Use of health information technology to reduce diagnostic errors

Robert El-Kareh,1,2 Omar Hasan,3 Gordon D Schiff4,5

ABSTRACT
Background  Health information technology (HIT) systems have the potential to reduce delayed, missed or incorrect diagnoses. We describe and classify the current state of diagnostic HIT and identify future research directions.

Methods  A multi-pronged literature search was conducted using PubMed, Web of Science, backwards and forwards reference searches and contributions from domain experts. We included HIT systems evaluated in clinical and experimental settings as well as previous reviews, and excluded radiology computer-aided diagnosis, monitor alerts and alarms, and studies focused on disease staging and prognosis. Articles were organised within a conceptual framework of the diagnostic process and areas requiring further investigation were identified.

Results  HIT approaches, tools and algorithms were identified and organised into 10 categories related to those assisting: (1) information gathering; (2) information organisation and display; (3) differential diagnosis generation; (4) weighing of diagnoses; (5) generation of diagnostic plan; (6) access to diagnostic reference information; (7) facilitating follow-up; (8) screening for early detection in asymptomatic patients; (9) collaborative diagnosis; and (10) facilitating diagnostic feedback to clinicians. We found many studies characterising potential interventions, but relatively few evaluating the interventions in actual clinical settings and even fewer demonstrating clinical impact.

Conclusions  Diagnostic HIT research is still in its early stages with few demonstrations of measurable clinical impact. Future efforts need to focus on: (1) improving methods and criteria for measurement of the diagnostic process using electronic data; (2) better usability and interfaces in electronic health records; (3) more meaningful incorporation of evidence-based diagnostic protocols within clinical workflows; and (4) systematic feedback of diagnostic performance.

INTRODUCTION

Unaided clinicians often make diagnostic errors. Vulnerable to fallible human memory, variable disease presentation, clinical processes plagued by communication lapses, and a series of well-documented ‘heuristics’, biases and disease-specific pitfalls, ensuring reliable and timely diagnosis represents a major challenge.1–3 Health information technology (HIT) tools and systems have the potential to enable physicians to overcome—or at least minimise—these human limitations.

Despite substantial progress during the 1970s and 1980s in modelling and simulating the diagnostic process, the impact of these systems remains limited. A historic 1970 article4 predicted that, by 2000, computer-aided diagnosis would have ‘an entirely new role in medicine, acting as a powerful extension of the physician’s intellect’.5 Revisiting this prediction in 1987, the authors conceded that it was highly unlikely this goal would be achieved and that ‘except in extremely narrow clinical domains (using computers for diagnosis) was of little or no practical value’.5 In 1990 Miller and Masarie noted that a fundamental issue with many of these systems was that they were based on a ‘Greek Oracle’ paradigm whereby clinical information was provided to the computer with the expectation that it will somehow magically provide the diagnosis.6 They suggested that a more useful approach would be to use computer systems as ‘catalysts’ to enable physicians to overcome hurdles in the diagnostic process rather than have the system become the diagnostician itself.

To understand and summarise how diagnostic accuracy can be enhanced, one needs a conceptual framework to organise HIT tools and their potential applications.
as ‘catalysts’ to known hurdles in the diagnostic process. Our objectives were to develop one such conceptual framework based on a review of published evidence and recent examples of HIT tools that have been used to improve diagnosis and to highlight particular areas in need of future research.

BACKGROUND

Early leaders in computer-aided diagnosis developed statistical methods8 9 and models9 10 to serve as underpinnings for diagnostic systems. Shortliffe and colleagues skillfully organised these approaches into categories including: clinical algorithms, databank analysis, mathematical modelling of physical processes, statistical pattern recognition, Bayesian techniques, decision theory approaches and symbolic reasoning.11 Additional summaries and categorisations of the various possible approaches are also well-described in other reviews.12–14 Several applications emerged to tackle medical diagnosis in a variety of contexts, including Present Illness Program (PIP),15 MYCIN,16 INTERNIST-1/Quick Medical Reference (QMR),17 18 Iliad,19 DXplain20 and several others. These pioneering efforts provided a foundation upon which much of the current work on diagnostic systems.

We describe recent contributions to the field, building upon the work and context provided by prior reviews of computerised diagnostic systems. In 1994 Miller summarised the work of diagnostic decision support21 and suggested that focused diagnostic systems such as those for ECG or arterial blood gas analysis were likely to proliferate. In order for more general diagnostic systems to succeed, he identified key steps which included: (1) development and maintenance of comprehensive medical databases; (2) better integration with HIT to avoid extensive data entry; and (3) improved user interfaces. Three subsequent reviews of computerised decision support22–24 identified a relatively small number of studies of diagnostic systems with only a handful showing improvement in clinician performance and only one demonstrating improved patient outcomes.25

METHODS

Article selection

We initially searched for studies related to diagnostic decision support systems and diagnosis-related HIT published since 2000 (see search strategy in online supplementary appendix). Because we found only modest advances during this time, we broadened the search to include some important work from earlier decades, largely obtained from previous reviews of computer-aided diagnosis.

Taxonomy development, data extraction and categorisation

We adapted models of the diagnostic process from Schiff et al.,1 26 Croskerry27 and Klein28 to create a model for categorising steps in the diagnostic process addressed by HIT and similar tools (figure 1) and linked each step with categories from the Diagnosis Error Evaluation and Research (DEER) taxonomy (figure 2).1 26 Based on this model, we created a condensed set of categories describing different steps or aspects of diagnosis targeted by HIT tools (box 1). During data abstraction, each study was linked to one or more of these categories.

We developed a customised data extraction form using Microsoft Access 2010. Following in-depth review, we determined the following information for each study: (1) whether the study met our inclusion criteria; (2) clinical problem/question addressed; (3) type of HIT system described; (4) whether it was evaluated in a clinical setting; (5) target of the HIT intervention/tool; (6) duration/sample size of the study; (7) study outcomes; and (8) results.

RESULTS

We summarised the main types of diagnostic HIT tools and mapped each type to steps in the diagnostic process that it currently or potentially targets (figure 3). Below we provide details of our findings in the 10 categories of interventions.

Tools that assist in information gathering

The value of a high-quality history and physical examination is well-recognised,29–31 but time pressures and reliance on clinician memory pose a major barrier to their performance. Beginning in the 1960s, various systems have been devised to assist history-taking through computer-based patient interviewing.32–34 Interestingly, these were mainly reported before the timeframe of our review, suggesting a loss of research interest for unclear reasons.35 Several recent studies have examined automated patient interviewing in specialised settings including home,36 emergency department waiting rooms37 38 and online visits in primary care.39 One study found that physician-acquired history and computer-based systems each elicited important information that the other missed,40 reaffirming the role of technology in complementing rather than replacing the physician-acquired history. To augment the clinician’s physical examination there have been systems designed to support interpretation of auscultation, both cardiac41–43 and pulmonary.44

The state of this research also remains underdeveloped with a paucity of recent or rigorous studies.

Cognition facilitation by enhanced organisation and display of information

The increasing volume of electronically available patient information creates significant challenges and necessitates tools to enable efficient review of patient information and pattern recognition. One logical direction to pursue is the graphical representation of numerical data.45 One usability study found that graphical
laboratory value displays led to reduced review times and that graphical and tabular representations were each more effective for answering different clinical questions. However, different clinical settings may benefit from differing data summary formats. For example, in a neonatal ICU, automatically generated textual summaries supported decision-making as well as graphical representations but not as well as their...
<table>
<thead>
<tr>
<th>Where in Diagnostic Process</th>
<th>What Went Wrong</th>
</tr>
</thead>
<tbody>
<tr>
<td>(~Anatomic localization)</td>
<td>(~Lesion)</td>
</tr>
<tr>
<td>1. Access/Presentation</td>
<td>A Failure/Delay in presentation</td>
</tr>
<tr>
<td></td>
<td>B Failure/denied care access</td>
</tr>
<tr>
<td></td>
<td>C Failure/delay in eliciting critical piece of history data</td>
</tr>
<tr>
<td>2. History</td>
<td>A Inaccurate/miss interpretation</td>
</tr>
<tr>
<td></td>
<td>B Failure in weighing</td>
</tr>
<tr>
<td></td>
<td>C Failure to follow-up</td>
</tr>
<tr>
<td>3. Physical Exam</td>
<td>A Failure/delay in eliciting critical physical exam finding</td>
</tr>
<tr>
<td></td>
<td>B Inaccurate/miss interpreted</td>
</tr>
<tr>
<td></td>
<td>C Failure in weighing</td>
</tr>
<tr>
<td></td>
<td>D Failure to follow-up</td>
</tr>
<tr>
<td>4. Tests (Lab/Radiology)</td>
<td>A Ordering</td>
</tr>
<tr>
<td></td>
<td>B Failure/delay in ordering needed test(s)</td>
</tr>
<tr>
<td></td>
<td>C Failure/delay in performing ordered test(s)</td>
</tr>
<tr>
<td></td>
<td>D Error in test sequencing</td>
</tr>
<tr>
<td></td>
<td>E Ordering of wrong test(s)</td>
</tr>
<tr>
<td></td>
<td>F Test ordered wrong way</td>
</tr>
<tr>
<td></td>
<td>G Sample mixup/mislabeled (eg wrong patient/test)</td>
</tr>
<tr>
<td></td>
<td>H Technical errors/poor processing of specimen/test</td>
</tr>
<tr>
<td></td>
<td>I Erroraneous lab/radiol reading of test</td>
</tr>
<tr>
<td></td>
<td>J Failed/delayed reporting of result to clinician</td>
</tr>
<tr>
<td></td>
<td>K Failed/delayed follow-up of (abnl) test result</td>
</tr>
<tr>
<td></td>
<td>L Error in clinician interpretation of test</td>
</tr>
<tr>
<td>5. Assessment</td>
<td>A Hypothesis Generation</td>
</tr>
<tr>
<td></td>
<td>B Suboptimal weighing/prioritizing</td>
</tr>
<tr>
<td></td>
<td>C Recognizing Urgency/Complications</td>
</tr>
<tr>
<td></td>
<td>D Failure to recognize/weight urgency</td>
</tr>
<tr>
<td></td>
<td>E Failure to recognize/weight complication(s)</td>
</tr>
<tr>
<td>6. Referral/Consultation</td>
<td>A Failure/Delay in ordering referral</td>
</tr>
<tr>
<td></td>
<td>B Failure/delay obtaining/scheduling ordered referral</td>
</tr>
<tr>
<td></td>
<td>C Error in diagnostic consultation performance</td>
</tr>
<tr>
<td></td>
<td>D Failed/delayed communication/follow-up of consultation</td>
</tr>
<tr>
<td>7. Follow-up</td>
<td>A Failure to refer patient to close/safe setting/monitoring</td>
</tr>
<tr>
<td></td>
<td>B Failure/delay in timely follow-up/rechecking of patient</td>
</tr>
</tbody>
</table>

Figure 2  Diagnosis Error Evaluation and Research (DEER) taxonomy.

human-generated counterparts. Overall, improved organisation and display of data might facilitate identification of temporal patterns as well as helping to ensure that items do not get overlooked, especially to offset electronic health record (EHR) data hyper-trophy, but the evidence base to date is quite limited.

Aids to generation of a differential diagnosis
One repeatedly demonstrated contributor to diagnostic errors is the lack of a sufficiently broad differential diagnosis. One suggested approach to support this process is to provide diagnostic checklists with common, ‘don’t miss’ or commonly missed diagnoses for various presenting symptoms and signs. This approach can be facilitated with computer-based differential diagnosis list generators. While work in this area has spanned decades, we focus on recent additions to the field. Four systems in current use (Isabel, DXplain, Diagnosis Pro and PEPID) were recently reviewed and evaluated on test cases. There have
also been various evaluations of these systems and earlier counterparts (e.g., QMR and Illiad) including retrospective\textsuperscript{52}–\textsuperscript{58} and simulated cases\textsuperscript{53} \textsuperscript{59} \textsuperscript{60} as well as pre–post\textsuperscript{61} and prospective\textsuperscript{62} studies. In general, these studies—although not always rigorously performed—demonstrate that the systems include the gold standard diagnosis within the output list of up to 30 diagnoses in 70–95% of cases. Whether undifferentiated lists of this kind are clinically helpful requires further evidence. One study found that using such a system led to a similar number of diagnoses changed from correct to incorrect as from incorrect to correct.\textsuperscript{58}

**Tools and calculators to assist in weighing diagnoses**

Once a differential diagnosis is generated, weighing the likelihood of candidate diagnoses is subject to various challenges and cognitive pitfalls.\textsuperscript{49} \textsuperscript{63} Several of the differential diagnosis generators described above provide rankings of their diagnostic suggestions.\textsuperscript{51} \textsuperscript{64} \textsuperscript{65} Another more quantitative approach is the use of ‘clinical prediction rules’, which are scoring systems to calculate the likelihood of diagnoses based on sets of clinical symptoms, signs or test results.\textsuperscript{66} \textsuperscript{67} Examples that have been recently evaluated in clinical settings include prediction rules for pulmonary embolism,\textsuperscript{68} \textsuperscript{69} deep vein thrombosis,\textsuperscript{70} paediatric appendicitis,\textsuperscript{71} meningitis,\textsuperscript{72}–\textsuperscript{74} cervical spinal injury,\textsuperscript{75} \textsuperscript{76} intra-abdominal injury after blunt trauma\textsuperscript{77} and osteoporosis.\textsuperscript{78}

EHRs can embed algorithms into the workflow to determine whether one condition is present or select one diagnosis from a small predetermined set of potential candidates. Examples of such embedded algorithms have evaluated patients for pneumonia,\textsuperscript{79} \textsuperscript{80} acute myocardial infarction,\textsuperscript{81} postoperative infections\textsuperscript{82}
and, in one broad effort, to diagnose general paediatric patients with one of 18 potential conditions. Although several systems showed promising results, the acute myocardial infarction system did not impact decision-making in the emergency department in a pre–post evaluation, and we were unable to find evaluations of the use of the other systems in clinical care.

Support for intelligent selection of diagnostic tests/plan
Diagnostic protocols can facilitate evidenced-based diagnostic strategies. Often they can be embedded or integrated into various electronic tools. One well-designed study found that a handheld diagnostic algorithm of evaluation of suspected pulmonary embolism with integrated clinical decision support improved the appropriateness of investigations, and a validation study of a chest pain protocol confirmed the safety of referring patients with low-risk chest pain to outpatient stress testing. We also found examples of protocols without impact, including a cluster randomised controlled trial of a protocol for evaluating skin lesions using instant cameras in primary care which failed to improve the proportion of benign lesions excised.

One targeted use of such algorithms and embedded electronic clinical decision support is to use the order entry function of EHRs to improve the appropriateness of diagnostic tests although, when tested, it failed to demonstrate an impact on the proportion of radiology tests with positive findings or improve patient outcomes.

One group designed a ‘Smart Form’ for acute respiratory illnesses to standardise and harness clinical documentation and integrate it with diagnostic decision support. Usage of the system was low, with resulting minimal impact on diagnostic decision-making or antibiotic prescription appropriateness.

Enhanced access to diagnostic reference information and guidelines
Simply providing access and time to review a medical textbook can support a diagnostician by avoiding exclusive reliance on memory. Various electronic approaches and products aim to support timely access to context-specific information, and these can be active (requiring the user to look up information) or more passive (information is automatically pushed to the user). One popular approach to make relevant reference information readily available is the ‘infobutton.’ This functionality provides context-specific links from clinical systems to reference systems and is often designed to anticipate clinicians’ information needs. Infobuttons have the potential to provide diagnosis-specific information without requiring clinicians to exit the EHR to perform a separate search. However, studies of infobuttons to date have focused mainly on medications, with little published evidence on how they might support the diagnostic process.

Tools to facilitate reliable follow-up, assessment of patient course and response
Patient follow-up and assessment of response over time is often a crucial part of ensuring an accurate diagnosis. An important related issue is follow-up of test results, especially those with long or variable turnaround times (eg, microbiology tests, pathology results, ‘send-out’ tests). Other studies have used tools to facilitate longitudinal automated assessments of asthma symptoms, visualise imaging for neuro-oncology patients over time, and an interactive voice response (IVR) system integrated into EHRs to provide systematic follow-up of walk-in clinic patients to screen for misdiagnoses.

To help improve the reliability of follow-up of the high volume of test results, electronic result managers have been created—both comprehensive systems as well as test-specific systems such as tools related to cancer screening or follow-up. Other approaches target test result follow-up for specific high-risk scenarios such as microbiology cultures pending at the time of discharge from the hospital or automatic gastroenterology consultations for positive faecal occult blood tests. When evaluated, these systems often showed improvements in process measures, although they have been insufficiently powered to show impact on clinical outcomes.

Tools/alerts that support screening for early detection of disease in asymptomatic patients
An important aspect of timely diagnosis is early disease detection via screening of appropriate populations, for which there is an extensive literature.

Here we highlight illustrative examples. One approach involves generation of unsolicited alerts informing providers of recommended or overdue screening tests. Studies have evaluated alerts designed to screen for a wide range of conditions including cancer, osteoporosis, diabetes, overdue vaccinations and others.

When studied in clinical settings, these alerts often show statistical improvements in provider performance. However, improvements are often surprisingly modest (typically 3–15% absolute improvement in screening rates). In addition to alerts targeting individual providers, population management informatics tools (eg, panel managers that list and facilitate contacting overdue patients) have been shown to be moderately effective in improving diagnostic screening rates.

Tools to facilitate diagnostic collaboration, particularly with specialists
Just as instantaneous access to information and reference resources is likely to improve diagnosis, timely expert consultations can support diagnosis quality. Driven mainly by desires to support more remote/rural clinicians in obtaining consultations, ‘telemedicine’ specialty consultation systems have been widely deployed and tested. Given the expanding
numbers of articles, we cite excellent reviews rather than detail individual published studies. An entire journal is now devoted to this approach, featuring uses such as tele-dermatology, tele-radiology and tele-pathology. The objectives are not necessarily to improve a specialist’s diagnosis but to achieve comparable accuracy for remote patients. ‘Store-and-forward’ (asynchronous) and real-time consultation technology have been reported to result in a more timely diagnosis for patients than a conventional referral process. An exciting, largely untapped potential for diagnostic support is facilitated collaboration and coordination among different members of the care team, including patients and their families, for facilitated access for concerning symptoms and collaborative diagnostic decision-making.

**Systems that facilitate feedback and insight into diagnostic performance**

Systematic provision of feedback (immediate or longer term) to individual providers (or organisations) represents a powerful potential for improving diagnosis. For generations, autopsies and/or ‘second opinions’ have been used for this purpose in selected patients. Automating systematic feedback, despite its great potential, is mostly non-existent, making current medical practice largely an ‘open loop’. While several examples of decision support to facilitate feedback of clinician diagnostic performance have been demonstrated in actual clinical practice due in part to well-described barriers common to decision support systems, rarely have studies been performed, benefits shown in retrospective, simulated or controlled environments have barely been demonstrated in actual clinical practice due in part to well-described barriers common to decision support systems in general. We found limited evidence to support diagnostic protocols to guide investigations and alerts and panel management tools to improve performance of screening tests. However, for the majority of the categories of HIT tools, the evidence base was too scant to determine their utility in clinical settings.

We believe that the field of diagnostic HIT research can move forward by focusing on a few areas. First, we need to develop the electronic ‘yardstick’ to measure the accuracy of the diagnostic process. Improved measurement will enable both targeted decision support as well as more robust and useful feedback to clinicians. Ideally, this needs to be done in a way that is well-integrated into the clinical workflow rather than requiring extensive manual data collection. Second, we should expand collaboration with cognitive science and human/computer interaction experts to improve the structure and interfaces of EHRs. Design and implementation of enhancements will need to be done thoughtfully to become useful in everyday practice. Third, there is an urgent need to integrate evidence-based diagnostic investigations more effectively into computerised order entry systems. The challenge is to create diagnostic protocols with enough flexibility to allow clinicians to exercise their clinical judgement but to avoid unnecessary or suboptimal diagnostic strategies as well as over-alerting. Fourth, support for systematic feedback of diagnostic performance is underdeveloped and warrants more attention. As this field evolves, evaluations of diagnostic HIT tools should assess the strength of the evidence behind them. We propose a five-level hierarchy based on the model of Fryback and Lusted as a way to approach such critical and evidence-based assessments (box 2).

**DISCUSSION**

The goals of this review were to provide an overview of the current state of diagnostic HIT tools and systems and to outline a conceptual framework that can serve to suggest areas for further exploration. We adapted prior models of the diagnostic process and reviewed the published literature to create a map showing steps of the diagnostic process targeted by each group of tools. Through this iterative process, we identified areas with gaps in evidence as well as common themes to guide future work.

Overall, we found that progress in diagnostic HIT has been slow and incremental with few significant ‘game-changing’ approaches emerging in the last decade. While there were representative studies in each of our 10 categories of tools, rigorous studies in clinical settings were very infrequent. When clinical studies were performed, benefits shown in retrospective, simulated or controlled environments have rarely been demonstrated in actual clinical practice due in part to well-described barriers common to decision support systems in general.

**Box 2 Proposed levels of evidence for evaluating diagnostic HIT tools*”**

- **Level I.** Appear useful for suggesting, weighing, or in other ways helping physicians in diagnosis-related tasks (face validity)
- **Level II.** Clinicians (or students) report they like and find helpful in directing them to correct diagnosis in a more timely, reliable, useful way (and ideally, regularly use them).
- **Level III.** Compared to not using these tools (ideally concurrent, or at least historical controls) physicians arrive at the correct diagnosis more often, sooner or more safely.
- **Level IV.** Improved outcomes in patients (ideally randomly assigned) for whom tools are used—fewer errors, more timely diagnosis, or more efficient or cost-efficient diagnostic evaluation process.
- **Level V.** Tools show both improved patient outcomes and produce sufficiently greater marginal benefit to justify investment resources expended (money, clinician time) on the tools vs. other places that those resources could be invested. (ie, ROI).

*Based on model of Fryback and Thornbury.
Our review has several limitations. We focused on recent work, largely excluding studies prior to 2000. While we reviewed a broad representation of tools and systems, we recognize the list was not exhaustive. Although we covered many types of interventions and approaches, we could not cover all because of time and space considerations and excluded various important domains such as computer-aided diagnostic tools for radiology studies, alarms and alerts built into monitoring equipment and support tools targeting patients and non-clinicians.

In conclusion, we found that the field of diagnostic health information technology is still in its early stages and there has been minimal development over the past decade in various promising realms. Many aspects of the diagnostic process have been targeted, but few tools and systems have been shown to improve diagnosis in actual clinical settings. We can move the field forward by developing and testing interventions in real-world settings using cross-disciplinary research and systematic feedback of diagnostic performance.

Contributors All three authors made significant contributions to the conception, design, drafting and revision of the manuscript, as well as providing final approval of the version to be published.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non-Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/

REFERENCES
10 Spitzer RL, Endicott J. Diagnosis—a computer program for psychiatric diagnosis utilizing differential diagnostic procedure. Arch Gen Psychiatry 1968;18:746.
105 Samson D, Schoelles KM. Chapter 2: Medical tests guidance (2) developing the topic and structuring systematic reviews of


Berner ES. Closing the Feedback Loop to Improve Diagnostic Quality: AHRQ, 2010.


