the disease of the biasing phase.

subjected-to-bias vignettes (respectively 0.39 vs 0.32; p=0.12; absolute difference in diagnostic accuracy 0.7 (95% CI −0.02 to 0.15)), nor between non-immunised and immunised physicians (0.33 vs 0.38; p=0.17; difference −0.06 (95% CI −0.14 to 0.02)), but there was a significant interaction effect (p=0.02). Post hoc analysis showed that the performance of immunised and non-immunised physicians only differed on subjected-to-bias vignettes. When diagnosis was preceded by exposure to a similar-looking disease (subjected-to-bias vignettes), physicians who had been immunised performed significantly better than those who had not (respectively 0.40 vs 0.24; p=0.004), with an absolute difference in diagnostic accuracy between the two groups of 0.16 (95% CI 0.05 to 0.27). On not-subjected-to-bias vignettes, immunised and non-immunised physicians did not significantly differ in accuracy (0.36 vs 0.41; p=0.45; difference −0.05 (95% CI −0.17 to 0.08)). Bias only hampered non-immunised physicians. They performed worse on subjected to bias than not subjected to bias on vignettes (p=0.005), with a difference in accuracy of −0.17 (95% CI −0.28 to −0.05), whereas the performance of immunised physicians did not significantly differ (p=0.56; difference 0.04 (95% CI −0.09 to 0.17)).

Figure 4 presents the frequency with which the diagnosis of the bias-inducing vignette was given as the diagnosis of similar-looking test vignettes. Overall, the frequency did not differ between subjected-to-bias and not-subjected-to-bias vignettes (respectively 0.45 vs 0.31; p=0.13; difference in frequency 0.14 (95% CI −0.04 to 0.33)), nor between immunised and non-immunised physicians (0.32 vs 0.41; p=0.21; difference 0.10 (95% CI −0.05 to 0.24)). However, the interaction was significant (p=0.02). Test vignettes diagnosed after exposure to a similar-looking disease were more frequently confused with this disease by non-immunised than by immunised physicians (respectively 0.60 vs 0.30; p=0.02), with a difference in frequency of 0.30 (95% CI 0.04 to 0.56). When vignettes were not preceded by a similar-looking disease in the biasing phase (not subjected to bias), non-immunised and immunised physicians did not significantly differ in how frequently they mentioned the related diagnosis (0.25 vs 0.36; p=0.31; difference −0.11 (95% CI −0.32 to 0.40)). Only among the non-immunised physicians the frequency with which the bias-inducing diagnosis was mentioned for similar-looking test vignettes increased on subjected-to-bias relative to not-subjected-to-bias vignettes (p=0.01), with a difference of 0.34 (95% CI 0.08 to 0.60). Among immunised physicians, this frequency did not significantly differ (p=0.61; difference −0.07 (95% CI −0.33 to 0.20)).

DISCUSSION

An immunisation intervention directed at increasing physicians' knowledge of a cluster of related diseases

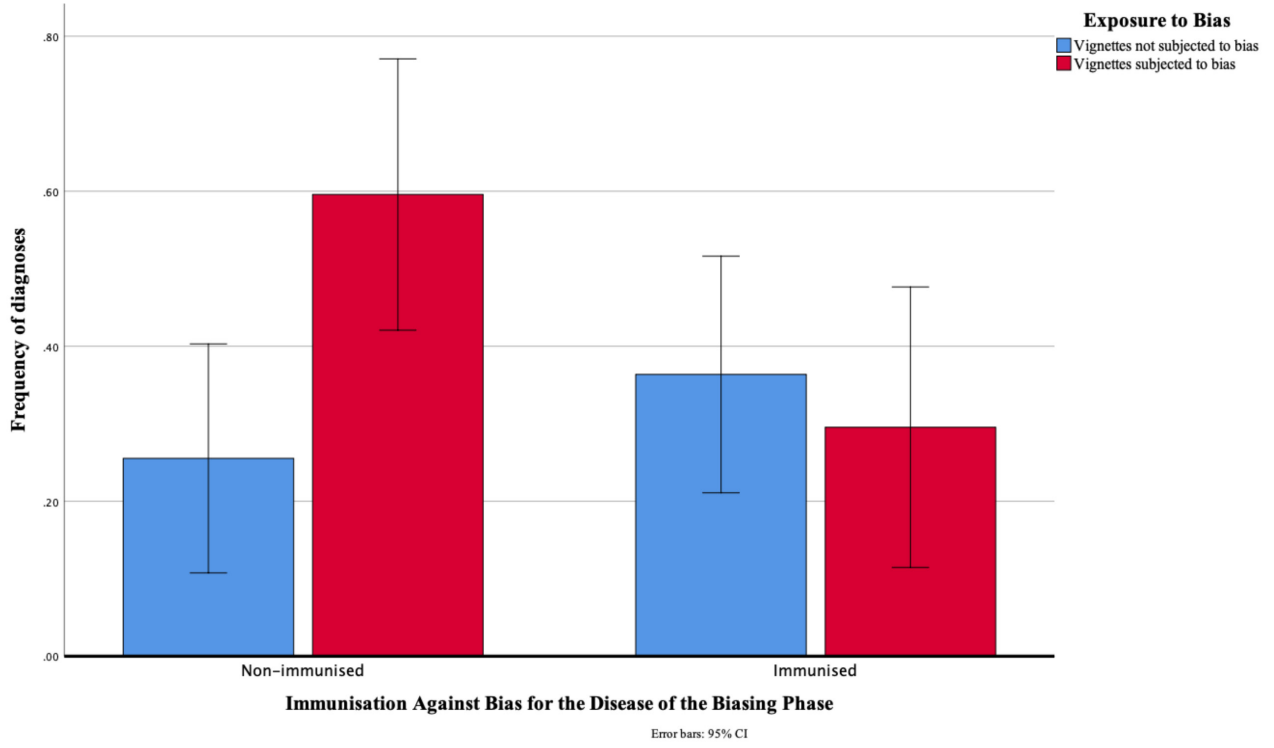


Figure 4 Frequency with which the diagnosis of the vignette of the biasing phase was incorrectly given to similar-looking test vignettes (range 0–3) as a function of exposure to a similar-looking disease in the biasing phase and immunisation against bias for the disease of the biasing phase.

decreased the rates of diagnostic error when physicians diagnosed new vignettes of these diseases 1 week later under circumstances that are known to induce bias.[13 14] After encountering one case of a disease, non-immunised physicians incorrectly gave that diagnosis to vignettes of different (though similar) diseases twice more frequently than immunised physicians. Consequently, diagnostic accuracy decreased 40% between immunised and non-immunised physicians. This difference in diagnostic accuracy was only observed on subjected-to-bias vignettes. Immunised and non-immunised physicians performed similarly on vignettes not preceded by exposure to a look-alike disease.

Taken together, these findings show that availability bias caused a substantial proportion of the diagnostic errors, and that the intervention counteracted the bias. The intervention required comparing and contrasting alternative diagnoses for look-alike diseases, focusing not on typical findings associated with a particular disease but on how that disease differs from other diseases that are frequent alternative explanations for a certain configuration of clinical findings. Psychological research[32] supports the expectation that juxtaposing the alternative diagnoses and drawing attention to discriminating features would strengthen in physicians' memory knowledge of critical features to be retrieved during differential diagnosis of these diseases. Robust knowledge of discriminating features would make a physician less likely to overlook them when irrelevant information, such as recent experiences with a similar-looking disease, brings an inappropriate diagnosis to mind. The findings suggest that this may have actually happened.

Although interventions exist that have been shown to reduce diagnostic errors[30 33 34] or to counteract bias,[13 14] all successful interventions up to now involve instructing physicians *while* they diagnose cases, such as priming them to review their initial diagnosis by engaging in deliberate reflection,[13 14] using checklists[35] or electronic support systems.[36] Whereas empirical evidence exists of the effectiveness of these 'workplace interventions', interventions carried out *prior* to the diagnostic moment with the aim of increasing physicians' resistance to bias in future situations have up to now shown no effect on rates of diagnostic errors.[23 27] In the present study, the intervention made physicians less vulnerable to availability bias when they diagnosed, without receiving any particular instruction, new cases *1 week later*. Contrary to process-oriented 'debiasing' strategies,[21 22] the intervention did not aim at recognition of bias-inducing cues but rather at recognition of critical diagnostic cues. Such intervention is therefore specific to sets of diseases that share a similar clinical presentation, consistently with the assumption that susceptibility to bias results primarily from lack of knowledge rather than from errors in reasoning. Note that the findings do not refute the potential influence of bias on reasoning, but they do show that specific disease knowledge counteracts such influence. Taken together with the hitherto limited effects of educational

interventions aimed at improving reasoning processes on rates of diagnostic errors, our findings call for a new perspective in the search for strategies to increase physicians' resistance to bias which gives attention to more knowledge-oriented interventions.[22]

The findings also reaffirm the potential of availability bias to cause diagnostic error. Exposure to only one case of a disease caused physicians to incorrectly provide this diagnosis to subsequent diseases that, though looking alike, were in fact different. The effect of the bias was not large but may increase when physicians encounter not one but several patients with similar presentations that are caused by different diseases, as it often happens in real settings such as primary care services or emergency rooms. Arguably, a wrong initial diagnosis generated under these circumstances may be repaired subsequently. However, the strongest predictor of final diagnostic accuracy is an accurate initial diagnosis,[37 38] possibly because the initial hypothesis heavily influences subsequent information seeking. Physicians who generate an inaccurate hypothesis are more likely to fail to gather critical diagnostic information or to accurately interpret it, overvaluing neutral information as supporting the hypothesis while ignoring contradictory evidence.[39 40] The studies showing lapses in physicians' reasoning to be implicated in most diagnostic errors indeed suggest that an incorrect initial diagnosis is not easily overthrown.[5 7 41]

The intervention tested in the study has potential for adoption in practice in medical education. Many undergraduate and postgraduate programmes already have regular activities aimed at providing advanced students and residents with opportunity to practise with clinical problems. Exercises such as the intervention could be integrated into these activities, with trainees engaging in comparing and contrasting alternative diagnoses for similar-looking diseases. It would require selection of a set of frequent, relevant complaints and their usual clinical presentation, and organisation of practice around clusters of diseases that are usually alternative diagnoses for them. Organising such practice would require teachers to invest time and effort particularly for the development of appropriate cases, which may involve costs. On the other hand, the exercises themselves can be carried out independently by the trainees, without any particular supervision.

In the present study, one single exercise was enough to counteract the influence of availability bias, but further research is needed to determine the frequency with which trainees need to practise with the same cluster of diseases to ensure that the effect lasts. More research is required also to examine whether other target groups would also benefit from the intervention. Our participants were residents, and though it is likely that the intervention could be useful to advanced undergraduate students, this demands further investigation. Finally, the intervention showed to be effective to counteract availability bias. Other cognitive biases have been described,[10 11] and though it is likely that they could also be counteracted by a knowledge-oriented intervention, this is still to be determined.

The study has limitations. First, the study was conducted in a simulated setting. The use of written vignettes, though shown by experimental research to be a good proxy for real settings performance,[42 43] restricts generalisation of findings to real practice, where other cues would be available for the clinicians. On the other hand, while the vignettes contained all the information required for the diagnosis, in real practice physicians would need to search for the information themselves, and such search tends to be hindered by a wrong initial diagnosis.[39 40] If bias caused error even when all the relevant information is given, the need to gather it would probably increase rather than reduce the damage. Second, our participants were residents with moderate experience with the diseases, and it is unclear if findings apply to experienced physicians. Whether experience per se makes physicians more or less susceptible to bias is unknown, as experienced physicians have more difficulties to revise initial hypotheses in light of disconfirming information,[44 45] and escaping bias may depend not only on experience but also on specific features of disease knowledge. Experienced physicians would probably have more of this knowledge, and the intervention may turn to be less useful to them. Third, we tested the effect of the intervention after 1 week, the effect was considerable in light of what is at stake, but it may not last. Indeed, a single 2-hour exercise would probably not beat the influence of many other experiences that participants will go through in the course of their training. However, the study was a test in a simulated environment of an intervention that in real settings would involve not a single session but a longitudinal programme with regular similar exercises, which would tend to amplify learning. As it is the case for many vaccination schedules, multiple 'doses' of exercises such as the one tested in the study would probably be necessary for resistance to bias to be maintained across time. Finally, we studied availability bias, which was shown to occur and cause diagnostic errors in experiments[13 14 46] and in retrospective reviews of errors,[47–49] and it is unclear whether other cognitive biases could also be counteracted but a similar intervention.

In conclusion, an intervention directed to increase knowledge of clusters of diseases that are usually alternative diagnoses for a particular configuration of clinical findings, especially knowledge of findings that help discriminate between these diagnoses, made physicians less susceptible to availability bias when they diagnosed new cases after 1 week, reducing diagnostic errors. These findings suggest that the search for approaches to increase physicians' resistance to bias, which are critical to minimise the burden of diagnostic error and improve patient safety, should

focus on the development of knowledge-oriented interventions. Future research should investigate the effectiveness of the intervention in counteracting other types of cognitive biases and its value for experienced physicians.

Author affiliations
[1]Institute of Medical Education Research Rotterdam, Erasmus MC, Rotterdam, Zuid-Holland, The Netherlands
[2]Psychology, Education and Child Studies, Erasmus University Rotterdam, Rotterdam, Zuid-Holland, The Netherlands
[3]Internal Medicine, State University of Campinas, Campinas, Brazil
[4]Center for Education Development and Research in the Health Professions, University of Groningen, Groningen, The Netherlands
[5]Propeudeutics, Federal University of Minas Gerais, Belo Horizonte, Brazil
[6]Education and Research Center, Santa Casa BH, Belo Horizonte, Minas Gerais, Brazil
[7]Internal Medicine, Universidade de São Paulo, Sao Paulo, Brazil
[8]Department of Medical Education Development, UNIFENAS Medical School, Belo Horizonte, Brazil

Twitter Laura Zwaan @laurazwaan81

Acknowledgements We are grateful to the residents who allocated time for the study and thank the following colleagues for their collaboration to organise the study sessions in their hospitals: Elaine de Andrade Azevedo, Miguel Angelo Pinto Farage, Tereza Cristina de Abreu Ferrari (Belo Horizonte); Andrea Rosa da Costa, Daniel de Castro, Rainardo Antonio Puster (Fortaleza); Thais Ditolvo da Costa Salinas (Manaus).

Contributors All authors had full access to all the study data and take responsibility for the integrity of the data and the accuracy of the data analysis. Study conception and design: SM, LZ, HGS. Development of study materials: SM, MACF, DF, MPTN, JB. Acquisition of data: SM, MACF, RMDF, DF, MPTN, LMCR, JB. Analysis or interpretation of data: All authors. Drafting of the manuscript: SM, HGS. Critical revision of the manuscript for important intellectual content: All authors. Statistical analysis: SM, HGS. Administrative, technical or material support: SM, MACF, RMDF, DF, MPTN, LMCR, JB. Supervision: SM, HGS.

Funding The universities and teaching hospitals involved in the study provided materials and facilities for the data collection.

Disclaimer The funding institutions had no role in the design and conduct of the study; collection, analysis and interpretation of data; preparation, review or approval of the manuscript; and decision to submit the manuscript for publication.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The study was approved by the University of São Paulo Institutional Review Board and registered on the National Research Ethics platform (#68745917.1.1001.0068) as a multicentre study.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Data availability is subject to the requirements of the existing policies for data management adopted by the institutions involved in the study.

ORCID iD
Sílvia Mamede http://orcid.org/0000-0003-1187-2392

REFERENCES

1 Institute of Medicine. *Improving diagnosis in health care*. Washington, DC: National Academies of Sciences, Engineering, and Medicine, 2015.
2 Graber ML. The incidence of diagnostic error in medicine. *BMJ Qual Saf* 2013;22(Suppl 2):ii21–7.
3 Shojania KG, Burton EC, McDonald KM, *et al*. Changes in rates of autopsy-detected diagnostic errors over time: a systematic review. *JAMA* 2003;289:2849–56.
4 McDonald KM, Matesic B, Contopoulos-Ioannidis DG, *et al*. Patient safety strategies targeted at diagnostic errors: a systematic review. *Ann Intern Med* 2013;158:381–9.
5 Poon EG, Kachalia A, Puopolo AL, *et al*. 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;27:1416–23.
6 Fenn P, Gray A, Rivero-Arias O, *et al*. *The epidemiology of error: an analysis of databases of clinical negligence litigation*. Manchester, U.K: Manchester Centre for Healthcare Management, University of Manchester, 2004.
7 Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* 2005;165:1493–9.
8 Singh H, Giardina TD, Meyer AND, *et al*. Types and origins of diagnostic errors in primary care settings. *JAMA Intern Med* 2013;173:418–25.
9 Croskerry P. From mindless to mindful practice--cognitive bias and clinical decision making. *N Engl J Med* 2013;368:2445–8.
10 Redelmeier DA. Improving patient care. The cognitive psychology of missed diagnoses. *Ann Intern Med* 2005;142:115–20.
11 Klein JG. Five pitfalls in decisions about diagnosis and prescribing. *BMJ* 2005;330:781–3.
12 Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. *Science* 1974;185:1124–31.
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;304:1198–203.
14 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* 2014;89:285–91.
15 Mamede S, van Gog T, van den Berge K, *et al*. Why do doctors make mistakes? A study of the role of salient distracting clinical features. *Acad Med* 2014;89:114–20.
16 Hatala R, Norman GR, Brooks LR. Influence of a single example on subsequent electrocardiogram interpretation. *Teach Learn Med* 1999;11:110–7.
17 Evans JSBT. Dual-processing accounts of Reasoning, judgment, and social cognition. *Annu Rev Psychol* 2008;59:255–78.
18 Evans JSBT. The heuristic-analytic theory of Reasoning: extension and evaluation. *Psychon Bull Rev* 2006;13:378–95.
19 Bond WF, Deitrick LM, Arnold DC, *et al*. Using simulation to instruct emergency medicine residents in cognitive forcing strategies. *Acad Med* 2004;79:438–46.
20 Ogdie AR, Reilly JB, Pang WG, *et al*. Seen through their eyes: residents' reflections on the cognitive and contextual components of diagnostic errors in medicine. *Acad Med* 2012;87:1361–7.
21 Reilly JB, Ogdie AR, Von Feldt JM, *et al*. Teaching about how doctors think: a longitudinal curriculum in cognitive bias and diagnostic error for residents. *BMJ Qual Saf* 2013;22:1044–50.

22 Norman GR, Monteiro SD, Sherbino J, *et al*. The causes of errors in clinical Reasoning: cognitive biases, knowledge deficits, and dual process thinking. *Acad Med* 2017;92:23–30.
23 Lambe KA, O'Reilly G, Kelly BD, *et al*. Dual-process cognitive interventions to enhance diagnostic Reasoning: a systematic review. *BMJ Qual Saf* 2016;25:808–20.
24 Sherbino J, Kulasegaram K, Howey E, *et al*. Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic Reasoning: a controlled trial. *CJEM* 2014;16:34–40.
25 Sherbino J, Dore KL, Siu E, *et al*. The effectiveness of cognitive forcing strategies to decrease diagnostic error: an exploratory study. *Teach Learn Med* 2011;23:78–84.
26 Smith BW, Slack MB. The effect of cognitive debiasing training among family medicine residents. *Diagnosis* 2015;2:117–21.
27 Ludolph R, Schulz PJ. Debiasing health-related judgments and decision making: a systematic review. *Med Decis Making* 2018;38:3–13.
28 Sperber D, Cara F, Girotto V. Relevance theory explains the selection task. *Cognition* 1995;57:31–95.
29 Cohen J. A power primer. *Psychol Bull* 1992;112:155–9.
30 Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Med Educ* 2008;42:468–75.
31 Mamede S, Van Gog T, Schuit SCE, *et al*. Why patients' disruptive behaviours impair diagnostic Reasoning: a randomised experiment. *BMJ Qual Saf* 2017;26:13–18.
32 Klayman J, Brown K. Debias the environment instead of the judge: an alternative approach to reducing error in diagnostic (and other) judgment. *Cognition* 1993;49:97–122.
33 Bass A, Geddes C, Wright B, *et al*. Experienced physicians benefit from analyzing initial diagnostic hypotheses. *Can Med Educ J* 2013;4:e7–15.
34 Mamede S, Schmidt HG, Rikers RMJP, *et al*. Conscious thought beats deliberation without attention in diagnostic decision-making: at least when you are an expert. *Psychol Res* 2010;74:586–92.
35 Sibbald M, de Bruin ABH, van Merrienboer JJG. Checklists improve experts' diagnostic decisions. *Med Educ* 2013;47:301–8.
36 Feyzi-Behnagh R, Azevedo R, Legowski E, *et al*. Metacognitive scaffolds improve self-judgments of accuracy in a medical intelligent tutoring system. *Instr Sci* 2014;42:159–81.
37 Barrows HS, Norman GR, Neufeld VR, *et al*. The clinical Reasoning of randomly selected physicians in general medical practice. *Clin Invest Med* 1982;5:49–55.
38 Kostopoulou O, Devereaux-Walsh C, Delaney BC. Missing celiac disease in family medicine: the importance of hypothesis generation. *Med Decis Making* 2009;29:282–90.
39 Kostopoulou O, Russo JE, Keenan G, *et al*. Information distortion in physicians' diagnostic judgments. *Med Decis Making* 2012;32:831–9.
40 Wallsten TS. Physician and medical student bias in evaluating diagnostic information. *Med Decis Making* 1981;1:145–64.
41 Gandhi TK, Kachalia A, Thomas EJ, *et al*. Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims. *Ann Intern Med* 2006;145:488–96.
42 Peabody JW, Luck J, Glassman P, *et al*. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA* 2000;283:1715–22.
43 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;141:771–80.
44 Eva KW. The aging physician: changes in cognitive processing and their impact on medical practice. *Acad Med* 2002;77:S1–6.
45 Eva KW, Cunnington JPW. The difficulty with experience: does practice increase susceptibility to premature closure? *J Contin Educ Health Prof* 2006;26:192–8.
46 Hatala R, Norman GR, Brooks LR. Impact of a clinical scenario on accuracy of electrocardiogram interpretation. *J Gen Intern Med* 1999;14:126–9.
47 Dan O, Hochner-Celnikier D, Solnica A, *et al*. Association of catastrophic neonatal outcomes with increased rate of subsequent cesarean deliveries. *Obstet Gynecol* 2017;129:671–5.
48 Riddell CA, Kaufman JS, Hutcheon JA, *et al*. Effect of uterine rupture on a hospital's future rate of vaginal birth after cesarean delivery. *Obstet Gynecol* 2014;124:1175–81.
49 Saposnik G, Redelmeier D, Ruff CC, *et al*. Cognitive biases associated with medical decisions: a systematic review. *BMC Med Inform Decis Mak* 2016;16:138.

ONLINE ONLY SUPPLEMENTARY FILES

Supplement 1 - Study protocol

Supplement 2 - Additional results

Supplement 3 – Example of a vignette used in the study (inflammatory bowel disease) in the immunisation intervention (Exercise 1)

Mamede S, *et al. BMJ Qual Saf* 2020;0:1–10. doi: 10.1136/bmjqs-2019-010079

Supplement 1

Study protocol

Title: Strategies for the development of physicians' clinical reasoning and reduction of diagnostic error

Authors:

1. Sílvia Mamede, MD, PhD, Institute of Medical Education Research Rotterdam, Erasmus MC, University Medical Centre Rotterdam; Department of Psychology, Erasmus University Rotterdam, The Netherlands
2. Marco Antonio de Carvalho Filho, MD, PhD, Centre for Educational Research and Development in Health Professions, University Medical Centre, Groningen, The Netherlands; and Internal Medicine Department, School of Medical Sciences, State University of Campinas, Brazil
3. Rosa Malena Delbone de Faria, MD, PhD, Department of Propedeutics, Faculty of Medicine, Federal University of Minas Gerais, Belo Horizonte, Brazil
4. Daniel Franci, MD, Internal Medicine Department, School of Medical Sciences, State University of Campinas, Brazil.
5. Maria do Patrocinio Nunes, MD, PhD, Faculty of Medicine, University of São Paulo, Brazil
6. Ligia Maria Cayres Ribeiro, MD, MHPE, Department of Medical Education Development, José do Rosário Vellano University (UNIFENAS) Medical School, Belo Horizonte, Brazil.
7. Julia Biegelmeyer, MD, Faculty of Medicine, University of São Paulo, Brazil
8. Laura Zwaan, PhD, Institute of Medical Education Research Rotterdam, Erasmus MC, University Medical Centre Rotterdam
9. Henk G. Schmidt, PhD, Institute of Medical Education Research Rotterdam, Erasmus MC, University Medical Centre Rotterdam; Department of Psychology, Erasmus University Rotterdam, The Netherlands

1. Background

Diagnostic errors have attracted increasing attention since the 1999 Institute of Medicine report showed that between 44,000 and 98,000 people die annually in the United States due to avoidable medical errors.[1] If one takes the lower figure, deaths due to adverse effects caused by medical errors would supplant deaths due to traffic accidents, breast cancer or AIDS. A large fraction of these errors refers to treatment, but a substantial proportion consists of diagnostic errors, which involve a high cost, are potentially preventable, and have a high impact for both physicians and patients.[2] Diagnostic errors are found in all medical specialties at rates ranging from 5% in specialties of a more perceptive nature (e.g. radiology and pathology) to 15% in specialties such as emergency medicine and internal medicine.[3] Many of these errors can be corrected in time or produce minor adverse effects, but a substantial proportion leads to serious consequences, as autopsy studies have shown.[4] The Institute of Medicine report is usually seen as a milestone in the history of diagnostic error research, but the problem is not restricted to the United States, as subsequent studies in several countries have shown.[5] Although there are no large-scale studies of diagnostic error in Brazil, there is no reason to assume that the problem is less serious.

The literature distinguishes between three types of diagnostic error.[2] The “no-fault error” occurs in situations where the correct diagnosis could hardly be expected, for example, an extremely atypical presentation of a disease. The second type of error, known as a "system-related error", stems from failures in health services that affect physician performance, such as communication flaws. Finally, the "cognitive error" is one that can be attributed directly to the physician, resulting from a lack of appropriate knowledge, inadequate information collection or interpretation, inadequate verification or poor reasoning. Although multiple factors may interact to produce a diagnostic error, it has been repeatedly demonstrated that most of the errors are cognitive in nature. For example, a study of diagnostic errors in internal medicine conducted in US university hospitals showed flaws in the cognitive processes of physicians in 74% of cases.[6] Most of these errors were produced not because of lack of knowledge but because of deviations or flaws in clinical reasoning. Research on causes of diagnostic error in primary care services has reached the same conclusions, attributing the majority of errors to failures in the physician's reasoning.[7]

The reasons that make a physician run into flaws in clinical reasoning even though he would have enough knowledge to solve the problem have been the subject of much speculation. Such failures have often been attributed to cognitive biases associated with which has been named "non-analytical reasoning".[8] As they gain experience, physicians tend to generate diagnostic hypotheses by rapidly recognizing similarities between the case in question and examples of previous patients (or prototypical scripts of diseases it has stored in memory), a process known as "pattern recognition".[9,10] What usually happens is that, in the first moments of a clinical encounter, characteristics of the patient "activate" in the doctor's memory scripts of one (or few) diseases, generating a diagnostic hypothesis. The elements of this illness script guide the physician in the subsequent process of seeking more information to verify whether the patient's findings are in fact compatible with the elements of the script.
This "pattern recognition" process occurs in a largely unconscious way, without involving effort, and is usually efficient. However, it seems to open the door to the occurrence of cognitive biases that can distort reasoning and lead to error.[11]

Many cognitive biases that may affect clinical reasoning have been described, but one of the most prevalent is the *availability bias*, which leads people to evaluate the likelihood of an event by the ease with which examples of this event come to mind.[12] Availability bias may produce diagnostic errors, for example, when exposure to media information or recent clinical experiences with a disease leads clinicians to diagnose similar (but in fact different) cases as the previously seen disease. By seeing, for example while on shift in an emergency room, a series of patients with influenza makes this diagnosis come to mind more easily when the physician encounters a close patient with similar symptoms, which can lead to error when the patient in fact has dengue fever. The literature on diagnostic error has suggested that availability bias is an important cause of cognitive diagnostic errors and at least two studies provide experimental evidence of this fact.[13,14] In one of these studies, internal medicine residents made more diagnostic errors in cases with similar clinical presentation (but different diagnosis) to cases they had encountered in a previous task.[13]

The recognition of the role of cognitive bias, such as the availability bias, in causing diagnostic errors has stimulated the search for interventions that make physicians less susceptible to such reasoning errors. One type of intervention that has been explored is to train physicians (or students) about possible biases, assuming that awareness of bias would reduce the diagnostic error.[15] Several formats of courses on clinical reasoning and cognitive

bias have been tried, for example, with residents of internal medicine or medical emergency, but the results have not been favourable. In the few studies in which diagnostic performance was assessed in subsequent tests, the intervention did not reduce the occurrence of diagnostic errors.[16] It has been questioned whether such intervention make sense, not only because of the results of these studies, which suggest a lack of effectiveness, but due to the very nature of cognitive bias. Because they derive from non-analytic reasoning, which takes place largely automatically, they are not subject to conscious control. Some authors have argued that cognitive biases must necessarily be related to the lack of sufficiently elaborate knowledge about the distinction between clinical presentations that look like but actually constitute different diseases.[16] Attention has therefore, it has been claimed, to be directed to the investigation of interventions focused on knowledge development, in particular the refinement of disease scripts that physicians have stored in memory and which are the basis of the diagnostic process.[15]

Better structured illness scripts, including knowledge of elements that allow differentiation between diagnoses that have similar clinical presentation, would make physicians less susceptible to bias and hence less prone to error. A deliberate reflection procedure on to-be-diagnosed cases, developed by Mamede et al.,[17,18] has been used to promote the refinement of illness scripts in a series of studies with medical students.[19,20] Briefly, the procedure involves comparing & contrasting different alternative diagnoses to the case in question, by means of a structured sequence of steps. In these studies, the students solved, during a learning session, the same set of cases, using the reflection procedure or making a differential diagnosis. The students who reflected on the cases in the learning session made fewer diagnostic errors when they resolved new cases a week later than the students who made the differential diagnosis. These findings suggest that, consistently with research in other fields,[21,22] the strategy of comparing & contrasting different scripts of alternative diagnoses for a to-be-solved problem leads to the refinement of illness scripts, making the clinician better able to distinguish between similar diseases in the future. If this effect also applies to more experienced trainees, such as residents, applying the deliberate reflection procedure during practice with clinical cases could contribute to prevent the occurrence of cognitive bias, such as availability bias, and reduce the occurrence of diagnostic error when physicians solve similar cases in the future.

The present study aims to investigate the effectiveness of an immunization intervention based on deliberate reflection on clinical cases to reduce the negative effect of availability bias during the diagnosis of clinical problems.

Based on the aforementioned studies of availability bias and diagnostic error and on the influence of structured reflection on the learning of clinical diagnosis, it is expected that an intervention based on deliberate reflection acts as an "immunization" against the occurrence of bias, leading the following primary hypotheses:
(1) The prior exposure to cases of a given clinical presentation would induce availability bias during subsequent resolution of cases with similar clinical presentation but different diagnoses, leading to diagnostic errors and, consequently, to a lower diagnostic accuracy when these cases are resolved after exposure to a similar-looking disease ("subject-to-bias cases") than when they are resolved without prior exposure ("not-subject-to-bias cases").
(2) Previous practice with deliberate reflection on cases that share similar clinical presentation during an “immunization” intervention will reduce the deleterious effect of availability bias during subsequent resolution of similar-looking cases, leading to higher diagnostic accuracy in subject-to-bias cases that were seen during the immunization intervention than in subject-to-bias cases that were not seen during the intervention (i.e., “immunized physicians” would be less susceptible to availability bias and made fewer mistakes than “non-immunized physicians” when solving subject-to-bias cases, but, consistently with (1), no difference between immunized and non-immunized physicians would be observed in not-subject-to-bias cases).

2. Methods

2.1. Design

The present study is an experiment with two phases: an immunization intervention phase (session 1) and a test phase (session 2), one week after the first phase (see diagram of the study design in Appendix 1). The test phase consists of two tasks, a biasing phase and a diagnostic performance test. In the biasing phase, the physicians will first perform a "confirmation task," which requires evaluating the accuracy of the diagnosis given for a clinical case, having as chief complaint either chronic diarrhoea or jaundice. These cases are presented mixed with cases of non-relevant diseases, for which the same task is performed. All residents will subsequently diagnose eight new cases, four of which are similar to the

cases of the case of one of the syndromes seen in the biasing task and four similar to the cases of the second syndrome seen in the biasing task, but all with different diagnoses. This confirmation task (biasing task) has been shown in a previous study to induce availability bias and, consequently, diagnostic errors [13] However, in the present study, the residents will participate, one week before the experiment, in an immunization intervention consisting of practice, based on deliberate reflection, with clinical cases of one of the two clinical syndromes (either chronic diarrhoea or jaundice).

2.2. Participants

Participants in the study will be 98 second-year internal medicine residents from teaching hospitals in São Paulo and other cities (Appendix 2). Second-year residents are considered eligible for the present study because, as previous studies suggest, these professionals have, at this stage of their training, a similar expertise in a medical specialty that deals with a broad spectrum of problems and sufficient clinical experience to have developed pattern-recognition based on pattern recognition. The sample size was determined estimating a dropout rate of 20% between the two sessions. A prior power analysis, using to-be-detected effect of medium size Cohen's $f = 0.25$ (previous studies with similar interventions are not available), $\alpha = 0.05$, $\beta = 0.80$, for a mixed ANOVA with immunization as between-subjects factor (immunized or non-immunized) and biasing condition (subject-to-bias and not-subject-to-bias) as within-subjects factor.

All residents attending the second year of the internal medicine residency in each hospital will be invited by the program director to voluntarily participate in a study on interventions for improvement of diagnostic reasoning (see Appendix 2). Those who accept the invitation will be registered as participants. A code system based on self-determined codes will be used to ensure that responses are treated anonymously but allowing the connection of each participant's data in the two phases.

Potential adverse consequences to participants. Although there is no risk of participating in the study, the two phases will be carried out outside regular working hours, and it is possible that some of the participants feel fatigued by the additional work, although the activities have short duration. To avoid this problem, the activities will be carried out at the end of the week and the possibility to leave the activity at any moment will be assured to all the participants.

Benefits. Practice with a diversity of clinical cases is recognized as the primary mechanism in the development of clinical reasoning,[10] and participation in the study is expected to

contribute to developing the diagnostic performance of residents. In addition, at the end of the first and second phase, participants will receive feedback that will demonstrate the rationale of an experienced internist to resolve each case. Each participant will have the opportunity to compare their own solution of the case with that of the expert, which should generate additional learning. Participants will also be informed about the theoretical basis of the study, during a lecture on the basics of clinical reasoning, cognitive bias and diagnostic error, performed immediately after the test phase. They will therefore have opportunity to gain additional knowledge about research findings in the area of medical expertise and clinical reasoning.

2.3. Materials and procedures

In total, 25 clinical cases will be used, 11 in the immunization phase and 14 in the test phase (5 for the confirmation task and 9 for the diagnostic task). Appendix 3 presents a breakdown of clinical cases according to each major complaint. Each case consists of a description (about 400 words) of a patient's medical history, history of the current problem, symptoms and findings of the physical examination, and diagnostic tests. The cases will be adapted from difficult cases used in previous studies,[13,17,18,23] all of them prepared by internists based on actual patients and with a confirmed diagnosis. In both phases of the study, the cases will be presented to participants in booklets, one by one.

In the immunization phase (session 1), the two versions of the booklets (either with diseases associated with chronic diarrhoea or with diseases associated with jaundice) will be randomly distributed to the registered participants (see Appendix 2); each participant will therefore practice with one of the two syndromes. The session is expected to last around 2 hours, and consists of two exercises. In the first exercise, for each case, participants are firstly asked to write the most likely diagnostic hypothesis for the case. The case is then re-presented on the subsequent page, and the participant is asked to reflect on the case by following a procedure that has been employed in previous studies, using a table provided to help register the results of reflection.[13,17] The participant is asked to 1) write the diagnosis previously given for the case; (2) read the case again and list the findings in the case description supporting this diagnosis, the findings that speak against it, and the findings that would have been expected to be present if that diagnosis were true but are not described in the case; (3) list alternative diagnoses that he/she would consider if the initial diagnosis generated for the case were found to be incorrect; (4) perform the same analysis (step 2) for each alternative diagnosis; (5) indicate his/her final decision about the most likely diagnosis; (6) and, finally, underline

in the table the findings that are shared by more than one of the diagnoses considered and circulate the findings that discriminate between these diagnoses. After completing the table, the participant is asked to list the findings that help discriminate between the alternative diagnoses because their presence (or absence) is strongly associated with only one of the diagnoses. After having solved all cases and completed the first exercise, participants move to the second exercise, which consists of comparing their reasoning on the case with that of an experienced internist. A new booklet presents each case again, one by one, together with the reflection table filled out by the internist, and the list of findings that discriminate between alternative diagnoses (tables will be prepared by the internists co-researchers, through a consensus model). The participant is asked to compare his / her diagnostic reasoning with the analysis of the case made by the internist.

In the test phase (session 2), lasting approximately 60 minutes, the participants will be randomly assigned to the biasing task either with a case of syndrome 1 or syndrome 2. Subsequently, all participants will diagnose the same 9 cases. The two tasks will be presented as two independent studies to minimize the chance that the confirmation task reveals the possibility of availability bias, consequently changing how residents would approach subsequent cases by inducing a more careful approach that does not reflect reasoning in practice situations. In the biasing task, 5 cases (1 bias-inducing case and 4 fillers) are presented one by one, with a diagnosis, and the participant is asked to indicate (in percentage) the probability that that diagnosis is correct. Subsequently, in what is presented to them as an independent study, all participants are asked to diagnose 9 cases, presented one by one. The participant should read the case and write the most likely diagnosis. Finally, participants are asked to provide background information (age, gender) and to indicate their clinical experience with the diseases used in the study by using a 5-point Likert-scale. Upon completion of the study, the participants receive feedback on the correct diagnosis of the cases.

2.4. Data analysis

The primary outcome of the study is the mean score of diagnostic accuracy in cases diagnosed under the following conditions: (1) "cases subject to bias and without previous immunization against bias for the disease of the biasing phase"; (2) "cases subject to bias with previous immunization against bias for the disease of the biasing phase " (3) "cases not subject to bias and with previous immunization against bias for the disease of the biasing phase "; (4) "cases not subject to bias and without previous immunization against bias for the

disease of the biasing phase". Besides diagnostic accuracy, for these types of cases, the frequency with which the diagnosis of the bias-inducing case in the biasing phase (i.e. the confirmation task) is mentioned on the similar-looking cases of the diagnostic task will be computed to evaluate the actual occurrence of availability bias. The comparison of the mean scores in each of the case types and the frequency of the diagnosis of the biasing task will allow us to evaluate, respectively, whether errors increase as a consequence of the bias-inducing task and are counteracted by the immunization, and whether these errors actually increased because availability bias occurred (and as counteracted by the immunization). This will examine whether the results observed in previous studies[13,14] are replicated and, in particular, if the immunization intervention was effective to prevent the occurrence of the bias and the resulting diagnostic errors.

For the computation of the diagnostic accuracy score, the accuracy of the diagnoses given by the participants will be evaluated considering the confirmed diagnosis of each case as a standard. All responses given by the participants to each case will be entered by a research assistant in a word file, without identification of the condition under which the response was provided thereby allowing for blind scoring. Two specialists in internal medicine will independently evaluate each diagnosis given by the participants, without knowing the condition under which they were given, as correct, partially correct or incorrect (assigning a score of 1, 0.5 or 0, respectively). A response will be considered correct whenever it mentions the core diagnosis of the case, and partially correct when the core diagnosis was not quoted, but a constituent element of the diagnosis was mentioned. This procedure has shown high levels of reliability in previous studies.[13,14,17] The interrater agreement will be assessed using a two-way mixed, absolute agreement, average-measures ICC. Differences in scores will be discussed by the two raters to reach a final score.

For each participant, mean diagnostic accuracy scores obtained on cases subject to bias and on cases not subject to bias will be computed. Similarly, the frequency with which the diagnosis of the bias-inducing case was given to similar-looking test cases will be computed on each type of case. Descriptive statistics will be computed for these two measures on subject-to-bias cases and not-subject-to-bias-cases. Two separate mixed ANOVAs with immunization against bias for the disease of the biasing phase (immunized or non-immunized) as between-subjects factor and exposure to bias diagnosing condition (subject to bias and not subject to bias) will be performed on these two outcome measures (i.e. diagnostic accuracy scores and frequency of the bias-inducing diagnosis). Significant effects

will be further explored by performing independent and paired t-tests. Finally, descriptive statistics will be computed for participants' background characteristics and experience with the diseases used in the study and compared by performing Chi-square (for gender) and t-tests for age and experience.

Ethical approval

This study protocol will be submitted to the Research Ethics Committee of the University of São Paulo (CAPPESQ) and subsequently to "*Plataforma Brasil",* where the study is to be registered as a multicentred study.

Mamede S, *et al. BMJ Qual Saf* 2020;0:1–10. doi: 10.1136/bmjqs-2019-010079

References

1. Kohn LT, Corrigan J, Donaldson MS. To err is human : building a safer health system. Washington, D.C.: National Academy Press; 2000. xxi, 287 p. p.

2. Graber M, Gordon R, Franklin N. Reducing diagnostic errors in medicine: what's the goal? *Acad Med* 2002;77(10):981-92.

3. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *The Am J M* 200;121(5 Suppl):S2-23.

4. Graber ML. The incidence of diagnostic error in medicine. *BMJ Qual Saf* 2013 Oct;22 Suppl 2:ii21-ii7.

5. Weingart SN, Wilson RM, Gibberd RW, Harrison B. Epidemiology of medical error. *BMJ* 2000;320(7237):774-7.

6. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* 2005;165(13):1493-9.

7. Singh H, Thomas EJ, Khan MM, Petersen LA. Identifying diagnostic errors in primary care using an electronic screening algorithm. *Arch Intern Med* 2007;167(3):302-8.

8. Croskerry P. From mindless to mindful practice--cognitive bias and clinical decision making. *N Engl J Med* 2013;368(26):2445-8.

9. Schmidt HG, Norman GR, Boshuizen HPA. A Cognitive Perspective on Medical Expertise - Theory and Implications. *Acad Med* 1990;65(10):611-21.

10. Schmidt HG, Rikers RM. How expertise develops in medicine: knowledge encapsulation and illness script formation. *Med Educ* 2007;41(12):1133-9.

11. Redelmeier DA. Improving patient care. The cognitive psychology of missed diagnoses. *Ann Intern Med* 2005;142(2):115-20.

12. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science* 1974;185(4157):1124-31.

13. Mamede S, van Gog T, van den Berge K, Rikers RM, van Saase JL, van Guldener C, et al. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA* 2010;304(11):1198-203.

14. 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(2):285-91.

15. Schmidt HG, Mamede S. How to improve the teaching of clinical reasoning: a narrative review and a proposal. *Med Educ* 2015;49(10):961-73.

16. 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 Jan;92(1):23-30.

17. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Med Educ* 2008;42(5):468-75.

18. 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. *Psychol Res* 2010;74(6):586-92.

19. Mamede S, van Gog T, Moura AS, de Faria RM, Peixoto JM, Rikers RM, et al. Reflection as a strategy to foster medical students' acquisition of diagnostic competence. *Med Educ* 2012;46(5):464-72.

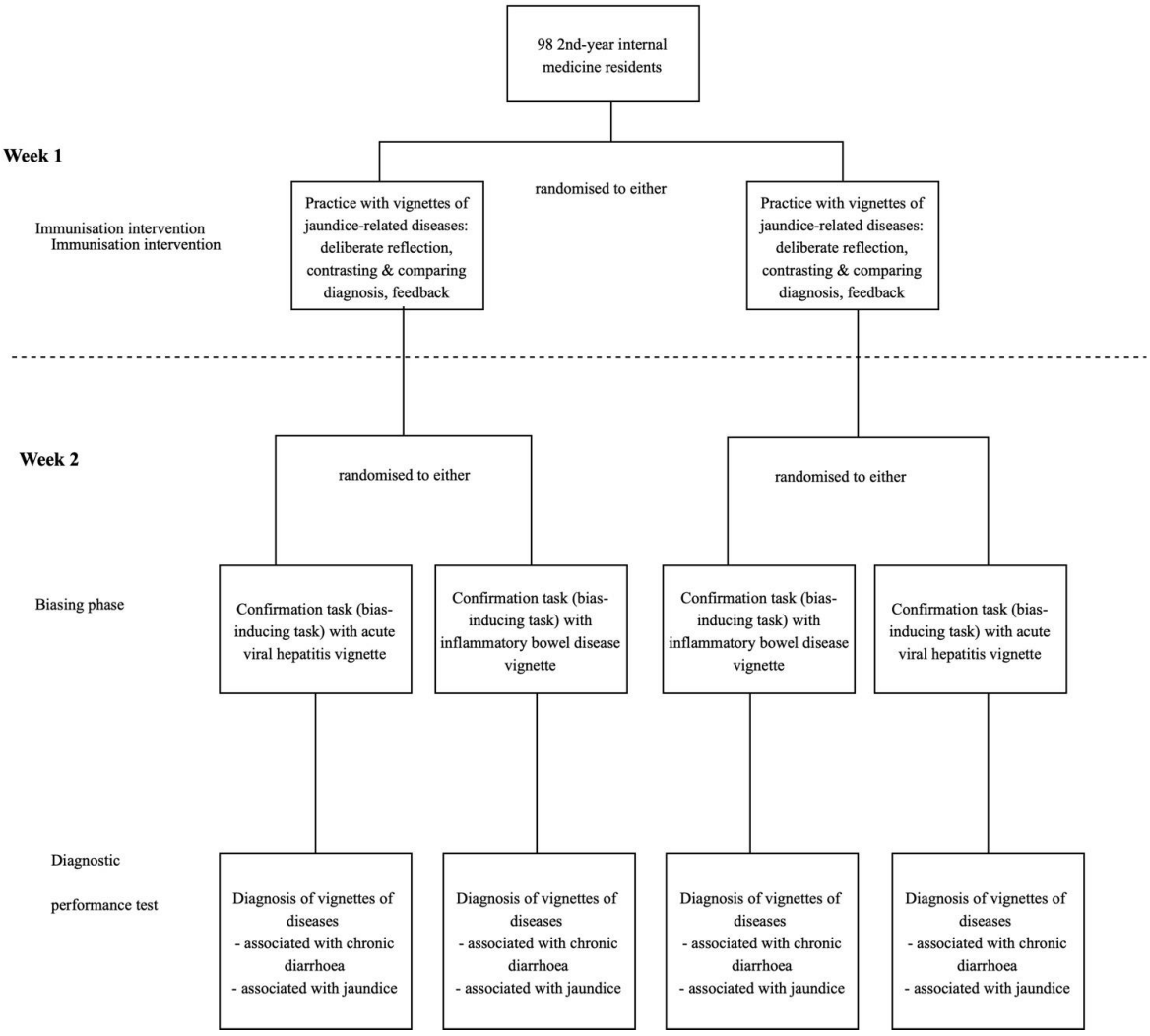
20. Mamede S, van Gog T, Sampaio AM, de Faria RM, Maria JP, Schmidt HG. How can students' diagnostic competence benefit most from practice with clinical cases? The effects of structured reflection on future diagnosis of the same and novel diseases. *Acad Med* 2014;89(1):121-7.

21. Gentner D, Loewenstein J, Thompson L. Learning and transfer: A general role for analogical encoding. *J Educ Psychol* 2003;95(2):393-408.

22. Catrambone R, Holyoak KJ. Overcoming Contextual Limitations on Problem-Solving Transfer. *J Exp Psychol Learn* 1989;15(6):1147-56.

23. Mamede S, Schmidt HG, Rikers RMJP, Penaforte JC, Coelho-Filho JM. Breaking down automaticity: case ambiguity and the shift to reflective approaches in clinical reasoning. *Med Educ* 2007;41(12):1185-92.

Appendix 1 - Diagram of the study design



Appendix 2 – Logistics and operational aspects

Cities and teaching hospitals involved in the study

City	Hospital	Eligible residents
Belo Horizonte	Santa Casa (41), Federal University of Minas Gerais (12), FHEMIG (40)	93
Campinas	UNICAMP	38
Fortaleza	HGF (13), Federal University of Ceará (16)	28
Manaus	Federal University of Amazonas	13
São Paulo	University of São Paulo	60
Total		232

An enrolment rate of 60 % is estimated, with variation expected due to local circumstances. A dropout rate of 20% between the first and the second session is expected.

Logistics

The study is estimated to be carried out in the course of one year, starting in the summer of 2017. Two sessions will be held in each hospital, by using local facilities regularly used for the training activities. The sessions will be booked by the director of the residency training considering the program schedule. The printing of booklets with be under the responsibility of the director of the residency training and will be carried out by using local regular printing schemas.

Procedure for randomization

Due to the difficulty to ensure that participants who accept to participate in the study actually attend the sessions and the need to maintain balance across the study, randomization will be ensure by having the four versions of the booklets prepared for each session randomly distributed to the attending residents. Prior to the session, the booklets to be used in the session will be put in individual envelopes and piled on blocks alternating the four versions of the booklets. After the residents are seated in the auditoriums, the envelopes will be handled to the attendees in the pre-arranged sequence, thereby ensuring that the distribution of participants within each condition remains balanced. The booklets will not contain any information that would identify the booklet as lined to a specific experimental condition.

Appendix 3 – Diagnoses of the vignettes to be used in the three phases of the study

Immunisation intervention	**Biasing phase**	**Test phase**
Jaundice-related set		
Acute viral hepatitis	Acute viral hepatitis	Alcoholic cirrhosis
Alcoholic cirrhosis		Primary sclerosis cholangitis
Primary sclerosis cholangitis		Pancreas carcinoma
Pancreas carcinoma		
Chronic diarrhoea-related set		
Inflammatory bowel disease	Inflammatory bowel disease	Celiac disease
Celiac disease		Pseudomembranous colitis
Pseudomembranous colitis		Chronic infectious diarrhoea
Chronic infectious diarrhoea		
Fillers		
Rheumatoid arthritis	Stomach cancer	Nephrotic syndrome
Hyperthyroidism	Meningoencephalitis	Heart failure
Acute pyelonephritis	Chronic pulmonary obstructive disease	Acute appendicitis

Supplement 2 – Additional results

Table 1 - Cities and teaching hospitals involved in the study

City	Hospital	Eligible residents	Enrolled	Completed
Belo Horizonte	Santa Casa, Federal University of Minas Gerais, FHEMIG	93	55	49
Campinas	UNICAMP	38	15	12
Fortaleza	HGF, Federal University of Ceará	28	18	10
Manaus	Federal University of Amazonas	13	11	10
São Paulo	University of São Paulo	60	13	10
Total		232	112	91

Table 2 – Initial and final diagnostic accuracy (range 0 – 1) in Exercise 1 of the immunization intervention on diseases associated with chronic diarrhoea and diseases associated with jaundice*

	Initial diagnostic accuracy Mean (SD)	Final diagnostic accuracy Mean (SD)	Statistics
Chronic diarrhoea	0·45 (0·30)	0·57 (0·30)	$t(43) = 3{\cdot}81$; $p < 0{\cdot}001$
Jaundice	0·46 (0·25)	0·61 (0·21)	$t(46) = 4{\cdot}73$; $p < 0{\cdot}001$

*Participants were randomly allocated to work either with vignettes with diseases associated with chronic diarrhoea or with vignettes with diseases associated with jaundice in the immunization intervention.

Supplement 3 – Example of a vignette used in the study (inflammatory bowel disease) in the immunisation intervention (Exercise 1)

Case 5

Read the following case and write down your initial diagnosis.

Male patient, 25-year-old, presents with complaints of diarrhoea over the last 4 weeks characterized by watery defecations around three times per day. He denies blood or mucus in the stools but complains of constant and uncomfortable abdominal pain in the lower abdomen on the left side. Since the beginning, he has fever which, despite having decreased intensity, is still present. He lost 8 kg in the period but is otherwise feeling well.

History: no significant pathologies were reported.

Physical examination: the patient is emaciated; weight 60 kg; height 1·65 m; BMI 20; Blood pressure 100 x 60 mmHg; heart rate = 90 bpm, regular pace. Head / neck: no abnormalities. Heart: regular heart rhythm with normal heart tones without heart murmurs. Lungs: normal and symmetric lung sounds; clear sounds on percussion. The abdomen is flaccid; the liver is palpable 1 cm below the right costal border with a smooth surface; the spleen is not palpable. The patient refers pain during palpation of the right side of the lower abdomen. There is no blood nor tumours on rectal examination. Extremities: necrotic, purulent lesion in the left ankle. The patient also complains of pain during the palpation of the left sacroiliac joint.

Laboratory tests: Haemoglobin = 78 g / L; White cells count: = 12.6 x 10^9/L; Eosinophils = 10%; Platelets = 160 x 10^9/L; ESR = 28 mm/hr; CRP = 5 mg/dL; TSH = 1.8 mU/L; AST = 30 U L; ALT = 25 U/L; Glucose (fasting) = 4.6 mmol/L; Faecal examination = no parasites.

What is the most likely diagnosis for this case?

Inflammatory Bowel Disease

Turn the page

Case 5

The case is described here again

Male patient, 25-year-old, presents with complaints of diarrhoea over the last 4 weeks characterized by watery defecations around three times per day. He denies blood or mucus in the stools but complains of constant and uncomfortable abdominal pain in the lower abdomen on the left side. Since the beginning, he has fever which, despite having decreased intensity, is still present. He lost 8 kg in the period but is otherwise feeling well.

History: no significant pathologies were reported.

Physical examination: the patient is emaciated; weight 60 kg; height 1·65 m; BMI 20; Blood pressure 100 x 60 mmHg; heart rate = 90 bpm, regular pace. Head / neck: no abnormalities. Heart: regular heart rhythm with normal heart tones without heart murmurs. Lungs: normal and symmetric lung sounds; clear sounds on percussion. The abdomen is flaccid; the liver is palpable 1 cm below the right costal border with a smooth surface; the spleen is not palpable. The patient refers pain during palpation of the right side of the lower abdomen. There is no blood nor tumours on rectal examination. Extremities: necrotic, purulent lesion in the left ankle. The patient also complains of pain during the palpation of the left sacroiliac joint.

Laboratory tests: Haemoglobin = 78 g / L; White cells count: = 12.6 x 10^9/L; Eosinophils = 10%; Platelets = 160 x 10^9/L; ESR = 28 mm/hr; CRP = 5 mg/dL; TSH = 1.8 mU/L; AST = 30 U L; ALT = 25 U/L; Glucose (fasting) = 4.6 mmol/L; Faecal examination = no parasites.

A) The table below presents three possible diagnoses for this case in the column "Diagnostic Hypothesis". If the diagnosis you wrote on the previous page is not among these hypotheses, write it in the last line of the table in the "Diagnostic Hypothesis" column.
B) For each diagnostic hypothesis, write in the respective column the findings of the case that speak in favour of the hypothesis, the findings that speak against it and the findings that you would expect to find in the patient if this hypothesis were correct, but which are not present in the case.
C) After this analysis, evaluate the likelihood of each diagnosis under consideration. Write, in the "Likelihood" column, 1 for the most likely diagnosis for this case, 2 for the second most likely diagnosis, and so on.
D) Now, underline the findings that are present in more than one diagnostic hypothesis and then circle those that appear in only one of the hypotheses. Do this in the column "Findings that speak in favour of the diagnostic hypothesis".

Diagnostic Hypothesis	*Findings that speak in favour of the diagnosis*	*Findings that speak against the diagnosis*	*Findings expected to be found were the diagnosis true, but not present in the case*	*Likelihood*
Inflammatory Bowel Disease	Young patient, persistent diarrhoea, abdominal pain, fever, weight loss, anaemia, elevation of ESR and CRP, leucocytosis Necrotic and purulent lesion in the left ankle (Pyoderma Gangrenous?) Pain during the palpation of left sacroiliac joint	Absence of blood and mucus in the stool	Stools with blood and mucus Inflammatory lesion present in imaging tests (colonoscopy, CT, MRI) Compatible biopsy	1
Infectious Gastroenteritis	Persistent diarrhoea, fever, abdominal pain, increased ESR; increased CRP, leucocytosis	Eosinophilia, marked anaemia, absence of blood and mucus in the stool	Jaundice secondary to transinfectious hepatitis Reactive arthritis Positive stools culture	2
Celiac disease	Young patient, persistent diarrhoea, weight loss, abdominal pain, anaemia	Extra-intestinal lesions	Anti-Endomysial Antibodies and Anti-tissue Transglutaminase Antibody (anti-tTG) Duodenal biopsy with villous atrophy	3

List the "discriminatory" findings, the most important ones to reach the diagnosis in this case: Persistent diarrhoea in a young man with positive inflammatory tests and abdominal pain, with extra intestinal lesions compatible with Pyoderma Ga

Appendix 1 -Diagnoses of the vignettes used in the three phases of the study

Inoculation intervention	Biasing phase	Test phase
Jaundice-related set		
Acute viral hepatitis	Acute viral hepatitis	Alcoholic cirrhosis
Alcoholic cirrhosis		Primary sclerosis cholangitis
Primary sclerosis cholangitis		Pancreas carcinoma
Pancreas carcinoma		
Chronic diarrhoea-related set		
Inflammatory bowel disease	Inflammatory bowel disease	Celiac disease
Celiac disease		Pseudomembranous colitis
Pseudomembranous colitis		Chronic infectious diarrhoea
Chronic infectious diarrhoea		
Fillers		
Rheumatoid arthritis	Stomach cancer	Nephrotic syndrome
Hyperthyroidism	Meningoencephalitis	Heart failure
Acute pyelonephritis	Chronic pulmonary obstructive disease	Acute appendicitis