Users' guides to the medical literature
Adrienne G Randolph; R Brian Haynes; Jeremy C Wyatt; Deborah J Cook; Gordon H...
JAMA; Jul 7, 1999; 282, 1; AMA Titles
pg. 67

Users' Guides to the Medical Literature

XVIII. How to Use an Article Evaluating the Clinical Impact of a Computer-Based Clinical Decision Support System

Adrienne G. Randolph, MD, MSc
R. Brian Haynes, MD, PhD
Jeremy C. Wyatt, MD
Deborah J. Cook, MD, MSc
Gordon H. Guyatt, MD, MSc

CLINICAL SCENARIO

It is 7 AM, and medical rounds are starting on university hospital ward 3B. In the past 24 hours of your residency, you have transferred 2 critically ill patients to the intensive care unit; accepted 11 patients to your medical service; examined and revised medication orders for 22 patients; placed 9 intravascular catheters; written 35 notes; and reviewed, categorized, and acted on more than 300 new pieces of laboratory and radiology data. You were planning to ask the infectious disease specialist about a patient, but he seems very busy, and the broad-spectrum antibiotic regimen you prescribed should suffice. You were just told that you ordered total parenteral nutrition for the wrong patient. While deciding which patient should receive parenteral nutrition, you realize that the calculations for the amino acid concentration are erroneous. After the first 5 minutes of your first patient presentation, the senior physician asks you details from the patient's past medical history. You wish you could refer to your admission note, but you couldn't access it before your rounds because a utilization review clerk had the chart.

The chair of medicine keeps promising to install computers to help manage all of this information, but she is limited by the budget squeeze. She needs proof that computerization will improve patient care to justify such a major expense. She asks you to help. You remember reading, in one of the many journals piled up at home, about how computers can be used to provide decision support leading to improved patient outcomes. If you can show that computers improve patient care, maybe the hospital administration will see the expense as an investment that could reduce costs.

THE SEARCH

When you get home that night, you connect to the Internet and decide to search the medical literature using Internet Grateful Med from the US National Library of Medicine. You type *http://igm.nlm.nih.gov/* into your browser and choose MEDLINE. You quickly realize that you don't know what search terms to use. You enter *decision* then click the button for *Find MeSH/Meta Terms*. From the 31 Medical Subject Headings terms offered, you choose *decision making, computer-assisted*; *therapy, computer assisted*; *diagnosis, computer-assisted*; *drug therapy, computer-assisted*, specifying that they are the major topics of the article. You limit your search to randomized controlled trials in English during the years 1995 to 1998. Browsing through the 45 abstracts from the search, you choose "A Randomized Trial of 'Corollary Orders' to Prevent Errors of Omission." The abstract of this article concludes that "physician work stations, linked to a comprehensive electronic medical record, can be an efficient means for decreasing errors of omissions and improving adherence to practice guidelines."[1]

You order the full article over the Internet from Loansome Doc. In this study[1] conducted on the inpatient general medical wards of an inner-city public hospital, 6 independent services (red service, green service, etc) cared for the inpatients. Each service included a faculty internist, a senior resident, and 2 interns. A different physician team rotated onto each service every 6 weeks, and during a year, 8 different teams worked on each service. At the beginning of the study, the investigators randomly allocated 3 of the 6 services to the intervention group,

Author Affiliations: Departments of Pediatrics and Anesthesia, Children's Hospital and Harvard Medical School, Boston, Mass (Dr Randolph); Departments of Clinical Epidemiology and Biostatistics and Medicine, McMaster University, Hamilton, Ontario (Drs Haynes, Cook, and Guyatt); and School of Public Policy, University College London, London, England (Dr Wyatt).

The original list of members (with affiliations) appears in the first article of this series (*JAMA*. 1993; 270:2093-2095). A list of new members appears in the 10th article of the series (*JAMA*. 1996;275:1435-1439). The following members contributed to this article: Anne Holbrook, MD, PharmD, MSc; Virginia Moyer, MD, MPH; W. Scott Richardson, MD; David L. Sackett, MD, MSc.

Corresponding Author and Reprints: Gordon H. Guyatt, MD, MSc, McMaster University Health Sciences Centre, 1200 Main St W, Room 2C12, Hamilton, Ontario, Canada L8N 3Z5.

Users' Guides to the Medical Literature Section Editor: Drummond Rennie, MD, Deputy Editor (West), *JAMA*.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

which had access to a computer-based clinical decision support system (CDSS); the other 3 services served as controls and did not have access to a CDSS. Teams were randomly assigned to the intervention and control services. The CDSS responded to trigger orders by suggesting corollary orders needed to detect or ameliorate adverse reactions and allowed physicians to accept or reject these suggestions. TABLE 1 shows examples of corollary orders and their trigger orders.

CLINICAL COMPUTER SYSTEMS

Clinicians who manage the care of patients are dependent on computers. Laboratory data management software, pharmacy information management systems, applications for tracking patient location through admission and discharge, mechanical ventilators, and oxygen saturation measurement devices are among the many types of computerized systems that have become an integral part of the modern hospital. These devices and systems capture, transform, display, or analyze data for use in clinical decision making. Using computers to search the medical literature or to improve the legibility, display, and accessibility of information in the patient's chart may produce benefits that can sometimes be related to the care of an individual patient. However, medical literature databases and ordinary patient charting systems do not filter and abstract information from detailed clinical data. We use the term *CDSS* to describe software designed to directly aid in clinical decision making about individual patients. Specifically, detailed individual patient data are input into a computer program that sorts and matches them using programs or algorithms in a knowledge base, resulting in the generation of patient-specific assessments or recommendations for clinicians.[2] TABLE 2 shows functions of decision support systems developed for the following medical purposes: alerting, reminding, critiquing, interpreting, predicting, diagnosing, assisting, and suggesting.[3]

Many alerting, reminding, and critiquing systems are based on simple *if-then* rules that tell the computer what to do when a certain event occurs. Alerting systems monitor a continuous signal or stream of data and generate a message (an alert) in response to items or patterns that might require action on the part of the care provider.[4] A simple example of an alert is the starred (*) or highlighted item (with H or L marking or with **BOLD** or changed colors on the screen) that alerts the clinician to values that are out of range on computerized laboratory printouts and display screens. Alerting systems draw attention to events as they occur. Reminder systems notify clinicians of important tasks that need to be done before an event occurs. An outpatient clinic reminder system may generate a list of immunizations that each patient on the daily schedule requires. Although the technical rules that generate alerts and reminders are often simple, alerting the right person in a timely fashion is quite complex.

When the clinician has made a decision and the computer evaluates that decision and generates an appropriateness rating or alternative suggestion, the decision support approach is called critiquing. The distinction between assisting and critiquing decision support programs is that assisting programs help formulate the clinical decision, whereas critiquing programs have no part in suggesting the order or plan but evaluate the plan, after it is entered, against an algorithm in the computer.[3] Critiquing systems are commonly applied to physician order entry. For example, a clinician entering an order for a blood transfusion may receive a message stating that the pa-

Table 1. Example Trigger and Corollary Orders

Trigger Orders	Corollary Orders
Heparin infusion	Platelet count once before heparin starts, then every 24 h
	Activated partial thromboplastin time at start, again 6 h after a dosage change
	Prothrombin time once before heparin started
	Hemoglobin at start of therapy, then every morning
	Test stools for occult blood while administering heparin
Intravenous fluids	Place a saline lock when intravenous fluids are discontinued
Narcotics (class II)	Docusate if not taking any other stool softener or laxative
Nonsteroidals	Creatinine level (if not 1 in previous 10 d); SMA-12,* blood urea nitrogen counted as equivalent
Aminoglycosides	Peak and trough levels after dosage changes and every week
	Creatinine level twice per week (every Monday and Thursday)
Warfarin sodium	Prothrombin time each morning
Amphotericin B	Creatinine level twice per week (every Monday and Thursday)
	Magnesium level (twice per week while receiving therapy)
	Electrolytes (twice per week while receiving therapy)
	Acetaminophen (650 mg by mouth 30 min before each dose)
	Diphenhydramine hydrochloride (50 mg 30 min before each amphotericin dose)

*SMA-12 indicates sequential multiple analyzer, measuring glucose, blood urea nitrogen, uric acid, calcium, phosphorus, total protein, albumin, cholesterol, total bilirubin, alkaline phosphatase, serum glutamic oxaloacetic transaminase, and lactate dehydrogenase.

Table 2. Functions of Computer-Based Clinical Decision Support Systems

Function	Example
Alerting	Highlighting out-of-range laboratory values
Reminding	Reminding the clinician to schedule a mammogram
Critiquing	Rejecting an electronic order
Interpreting	Interpreting the electrocardiogram
Predicting	Predicting risk of mortality from a severity-of-illness score
Diagnosing	Listing a differential diagnosis for a patient with chest pain
Assisting	Tailoring the antibiotic choices for liver transplantation and renal failure
Suggesting	Generating suggestions for adjusting the mechanical ventilator

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

tient's hemoglobin level is above the transfusion threshold, and the clinician must justify the order by stating an indication, such as active bleeding.[5] Getting the attention of the person who can take action is one of the most difficult aspects of making alerting, reminding, and critiquing systems effective.

The automated interpretations of electrocardiogram readings[6] and the outcome predictions generated by severity-of-illness scoring systems[7] are examples of decision support systems used for interpreting and predicting, respectively. These systems filter and abstract detailed clinical data and generate a report characterizing the meaning of the data (eg, anterior myocardial infarction).[6]

Computer-aided diagnostic systems assist the clinician with the process of differential diagnosis.[8] When the electrocardiogram results are not definitive, computer systems that try to distinguish between myocardial infarction and other sources of chest pain can sometimes outperform a clinician.[9] These types of systems require pertinent patient information, such as signs, symptoms, past medical history, laboratory values, and demographic characteristics. The programs start generating hypotheses, often prompt the user for more information, and ultimately provide a diagnosis or a list of possible diagnoses ranked probabilistically.

Computerized patient management systems are complex programs that make suggestions about the optimal decision based on the information currently known by the system. These types of systems are often integrated into the physician ordering process. After collecting information on specific patient variables, the assistant program tailors the order to the patient based on prior information in the database regarding appropriate dosages or by implementing specified protocols. The Antibiotic Assistant[10] is a CDSS that implements guidelines to assist physicians with ordering antibiotics. This system recommends the most cost-effective antibiotic regimen taking into account the patient's renal function, drug allergies, the site of infection, the epidemiology of organisms in patients with this infection at this hospital over many years, the efficacy of the antibiotic regimen, and the cost of therapy. A system that instructs caregivers about how to manage the ventilation of patients with adult respiratory distress syndrome[11] is another example.

The primary reason to invest in computer support is to improve quality of care. If a computer system purports to aid clinical decisions, enhance patient care, and improve outcomes, then it should be subject to the same rules of testing as any other health care intervention with similar claims. In this article, we describe how to use articles that evaluate the clinical impact of a CDSS. While the focus of a CDSS may be restricted to diagnosis or prognosis, we will limit our discussion to the situation in which the CDSS is designed to change clinician behavior and patient outcome. Many iterative steps are involved in developing, evaluating, and improving a CDSS before it can progress beyond the laboratory environment and pilot-testing phase and be allowed to have a wider impact on physicians and patients. These evaluations involve social science methods for evaluating human behavior and computer science methods for evaluating technological safety and robustness.[4] We limit our discussion to mature systems that have surpassed initial evaluation and are being implemented to change physician behavior and patient outcome.

Are the Results of the Study Valid?

When clinicians examine the effect of a CDSS on patient management or outcome, they should use the same criteria appropriate for any other intervention (TABLE 3), whether it be a drug, a rehabilitation program, or an approach to diagnosis or screening.[12] In our Users' Guide to prevention and therapy,[13] the importance of random assignment, blinding of patients and outcome assessors, and complete follow-up were explained. The purpose of our discussion in this article is to highlight issues of particular importance in the evaluation of a CDSS.

Table 3. Using Articles Describing Computer-Based Clinical Decision Support Systems (CDSSs)

Are the results of the study valid?
Was the method of participant allocation appropriate?
Was the control group uninfluenced by the CDSS?
Aside from the CDSS, were the groups treated equally?
What were the results?
What was the effect of the CDSS?
Can you apply the computer-based CDSS in your clinical setting?
What elements of the CDSS are required?
Is the CDSS exportable to a new site?
Is the CDSS likely to be accepted by clinicians in your setting?
Do the benefits of the CDSS justify the risks and costs?

Was the Method of Participant Allocation Appropriate? The validity of the observational study designs often used to evaluate a CDSS is limited. The most common observational design is the before-after study design, in which investigators compare outcomes before a technology is implemented (using a historic control group) with those after the system is implemented. The validity of this approach is threatened by the possibility that changes over time (called secular trends) in patient mix or in aspects of health care delivery may result in changes in behavior that appear to be attributable to the CDSS. Consider a CDSS that assisted physicians with antibiotic ordering[10] in the late 1980s and was associated with improvements in the cost-effectiveness of antibiotic ordering over the next 5 years. Changes in the health care system, including the advent of managed care, were occurring simultaneously during that time. To control for secular trends, the computerized antibiotic practice guideline study investigators[10] compared antibiotic prescribing practices with those of other nonfederal US acute care hospitals for the duration of the study.

One type of time-series design, in which the intervention is turned on and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

off multiple times, has been used to control for potential secular trends. Although this provides some protection against bias, random allocation of patients to a concurrent control group remains the strongest study design for evaluating therapeutic or preventive interventions.[13] Use of historical controls may lead to a higher tendency to see positive results. A comparison of the 2 types of studies used to evaluate the same antihypertensive drugs revealed that 80% of historically controlled studies suggested that the new drugs were effective, whereas only 20% of randomized controlled trials confirmed this result.[14] Randomized controlled trials have been successfully used to evaluate more than 70 CDSSs.[2,15-17]

An important issue for CDSS evaluation is the unit of allocation. Investigators in clinical trials usually randomize patients. When evaluating the effect of a CDSS on patient care, the intervention is usually aimed at changing the decision making of the clinician, so investigators may randomize individual clinicians or clinician clusters such as health care teams, hospital wards, or outpatient practices.[18] A common mistake made by investigators is to analyze their data as if they had randomized patients rather than clinicians. This is called a *unit of analysis error*.[19]

To highlight the problem, we will use an extreme example. Investigators randomize study participants to ensure that treatment and control groups are balanced with respect to important predictors of outcome. Randomization often fails to balance groups if sample size is small. Consider a study in which an investigator randomizes one team of clinicians to a CDSS and another to standard practice. During the course of the study, each team sees 10 000 patients. If the investigator analyzes the data as if patients were individually randomized, the sample size appears huge (the unit of analysis error[19]). However, it is very plausible, perhaps even likely, that the 2 teams' performance differed at the start and that this difference persisted through the study independent of the CDSS. Because the base sample size in this study is only 2 (2 teams), the likelihood of imbalance despite randomization is very large.

When investigators randomize physicians and health care teams, obtaining a sample of sufficient size can be difficult. If only a few health care teams are available, stratification of these teams according to important prognostic factors can reduce potential imbalances. If there are many known risk factors, investigators can pair health care teams according to their similarities and randomly allocate the intervention within each matched pair.[20] In addition, investigators can use statistical methods developed specifically for analyzing studies using cluster randomization.[21]

There is one other issue regarding randomization to which clinicians should attend. If some clinicians assigned to CDSS fail to receive the intervention, should these clinicians be included in the analysis?

The answer, counterintuitive to some, is yes. Randomization can accomplish the goal of balancing groups with respect to both known and unknown determinants of outcome only if patients (or clinicians) are analyzed in the groups to which they are randomized. Deleting or moving patients after randomization compromises or destroys the balance that randomization is designed to achieve. An analysis in which patients are included in the groups to which they were randomized, whether or not they received the intervention, is called *intention to treat*.[13]

In the study by Overhage et al,[1] during the course of a year, there were 36 teams randomly assigned to 18 CDSSs and 18 control services. House staff were required to write all orders and were used as the unit of analysis. Each service admitted patients in sequence, so that all 6 services received equal numbers of patients. A total of 86 house staff physicians who each received more than 5 corollary orders during the study cared for 2181 different patients during 2955 different admissions.

Random assignment of teams to CDSS and non-CDSS services increases our belief that the results are valid. However, although investigators did not randomly assign house staff to services, they conducted their analysis at the individual house staff level, comparing 45 intervention physicians with 41 control physicians. They took no steps to ensure that the characteristics of house staff on the intervention and control teams were similar, leaving the study open to biases from baseline differences in house staff performance. Moreover, the use of individual house staff instead of the team as the unit of analysis may have led to false precision in estimating the impact of the intervention because of a falsely inflated sample size.

In the study by Overhage et al,[1] investigators excluded 6 physicians from the intervention group because those physicians received fewer than 5 suggestions about corollary orders. This decision violates the intention-to-treat principle and risks introducing bias, because physicians on the control side who received fewer than 5 suggestions were included. Fortunately, the small number of excluded physicians were mostly off-service physicians covering night calls for 1 or 2 nights and not actually service team members, so the contribution of such physicians to the comparison of CDSS and control is small.

Was the Control Group Uninfluenced by the CDSS? One problem with performing a controlled trial randomizing a CDSS across patients is the difficulty in controlling for contamination of the control group by the intervention. Strickland and Hasson[22] randomly allocated patients to have changes in their level of mechanical ventilator support either directed by a computer protocol and implemented through a physician or directed by the physician independently. Because the same physicians and respiratory therapists who used the computer protocol managed the care of patients not assigned to the protocol, it is possible clinicians remembered and applied protocol algorithms in control patients. When the control group is influenced by the intervention, the effect of the CDSS may be diluted. Contamination

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

may spuriously decrease, or even eliminate, a true intervention effect.

One method of preventing exposure of the control group to the CDSS is to assign individual clinicians to use or not use the CDSS. This is often problematic because of cross-coverage of patients. Comparing the performance of wards or hospitals that do or do not use the CDSS is another possibility. Unfortunately, it usually is not feasible to enroll a sufficient number of hospitals in a study to avoid the problem we described earlier—when sample size is small, randomization may fail to ensure prognostically similar groups.

In the study by Overhage et al,[1] physicians whose teams were assigned to a control service had the CDSS guidelines available on paper but did not receive assistance when ordering. To control for the risk that cross-coverage of patients could expose the control group to the CDSS, the investigators had the chief medical resident construct the residents' evening call schedule to separate coverage for patients based on patients' study status. If switches in the schedule were made, control physicians provided call coverage only for non-CDSS patients, and intervention physicians covered only CDSS patients. Furthermore, to avoid contamination that could occur if intervention physicians cared for control patients, the computer suggested orders only when the patient had been assigned to a physician in the CDSS group, and corollary order suggestions were suppressed if the patient was assigned to the control group. If physicians returned for a second rotation and changed study status, the investigators excluded data from their second rotation. All of these efforts were to prevent contamination of the control group by the CDSS.

Aside From the CDSS, Were the Groups Treated Equally? The results of studies evaluating interventions aimed at therapy or prevention are more believable if patients, their caregivers, and study personnel are blind to the treatment.[13] Unblinded study personnel who are measuring outcomes may provide different interpretations of marginal findings or differential encouragement during performance tests.[23] Blinding also diminishes the placebo effect,[13] which, in the case of CDSS, may be the tendency of patients or clinicians to ascribe positive attributes to use of a computer workstation.[4] Although blinding the clinicians, patients, and study personnel to the presence of the computer-based CDSS may prevent this type of bias, blinding is sometimes not possible.

Interventions other than the treatment being studied that can influence the outcome are called *cointerventions*. They frequently occur because most patients receive multiple therapies aimed at improving their outcome. A problem arises when cointerventions are differentially applied to the treatment and control groups. This situation is more likely to arise in unblinded studies, particularly if the use of very effective nonstudy treatments is permitted at physicians' discretion.[13] Clinicians' concerns regarding lack of blinding are ameliorated if investigators describe permissible cointerventions and their differential use and/or standardize cointerventions[24] to ensure that their application is similar in both treatment and control groups.

It is also important to ensure that the evaluation of the outcome for each group is not biased. In some studies, the computer system may be used as a data collection tool to evaluate the outcome in the CDSS group. The "data completeness bias" can occur when the information system is used to log episodes in the treatment group and a manual system is used to log episodes in the non-CDSS group.[4] Because the computer may log more episodes than the manual system, it may appear that the CDSS group had more events, which could bias the outcome in favor of or against the CDSS group. To prevent this bias, outcomes should be logged similarly in both groups.

In the study by Overhage et al,[1] although faculty were proscribed from writing orders except during emergencies, physicians practiced within teams, and the faculty influenced the residents through their teaching. Faculty could rotate with different house staff on different rotations during the study, further complicating this situation. To allow for this clustering of physicians within teams, the investigators used generalized estimating equations to control for potential cointervention.

What Are the Results?

What Is the Effect of the CDSS? A CDSS is often aimed at preventing adverse events or health outcomes or at improving compliance with a treatment regimen. (See our Users' Guide for prevention or therapy[13] for a discussion of relative risk and relative risk reductions, risk differences and absolute risk reductions, and confidence intervals.) In the study by Overhage et al,[1] intervention physicians ordered the corollary orders suggested by the CDSS much more frequently than control physicians spontaneously ordered them. This was true when measured by immediate compliance (46.3% vs 21.9%; relative increase, 2.11; $P<.0001$), 24-hour compliance (50.4% vs 29.0%; relative increase, 1.74; $P<.0001$), or hospital-stay compliance (55.9% vs 37.1%; relative increase, 1.51; $P<.0001$). Because the numerators and denominators are not reported for the total numbers of corollary orders complied with and not complied with for each group, we cannot calculate the confidence intervals for the risk difference for the increase in compliance. However, because the P values are very small, we know that the lower boundary of the confidence interval is appreciably greater than 1, and the confidence interval is therefore relatively narrow.

Length of stay and hospital charges did not differ significantly. Pharmacists made 105 interventions with the CDSS group of physicians and 156 with control physicians (2-tailed $P = .003$) for errors considered to be life-threatening, severe, or significant.

Can You Apply the CDSS in Your Clinical Setting?

What Elements of the CDSS Are Required? Investigators should specify the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

intervention that they are evaluating. Two of the major elements of a CDSS, the logic and the computer interface used to present the logic, could each be evaluated as a separate intervention. However, sometimes it is not possible to separate these 2 elements and achieve the same result. For example, we mentioned a randomized controlled trial that compared a computerized protocol for managing patients diagnosed as having adult respiratory distress syndrome with standard clinical care using extracorporeal carbon dioxide removal as rescue therapy.[11] The computerized protocol group without rescue therapy did as well as the rescue therapy group. Was this due to the logic in the protocol, the use of the computer, or both interacting together?

To test whether the computer is needed requires that one group apply the protocol logic as written on paper and the other group use the same logic implemented by the computer. Sometimes the protocol logic is so complex that use of a computer may be required for implementation.

The CDSS may have a positive impact for unintended reasons. The impact of structured data collection forms and performance evaluations (the Checklist Effect and the Feedback Effect,[4] respectively) on decision making can equal that of computer-generated advice.[25] The CDSS intervention itself may be administered by research personnel or by paid clinical staff who receive scant mention in the published report but without whom the impact of the system is seriously undermined.

The CDSS in the study by Overhage et al of corollary orders[1] and in the adult respiratory distress syndrome study[11] had 3 components: a knowledge base defining which corollary orders were required for each trigger order, a database that stored the trigger orders, and an inference engine that compared the database with the knowledge base when a trigger order was received and sent a list of suggested corollary orders to the computer terminal for display.

Is the CDSS Exportable to a New Site? For a CDSS to be exported to a new site, it has to be able to integrate with existing software, users at the new site must be able to maintain the system, and users must accept the system. Double-charting occurs when systems require staff (usually nurses) to enter the data twice—into the computer and again on a flow sheet. Systems that require double-charting increase staff time devoted to documentation, frustrate users, and divert time that could be devoted to patient care. In general, such systems fail in clinical use.

Successful systems usually have automatic electronic interfaces to existing data-producing systems. Unfortunately, building interfaces to diverse computer systems is often challenging and sometimes impossible.

The program described in the study by Overhage et al[1] was implemented using the Regenstrief Medical Record System developed at Indiana University School of Medicine. This system provides an electronic medical record system and a physician order entry system. While it may be possible to use the knowledge built into the system in a health care environment in which the patient population is similar, the inference engine used to compare the rules with the order entered into the database is not easily exported to other locations. If, after critically appraising the article, you are convinced that a CDSS for implementing guidelines would be useful, you would need sufficient resources to rebuild the system at your own site.

Is the CDSS Likely to Be Accepted by Clinicians in Your Setting? A CDSS may not be accepted if the clinicians differ in important ways from those who participated in the study. The choice of evaluative group may limit the generalizability of the conclusions if recruitment is based upon enthusiasm, demographics, or a zest for new technology. Clinicians in a new setting may be surprised when their colleagues do not use a CDSS with the same avidity as the original participants.

The user interface is an important component of the effectiveness of a CDSS. The CDSS interface should be developed on the basis of potential users' capabilities and limitations, the users' tasks, and the environment in which those tasks are performed.[26] One of the main difficulties with alerting systems is notifying the individual with decision-making capability as rapidly as possible that there is an abnormal laboratory value or other potential problem. A group of investigators tried a number of different alerting methods, from a highlighted icon on the computer screen to a flashing yellow light placed on the top of the computer.[27] These investigators later gave the nurses pagers to alert them to abnormal laboratory values.[28] The nurses could then decide how to act on the information and when to alert the physician.

To ensure user acceptance, users must feel that they can depend on the system to be available whenever they need it. The amount of downtime needed for data backup, troubleshooting, and upgrading should be minimal. The response time must be fast, data integrity must be maintained, and data redundancy must be minimized. If systems have been functioning at other sites for a period of time, major problems or software bugs may have been eradicated, decreasing downtime and improving acceptance. Investigators should also assess the amount of training required for users to feel comfortable with the system. If users become frustrated, system performance will be suboptimal.

Many computer programs may function well at the site where the program was developed; unfortunately, the staff at your own institution may have objections to the approaches taken elsewhere. For example, an expert system for managing ventilated patients who have adult respiratory distress syndrome may use continuous positive airway pressure trials to wean patients off the ventilator, whereas clinicians at your institution may prefer pressure-support weaning. Syntax, laboratory coding and phrasing of diagnoses, and therapeutic interventions can vary markedly among institutions. Customizing the application to the environ-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ment may not be feasible, and additional expense may be invoked when mapping vocabulary to synonyms unless a mechanism to do so is already programmed into the system. To ensure user acceptance, the needs and concerns of users should be considered, and users should be included in decision making and implementation stages.

The logic in the Regenstrief Order Entry system[1] was based on the expertise of a hospital committee of staff physicians and pharmacists. Although the investigators used reference texts, the degree to which they applied an evidence-based approach is unclear. Use of solid evidence[29] from the literature could enhance clinician acceptance by convincing physicians that the rules positively affect patient outcomes. However, gaining consensus even with evidence-based practices can be difficult and a method for gaining consensus must be integrated into the local processes and culture of care. Furthermore, physicians will need some time to become acquainted with any new system, especially an order entry system.

When the study by Overhage et al began, all physicians on the medical wards had been entering all inpatient orders directly into physician workstations for 12 months. Because the order entry program was developed over time and refined by user input, it was tailored to the needs of the clinicians at that hospital. Whether this system would be easily accepted in a new environment by clinicians who had nothing to do with its development is open to question.

Do the Benefits of the CDSS Justify the Risks and Costs? Does the report reveal the behind-the-scenes costs? The real cost of the CDSS is usually much higher than the initial hardware, software, interface, training, maintenance, and upgrade costs (which may not be in the report). Often the CDSS is designed and maintained by staff whose actions are critical to the success of the intervention. An institution might not want to pay for the time of such people in addition to the cost of the computer software and hardware. Indeed, it can be very difficult to estimate the costs of purchasing or building and implementing an integrated CDSS.

Are CDSSs Beneficial? Human performance may improve when participants are aware that their behavior is being observed (the Hawthorne effect[30]); the same behavior may not be exhibited when the monitoring of outcomes has stopped. Taking into account the influence of a study environment, a recently updated[17] published systematic review of studies assessing CDSSs used in inpatient and outpatient clinical settings by health care providers[2] showed that the majority of CDSSs studied were beneficial. The review assessed patient-related outcomes (eg, mortality, length of hospital stay, decrease in infections) or health care process measures (eg, compliance with reminders or with evidence-based processes of care). A total of 68 prospective trials using concurrent control groups have reported the effects of using CDSSs on drug dosing, diagnosis, preventive care, and active medical care. Forty-three (66%) of 65 studies showed that CDSSs improved physician performance. These included 9 of 15 studies on drug dosing systems, 1 of 5 studies on diagnostic aids, 14 of 19 preventive care systems, and 19 of 26 studies evaluating CDSSs for active medical care. Six (43%) of 14 studies showed that CDSSs improved patient outcomes, 3 studies showed no benefit, and the remaining studies lacked sufficient power to detect a clinically important effect.

Health care processes are more often evaluated than patient health outcomes because process events occur more frequently. For example, a trial designed to show a 25% improvement (from 50% to 62.5%) in the proportion of patients who are compliant with a certain medication regimen would need to enroll 246 patients per group. A trial designed to show that this medication reduces mortality by 25% (from 5% to 3.75%) would need to enroll 4177 patients per group. Furthermore, long follow-up periods are required to show that preventive interventions improve patient health outcomes.

Fortunately, evaluation of health care processes will adequately infer benefit if the care processes being monitored are already known to improve outcomes.[31] We could conclude that a CDSS that increased the frequency with which aspirin, β-blockers, and angiotensin-converting enzyme inhibitors were administered to appropriate patients after myocardial infarction was beneficial, because large, well-designed randomized trials have demonstrated the benefit of these 3 interventions. Unfortunately, the link between processes and outcomes is often weak or unknown.

The study by Overhage et al[1] was able to demonstrate that a physician workstation, when linked to an order entry system able to run a series of rules, is an efficient means for decreasing errors of omission and improving adherence to practice guidelines. It is unclear how many of the rules in the system were based on solid evidence and thus how likely it is that compliance with rules will improve outcomes. Therefore, it is unclear whether the benefits are worth the cost of purchasing, configuring, installing, and maintaining the CDSS.

RESOLUTION OF THE SCENARIO

A computer-based CDSS evaluation involves the interplay between 3 elements: 1 or more human intermediaries, an integrated computerized system and its interface, and the knowledge in the decision support system. This makes evaluation of a computer-based CDSS a complex undertaking. Systematic reviews[32] of the impact of a CDSS on provider behavior and patient outcome have shown evidence of benefit.[2,15-17] Because the evaluation process for these reviews was not standardized, it is difficult to compare the results.

We have described a process of evaluating articles that aims to measure the impact of a computer-based CDSS on provider decisions or patient out-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

comes. Despite the complexity of evaluation, clinicians can use basic principles of evidence-based care to evaluate CDSSs. A study evaluating a CDSS is more believable if there is a concurrent control group with a random allocation of subjects. Randomization of clinicians by clusters can prevent the cross-contamination of the control group by the intervention that could mask the effect of the CDSS. When using multilevel designs (composed of the physician or physician group and their respective patients) investigators should treat the physician or group, not the patients, as the unit of analysis. Because most studies evaluating CDSSs are not blinded, we stressed the importance of controlling for cointerventions that could bias the outcome.

Even if the study is valid and a positive effect is shown, CDSSs have special applicability issues that must be considered. Is the computer essential to deployment of the knowledge in the CDSS? Can the CDSS be exported to a new site? Will clinicians accept the CDSS? And, finally, is it possible to accurately evaluate the cost of the CDSS when assessing risks and benefits?

REFERENCES

1. Overhage JM, Tierney WM, Zhou XH, McDonald CJ. A randomized trial of "corollary orders" to prevent errors of omission. *J Am Med Inform Assoc*. 1997; 4:364-375.
2. Johnston ME, Langton KB, Haynes RB, Mathieu A. Effects of computer-based clinical decision support systems on clinician performance and patient outcome: a critical appraisal of research. *Ann Intern Med*. 1994; 120:135-142.
3. Pryor TA. Development of decision support systems. *Int J Clin Monitoring Computing*. 1990;7:137-146.
4. Friedman CP, Wyatt JC. *Evaluation Methods in Medical Informatics*. New York, NY: Springer-Verlag; 1997.
5. Lepage EF, Gardner RM, Laub RM, Golubjatnikov OK. Improving blood transfusion practice: role of a computerized hospital information system. *Transfusion*. 1992;32:253-259.
6. Weinfurt PT. Electrocardiographic monitoring: an overview. *J Clin Monitoring*. 1990;6:132-138.
7. Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system: risk prediction of hospital mortality for critically ill hospitalized adults. *Chest*. 1991;100:1619-1636.
8. Berner ES, Webster GD, Shugerman AA, et al. Performance of four computer-based diagnostic systems. *N Engl J Med*. 1994;330:1792-1796.
9. Kennedy RL, Harrison RF, Burton AM, et al. An artificial neural network system for diagnosis of acute myocardial infarction (AMI) in the accident and emergency department: evaluation and comparison with serum myoglobin measurements. *Comput Methods Programs Biomed*. 1997;52:93-103.
10. Evans RS, Pestotnik SL, Classen DC, et al. A computer-assisted management program for antibiotics and other antiinfective agents. *N Engl J Med*. 1998;338:232-238.
11. Morris AH, Wallace CJ, Menlove RL, et al. Randomized clinical trial of pressure-controlled inverse ratio ventilation and extracorporeal CO_2 removal for adult respiratory distress syndrome. *Am J Respir Crit Care Med*. 1994;149(2 pt 1):295-305.
12. Spiegelhalter DJ. Evaluation of clinical decision-aids, with an application to a system for dyspepsia. *Stat Med*. 1983;2:207-216.
13. Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, II: how to use an article about therapy or prevention, A: are the results of the study valid? *JAMA*. 1993;270:2598-2601.
14. Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med*. 1982;72:233-240.
15. Shea S, DuMouchel W, Bahamonde L. A meta-analysis of 16 randomized controlled trials to evaluate computer-based clinical reminder systems for preventive care in the ambulatory setting. *J Am Med Inform Assoc*. 1996;3:399-409.
16. Balas EA, Austin SM, Mitchell JA, Ewigman BG, Bopp KD, Brown GD. The clinical value of computerized information services: a review of 98 randomized clinical trials. *Arch Fam Med*. 1996;5:271-278.
17. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *JAMA*. 1998;280:1339-1346.
18. Cornfield J. Randomization by group: a formal analysis. *Am J Epidemiol*. 1978;108:100-102.
19. Whiting-O'Keefe QE, Henke C, Simborg DW. Choosing the correct unit of analysis in medical care experiments. *Med Care*. 1984;22:1101-1114.
20. Klar N, Donner A. The merits of matching in community intervention trials: a cautionary tale. *Stat Med*. 1997;16:1753-1764.
21. Thompson SG, Pyke SD, Hardy RJ. The design and analysis of paired cluster randomised trials: an application of meta-analysis techniques. *Stat Med*. 1997; 16:2063-2079.
22. Strickland JH Jr, Hasson JH. A computer-controlled ventilator weaning system: a clinical trial. *Chest*. 1993;103:1220-1226.
23. Guyatt GH, Pugsley SO, Sullivan MJ, et al. Effect of encouragement on walking test performance. *Thorax*. 1984;39:818-822.
24. Morris AH, East TD, Wallace CJ, et al. Standardization of clinical decision making for the conduct of credible clinical research in complicated medical environments. *Proc AMIA Annu Fall Symp*. October 1996:418-422.
25. Adams ID, Chan M, Clifford PC, et al. Computer aided diagnosis of acute abdominal pain: a multicentre study. *BMJ*. 1986;293:800-804.
26. Salvemini AV. Improving the human-computer interface: a human factors engineering approach. *MD Comput*. 1998;15:311-315.
27. Bradshaw KE, Gardner RM, Pryor TA. Development of a computerized laboratory alerting system. *Comput Biomed Res*. 1989;22:575-587.
28. Tate KE, Gardner RM, Scherting K. Nurses, pagers, and patient-specific criteria: three keys to improved critical value reporting. *Proc Annu Symp Comput Applications Med Care*. October 1995:164-168.
29. Guyatt GH, Sackett DL, Sinclair JC, et al. Users' guides to the medical literature, IX: a method for grading health care recommendations. *JAMA*. 1995;274: 1800-1804.
30. Roethligsburger FJ, Dickson WJ. *Management and the Worker*. Cambridge, Mass: Harvard University Press; 1939.
31. Mant J, Hicks N. Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction. *BMJ*. 1995;311:793-796.
32. Oxman AD, Cook DJ, Guyatt GH, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, VI: how to use an overview. *JAMA*. 1994;272:1367-1371.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.