Michael Kilian, Jonathan Sherbino, Christopher Hicks and Sandra D. Monteiro*

Understanding diagnosis through ACTion: evaluation of a point-of-care checklist for junior emergency medical residents

<https://doi.org/10.1515/dx-2018-0073>

Received August 13, 2018; accepted March 22, 2019; previously published online April 16, 2019

Abstract

Background: Avoiding or correcting a diagnostic error first requires identification of an error and perhaps deciding to revise a diagnosis, but little is known about the factors that lead to revision. Three aspects of reflective practice, seeking Alternative explanations, exploring the Consequences of missing these alternative diagnoses, identifying Traits that may contradict the provisional diagnosis, were incorporated into a three-point diagnostic checklist (abbreviated to ACT).

Methods: Seventeen first and second year emergency medicine residents from the University of Toronto participated. Participants read up to eight case vignettes and completed the ACT diagnostic checklist. Provisional and final diagnoses and all responses for alternatives, consequences, and traits were individually scored as correct or incorrect. Additionally, each consequence was scored on a severity scale from 0 (not severe) to 3 (very severe). Average scores for alternatives, consequences, and traits and the severity rating for each consequence were entered into a binary logistic regression analysis with the outcome of revised or retained provisional diagnosis.

Results: Only 13% of diagnoses were revised. The binary logistic regression revealed that three scores derived from the ACT tool responses were associated with the decision

to revise: severity rating of the consequence for missing the provisional diagnosis, the percent correct for identifying consequences, and the percent correct for identifying traits ($\chi^2=23.5$, $df=6$, $p<0.001$). The other three factors were not significant predictors.

Conclusions: Decisions to revise diagnoses may be cued by the detection of contradictory evidence. Education interventions may be more effective at reducing diagnostic error by targeting the ability to detect contradictory information within patient cases.

Keywords: diagnosis; diagnostic checklist; education strategy; error detection.

Introduction

The dual process model of reasoning has gained popularity within medical education [1, 2]. This model distinguishes between rapid and slow cognitive processes referred to as System 1 (rapid) and System 2 (slow) [1, 2]. This model is associated with the heuristics and biases literature where System 1 is often characterized as error prone [3]. Although research provides evidence against this characterization [4–7], the interpretation remains popular. This proposition has sparked research into diagnostic aids to support the integration of both analytic (System 1) and non-analytic (System 2) processes. Accordingly, two classes of strategies have arisen with the explicit goal of engaging both systems to examine available data.

One strategy focuses on debiasing through the recruitment of metacognition to *enhance* cognitive processing (i.e. be more cautious or attentive), and consider sources of bias [8, 9]. This approach implies that System 2 can act as a rational corrective agent against diagnostic errors. Attempts to evaluate the effectiveness of metacognitive debiasing strategies have been disappointing; although participants in these studies are able to learn the terminology and the related concepts well, this new knowledge does not improve their diagnostic accuracy [10–12].

The second set of strategies aid in *mobilizing* existing knowledge [13–20]. These strategies link specific aspects of the clinical problem to prior medical knowledge using prompts in a checklist format. Experimentally, checklists

*Corresponding author: Sandra D. Monteiro, PhD, Department of Health Research Methods, Evidence and Impact, McMaster University, 1280 Main Street West, 5th Floor, 5002 A/E, Hamilton, ON L8S 4L8, Canada; and McMaster Faculty of Health Sciences Education Research, Innovation and Theory (MERIT) Program, McMaster University, Hamilton, ON, Canada, Phone: +905 525 9140, Fax: +905-572-7099, E-mail: monteisdmcmaster.ca

Michael Kilian: Division of Emergency Medicine, North York General Hospital, Toronto, ON, Canada, E-mail: kilianmr@gmail.com

Jonathan Sherbino: Division of Emergency Medicine, Department of Medicine, McMaster University, Hamilton, ON, Canada; and McMaster Faculty of Health Sciences Education Research, Innovation and Theory (MERIT) Program, McMaster University, Hamilton, ON, Canada, E-mail: sherbino@mcmaster.ca

Christopher Hicks: Division of Emergency Medicine, Department of Medicine, St Michael's Hospital, University of Toronto, Toronto, ON, Canada, E-mail: chrismikehicks@gmail.com

have led to small improvements in diagnostic accuracy [15–17, 20]. The benefit seems to come from guiding the clinician through a detailed search of the patient case for clinical features that might support their provisional diagnosis or infer alternatives [15, 21]. These cognitive checklists do not focus on actions or behaviors, as in a surgical context [22, 23], but rather on search strategies and consideration of specific attributes. Key features of effective reflective protocols include an initial or provisional diagnosis followed by identification of important case features, alternative diagnoses, and evidence to confirm or contradict the provisional diagnosis [17].

To date, research has demonstrated not only that diagnostic accuracy may be improved through enforced (i.e. not self-guided) analysis of available data [14, 15], but also that these effects are fragile [13, 14, 24] and may depend on having the requisite knowledge to decide when revisions to a provisional diagnosis are necessary [25]. One aspect of error identification and correction that remains unexplored is the mechanism that leads to the decision to change a diagnosis. What factors within the exploration of a patient case may influence whether a clinician decides to take another look or change their diagnosis?

Graber and colleagues [8, 26] synthesized the essential aspects of an analytic approach into four distinct elements: seeking alternative explanations, exploring the consequences of missing these alternative diagnoses, being open to tests that would help differentiate between different differential diagnoses, and being open to uncertainty. We translated these elements of reflective practice into a simple diagnostic checklist (the ACT diagnostic checklist) that asks three questions of the clinician:

1. Are there Alternative diagnoses that have not been considered?
2. What are the Consequences of missing these diagnoses?
3. Is there contradictory evidence from Traits (features of the case) that do not support my provisional diagnosis?

We studied whether clinicians' decisions to revise a diagnosis is related to their ability to identify case features using the ACT diagnostic checklist.

Materials and methods

Materials

A selection of eight written medical cases were adapted from previous studies [27, 28]. These cases represented real-life scenarios

taken from the experiences of emergency physicians in Hamilton health care centers. Based on previous studies, these cases were deemed appropriately challenging for residents in their first year of training. All case vignettes presented information about patient history, primary complaint, physical exam results, and test results (if relevant) in a standard sequence. A sample case that was used in the study is presented in Supplementary material, Appendix A.

Participants

Emergency medicine residents in their first and second year of training at the University of Toronto (UofT) were approached by the principal investigator (MK) and asked to volunteer for the study. Initial recruitment occurred through email linking volunteers to an online survey. However, due to lack of success, additional recruitment was conducted in person with a paper-based survey.

Ethics

In accordance with the World Medical Association Declaration of Helsinki regarding ethical conduct of research involving human subjects, the study was approved by the Ethics Board at St. Michael's Hospital (#15-228).

Procedure

Participants were asked to read a patient case, select an initial diagnosis, use the ACT diagnostic checklist, and select a final diagnosis for each case; participants could diagnose up to eight cases. All participants were first oriented to the ACT diagnostic checklist and then reviewed each case in the same order. While completing each task, the participant was able to see the entire case vignette.

Task one: Participants reviewed the written case vignette and made a provisional diagnosis using five key words or less in an open field text box.

Task two: Participants then completed the ACT tool. This appeared to them as a series of three questions with free text boxes in which the participant could identify up to two alternative diagnoses, identify one consequence of missing the provisional and each alternative diagnosis (e.g. a consequence of missing a diagnosis of aortic dissection would be death), and identify any missing evidence or inconsistent information (traits) related to the provisional diagnosis.

Task three: Participants were asked to state a final diagnosis, which could be the initial (provisional) diagnosis, any of the alternative diagnoses, or a diagnosis that was not previously listed.

Data collection and scoring

Subject matter experts led by JS developed a scoring rubric which was applied in four previous studies using written cases [25, 27–29]. Using this same rubric, the provisional and final diagnoses were scored as correct or incorrect (1 or 0) by SM and MK. All responses using the ACT diagnostic checklist were scored by MK (senior emergency

medicine resident) based on clinical appropriateness (see Supplementary material, Appendix B).

Alternatives were scored 0 for no alternatives offered, and 1 for each plausible alternative.

Consequences were scored overall between 0 and 3. A score of 3 could be achieved by providing three reasonable consequences related to missing the provisional diagnosis and missing the two alternative diagnoses. Consequences and alternatives were scored independently.

The traits were scored 0 for no evidence provided and a score of 2 for providing two plausible data that were missing or contradicted the provisional diagnosis.

An additional score was assigned for the severity of each listed consequence ranging from 0 (not severe) to 3 (very severe). A score of 3 was given to any condition that is acutely life or limb threatening, requires emergent medical or surgical intervention, or results in a permanent disability. A score of 1 was given to any condition that could be managed non-urgently and would not require immediate attention by a physician. A score of 2 was assigned to anything that fell in between these two definitions.

To summarize, this study produced the following sets of responses for analysis: (a) accuracy scores for provisional and final diagnoses (out of 1), (b) accuracy scores for ACT responses: alternative diagnoses (out of 2), consequences (out of 3) and traits (out of 2), and (c) severity rating per consequence (out of 3).

Analysis

To determine if the decision to revise a diagnosis was related to responses to the ACT protocol questions, the average ACT scores and the severity scores of the consequences were submitted to a binary logistic regression analysis with the binary outcome of a decision to revise or retain a provisional diagnosis. Additionally, we explored if average ACT scores varied with the decision to revise, or the accuracy of the provisional diagnosis. We conducted two separate multivariate analysis of variance (ANOVA). In the first analysis, the accuracy of the provisional diagnosis was a between-subjects factor and the three ACT scores were dependent variables. In the second analysis, the decision to revise or retain was a between-subjects factor and the three ACT scores were dependent variables.

Results

Recruitment was stopped after 17 participants; not all participants completed all eight cases. Eight participants were recruited online and nine were recruited in person. Only scores for completed cases were included in the analysis. A total of 105 cases were completed (Table 1).

Relationship between ACT protocol and the decision to revise

The binary logistic regression analysis revealed that three scores derived from the ACT protocol were associated with the decision to revise: severity rating of the consequence for missing the provisional diagnosis, the overall percent correct for identifying consequences, and the overall percent correct for identifying missing evidence or traits. The other three factors were not significant predictors. The overall model was significant ($\chi^2=23.5$, $df=6$, $p<0.001$). The Hosmer-Lemeshow test revealed that the model fit the data ($\chi^2=11$, $p=0.2$).

The odds ratio (OR) for severity was 0.27 [95% confidence interval (CI)=0.09–0.83], with a β of -1.3 suggesting an inverse relationship between perceived severity and decision to revise (Wald=5.17, $p<0.05$). The decision to revise the provisional diagnosis was associated with a lower perceived severity of consequence for missing the provisional [1.85/3, standard deviation (SD)=0.8] than the decision to retain the provisional (2.5/3, SD=0.7) [$F(1102)=4.78$, $p<0.01$].

The average score for consequences was also a significant factor (OR=0.04, 95% CI=0.0–0.56), with a $\beta=-3.15$ suggesting that the overall ability to identify correct consequences was more strongly related to the decision to *retain* a diagnosis (Wald=5.76, $p<0.05$).

Table 1: Count of provisional diagnoses changed, showing the number that was correct and the number that was incorrect as a result of the change.

	Case diagnosis	Participants who diagnosed this case (n)	Count of correct provisional diagnoses	Number of provisional diagnoses changed	Count of correct final diagnoses
1	Ascending aortic aneurysm	17	5	4	6
2	Perforated viscus	15	15	2	13
3	Addison's disease	13	4	2	4
4	Thoracic aortic dissection	12	11	0	11
5	Ascending cholangitis	12	8	2	10
6	Malignant otitis externa	12	4	2	3
7	Lingular pneumonia	12	8	0	8
8	Hemorrhagic cystitis	12	1	2	1

These findings replicate those of Monteiro et al. [29].

Finally, the average score for traits was a significant factor (OR=14.51, 95% CI=1.95–107.98), with a $\beta=2.68$ suggesting that the identification of correct missing or inconsistent evidence in the case was associated with decisions to revise (Wald=6.82, $p<0.01$).

Analysis of individual ACT tool sections

Alternatives

Two alternative diagnoses were provided in 90% of cases, but they were not always appropriate. Only 47 diagnoses were accompanied by two *correct* alternatives. This translates into an overall percent correct score for alternatives of 66%, SD=34%. There was no relationship detected between the score for alternatives and the decision to revise ($p>0.8$). The low score for alternatives indicated a limited ability to consider multiple unrelated diagnoses or consequences of missing them. When provisional diagnoses were revised, the percent correct for alternatives was about the same (67.9%) as when the provisional was retained (65.9%). When provisional diagnoses were correct, the alternatives identified were slightly more accurate (70.5%) compared to when they were incorrect (61.2%), but this was not significant.

Consequences

The percent correct for all three consequences was 84%, SD=28%. The percent of accurate consequences was higher (87.2%) when provisional diagnoses were retained than when revised [66.7%, $F(1103)=6.8$, $p<0.01$]. We explored this further using descriptive statistics. On 75 out of 105 responses, the score for consequences was 100% (that is participants indicated three correct consequences for the provisional and two alternatives). When participants had a *perfect* score for consequences, 68 out of 75 diagnoses were retained.

Traits

The percent correct for contradictory evidence (T) was 32%, SD=36%. Percent correct for traits was low (29%) when diagnoses were retained, compared to 57% when diagnoses were revised [$F(1103)=8.2$, $p<0.01$], suggesting this evidence may have served to cast doubt on the provisional diagnosis. Conversely, accuracy for detecting contradictory evidence was higher when the provisional

diagnosis was incorrect (38%) than when it was correct (28%), but this was not significant ($p=0.2$).

Relationship between ACT protocol and accuracy of the final diagnosis

A binary logistic regression examining the ACT scores and the final diagnostic accuracy did not identify a predictive factor as none of the variables in the equation were significant. However, there was a marginal relationship between correctly detecting contradictory evidence (i.e. traits) and providing a correct diagnosis ($p=0.07$).

The three-point ACT checklist was ineffective as an error reduction tool. The accuracy of the diagnosis was identical for both provisional and final diagnoses (0.53, SD=0.50). Overall, 49 out of 105 provisional diagnoses were incorrect, with only 14 attempted corrections and no reduction of errors. The number of revisions that resulted in increased accuracy was equal to the number that resulted in decreased accuracy. Table 1 describes the count of changes made and the direction of the change (i.e. correct or incorrect).

Discussion

It has been suggested that taking the time to reflect on a diagnosis in a structured manner can improve self-correction of diagnostic error [8, 26]. To date, there has been little evidence in support of this proposal [11]. One reason for the low success is that studies also demonstrate a reluctance to revise a provisional diagnosis [25, 29]. Specifically, Monteiro and colleagues demonstrated that the decision to revise occurs on an individual case basis, is not determined by individual physician's experience or knowledge, and does not typically lead to improved diagnostic accuracy [25].

The goal of the current study was to evaluate whether an abbreviated diagnostic checklist (ACT) can illuminate the relationship between features of a patient case and the decision to revise a provisional diagnosis. Critically, the goal of the study was not to evaluate individual physician's diagnostic accuracy or tendencies to revise a diagnosis when applying analytic reasoning; these have already been explored in other studies [25, 27–29]. The goal was to evaluate the impact of specific items in the diagnostic checklist on the decision to revise as well as on overall accuracy.

The scores derived from the ACT tool ranged from around 33% on the detection of traits (indicating junior residents may need training on detecting contradictory clinical features) to 84% on the identification of consequences (suggesting participants had much stronger knowledge about the clinical attributes of competing conditions). Notably, there were relationships detected between the participants' accuracy for identifying contradictory information and the decision to revise a provisional diagnosis. This finding highlights the potential beneficial characteristic of reflective protocols that specifically instruct participants to identify inconsistent information within patient cases [13–15, 21].

When a provisional diagnosis was incorrect, participants were more likely to accurately detect contradictory evidence, and more likely to attempt a revision. However, the lack of improvement in diagnostic accuracy would suggest that participants were not able to self-correct even in the face of this evidence. An important aspect of diagnostic reasoning that still needs exploration is the ability to re-examine a case or seek additional input once doubt has been identified. The somewhat low scores for identifying alternatives (66%) offer a possible explanation – that participants were unable to identify a reasonable alternative 33% of the time. These missed opportunities may be a reasonable estimate of the source of diagnostic error; a limited ability to generate accurate diagnostic hypotheses restricts the potential for self-correction [24, 25, 29].

This study is best understood in the context of process-based strategies to reduce diagnostic error. Providing diagnosticians with a structured process, whether simple as in this study, or more complex as used by Mamede and colleagues, may provide little to no advantage at point of care. What is most promising about the results of this study, however, is the potential to use this simplified diagnostic checklist as an assessment or training protocol. Trainees may be encouraged to consider using the ACT protocol with practice cases to help activate and structure their knowledge as it pertains to medical diagnosis. Additionally, trainees may benefit from using the protocol in the workplace to identify potential sources of doubt, prompting them to seek a second opinion or additional advice.

There was also a positive relationship between the perceived severity of consequences for missing the provisional diagnosis and the decision to revise the provisional diagnosis; for increasing perceived severity, more diagnoses were revised. This finding may be consistent with medical training that emphasizes attention to “can’t miss” diagnoses; if participants felt that the more severe diagnosis had to be pursued in order to rule it out, this may have

also affected the decision to revise. It is possible that the perception that serious consequences are more likely for missing one diagnosis than others may encourage a physician to commit more strongly to that diagnosis.

Limitations

There are several limitations to this study. The major limitation, which is true for any point-of-care diagnostic checklist, is that there is no reasonable way to force each individual to reason in a standard manner. We cannot be assured that our participants were answering the ACT questions in the same order all the time and whether that impacted their answers at all. However, this would be a limitation of real practice as well; *thinking* cannot be standardized. A second major limitation of the paper is that the scores of the ACT responses were done by one provider with broadly specified criteria (“based on clinical appropriateness”); these factors limit the replicability of these results. In addition, there was no control group and it is possible that some other factor besides the ACT tool led to participants' decisions. Finally, it was not possible to force all participants to spend an equal amount of time on each case or to complete all the cases.

Conclusions

Simplistic or brief point-of-care error reduction strategies do not appear to be effective. Future work should focus on improving the acquisition and integration of knowledge and experience [11, 12, 24, 25, 29, 30]. This is particularly important during the transition toward more independent practice for junior emergency medicine residents. Although the ACT tool did not affect diagnostic accuracy, it may be a useful teaching tool at the bedside, particularly when staff are reviewing with junior residents and developing an assessment and plan for their patients.

Acknowledgments: We would like to acknowledge the assistance of Melissa McGowan in the development and organization of this study.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

Employment or leadership: None declared.

Honorarium: None declared.

Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Evans JS. In two minds: dual-process accounts of reasoning. *Trends Cogn Sci* 2003;7:454–9.
2. Evans JS, Stanovich KE. Dual-process theories of higher cognition: advancing the debate. *Perspect Psychol Sci* 2013;8:223–41.
3. Croskerry P. Clinical cognition and diagnostic error: applications of a dual process model of reasoning. *Adv Health Sci Educ Theory Pract* 2009;14:27–35.
4. Norman G, Young M, Brooks L. Non-analytical models of clinical reasoning: the role of experience. *Med Educ* 2007;41:1140–5.
5. Norman G. Research in clinical reasoning: past history and current trends. *Med Educ* 2005;39:418–27.
6. Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ* 2005;39:98–106.
7. Eva KW, Hatala RM, LeBlanc VR, Brooks LR. Teaching from the clinical reasoning literature: combined reasoning strategies help novice diagnosticians overcome misleading information. *Med Educ* 2007;41:1152–8.
8. Graber ML. Educational strategies to reduce diagnostic error: can you teach this stuff? *Adv Health Sci Educ Theory Pract* 2009;14:63–9.
9. Croskerry P. A universal model of diagnostic reasoning. *Acad Med* 2009;84:1022–8.
10. De Bruin AB, Dunlosky J, Cavalcanti RB. Monitoring and regulation of learning in medical education: the need for predictive cues. *Med Educ* 2017;51:575–84.
11. Norman GR, Monteiro SD, Sherbino J, Ilgen JS, Schmidt HG, Mamede S. The causes of errors in clinical reasoning: cognitive biases, knowledge deficits, and dual process thinking. *Acad Med* 2017;92:23–30.
12. Shimizu T, Matsumoto K, Tokuda Y. Effects of the use of differential diagnosis checklist and general de-biasing checklist on diagnostic performance in comparison to intuitive diagnosis. *Med Teach* 2013;35:e1218–29.
13. Mamede S, Schmidt HG. Reflection in medical diagnosis: a literature review. *Health Prof Educ* 2017;3:15–25.
14. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Med Educ* 2008;42:468–75.
15. 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. *J Am Med Assoc* 2010;304:1198–203.
16. Sibbald M, de Bruin AB, van Merriënboer JJ. Checklists improve experts' diagnostic decisions. *Med Educ* 2013;47:301–8.
17. Sibbald M, de Bruin AB, Cavalcanti RB, van Merriënboer JJ. Do you have to re-examine to reconsider your diagnosis? Checklists and cardiac exam. *BMJ Qual Saf* 2013;22:333–8.
18. Sibbald M, De Bruin AB, van Merriënboer JJ. Finding and fixing mistakes: do checklists work for clinicians with different levels of experience? *Adv Health Sci Educ Theory Pract* 2014;19:43–51.
19. Sibbald M, de Bruin AB, Yu E, van Merriënboer JJ. Why verifying diagnostic decisions with a checklist can help: insights from eye tracking. *Adv Health Sci Educ Theory Pract* 2015;20:1053–60.
20. Ely JW, Graber ML, Croskerry P. Checklists to reduce diagnostic errors. *Acad Med* 2011;86:307–13.
21. Mamede S, Schmidt HG. The structure of reflective practice in medicine. *Med Educ* 2004;38:1302–8.
22. Levy SM, Senter CE, Hawkins RB, Zhao JY, Doody K, Kao LS, et al. Implementing a surgical checklist: more than checking a box. *Surgery* 2012;152:331–6.
23. Van Klei WA, Hoff RG, Van Aarnhem EE, Simmermacher RK, Regli LP, Kappen TH, et al. Effects of the introduction of the WHO “Surgical Safety Checklist” on in-hospital mortality: a cohort study. *Ann Surg* 2012;255:44–9.
24. Ilgen JS, Bowen JL, McIntyre LA, Banh KV, Barnes D, Coates WC, et al. Comparing diagnostic performance and the utility of clinical vignette-based assessment under testing conditions designed to encourage either automatic or analytic thought. *Acad Med* 2013;88:1545–51.
25. 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;30:1270–4.
26. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* 2005;165:1493–9.
27. Sherbino J, Dore KL, Wood TJ, Young ME, Gaissmaier W, Kreuger S, et al. The relationship between response time and diagnostic accuracy. *Acad Med* 2012;87:785–91.
28. Norman G, Sherbino J, Dore K, Wood T, Young M, Gaissmaier W, et al. The etiology of diagnostic errors: a controlled trial of system 1 versus system 2 reasoning. *Acad Med* 2014;89:277–84.
29. Monteiro SD, Sherbino JD, Ilgen JS, Dore KL, Wood TJ, Young ME, et al. Disrupting diagnostic reasoning: do interruptions, instructions, and experience affect the diagnostic accuracy and response time of residents and emergency physicians? *Acad Med* 2015;90:511–7.
30. Zwaan L, de Bruijne M, Wagner C, Thijs A, Smits M, van der Wal G, et al. Patient record review of the incidence, consequences, and causes of diagnostic adverse events. *Arch Intern Med* 2010;170:1015–21.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/dx-2018-0073>).