John W. Ely* and Mark A. Graber

Checklists to prevent diagnostic errors: a pilot randomized controlled trial

Abstract

Background: Many diagnostic errors are caused by premature closure of the diagnostic process. To help prevent premature closure, we developed checklists that prompt physicians to consider all reasonable diagnoses for symptoms that commonly present in primary care.

Methods: We enrolled 14 primary care physicians and 100 patients in a randomized clinical trial. The study took place in an emergency department (5 physicians) and a same-day access clinic (9 physicians). The physicians were randomized to usual care vs. diagnostic checklist. After completing the history and physical exam, checklist physicians read aloud a differential diagnosis checklist for the chief complaint. The primary outcome was diagnostic error, which was defined as a discrepancy between the diagnosis documented at the acute visit and the diagnosis based on a 1-month follow-up phone call and record review.

Results: There were 17 diagnostic errors. The mean error rate among the seven checklist physicians was not significantly different from the rate among the seven usual-care physicians (11.2% vs. 17.8%; p=0.46). In a post-hoc subgroup analysis, emergency physicians in the checklist group had a lower mean error rate than emergency physicians in the usual-care group (19.1% vs. 45.0%; p=0.04). Checklist physicians considered more diagnoses than usual-care physicians during the patient encounters (6.5 diagnoses [SD 4.2] vs. 3.4 diagnoses [SD 2.0], p<0.001).

Conclusions: Checklists did not improve the diagnostic error rate in this study. However further development and testing of checklists in larger studies may be warranted.

Keywords: checklist; diagnostic errors; emergency medicine; primary health care; randomized controlled trial.

Introduction

The incidence of diagnostic errors in emergency departments and primary care clinics is unknown but estimated to be 5%–10% [1, 2]. Patient outcomes following diagnostic errors range from trivial to devastating, but outcome data are limited. The causes of diagnostic errors have been grouped into system errors, such as lost lab results, and thinking errors, such as cognitive biases and failed heuristics [3]. The causes have been further classified by Schiff and colleagues, who also gathered data on their frequency [4]. The most common cause in their study of 583 diagnostic errors was the physician's failure to consider the correct diagnosis as a possibility. This finding was supported in a separate study of 100 diagnostic errors in which premature closure of the diagnostic process was the most common cause [5].

Efforts to prevent diagnostic errors have also been grouped into system interventions and cognitive interventions [6–8]. To help avoid premature closure, physicians can employ de-biasing strategies, such as diagnostic "time-outs" [8, 9] and decision support tools, such as "Isabel" (Isabel Healthcare Inc., USA), to help broaden the differential diagnosis [8, 10]. However, these interventions have not been adequately tested in clinical trials involving real patients. A recent review by McDonald and colleagues found only eight clinical trials in which diagnostic accuracy was the primary outcome and none of them tested a comprehensive de-biasing strategy or decision support tool [7].

The purpose of this study was to test a set of diagnostic checklists for common presenting symptoms in primary care. The checklists provided the differential diagnosis for

*Corresponding author: John W. Ely, MD, MSPH, Carver College of Medicine, Department of Family Medicine, University of Iowa, Iowa City, IA 52242, USA, Phone: 319-384-7533,
E-mail: john-ely@uiowa.edu
Mark A. Graber: Carver College of Medicine, Department of Emergency Medicine and Family Medicine, University of Iowa, Iowa City, IA, USA

common presenting symptoms, such as abdominal pain and dizziness. If such checklists were found to be effective and practical, they could potentially change the diagnostic process at the point of care.

Materials and methods

Study design

We conducted a randomized clinical trial with two parallel arms, comparing usual care with diagnostic checklists. A cluster design was used in which physicians, rather than patients, were randomly assigned to each study arm. Each cluster comprised one physician and three to 10 patients seen by that physician. The cluster design was used to help avoid contamination between study arms in which the physician might remember a more complete differential diagnosis with a usual-care patient who followed a checklist patient. Study procedures took place at the University of Iowa Hospitals and Clinics, in the emergency department and the family medicine same-day access clinic. The primary outcome was diagnostic error, which we defined as a meaningful discrepancy between the impression in the medical record and a 1-month follow-up diagnosis. A "meaningful" discrepancy was defined as one that would have likely changed the patient's management as judged by the investigators. The follow-up diagnosis was based on a medical record review and a phone call to the patient to determine the clinical course and any subsequent medical encounters occurring outside the University of Iowa.

Study sample

All attending family physicians and emergency physicians at the University of Iowa were eligible for the study. Patients were eligible if they were over age 18 years and presented with one of the 63 symptoms addressed by the checklists (Appendix). The study was
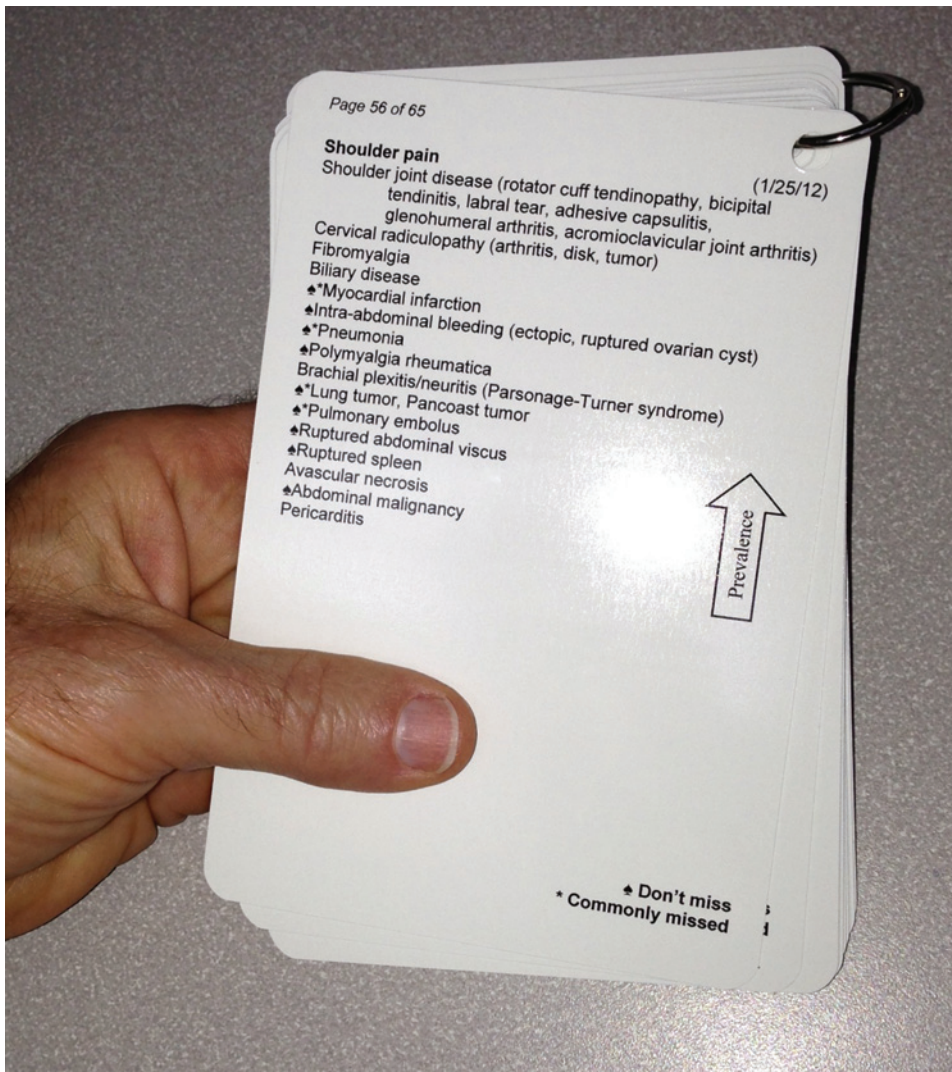

Figure 1: Diagnostic checklist in use.

approved by the University of Iowa Institutional Review Board and all physicians and patients signed informed consent documents before participating. The consent documents included the risk that the checklist could lead the physician away from an initially correct diagnosis, and the risk that it could prompt unnecessary testing. We continued to recruit patients until 100 completed the trial. We did not perform a sample-size calculation because the trial was primarily designed to troubleshoot study procedures and provide variance estimates for a future larger study.

Intervention

Physicians in the checklist group were asked to review a differential diagnosis checklist in the presence of the patient (Figure 1 and Supplemental Material that accompanies the article at http://www.degruyter.com/view/j/dx.2015.2.issue-3/dx-2015-0008/dx-2015-0008.xml?format=INT). The checklists were printed on 4×6 cards with one symptom per card. Each symptom included an average of 22 possible diagnoses (SD 10, range 8–59). Commonly missed diagnoses were marked with an asterisk. "Don't-miss" diagnoses (those with potentially serious consequences if missed) were marked with an ace of spades. These designations were based on the investigators' experience and previously published data [4, 11]. The diagnoses were ordered according to their prevalence in primary care, which was also based on the investigators' clinical experience and a focused literature review [12–15]. The checklists used in this study were slightly modified from a previously published version [16].

Procedures

Eligible physicians received an introductory letter followed by a phone call or face-to-face encounter with the principal investigator to invite them into the trial. Eligible patients were approached by the principal investigator at the acute visit. Physicians were randomly assigned to usual care vs. checklist using a randomized block design.

In both study arms, physicians conducted the history and physical exam as usual, unobserved by the investigator, who stood outside the exam room. When usual-care physicians exited the room, the investigator asked for the primary diagnosis and the differential diagnosis along with the management plan. When checklist physicians exited, the investigator and physician immediately re-entered the exam room, and the physician verbally reported the primary diagnosis, the differential diagnosis, and the management plan to the investigator and patient. The physician then read aloud the diagnostic checklist in the presence of the patient and reported any revisions to the primary diagnosis, the differential diagnosis, and the management plan. During these encounters the investigator used pencil and paper to record physician responses and the time required to review the checklist.

All study procedures took place between June 2010 and May 2013. The checklist review process is illustrated in a 6-min video, available online (http://www.youtube.com/watch?v=uHpieuyP1w0). We asked physicians to read the checklists in the patient's presence rather than outside the exam room because of anecdotal experience in which this practice prompted further history questions from the physician and valuable information from the patient.

One month after the acute visit, the principal investigator reviewed the medical record to obtain the diagnosis documented in the physician's note for the acute encounter. Follow-up visits and hospitalizations were also reviewed to detect any evidence of a missed or delayed diagnosis. In addition, the investigator called the patient to ask about the subsequent course, any medical visits to other facilities, and any evidence of discrepancy with the initial diagnosis. The patients were aware of the purpose of the study, and during the phone call the investigator said, "We just want to make sure we made the correct diagnosis at the time of your visit". These phone calls were not made if the medical record was sufficient to determine the accuracy of the initial diagnosis.

The principal investigator recorded the chart diagnosis and final diagnosis in an unblinded manner during the course of the study. After the last subject completed all study procedures, a copy of the data file was created in which all subject identifiers and study arm information were deleted. The two investigators then independently reviewed this file, blinded to physician identity and study arm to determine the existence of a meaningful discrepancy between the chart diagnosis and the final diagnosis. These judgments were subjective and dichotomous (meaningful discrepancy vs. no meaningful discrepancy). For example, a chart diagnosis of viral upper respiratory infection vs. a final diagnosis of sinusitis was considered meaningful because sinusitis could prompt antibiotic treatment.

Data analysis

We used descriptive statistics to determine the frequency of diagnostic errors, the number of diagnoses in the physician's differential diagnosis, the frequency of symptoms, and the time required to review the checklist. The analysis of the trial results required methods designed for cluster randomized trials rather than simple odds ratios. We used a cluster-level analysis in which the physician was the unit of analysis and the outcome was a summary measure for each physician (the proportion of patients seen by that physician who sustained a diagnostic error). Because the number of physicians was small, we used the t-test to compare these error rates as recommended by Hayes and Moulton [17]. The t-test is robust to violations of assumptions related to normal distributions and equal variances [18–20], and it can be applied to comparisons involving extremely small samples [21]. All analyses were performed with Stata Version 12 (StataCorp).

Results

All 14 invited physicians agreed to participate. We invited 104 patients to participate and 103 (99%) agreed. After signing the consent document, one patient became ineligible when the study physician asked a different physician, not in the study, to see her instead. Two patients withdrew from the study without explanation after signing the consent document. Thus the final sample included 14 physicians (6 females and 8 males) and 100 patients (67 females and 33 males). The emergency room portion of the sample included five physicians and 42 patients. The same-day access portion of the sample included nine physicians and

58 patients. The usual-care group comprised seven physicians and 47 patients. The checklist group comprised 7 physicians and 53 patients. We did not use imputation methods to perform an intention-to-treat analysis because the number of withdrawals was so small (n=3).

There were 17 diagnostic errors. The mean error rate among checklist physicians was not statistically different from the error rate among usual-care physicians (11.2% vs. 17.8%, p=0.46) (Table 1). In a post-hoc subgroup analysis, emergency physicians in the checklist group had a lower error rate than emergency physicians in the usual-care group (19.1% vs. 45.0%, p=0.04). The error rates in the same-day access clinic did not differ significantly between the checklist and usual-care groups (5.3% vs. 6.9%, p=0.77).

There were no important discrepancies between the two investigators in their independent judgments about the existence of a meaningful discrepancy between the initial diagnosis and the 1-month final diagnosis. The final diagnosis was evident from the medical record in 28 of the 100 patients. Of these 28 patients, 9 were hospitalized. The hospital record confirmed a correct initial diagnosis in 5 of these 9 and a diagnostic error in the remaining 4. Among the 19 non-hospitalized patients, 18 had follow-up clinic visits that confirmed a diagnostic error in 3 cases and a correct initial diagnosis in 15 cases. The remaining patient had no clinic follow-up but was not called because of an unstable psychiatric condition.

We attempted to reach the remaining 72 patients by phone and were successful with 62 of them (86%). We arbitrarily assumed that the remaining 10 patients

Table 1: Diagnostic error rates among checklist vs. usual-care physicians.

Variable	Checklist	Usual care	p-Value[a]
Combined same-day access clinic and emergency department			
Physicians, n	7	7	
Patients, n	53	47	
Mean Error Rate (SD)[b]	11.2% (9.7%)	17.8% (20.4%)	0.46
Same-day access clinic			
Physicians, n	4	5	
Patients, n	26	32	
Mean error rate (SD)[b]	5.3% (6.1%)	6.9% (9.6%)	0.77
Emergency department			
Physicians, n	3	2	
Patients, n	27	15	
Mean error rate (SD)[b]	19.1% (8.0%)	45.0% (7.1%)	0.04

[a]p-Values are based on the t-test of differences between mean error rates. [b]The error rate for each physician equals the number of patients with diagnostic errors seen by that physician divided by the total number of patients seen by that physician.

received the correct diagnosis at the initial visit. These 10 patients included 8 from the usual care group and 2 from the checklist group.

Only one physician was prompted by the checklist to change the diagnosis from incorrect to correct during the acute visit. The patient presented with leg edema and was initially diagnosed with edema secondary to renal failure. After reviewing the leg-edema checklist, the physician changed the diagnosis to edema secondary to medication (a non-steroidal anti-inflammatory drug and a beta blocker). After both medications were stopped, the leg edema resolved. There were no cases in which an initially correct diagnosis was changed to an incorrect diagnosis after reading the checklist.

The number of diagnoses in the physician's reported differential diagnosis was larger in the checklist group than the usual-care group [6.5 diagnoses (SD 4.2) vs. 3.4 diagnoses (SD 2.0), p<0.001]. Among checklist physicians, the number of diagnoses increased from a mean of 4.3 (SD 3.2) before reading the checklist to 6.5 (SD 4.2) afterward (p<0.001). The checklist prompted additional testing in only one patient: a thyroid stimulating hormone in the previously described patient with leg edema. The average time to read the checklist was 80 seconds (SD 41 seconds, range 15–180 seconds).

Physicians requested that the checklist review occur outside the exam room for 14 of the 53 checklist patients (26%). We did not systematically document the reasons for these requests, but they generally arose from the physician's anticipation that the patient might be upset after hearing possible causes for their presenting symptom. However, one physician requested that all checklist reviews occur outside the exam room. Among the remaining 39 patients who heard the checklist review, we had no indication that any became upset. The two patients who withdrew from the study did not hear the checklist reading.

The 100 study patients presented with 25 unique symptoms. The most common were abdominal/pelvic pain (total=17; checklist=10, usual care=7), back pain (total=10; checklist=5, usual care=5), cough (total=10; checklist=4, usual care=6), chest pain (total=7; checklist=6, usual care=1), generalized rash (total=7; checklist=4, usual care=3), and ear pain (total=6; checklist=2, usual care=4). There were 60 unique final diagnoses. The three most common were musculoskeletal back pain (n=9), urinary tract infection (n=7) and viral upper respiratory infection (n=6).

During the analysis, the investigators became aware of potential misclassifications in which "meaningful" discrepancies that met our definition of diagnostic error

Table 2: Potential misclassification of diagnostic errors.

Type of misclassification	Example
Evolution of disease	The diagnosis of viral upper respiratory infection (common cold) may have been correct at the time of the initial visit but progressed to bacterial sinusitis at a follow-up visit.[a]
Nomenclature discrepancy	The initial diagnosis was "viral illness". The final diagnosis was "bronchitis" without designating the etiology (which was likely viral).[a]
Diagnostic evaluation in process	The initial diagnosis was nausea of unknown etiology. Subsequent testing, which was ordered at the initial visit, revealed diabetic gastroparesis.[a]
Vague initial impression but specific initial plan[b]	An initial impression of "abdominal pain" accompanied by a prescription for a proton-pump inhibitor, followed by a final diagnosis of "peptic gastritis".[a]
One patient with two diseases	The chest pain in the emergency room was caused by reflux esophagitis even though the patient died of a myocardial infarction at home 6 h later.[c]
Discrepancy in disease severity	A patient with dehydration, initially treated and released from the emergency room, required admission for dehydration several hours later.[a]

[a]Example occurred in the present study. [b]Physicians may leave the final diagnosis vague in the medical record (e.g., "abdominal pain") for medicolegal reasons. [c]Theoretical example.

might not have actually been errors (Table 2). With the exception of nomenclature discrepancies, we treated these potential misclassifications as diagnostic errors in the analysis. For example, one case involved an initial diagnosis of viral upper respiratory infection paired with a final diagnosis of bacterial sinusitis. This was a "meaningful" discrepancy that prompted an antibiotic prescription at a return visit, but it may have simply represented progression of disease rather than an incorrect initial diagnosis. In addition, some emergency physicians used the presenting symptom for the final impression (e.g., abdominal pain) rather than documenting a diagnosis that would have supported the treatment plan (e.g., a proton-pump inhibitor). During the study we learned, anecdotally, that this practice may be based on medicolegal considerations in emergency medicine. Although our judgments about misclassification were subjective, at least 3 of the "errors" in the emergency room could have represented this type of misclassification, whereas none of the same-day access errors did.

Among the 17 patients with diagnostic errors, there were no deaths, 10 patients whose treatment was delayed, and seven with no apparent adverse consequences. Among the 10 patients with delayed treatment, three required hospitalization.

Discussion

Main findings

In this study, physicians who reviewed a diagnostic checklist generated a broader differential diagnosis than physicians who did not review the checklist, but there was no significant improvement in their diagnostic error rates. The checklist review usually took <2 min and appeared to be acceptable to both patients and physicians. Concerns that the checklist might prompt unnecessary testing or lead the physician away from a correct initial diagnosis proved to be unfounded in this study. We found six types of misclassification that could falsely elevate diagnostic error rates in studies like this one, where the patient's clinical course was used to detect errors.

Comparison with other studies

We searched PubMed to find studies of checklists to prevent diagnostic errors and retrieved 442 articles on November 21, 2014 using the Mesh term "Diagnostic Errors" combined with the text word, "checklists". Among these 442 articles, six addressed the use of checklists to expand the differential diagnosis [22–27], but none evaluated their efficacy in preventing diagnostic errors. Computerized decision support tools, such as Isabel and DXplain offer potential advantages over the simple checklists used in this study because they can narrow the differential diagnosis to fit the patient's clinical data. They have performed well in simulated [10, 28–32] and real settings [28, 29] but have not been tested against usual care in randomized controlled trials.

Limitations

Our findings should be viewed in light of several limitations. The study was not powered to find important differences between checklists and usual care. An adequately

powered trial would have required 230 patients in each arm, assuming a 20% error rate in the control group, a 10% error rate in the checklist group, a cluster size of 10 patients per physician, a significance level (alpha) of 0.05, a power of 80%, and an intracluster correlation coefficient of 0.019, which was the actual correlation coefficient in the current study [18]. Other limitations included our inability to blind subjects to the intervention, the lack of blinding of the investigator who recorded the chart diagnosis and final diagnosis, the lack of a 1-month follow-up diagnosis for 10 patients, the potential selection bias resulting from recruitment at a single academic center, and our inability to control for potential confounding factors such as the discussion about the differential diagnosis that occurred in the patient's presence for checklist physicians but not for usual-care physicians. Checklist physicians, knowing that they would soon be reading a diagnostic checklist, may have given more critical thought to the diagnosis even before they reviewed the checklist. We sampled only faculty physicians and speculate that trainees might derive more benefit from checklists. The post hoc nature of the subgroup analyses in Table 1 and the borderline statistical significance among emergency physicians could generate hypotheses for future studies, but the conclusion from this study is that there is no proven benefit from diagnostic checklists. Strengths of the study include its randomized controlled design and the inclusion of real patients rather than simulated patients.

Implications

Despite the negative findings in this study, diagnostic checklists may deserve further development and evaluation. The checklists in this study might be viewed as a low-tech alternative to more sophisticated decision support tools such as Isabel and DXplain, but we envision a different role for checklists. We designed the checklists for use every time, before the patient leaves the clinic or emergency department, similar to an airline pilot's checklist, which is used every time before take-off and before landing. In contrast, computerized decision support tools may be more appropriate for challenging cases and for use after clinic hours when the physician has more time to enter data into a computer. However, studies have found that physicians are not good judges of when they need help with diagnosis [33, 34]. Although most checklist reviews in this study took <2 min, even shorter times may be more realistic in busy outpatient settings. This could be accomplished by reading only the first 10 or 15 diagnoses, which, because of their ordering by prevalence, would

likely include the vast majority of patients in primary care. But before checklists can be recommended for routine use, they must be shown to be practical and beneficial in real-world settings. Future studies will require larger samples, more rigorous attempts to blind the investigators, and more systematic methods for addressing the different types of error misclassification found in this study.

References

1. Graber ML. The incidence of diagnostic error in medicine. BMJ Qual Saf 2013;22(Suppl 2):ii21–7.
2. Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual Saf 2014;23:727–31.
3. Croskerry P, Cosby KS, Schenkel SM, Wears RL. Patient safety in emergency medicine. Philadelphia: Lippincott Williams & Wilkins, 2009.
4. Schiff GD, Hasan O, Kim S, Abrams R, Cosby K, Lambert BL, et al. Diagnostic error in medicine: analysis of 583 physician-reported errors. Arch Intern Med 2009;169:1881–7.
5. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. Arch Intern Med 2005;165:1493–9.
6. Singh H, Graber ML, Kissam SM, Sorensen AV, Lenfestey NF, Tant EM, et al. System-related interventions to reduce diagnostic errors: a narrative review. BMJ Qual Saf 2012;21:160–70.
7. McDonald KM, Matesic B, Contopoulos-Ioannidis DG, Lonhart J, Schmidt E, Pineda N, et al. Patient safety strategies targeted at diagnostic errors: a systematic review. Ann Intern Med 2013;158:381–9.
8. Graber ML, Kissam S, Payne VL, Meyer AN, Sorensen A, Lenfestey N, et al. Cognitive interventions to reduce diagnostic error: a narrative review. BMJ Qual Saf 2012;21:535–57.
9. Trowbridge RL. Twelve tips for teaching avoidance of diagnostic errors. Med Teach 2008;30:496–500.
10. Graber ML, Mathew A. Performance of a web-based clinical diagnosis support system for internists. J Gen Intern Med 2008;23(Suppl 1):37–40.
11. 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.

12. Cherry DK, Hing E, Woodwell DA, Rechsteiner EA. National ambulatory medical care survey: 2006 summary. National Health Stat Report 2008;1–39. PMID: 18972720.

13. Burger FL, Lang CM. Diagnoses commonly missed in childhood. Long-term outcome and implications for treatment. Psychiatr Clin North Am 1998;21:927–40, viii.

14. Singh H, Giardina TD, Meyer AN, Forjuoh SN, Reis MD, Thomas EJ. Types and origins of diagnostic errors in primary care settings. J Am Med Assoc Intern Med 2013;173:418–25.

15. Phillips RL Jr., Bartholomew LA, Dovey SM, Fryer GE Jr., Miyoshi TJ, Green LA. Learning from malpractice claims about negligent, adverse events in primary care in the United States. Qual Saf Health Care 2004;13:121–6.

16. Ely JW, Graber ML, Croskerry P. Checklists to reduce diagnostic errors. Acad Med 2011;86:307–13.

17. Hayes RJ, Moulton LH. Cluster Randomized trials. Boca Raton, FL: Chapman & Hall/CRC, 2009.

18. Donner A, Klar N. Design and analysis of cluster randomization trials in health research. London: Arnold, 2000.

19. Donner A, Klar N. Statistical considerations in the design and analysis of community intervention trials. J Clin Epidemiol 1996;49:435–9.

20. Heeren T, D'Agostino R. Robustness of the two independent samples t-test when applied to ordinal scaled data. Stat Med 1987;6:79–90.

21. de Winter JC. Using the student's t-test with extremely small sample sizes. Pract Assessment Res Evaluat 2013;18:1–12.

22. Winters BD, Aswani MS, Pronovost PJ. Commentary: reducing diagnostic errors: another role for checklists? Acad Med 2011;86:279–81.

23. Sibbald M, de Bruin AB, van Merrienboer JJ. Checklists improve experts' diagnostic decisions. Med Educ 2013;47:301–8.

24. Schiff GD, Leape LL. Commentary: how can we make diagnosis safer? Acad Med 2012;87:135–8.

25. Schiff GD, Bates DW. Can electronic clinical documentation help prevent diagnostic errors? N Engl J Med 2010;362:1066–9.

26. Henriksen K, Brady J. The pursuit of better diagnostic performance: a human factors perspective. BMJ Qual Saf 2013;22(Suppl 2):ii1–5.

27. Graber ML, Wachter RM, Cassel CK. Bringing diagnosis into the quality and safety equations. J Am Med Assoc 2012;308:1211–2.

28. Ramnarayan P, Kapoor RR, Coren M, Nanduri V, Tomlinson AL, Taylor PM, et al. Measuring the impact of diagnostic decision support on the quality of clinical decision making: development of a reliable and valid composite score. J Am Med Inform Assoc 2003;10:563–72.

29. Ramnarayan P, Roberts GC, Coren M, Nanduri V, Tomlinson A, Taylor PM, et al. Assessment of the potential impact of a reminder system on the reduction of diagnostic errors: a quasi-experimental study. BMC Med Inform Decis Mak 2006;6:22.

30. Carlson J, Abel M, Bridges D, Tomkowiak J. The impact of a diagnostic reminder system on student clinical reasoning during simulated case studies. Simul Healthc 2011;6:11–7.

31. Bond WF, Schwartz LM, Weaver KR, Levick D, Giuliano M, Graber ML. Differential diagnosis generators: an evaluation of currently available computer programs. J Gen Intern Med 2012;27:213–9.

32. Sherbino J, Kulasegaram K, Howey E, Norman G. Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: a controlled trial. CJEM 2014;16:34–40.

33. Friedman CP, Gatti GG, Franz TM, Murphy GC, Wolf FM, Heckerling PS, et al. Do physicians know when their diagnoses are correct? Implications for decision support and error reduction. J Gen Intern Med 2005;20:334–9.

34. Meyer AN, Payne VL, Meeks DW, Rao R, Singh H. Physicians' diagnostic accuracy, confidence, and resource requests. J Am Med Assoc Intern Med 2013;173:1952–8.