Comparative outcome studies of clinical decision support software: limitations to the practice of evidence-based system acquisition

Gaurav Jay Dhiman1, Kyle T Amber2, Kenneth W. Goodman3

ABSTRACT

Clinical decision support systems (CDSSs) assist clinicians with patient diagnosis and treatment. However, inadequate attention has been paid to the process of selecting and buying systems. The diversity of CDSSs, coupled with research obstacles, marketplace limitations, and legal impediments, has thwarted comparative outcome studies and reduced the availability of reliable information and advice for purchasers. We review these limitations and recommend several comparative studies, which were conducted in phases; studies conducted in phases and focused on limited outcomes of safety, efficacy, and implementation in varied clinical settings. Additionally, we recommend the increased availability of guidance tools to assist purchasers with evidence-based purchases. Transparency is necessary in purchasers’ reporting of system defects and vendors’ disclosure of marketing conflicts of interest to support methodologically sound studies. Taken together, these measures can foster the evolution of evidence-based tools that, in turn, will enable and empower system purchasers to make wise choices and improve the care of patients.

Key words: clinical decision support systems, comparative study, medical ethics, medical economics, marketing

INTRODUCTION

Clinical decision support systems (CDSSs) aid clinicians with decision-making tasks, including patient diagnosis and treatment.1 They are conducted either on or off site via personal computers, the Internet, or handheld devices, or as components of electronic health record (EHR) systems. As components of health information technology (HIT),2 certain CDSSs provide reference material, drug interaction alerts (DIAs), medical calculators, or clinical diagnoses. Others create preventive screening reminders, deliver guidelines for treatment options, or improve communication and recordkeeping.3 Table 1 itemizes different CDSS capabilities.

Programs that emphasize clinical treatment and outcomes (i.e., preventive care, diagnosis, and treatment plan programs) rather than providing reference material or assisting with institutional efficiency will be designated as preventive, diagnostic, and treatment (PDT) CDSSs. With a large selection of products, prospective users are faced with a significant challenge when deciding among PDT programs for purchase. Although past normative analyses have addressed regulation of CDSSs,6 alert fatigue,7 and drawbacks to increased liability for users,8 insufficient attention has been paid to the consumer decision process behind selecting a PDT-CDSS. We will examine decision-making in purchasing PDT-CDSS; discuss research, marketplace, and legal limitations restricting the body of knowledge for purchasers; and provide recommendations for more realistic and meaningful comparative outcome studies.

PDT-CDSS Purchasing Decision-Making

Recognizing the potential benefits and pitfalls of such programs, the first and perhaps most basic question that purchasers ask when deciding among PDT-CDSSs is “What do we expect from our software?” Thus, purchasers must evaluate their needs and expectations for functional capabilities.9 Understandably, committing to a single program, which often must be adopted into a larger, preexisting information workflow,10 requires considerable capital and time, and represents a significant long-term investment. As such, making a proper decision requires various individual and institutional considerations to be defined and weighed. Certain academic sources9,11–14 explicitly detail some of these considerations.
Table 1: Summary of Major CDSS Capabilities

<table>
<thead>
<tr>
<th>Purpose of CDSS</th>
<th>Examples of Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institutional efficiency</td>
<td>Order sets that organize physician directions and develop individualized stay and treatment plans</td>
</tr>
<tr>
<td>Healthcare costs</td>
<td>Duplicate testing and drug availability notifications</td>
</tr>
<tr>
<td>Preventive care</td>
<td>Screening, immunization, and disease management suggestions</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Lists of ranked differential diagnoses</td>
</tr>
<tr>
<td>Treatment plans</td>
<td>Treatment guidelines, drug dosing recommendations, and DIAs</td>
</tr>
<tr>
<td>Reference</td>
<td>Searchable clinical information catalogues</td>
</tr>
</tbody>
</table>

Source: Berner. These are the foci of this article.

Some consumer websites, such as Healthcare Information and Management Systems Society, Leapfrog Group, Health Level Seven International, ECRI Institute, and InformationWeek HealthCare, provide guidelines to help individuals and institutions define their software needs and expectations when making purchases. Although such resources for purchaser education remain available, their often-general foci and missions may not adequately account for very real differences among specific systems.

Another important question for consumers is “What can each program do?” CDSSs do not necessarily provide the same capabilities, and the same data may be used for different applications. Several authors have assessed the leading CDSSs for technical and performance capabilities. Sittig et al. likewise assessed the clinical decision capacities of commercial EHR programs. Although this limited literature can provide some evidence-based information regarding program capabilities for potential consumers, it does not adequately meet the needs of an informed consumer. As Rosshayn and colleagues noted, important concerns for clinician consumers, such as “cost, user satisfaction, system interface and feature sets, unique design and deployment characteristics, and effects on user workflow,” were not frequently studied.

Still, the basic usability of a program may be a large concern for purchasers. A PDT-CDSS may remain difficult to learn and use, or may engender user input or system errors. It may create excessive alert notifications and provide unclear or periphrastic directions. In some cases, within institutions, disagreements may arise over dosage limits, order sets, and even alert language. Many concerns ultimately center on fostering extraneous and therefore unproductive effort while delaying healthcare treatment, making some purchasers wary of PDT-CDSSs.

Program modifiability may also be a concern. Clinicians and hospitals must determine whether to opt for a commercially produced system or build a customized one, how to smoothly implement their choice, and which metrics should be used to assess successful implementation. Users may wish to modify a commercial program’s settings according to personal or institutional preferences, but this may be difficult or time-consuming. This usually (but infrequently) applies to medication programs that provide DIAs, specify drug contraindications for allergies, and track formulary options.

Purchasers should also care about the safety and efficacy of a PDT-CDSS. Metzger et al. found that only 44% of harmful medication orders entered by physicians were later detected by CDSS in different hospitals, suggesting that systems might not adequately detect errors. However, in certain cases, this stems from entering incorrect patient values, including age and weight. Other studies suggest that PDT-CDSS-mediated harm to patients often stems from implementation problems rather than intrinsic system flaws.

Purchasers particularly weigh a CDSS in comparison to others. Despite some examples of comparative studies, there is no comprehensive body of evidence. Data from such studies are essential because purchasers remain wary of the potential for harm to patients and potential liability, even when acting in good faith. We have found few outcome-oriented studies that comprehensively evaluate CDSSs. For instance, some small yet ambitious studies explore the effect of program usage on clinician decision-making. Relatively few studies of CDSSs are randomized controlled trials, and most focus on the effect of the CDSS on decision making rather than the outcome. The literature on the safety and efficacy of a single program versus the null condition (i.e., having no program) is extensive. Some studies review multiple programs against the null condition, and most focus on the effect of the CDSS on decision making rather than the outcome. The literature on safety and efficacy of a single program versus the null condition is extensive. Some studies review multiple programs against the null condition, and most focus on the effect of the CDSS on decision making rather than the outcome. The literature on safety and efficacy of a single program versus the null condition is extensive.

Other studies review multiple programs against the null condition. Few studies, however, compare several programs against one another. In an analysis by Berner et al., a panel of expert clinicians compared the accuracy of four programs in producing diagnoses or differentials for cases based on actual patients. Bright et al. systematically reviewed randomized trials of 148 CDSSs to determine the frequency of studies focusing on the effect of CDSSs on the healthcare process, workflow, and cost. They found that 76% of studies measured against the null condition instead of a specific comparator, 86% evaluated process of care, and only 20% evaluated clinical outcomes, with even fewer assessing adverse outcomes or unintended events.
It is not for lack of interest in studying outcomes that adequate, useful data are lacking. There are numerous barriers to such research, and they can hinder comparative outcome research.

Purchaser Barriers
Obstacles to Comparative CDSS Studies
It is important to understand that comparative CDSS outcome studies often include systems that address a disparate variety of medical conditions and clinical tasks:

a. Many systems focus on single medical conditions, and studies with large sample sizes or across several sites are rare.
b. Programs may grow outdated and thus be updated or phased out, and purchasers may modify the settings of programs to suit personal needs.
c. Inpatient and larger academic settings are most frequently studied, making comparison with outpatient or smaller private settings less translatable.

The designs of PDT-CDSS studies limit proper evaluation and comparison of programs for limited outcomes. Because no two programs with generally similar purposes have identical uses, and the capabilities of a single program may be emphasized differently by practice by different clinicians, attempting to run a controlled trial, even a randomized controlled trial for limited outcomes, may control for certain variables but will never allow fully sufficient comparison. Many comparative studies also investigate usage for a single medical condition, utilize small sample sizes, and apply unevenly defined markers of safety and efficacy. This reduces generalizability to other clinical environments when using different parameters. This frequent choice of study design is partially understandable when measuring outcomes for programs intended to address a large array of conditions, but it can prove complicated, slow, and costly.

PDT-CDSSs may become outdated. Future systems will evolve, change, or improve, complicating both prospective and retrospective studies of usage patterns. Whether incorporated in an EHR or as a standalone system, PDT-CDSSs are complex programs and any change will confound attempts to compare them. For this reason, a well-formulated past study might not apply to more recent systems.

The fusion of EHR with CDSS has provided an additional challenge in study design. Many vendors’ EHRs and computerized provider-order entry programs possess CDSS units that users must configure for use in clinical decision support, and certain EHRs are modified by purchasers. Such individual modifications complicate comparative studies because differences in clinical vocabulary, representation of standard laboratory values, outcome variables, and pharmaceutical formularies thwart comparison among individual systems or even sites with the same system.

Finally, even when following a well-designed study, comparative PDT-CDSS studies are limited by their choice of clinical setting. For instance, many studies tend to focus on academic settings with well-established HIT personnel. Smaller, private practices may lack staff with CDSS experience. Thus, the demonstration of a system’s safety or efficacy in a large setting might not be generalizable to smaller practices. Notably, CDSS studies also tend to focus on implementation in inpatient settings. The typical conditions encountered in these settings tend to differ from those in outpatient clinics. Although some studies have described CDSS implementation in multisite, nonacademic locations, this is more the exception than the norm.

Barriers to CDSS Performance Transparency
Challenges related to transparency might explain the difficulty in conducting comparative studies:

a. Although the US government publishes online reports of CDSS adverse outcomes, submissions remain voluntary and rare.
b. Vendors may offer remuneration to previous purchasers for successfully recommending a new purchaser to their product.
c. Others, though not always legally liable for damages, may contractually obligate purchasers from publically disclosing adverse CDSS outcomes.

It is difficult to identify the adverse outcomes of PDT-CDSSs. The US Food and Drug Administration maintains an online database for adverse outcomes of medical devices, including HITs and CDSSs: the Manufacturer and User Facility Device Experience (MAUDE). The MAUDE website provides voluntary reports from 1993 onward, but specifies that the data should not be used in comparative studies or to represent the frequency of adverse outcomes. Of the nearly 900 000 MAUDE reports from January 2008 to July 2010, only 0.1% of the reports involved an HIT incident. Eleven percent of the 436 relevant HIT reports related to patient harm. Only 1% involved a patient death attributable to HIT, but HIT includes more than just CDSSs. MAUDE reporting remains voluntary; notably, system vendors must opt to include their CDSS on MAUDE, limiting the reporting of some programs and underreporting the actual occurrence of negative incidents. Certain research indicates that incident reports do not provide actual frequencies of errors and adverse outcomes, and so will inevitably fail to portray a complete picture. Moreover, incomplete and poor informatics data may frustrate or impede health decision making, both on clinical and public health levels. It follows that larger collections of incident reports are needed to identify and explain errors made by these systems.

Program vendors sometimes play a large role in disseminating product information. Even though there have been academic and government-funded studies in comparative effectiveness, purchasers may come to rely on more subjective and less evidence-based guidance—deferring to colleagues, personal websites, or HIT consultants. However, in some cases, it has been alleged but not publicly confirmed that...
Table 2: Limitations to Performing Comparative Studies

<table>
<thead>
<tr>
<th>Limitations to Performing Comparative Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDSSs not often performing the same task</td>
</tr>
<tr>
<td>Very narrow study designs (e.g., single conditions)</td>
</tr>
<tr>
<td>Inpatient and larger academic settings more frequently studied</td>
</tr>
<tr>
<td>Programs growing outdated</td>
</tr>
<tr>
<td>CDSS modified by individual purchasers</td>
</tr>
</tbody>
</table>

Table 3: Transparency Barriers to Disclosing Information

<table>
<thead>
<tr>
<th>Transparency Barriers to Disclosing Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incomplete and often voluntary government reports of adverse events</td>
</tr>
<tr>
<td>Financial remuneration for purchasers successfully recommending others</td>
</tr>
<tr>
<td>“Hold harmless” clauses in contracts providing vendors with limited liability</td>
</tr>
<tr>
<td>“Gag clauses” preventing public disclosure of CDSS incidents</td>
</tr>
</tbody>
</table>

Certain institutions that purchase EHRs may receive referral fees if they successfully refer another institution to the vendor. Such a vendor–purchaser relationship would constitute a conflict of interest (COI). One way to manage or mitigate the conflict might be to disclose such agreements at the outset, although a strong case can be made that such payments are inherently wrongful.

Further, it has been reported that vendors may attempt to insulate themselves from liability by insisting on contractual provisions for vendor–limited liability; that is, “hold harmless clauses” and disclaimers of warranty. Other clauses limit the disclosure of software glitches, mistakes, and design flaws to anyone but the vendor, including reports in publications; such clauses have been termed “gag clauses.” Whereas “no court has applied product liability standards to computer software” and vendors are encouraged to provide adequate training and warnings to purchasers, vendors and physicians or a hospital may be successfully sued for negligence. This concern has motivated certain vendors to provide overly inclusive DIAs and to prevent system users from modifying emergency levels attached to different notifications, although litigation has not yet hinged on the act of overriding these alerts. One consequence of such efforts to minimize liability is that consumer interest groups, government agencies, academia, and even patients may be blocked from access to information needed to support an informed opinion. At ground, potential PDT-CDSS customers have very limited access to any reliable “consumer reports.”

Recommendations

Recommendations for Guidance Tools

CDSSs have become fixtures in healthcare. Clinicians and institutions must make well-informed decisions before purchasing CDSS systems. Government, nonprofit, and private organizations are encouraged to provide even more guidance tools for potential CDSS purchasers: The Agency for Healthcare Research and Quality’s Health IT Evaluation Toolkit is meant to counsel nonacademic HIT purchasers. However, it provides advice other users on developing thorough plans for appraising purchases by various measures, each chosen for individual needs and expectations. Criteria include clinical outcomes, clinical processes, provider adoption and attitudes, patient adoption, knowledge and attitudes, workflow compact, and financial impact.

Recommendations for Comparative Studies

It is no small challenge to call for comprehensive studies when measuring limited outcomes has already proven difficult (Table 2). Nevertheless, recognition of these limitations should not preclude efforts to conduct further comparative evaluations. Well-informed PDT-CDSS consumers require more and better comparative studies of system safety and efficacy, which should feature varied study designs and clinical settings. At the least, a larger number of independently funded studies focused on limited outcomes are needed, given the lack of financial incentives for software and system comparisons. This is similar to the “small ball” mentality in HIT adoption, based on norms of small ball baseball, in which “narrower studies [are] conducted over the life-cycle of the project,” rather than “randomized experiments conducted only at the project’s conclusion (‘powerball’ studies).” Such studies would be phased in, rather than conducted simultaneously. Phased studies can analyze institutional readiness, normative and quantitative workflow efficiencies, usability, and post-discharge and follow-up audits (e.g., for patient readmissions). This approach allows vendors and users to determine strengths and weaknesses at the end of each phase and inform and guide improvements in subsequent phases.

Recommendations for Transparency

Finally, elimination of barriers to unbiased disclosure is needed to allow for meaningful studies (Table 3). The voluntary nature of submitting adverse CDSS events to government agencies remains tied to larger debates over government regulation of EHRs and CDSSs.

1. Although a previous purchaser may receive remuneration for successfully recommending a new purchaser to a system, proper COI recommending a new purchaser to a system—proper COI disclosure must be required to ensure that purchasers make informed, evidence-based decisions in purchases.

2. Moreover, whereas system developers might be concerned about liability for adverse events, this should never
preclude, discourage, or impede public disclosure, whether to consumer websites or government databases. So-called gag clauses in purchaser-vendor contracts are illicit and should never be condoned. The disclosure of each and every trivial defect may be too cumbersome for companies to report, and so is not necessary; however, adequate disclosure done in good faith should be encouraged, and standards should be developed to support and guide it. This requirement and corresponding guidance will foster improvements in current technology, advance public awareness of patient safety issues, and increase consumer confidence.

Hurdles to transparency must be overcome to allow for the kind of comparative effectiveness studies required to improve system performance and therefore patient care.

CONCLUSION
An appropriate first step in broadening the availability of advice for making CDSS purchases may be increased guidance tools from government, nonprofit, and private organizations. Even then, there remain numerous methodological barriers to conducting comparative outcome studies of CDSS used for prevention, diagnosis, and treatment. Supporting such studies will provide system purchasers with more useful knowledge and lead to improvements in PDT-CDSSs. Questions such as “How is one PDT-CDSS better than another?” and “How can this information be used to improve patient outcomes?” cannot be answered in the absence of such studies. Reliable data and information for conducting comparative studies require proper COI disclosure from vendors who compensate purchasers for referring others. CDSS users must also be allowed to disclose program flaws to government databases and consumer websites. Together, providing more guidance tools, supporting comparative studies, and removing barriers to unbiased disclosure can foster the evolution of evidence-based tools that, in turn, will enable and empower system purchasers to make better decisions and improve the care of patients.

ACKNOWLEDGEMENTS
We would like to acknowledge Dr. Eta Berner, Director of the Center for Health Informatics for Patient Safety/Quality in the School of Health Professions at the University of Alabama at Birmingham, for critiquing and commenting on an earlier version of this article.

CONTRIBUTORS
Mr. Dhiman conceived the concept of this submission, performed a majority of the literature search, and drafted and revised the manuscript. He was actively engaged in review, drafting, and final approval of the manuscript. He is accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. As corresponding author, he takes primary responsibility for communication with the journal during the manuscript submission, peer review, and publication process and is responsible for completing the journal’s administrative requirements. He, as guarantor, also accepts full responsibility for the work and controlled the decision to publish.

Dr. Amber performed literature searches for specific parts of each section, and helped draft and revise the manuscript. He was actively engaged in review, drafting, and final approval of the manuscript. He is accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Dr. Goodman helped conceive the concept of this submission, contributed to the literature search of ethical issues, and helped draft and revise the manuscript. He was actively engaged in review, drafting, and final approval of the manuscript. He is accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

COMPETING INTERESTS
None.

REFERENCES
56. Dixon BE, Granniss SJ. Why “What data are necessary for this project?” and other basic questions are important to address in public health informatics practice and research. J Public Health Inform. 2011;3:1–40.


**AUTHOR AFFILIATIONS**

1. University of Miami Miller School of Medicine, Miami, FL, USA
2. University of Miami Miller School of Medicine, Miami, FL, USA
3. Bioethics Program, University of Miami Miller School of Medicine, Miami, FL, USA