Checklists to Reduce Diagnostic Errors

John W. Ely, MD, Mark L. Graber, MD, and Pat Croskerry, MD, PhD

Abstract

Diagnostic errors are common and can often be traced to physicians’ cognitive biases and failed heuristics (mental shortcuts). A great deal is known about how these faulty thinking processes lead to error, but little is known about how to prevent them. Faulty thinking plagues other high-risk, high-reliability professions, such as airline pilots and nuclear plant operators, but these professions have reduced errors by using checklists. Recently, checklists have gained acceptance in medical settings, such as operating rooms and intensive care units. This article extends the checklist concept to diagnosis and describes three types of checklists: (1) a general checklist that prompts physicians to optimize their cognitive approach, (2) a differential diagnosis checklist to help physicians avoid the most common cause of diagnostic error—failure to consider the correct diagnosis as a possibility, and (3) a checklist of common pitfalls and cognitive forcing functions to improve evaluation of selected diseases. These checklists were developed informally and have not been subjected to rigorous evaluation. The purpose of this article is to argue for the further investigation and revision of these initial attempts to apply checklists to the diagnostic process. The basic idea behind checklists is to provide an alternative to reliance on intuition and memory in clinical problem solving. This kind of solution is demanded by the complexity of diagnostic reasoning, which often involves sense-making under conditions of great uncertainty and limited time.

Editor’s Note: A commentary on this article appears on page 279.

Diagnostic errors occur in medicine at an appreciable, though unknown, rate, estimated to be in the range of 10% to 15%. Many of these errors are consequential, but others result in substantial harm to patients. Diagnostic errors are more likely to be preventable and more likely to result in patient harm than are other types of medical errors.

Diagnostic errors reflect breakdowns in our health care systems, our clinical reasoning, or both. Solutions for the system-based problems are relatively easy to envision, but few interventions to reduce cognitive errors have been implemented or even proposed. Decision support tools can be helpful, but unless they are well integrated in the workflow, they tend to be underused. Other suggestions include reflective practice and training in metacognitive skills to recognize flaws in the intuitive “thinking” that underlies a substantial fraction of our diagnoses.

Given their success in other settings, it is reasonable to suggest that checklists might help reduce diagnostic errors. Checklists are used by airline pilots in all aspects of their work, but were not used routinely until the crash of a Boeing 299 bomber in 1935, which resulted from a pilot’s simple oversight—failure to release the elevator locks. Checklists are used by other high-risk, high-reliability professions, such as submarine crews and nuclear plant operators, to ensure safety. Recently, physicians and nurses have developed checklists to ensure the completion of critical procedures in hospitals. For example, intensive care unit staff use checklists to help prevent bloodstream infections and ventilator-associated pneumonia and a recent international project cut surgical deaths by half after introducing a 19-item checklist for operating rooms. The purpose of this article is to describe a potential role for checklists in avoiding diagnostic errors and to argue for the further development and evaluation of checklists in hospitals, clinics, and emergency rooms.

Cognitive Processes in Diagnosis

Some insights on how checklists work come from studies in cognitive psychology related to the “dual-process” model of thinking and reasoning (Figure 1). This model proposes two basic modes of thinking. Type 1 processes are fast, reflexive, intuitive, and may operate at a subconscious level. We perform many tasks that involve complex decision making without giving them much attention or thought, such as driving a car or performing a neurological exam. Provided they are repeated on a regular basis, these tasks are relegated to an automatic subconscious level, and if everything is as it seems, we perform well. In contrast, Type 2 processes are analytic, slow, and deliberate. They require focused attention.

Clinical work involves many behaviors, but most are overlearned and executed through Type 1 processes. However, as useful as Type 1 thinking can be, it is vulnerable to error. When we are in clinical situations that seem familiar, we are comfortable with our thoughts and may become overconfident. It is exactly under these circumstances that checklists prove effective. For diagnosis, generic checklists could force a reflective check, and specific checklists could force consideration of “must-not-miss” diagnoses.
Checklists could help us resist the biases and failed heuristics that lead to diagnostic errors\(^2\) (Table 1), and they could facilitate proposed techniques for improving diagnostic reasoning.\(^2,10\) Using generic and specific checklists, we are encouraged to

- decrease reliance on memory,
- consider a comprehensive differential diagnosis for common symptoms,
- step back from the immediate problem to examine our thinking process (metacognition),
- develop strategies to avoid predictable bias (cognitive forcing functions),
- recognize our altered mood states that arise from fatigue, sleep deprivation, or other conditions and develop strategies to reduce their negative consequences (affective forcing functions).

**Diagnostic Checklists**

Here, we describe three types of checklists that could potentially reduce diagnostic errors in hospitals, clinics, and emergency rooms. The content of these checklists will seem familiar and possibly even insultingly obvious (e.g., “Obtain your own complete history”), but their routine use in practice would be a major change for most physicians. After all, pilots no longer feel insulted when reminded by their copilots to release the elevator locks.

**The general checklist**

A general checklist provides a reproducible approach to diagnosis.\(^{21}\) List 1 offers an example of such a checklist. Some of the items may seem overly basic, but many errors result from failures in these areas.\(^{22}\) We sometimes forget the “dumb steps” in our work, precisely because they are dumb—we do not articulate them, and we take them for granted.\(^{11}\) Each of the steps in this checklist is discussed in detail in this section.

**Obtain your own complete medical history.** There is no substitute for obtaining your own history because diagnostic errors often result from a previous incomplete or misleading history. They can also result from “upstream” problems—those involving previous encounters—such as succumbing to the framing bias imparted by a previously suggested diagnosis. A diagnosis acquires enormous inertia once it is proposed and communicated, to the extent that subsequent physicians may discount or fail to consider other possible diagnoses. A related problem involves “groupthink,” in which the chances of error increase when the impressions of one member of a group are too quickly adopted by the others.\(^{23}\) Although there may be occasions when an excess of facts and data can be deleterious,\(^{24}\) the more common problem for busy clinicians is insufficient time to obtain a comprehensive medical history, which remains the foundation of reliable diagnosis.

Perform a **focused and purposeful physical exam.** The initial hypotheses that inevitably come to mind during the first moments of the patient encounter should identify elements of the subsequent physical exam that need special attention. However, we must also look for signs that might suggest alternate diagnoses.\(^{25}\)

Generate and differentiate initial hypotheses with further history, physical exam, and diagnostic tests. Diagnostic errors commonly involve problems related to diagnostic testing,\(^{26}\) and in a recent study testing-related problems were a factor in over half the cases.\(^{27}\) These problems can result from errors in the laboratory or radiology department itself, occurring at rates of 2% to 4%, or an error in the pre- or posttest period, occurring at rates of 10% to 20%.\(^{28}\) For example, the wrong test was ordered, the result was lost, or the physician misinterpreted the result.\(^{28}\)

Pause to reflect—take a diagnostic "time-out." Short of seeking a second opinion in every case, reflecting on the plausibility of the working diagnosis may be our best tool to avoid error.\(^{8,9}\) The two most common cognitive errors are context errors and premature closure.\(^{5,26}\) Context errors arise when a critical signal is distorted by the background against which it is perceived.\(^{24}\) A typical context error would be the assumption that abdominal pain reflects a problem with...
The gastrointestinal tract without considering other possibilities, such as pneumonia, lead poisoning, or diabetic ketoacidosis. Premature closure is our tendency to stop considering problems after we find an apparently adequate solution.29 Taking a diagnostic time-out would provide the opportunity to

- Consider the opposite: “Why can’t this be something else?” Tests that rule out alternative possibilities are often more valuable than tests that confirm our original suspicion.29

- Use “prospective hindsight”: Derived from military planners, this technique asks us to look into the future and see what would happen if our diagnosis was wrong. What did we miss, and what else should we have considered?30,31

- Apply decision support tools: A growing number of Web-based differential diagnosis generators are available, such as DXplain (http://dxplain.org/dxp/dxp.pl), Isabel (http://www.isabelhealthcare.com), VisualDx (http://www.visualdx.com), and PEPID (http://www.pepid.com). The low-tech counterpart is to employ a systematic approach, which might include a checklist.

An appropriate step at this point is to consider whether a diagnosis needs to be made at all, or if it can wait, because other decisions may take priority, such as empiric therapy or hospital admission. A patient’s presentation often changes over time as the symptoms evolve. It may be wise to hold off making a diagnosis32,33 and write “NYD” (not yet diagnosed) in the record after the presenting symptom.34 We should avoid any diagnostic label until our certainty is high because dialogue and thinking often stop the instant a label is applied.35,36

Embrace on a plan, but acknowledge uncertainty and ensure a pathway for follow-up.37,38 We often just play the odds when we make a diagnosis. Certainty is not a realistic possibility. The correct diagnosis often emerges over time as test results become available or as the patient’s symptoms and signs evolve. This longitudinal aspect of diagnosis mandates that we reconsider an initial diagnosis at later points in time.39 We strongly advocate including the patient in this process. We should tell the patient our initial thoughts, make clear any uncertainties, and lay out a concrete plan for follow-up.37 Closing this loop by ensuring follow-up is a strategy that can help improve the reliability of diagnosis and provide key feedback to help improve our “calibration”—the correlation between our perceived and actual diagnostic accuracy.

**Differential diagnosis checklists**

The final common pathway for most diagnostic errors is our failure to consider the correct diagnosis.5,26 We argue that this can be addressed by using a set of differential diagnosis checklists. The differential diagnosis checklists

**List 1**

**Proposed General Checklist for Diagnosis**

- Obtain your own complete medical history.
- Perform a focused and purposeful physical exam.
- Generate initial hypotheses and differentiate these with additional history, physical exam, and diagnostic tests.
- Pause to reflect—take a diagnostic “time out.”
  - Was I comprehensive?
  - Did I consider the inherent flaws of heuristic thinking?27
  - Was my judgment affected by any other bias?
  - Do I need to make the diagnosis now, or can I wait?
  - What is the worst-case scenario?
- Embrace on a plan, but acknowledge uncertainty and ensure a pathway for follow-up.
highlighted in List 2 and detailed in the Supplemental Digital Appendix (http://links.lww.com/ACADMED/A38) have a single purpose: to prompt the physician to consider a comprehensive list of causes for the complaints that commonly present diagnostic challenges. The checklists highlight diagnoses that should not be missed and those that are, in fact, commonly missed. The development and focus of the differential diagnosis checklists were based on published data and the authors’ experiences.

One of the authors (J.E.) used published differential diagnoses to develop checklists for 46 presenting complaints, such as chest pain, fatigue, cough, dizziness, and so on. The checklists were revised during two years of use in clinic. A six-minute video that demonstrates use of the differential diagnosis checklist is available on YouTube (http://www.youtube.com/watch?v=uHpieuyP1w0). The diagnoses are ordered according to prevalence in primary care, despite the lack of supporting data, because prevalence may provide more diagnostically helpful information than more traditional variables such as anatomy, pathophysiology, body system, or medical specialty.

It would require thousands of checklists to cover 100% of presenting complaints. Instead, we aimed to cover 99% of those patients who present diagnostic challenges with a small number of checklists. And within each checklist, our goal was to cover at least 99% of patients with a short list of causes for the complaint. We excluded complaints in which the focus is more on treatment than diagnosis, such as diabetes and hypertension, and we excluded complaints for which a list of causes would be unlikely to benefit clinicians, such as constipation and breast lumps.

We lumped diagnoses into clinically relevant groups rather than splitting them into distinct pathologic entities (e.g., “pneumonia” rather than “pneumococcal pneumonia,” “klebsiella pneumonia,” and so on). We also grouped presenting problems (e.g., “abdominal/pelvic pain” rather than “right-upper-quadrant pain,” “right-lower-quadrant pain,” and so on) because we wanted to avoid redundancy. For example, if we did not group presenting problems in this way, pneumonia would have to appear on the right-upper-quadrant-pain checklist, the right-lower-quadrant-pain checklist, and many others.

Although the checklists were developed in the outpatient setting, they may also improve diagnostic accuracy for inpatients. Admitted patients generally come with “admitting diagnoses,” but hospitalist physicians could review the checklist at the time of admission to help determine whether further history taking, physical exam, or diagnostic testing is indicated. They also might find it beneficial to review the checklist for patients who do not respond to initial treatment.

We have not formally evaluated the differential diagnosis checklists, but one of the authors (J.E.) has noted anecdotal success from two years of using the checklists in practice. For example, a 90-year-old woman with chronic obstructive pulmonary disease, coronary artery disease, and metastatic ovarian cancer presented to clinic with dyspnea. The resident noted wheezes which cleared after two albuterol nebulizer treatments, but the patient continued to complain of dyspnea. She had been seen four days earlier with a “COPD exacerbation” and was discharged from clinic after symptomatic improvement with a single albuterol treatment. The attending physician reviewed the dyspnea checklist with the resident, and this prompted a d-dimer test. The d-dimer was 13.89 μg/mL (normal: <0.50 μg/mL). A computerized tomographic angiogram showed pulmonary emboli, and the patient was admitted to the hospital and started on heparin.

This example should be viewed cautiously because it occurred against a background of many checklist reviews that did not alter the initial diagnosis and many that led to further testing with negative results.

Cognitive forcing checklists for specific diseases

Checklists can serve as cognitive forcing functions—critical elements in the execution of a process to ensure that a correct procedure is followed, or to prevent an untoward event. For example, a customer using an automatic teller machine cannot withdraw cash until the card is removed. Thus, the error of leaving the card in the machine is avoided. If the checklist is always built into diagnostic thinking, then it becomes a forcing function—the final diagnosis cannot be made until the checklist has been reviewed. Cognitive forcing can be generic or specific. In the generic sense, an overarching planning principle is applied (List 1). For example, the “ROWS” (rule out worst-case scenario) strategy ensures that the worst possibilities always receive consideration. In the specific sense, checklists may help avoid predictable pitfalls for particular diseases (List 3). Although errors of commission are typically more visible.
List 3
Example of a Disease-Specific Cognitive Forcing Checklist

Ankle injury
• Differential diagnosis
  o Ankle sprain
  o Delayed onset muscle soreness
  o Achilles tendon injury (partial or complete) and tendinitis
  o Ankle or foot fracture
  o Acute gout
  o Peroneal tendon syndromes (tendinitis, subluxation, tears)

• Forcing functions
  o Assess for neurovascular compromise (cold foot or paresthesia)
  o Consider stress films for ankle stability
  o Ankle and foot X-rays if indicated
  o Anterior drawer test
  o Talar tilt test
  o Squeeze test
  o Thompson test
  o Peroneal tendon stability test

• Pitfalls
  o Missed neurovascular injury (suspect if cold foot or paresthesias)
  o Underappreciated ankle instability
  o Missed associated fracture (especially navicular or metatarsal stress fracture)
  o Missed Maisonneuve fracture (proximal fibula)
  o Missed Achilles tendon rupture (partial or complete)
  o Missed complex regional pain syndromes
  o Missed peroneal tendon syndromes

and detectable than errors of omission, the latter tend to predominate,22,50 and forcing strategies will inevitably focus on them.

Further Considerations and Cautions

Previous investigators have proposed checklists as a concept that might reduce diagnostic errors.10,21,22 To move this concept forward, we developed three kinds of checklists, which we have used in our own practices. Each checklist has a different function, and each requires further development and evaluation.

Related studies
Diagnostic support tools include practice guidelines, clinical algorithms, differential diagnosis textbooks, and computerized decision support. However, most evidence-based guidelines address treatment rather than diagnosis. Diagnostic algorithms help physicians make rapid testing decisions, but they usually do not provide comprehensive differential diagnoses. Differential diagnosis textbooks contain more than simple lists, and their purpose goes beyond simple prompting. Commercial decision support tools, such as Isabel (www.isabelhealthcare.com) and Problem-Knowledge Couplers (www.pkc.com), use patient-specific data to provide patient-specific differential diagnoses.6 These tools seem superior to generic checklists because they narrow the list of diagnoses to those that are most likely for a particular patient.

However, decision support systems have not been widely adopted in practice,52,53 because they suffer from an inadequate knowledge base,6 they can be difficult to incorporate into the workflow,6,54,55 and their ability to improve diagnostic performance is promising but still being evaluated.6,57,58

Other interventions similar to checklists include chart reminders,59 preventive care prompts,60 medical record templates,61 and mnemonic devices.62 These interventions have various purposes, formats, and organizational structures that differ from diagnostic checklists.

Limitations of checklists
Recent success in adapting preflight-style checklists to medical procedures has received justifiable interest,14,15 but checklists for diagnosis may be a “bridge too far.” The analogy between actionable procedures in aviation and cognitive procedures in diagnosis is not tight. Thoughts are less tangible than actions, and it is more difficult to determine whether they have been completed. In both medical and nonmedical settings, checklists are read aloud by teams rather than silently by individuals.11 But diagnosis is usually silent, lonely work, and a natural pause point to review the checklist, such as before takeoff or before incision, does not exist in diagnosis, which can stretch over hours, days, or even months.

Diagnostic checklists have not been formally tested in practice to determine whether they are beneficial. The checklists in this article were not derived using rigorous or reproducible methods, and we are not promoting them for wider use before further revision based on rigorous methods. Instead we are promoting the need to study and test checklists as a potential method for preventing diagnostic errors. Checklists of actionable procedures might have enough face validity to make such testing unnecessary or even unethical.60,63 For example, Balas and colleagues asked the ethics of allowing patients to participate in a usual-care arm (i.e., no safety intervention) in clinical trials of safety innovations.60 Similarly, airline pilots did not formally test their checklists before adopting them. Instead, they learned from their mistakes and made thousands of incremental changes to prevent them.61 However, diagnostic checklists may have a greater potential for harm than preflight or surgical checklists. For example, they could lead to excessive consultation or needless testing (although most serious errors result from doing too little rather than too much).61

For most patients, diagnostic checklists seem unnecessary. Preflight checklists also seem unnecessary in most cases because experienced pilots could recite them from memory. But pilots have learned not to rely on their memories. In contrast, physicians value superior recall and shoot-from-the-hip decisions more than mental crutches, reflective thought, or disciplined task performance. Diagnostic expertise defines the medical profession. But as Donald Berwick said, “Genius diagnosticians make great stories, but they don’t make great health care.”64 Checklists were not adopted without struggle in operating rooms, intensive care units, or even airplanes.

Checklists could produce a false sense of reassurance that leads to complacency, evades the cognitive work required to make a correct diagnosis, neglects patient-specific factors, and obscures aspects of care unrelated to diagnosis. Similar concerns were raised with clinical algorithms. It was feared that physicians would rigidly follow algorithms without accounting for individual patient differences, but investigators found few data to support these concerns when algorithms were studied in practice.65,66

The key to reducing diagnostic errors may be less tied to checklists than to a diagnostic time-out—a brief pause to reflect on our diagnostic reasoning and affective state. But rather than unfocused
attempts to think harder or recognize a distracting mood, we could review a diagnostic checklist and document this procedure in the medical record: During an active diagnostic time-out, I reviewed each item in the general checklist and considered each item in the chest pain checklist. I considered but rejected pulmonary embolus because I judged the risk of harm from excessive testing and pursuit of false-positive results to be greater than the risk of missing that diagnosis in this patient.

Conclusions

We should ask many questions before adopting diagnostic checklists: (1) Will they prevent diagnostic errors? Could they do more harm than good? (2) What is the optimal content and organization for checklists? (3) Who should review the checklist: physicians, nurses, patients, family members, dedicated staff? (4) Will checklists be valued or even accepted by busy physicians? How should they be assimilated into the workflow? (5) How should checklists be presented: card in the pocket, poster on the wall, computer on the desk, or computer in the pocket? (6) When should checklists be reviewed—before, during, after, or remote from the patient encounter? (7) Should we use checklists routinely or selectively? If selectively, what should trigger their use?

Checklists are mandatory for pilots. Should they be mandatory for physicians? Diagnostic errors are common enough that mandatory checklists might be reasonable if they can be shown to work. Pilots do not have the option of skipping their checklists when the risk is low (sunny day, familiar airport, experienced crew). However, any recommendation to physicians to “use this checklist exactly when you think you don’t need it” will likely be met with skepticism. It would be tempting to use checklists only when we lack confidence in our diagnoses, but confidence is a poor predictor of diagnostic accuracy. Future studies might identify “red flags” that should prompt a time-out and checklist review. Generous red flags might include failure to respond to initial treatment, second visit to the emergency department for the same problem, or presenting symptoms that are commonly associated with diagnostic errors. Complaint-specific red flags for headache might include “thunderclap” headache, “worst-ever” headache, and stiff neck.

Most missed diagnoses result from our failure to consider the correct diagnosis as a possibility. Checklists could potentially help us avoid this and other common errors that lead to missed diagnoses. We should feel a sense of urgency to explore this potential in practice because harmful diagnostic errors are common, and they are commonly preventable.

Acknowledgments: The authors are indebted to Amy Miranda, Grace Garey, Mary-Lou Glazer, and Wendy Isser for their expert administrative and bibliographic support.

Funding/Support: None.

Other disclosures: None.

Ethical approval: Not applicable.

References

24. Croskerry P. Context is everything or how could I have been that stupid? Healthc Q. 2009;12 Spec No Patient:e171–e176.