The Effects of Computerized Clinical Decision Support Systems on Laboratory Test Ordering

A Systematic Review

Nicolas Delvaux, MD; Katrien Van Thienen, MD; Annemie Heselmans, PhD; Stijn Van de Velde, PhD; Dirk Ramaekers, MD, PhD; Bert Aertgeerts, MD, PhD

Objective. To evaluate the effect of computerized clinical decision support systems on appropriateness of laboratory test ordering.

Data Sources. We used MEDLINE, Embase, CINAHL, MEDLINE In-Process and Other Non-Indexed Citations, ClinicalTrials.gov, Cochrane Library, and Inspec through December 2015. Investigators independently screened articles to identify randomized trials that assessed a computerized clinical decision support system aimed at improving laboratory test ordering by providing patient-specific information, delivered in the form of an on-screen management option, reminder, or suggestion through a computerized physician order entry using a rule-based or algorithm-based system relying on an evidence-based knowledge resource. Investigators extracted data from 30 papers about study design, various study characteristics, study setting, various intervention characteristics, involvement of the software developers in the evaluation of the computerized clinical decision support system, outcome types, and various outcome characteristics.

Conclusions. Because of heterogeneity of systems and settings, pooled estimates of effect could not be made. Data showed that computerized clinical decision support systems had little or no effect on clinical outcomes but some effect on compliance. Computerized clinical decision support systems targeted at laboratory test ordering for multiple conditions appear to be more effective than those targeted at a single condition.


A fter years of persistent increase, laboratory test ordering has become the highest-volume medical act. In the United States and Europe, the annual increase in the use of laboratory tests averaged around 5% during the last decade. Medicare spending for clinical laboratory testing has peaked at almost $10 billion, accounting for 1.7% of the total health care budget. However, it has been estimated that when including non-Medicare spending, the figures for the United States are up to 7 times higher. In Europe, laboratory spending accounted for 1.8% of total health care spending in 2012, totaling €17 billion. Overuse of laboratory test ordering has been shown to be at 20%. Even though reducing the overuse of laboratory testing may not seem very important in relation to the whole of health care spending, the true benefits are to be found in reducing downstream costs. These downstream activities include additional diagnostics such as repeat testing or imaging, prescriptions, surgeries, and hospital stays. Besides reducing costs, improving appropriate laboratory test ordering improves the quality of care by avoiding false-positive results and accidental findings of unknown significance, and is not associated with an increase in adverse effects, morbidity, or mortality.

Computerized clinical decision support systems (CCDSSs) are believed to be effective in reducing unnecessary diagnostic testing. Computerized clinical decision support systems are information technology–based systems that use patient-specific characteristics and match these to a knowledge base using rule-based algorithms. These systems can then generate patient-specific reminders or recommendations prompting more appropriate care. A recent systematic review comparing strategies for changing laboratory testing behavior of physicians concluded that several interventions, such as educational strategies, feed-
back, changing test order forms, and reminders (a form of CCDSS) may be effective.10 The influence of CCDSSs on diagnostic test ordering has been evaluated in 3 previous systematic reviews.11–13 These reviews concluded that CCDSSs have not yet been able to show an effect on clinical outcomes but appear to have a positive effect on process outcomes for diagnostic test ordering. One of the practical barriers to measuring the effect of CCDSSs is inadequate compliance by health care professionals with their recommendations, which has been shown to be as low as 1%.14 However, reviews evaluating CCDSSs on diagnostic test ordering have included a very wide variety of systems aimed at diverse types of diagnostic tests or procedures. To date, no review has been undertaken to evaluate CCDSSs aimed specifically at laboratory test ordering. We conducted this review to evaluate the effect of CCDSSs on the appropriateness of laboratory test ordering of health care professionals.

METHODS
The details of our protocol have been published elsewhere.15 Our methods are summarized here, including post hoc details of our review process.

Research Question
What is the effect of CCDSSs integrated in the electronic health record (EHR) on appropriate laboratory test ordering?

Search Strategy
We searched the MEDLINE, CINAHL, Embase, MEDLINE In-Process and Other Non-Indexed Citations, Clinicaltrials.gov, Cochrane Library, and Inspec databases through December 18, 2015, for relevant citations. We included terms for the following key concepts: decision making, computer-assisted, decision support systems, order sets, laboratory, and diagnosis. We combined these search terms with terms for the following study designs: systematic reviews, meta-analyses, randomized clinical trials, cluster randomized trials, nonrandomized controlled trials, controlled before-after studies, and time series. Details on our search strategy can be found in the supplemental digital content (see supplemental digital content at www.archivesofpathology.org in the April 2017 table of contents). Post hoc we also searched the National Institute for Health and Clinical Excellence Evidence Search database for relevant systematic reviews and health technology assessments. We also hand searched the reference lists of included articles and included systematic reviews and health technology assessments.

Study Selection
Each study was independently assessed for eligibility by 2 reviewers and a third reviewer was consulted in case of disagreement. Initially, studies were assessed on title and abstract and subsequently full-text screening for eligibility was performed. We included studies that assessed a CCDSS providing patient-specific information, delivered in the form of an on-screen management option, reminder, or suggestion through a computerized physician order entry using a rule-based or algorithm-based system relying on an evidence-based knowledge resource. We excluded CCDSSs that were not primarily focused on laboratory test ordering and did not directly communicate with the EHR. This meant that we excluded systems where patient characteristics needed to be manually introduced and systems that used paper-based reminders even if the data were processed electronically. We also limited our review to systems studied in real-life settings, used by clinicians (not students) in hospital, outpatient, or primary care settings. We excluded studies that focused uniquely on laboratory testing in a preventive care setting, such as Papanicolaou tests or fecal occult blood tests. We did not include papers studying systems that provided reminder messages for attendance at appointments or examinations. We included the following study designs as defined by the Cochrane Effective Practice and Organization of Care Group: randomized clinical trials, nonrandomized controlled trials, controlled before-after studies, and interrupted time series.16

Assessment of Study Quality
Only randomized clinical trials were assessed on quality using the Cochrane checklist for risk of bias instrument because all other included designs were considered to have a high risk of bias.17 For each trial 7 areas for potential risk of bias were assessed, including random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, addressing of incomplete outcome data, selective reporting, and other potential sources of bias. Each item was scored as high, low, or unclear as proposed by the Cochrane risk of bias assessment tool.

Data Extraction
Data from all eligible trials were extracted by one reviewer and assessed by a second reviewer. Disagreements were resolved by consulting a third reviewer. The data extraction included elements such as study design, various study characteristics, study setting, various intervention characteristics, involvement of the software developers in the evaluation of the CCDSS, outcome types, and various outcome characteristics.

Data Synthesis and Analysis
We inventoried all the extracted data and classified these according to CCDSS type and outcome type. Most studies could be classified as using a system aimed at improving appropriateness of testing for drug monitoring (eg, suggesting next international normalized ratio examination for maintaining anticoagulant therapy), disease monitoring (eg, recommending glycated hemoglobin A1c [HbA1c] tests in patients overdue for this examination) or diagnosis (eg, recommendations on laboratory tests for the assessment of patients with a sore throat). For each study we inventoried the effect on the primary outcome and evaluated effectiveness using a strategy for dichotomizing this outcome used in previous studies.9,11 If a statistically significant effect was found for the primary outcome, we considered the CCDSS to be effective. If more than one primary outcome was measured, we considered the CCDSS to be effective if a statistically significant effect was found for half or more of the primary outcomes. Post hoc, we defined that if the interventions, setting, and outcomes were deemed homogenous, a meta-analysis using a random-effects Mantel-Haenszel risk ratio was to be performed using Review Manager.18 Pooled results with signs of statistical heterogeneity, defined as I² > 50%,19 were not reported. Other results were narratively synthesized and reported. The quality of the evidence was appraised using GRADE.19

RESULTS
Our search yielded 5455 citations, of which, based on title and abstract, 377 were selected for full-text screening. We excluded another 334 papers and performed a quality assessment of 43 papers. With the high yield of high-quality studies and in contrast with our protocol, we chose to exclude those studies without a randomized controlled design because of their intrinsic high risk of bias. On this basis we excluded 15 studies and ultimately included 28 papers reporting on 23 studies in our review. Figure 1 shows the flow of our process and the Table lists the included studies.

Included Studies
Risk of Bias in Included Studies.—Of the 23 studies, 7 (30%) were randomized clinical trials with randomization on the patient level and 16 (70%) were randomized clinical effects of CCDSS on Laboratory Test Ordering—Delvaux et al

586 Arch Pathol Lab Med—Vol 141, April 2017
trials with cluster randomization on the level of the health care provider or service. Figure 2 summarizes the risk of bias assessment for each included paper. Almost all studies could not sufficiently blind the health care providers to the intervention. Four studies (17%) attempted to blind the participants by introducing a CCDSS in both the intervention and the control group, but aimed at a different pathology or with a different design. Trials with randomization on the patient level could not guarantee the absence of performance bias because health care providers were not fully blinded to the intervention and cared for patients in both the intervention and control groups.

Setting and Participants.—Fourteen studies (61%) were conducted in the United States and Canada. Eight studies (35%) were conducted in Europe. One study (4%) was conducted in Europe and Australia. Thirteen studies (57%) evaluated CCDSSs in primary care, 5 (22%) in hospital outpatient ambulatory care, and 5 (22%) in hospital inpatient care.

Intervention.—Fifteen different CCDSSs were studied, of which 6 (40%) were focused solely on laboratory testing. 5 (33%) also included other reminders such as treatment options, and 4 (27%) were not stand-alone systems, but were developed by an EHR software vendor.

Of the CCDSSs focused solely on laboratory testing, 3 (50%) were targeted at regulating blood clotting in patients using anticoagulants. Other targeted diseases or situations included diabetes, hyperlipidemia, human immunodeficiency virus (HIV) infection, initiation or management of specific drug therapies, sore throat, urinary tract infections, recommendations regarding multiple conditions, and redundancy of laboratory tests. All studies targeted at anticoagulation used either the PARMA 5 (Instrumentation Laboratory SpA, Milan, Italy), DAWN AC (4S Information Systems Ltd, Cumbria, England), or the Anticoagulation Management Support System (Softop Information Systems, Warwick, England).

Noteworthy is the finding that the 8 studies reporting on a CCDSS integrated within the EHR used 1 of only 3 systems: the Longitudinal Medical Record as implemented at Brigham and Women's Hospital (Boston, Massachusetts), the Regenstrief Medical Record System (Indianapolis, Indiana), or the Kaiser Permanente EHR (Oakland, California). Besides these 3 early adopters that have integrated CCDSSs during the last 2 decades, no other systems with integrated CCDSSs were identified.

In 14 studies, the developer of the intervention was involved in the evaluation of its effect. In the studies where the developer was involved, 8 (57%) showed a significant effect on outcomes. In the 9 studies where the developers were not involved, only 2 (22%) showed a positive effect. Albeit not statistically significant, there appears to be a tendency that involving the evaluator of the CCDSS will generate a positive effect (risk ratio, 11.33; 95% CI, 0.73–175.10).

**Outcomes**

**Clinical Outcomes.**—Diabetes.—Three studies reported HbA1c as a primary outcome for glycemic control in their studies. MacLean et al studied the Vermont Diabetes Information System (Burlington, Vermont), which generates reminders based on the results from independent laboratories and sends these to the primary care physicians either through the EHR, by fax, or by mail, depending on their network capabilities. They could not show a significant effect on HbA1c values. The Diabetes Disease Management Application implemented in the Harvard University Adult Medicine Clinic (Cambridge, Massachusetts) includes reminders on HbA1c and cholesterol tests. During their 12-month trial, Meigs et al could not show a significant effect on glycemic control. Of the 3 studies reporting on glycemic control, not a single one could demonstrate a significant effect on HbA1c levels. Hetyevik et al studied a CCDSS service for the implementation of guidelines on diabetes and hypertension, including recommendations on diagnostics and laboratory tests. They noticed no significant effect on HbA1c values after 18 months. Overall, CCDSSs aimed at laboratory testing for diabetes appear not to have an effect on glycosylated hemoglobin, though this evidence is graded as weak because of important heterogeneity and imprecision (grade C).

Anticoagulation.—Of the studies evaluating a CCDSS service aimed at improving anticoagulation, reported the surrogate clinical outcome time in therapeutic range (ITR) and three reported clinical (bleeding and thrombotic) events. The study by Poller et al, the only study reporting clinical events as primary outcome, did not show any effect. The 2 other studies reporting clinical events, albeit not as a primary outcome, were also unable to show a significant effect on clinical events (see Figure 3 for the pooled results). Computerized clinical decision support systems aimed at anticoagulation testing have not been able to show a significant effect on adverse effects (grade B).
<table>
<thead>
<tr>
<th>Source, y</th>
<th>Setting</th>
<th>No. of Patients/ No. of Professionals</th>
<th>Intervention</th>
<th>Software Developer Involved</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Process</td>
<td>Clinical</td>
</tr>
<tr>
<td>Bates et al,25 1999</td>
<td>Hospital inpatient</td>
<td>11 586/NR</td>
<td>LMR: reminds physician that a test has previously been ordered within test-specific interval</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Claes et al,38 2005</td>
<td>Primary care</td>
<td>834/96</td>
<td>DAWN AC: calculates dosage anticoagulant based on INR values and proposes next test date</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Eccles et al,39 2002</td>
<td>Primary care</td>
<td>2335/279b</td>
<td>PRODIGY: triggers guidelines, presents patient scenarios, and offers suggestions for management</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Feldstein et al,20 2006; Smith et al,21 2009</td>
<td>Primary care</td>
<td>961/253</td>
<td>KP EHR: electronic messages recommending laboratory monitoring</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Fitzmaurice et al,40 2000</td>
<td>Primary care</td>
<td>367/NR</td>
<td>AMSS: calculates dosage anticoagulant based on INR values and proposes next test date</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Flottorp et al,41 2002</td>
<td>Primary care</td>
<td>16 939 (sore throat), 9887 (UTI)/620</td>
<td>Mediata: reminders on laboratory tests and management of sore throat and UTI</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Hetlevik et al,42,43 1999, 2000</td>
<td>Primary care</td>
<td>2239 (hypertension), 1034 (diabetes)/53</td>
<td>InfoTech: guides in diagnostics, history taking, physical examination, additional tests, and treatment</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Holmes et al,22 2015</td>
<td>Primary care</td>
<td>19 157/28</td>
<td>Alert provided to triage nurse with recommendations to follow triage protocol on radiology and laboratory test</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Lo et al,26 2009</td>
<td>Primary care</td>
<td>2757/366</td>
<td>LMR: reminders on baseline tests for ordered medication</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>MacLean et al,27,28 2009, 2006; Khan et al,29 2010</td>
<td>Hospital outpatient</td>
<td>7412/132</td>
<td>VDIS: reminders for overdue laboratory tests recommended for monitoring diabetes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Manotti et al,44 2001</td>
<td>Hospital outpatient</td>
<td>1358/NR</td>
<td>PARMA 5: calculates dosage anticoagulant based on INR values and proposes next test date</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Source, y</td>
<td>Setting</td>
<td>No. of Patients/No. of Professionals</td>
<td>Intervention*</td>
<td>Software Developer Involved</td>
<td>Outcomes</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------</td>
<td>-------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>-----------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Matheny et al,30</td>
<td>Primary care</td>
<td>1922/303</td>
<td>LMR: reminders on laboratory monitoring of drugs</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Meigs et al,31</td>
<td>Hospital outpatient</td>
<td>598/66</td>
<td>DDMA: reminders on diagnosis and management of diabetes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mitra et al,32</td>
<td>Hospital inpatient</td>
<td>30/NR</td>
<td>DAWN AC: calculates dosage anticoagulant based on INR values and proposes next test date</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Nieuwlaat et al,33</td>
<td>Hospital outpatient</td>
<td>1068/NR</td>
<td>DAWN AC: calculates dosage anticoagulant based on INR values and proposes next test date</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Overhage et al,34</td>
<td>Hospital inpatient</td>
<td>2181/89</td>
<td>RMRS: system suggests corollary orders based on patient characteristics</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Palen et al,35</td>
<td>Primary care</td>
<td>26 586/207</td>
<td>KP EHR: reminders on laboratory monitoring of drugs at initiation of therapy</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Poller et al,36,37</td>
<td>Hospital outpatient</td>
<td>13 219/NR</td>
<td>DAWN AC and PARMA 5: calculates dosage anticoagulant based on INR values and proposes next test date</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Robbins et al,36</td>
<td>Hospital outpatient</td>
<td>1011/33</td>
<td>Virology FasTrack: generates alerts of potential adverse events</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Sequist et al,37</td>
<td>Primary care</td>
<td>6243/194</td>
<td>LMR: reminders on diagnosis and management of diabetes and coronary heart disease</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>van Wijk et al,33</td>
<td>Primary care</td>
<td>NR/60</td>
<td>Bloodlink: proposes order sets based on national guidelines on laboratory testing</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Of the 5 studies reporting TTR as a primary outcome, only the studies by Manotti et al44 and Mitra et al32 showed a significant effect. Poller et al46 reported TTR as a secondary outcome and found a small but statistically significant change of 0.7% difference ($P = .02$) in favor of the CCDSS arm. We did not attempt to meta-analyze these results despite similar CCDSS services and outcomes because of large heterogeneity in settings and means of reporting outcomes. Computerized clinical decision support systems aimed at anticoagulation testing may have a positive, but small, effect on TTR, but this evidence is considered of low quality because of heterogeneity of results and imprecision of results (grade C).

**Human Immunodeficiency Virus.**—We identified 1 study26 that used a CCDSS for guiding laboratory testing in patients with HIV. This large study, following 1001 HIV-infected patients, reported the mean CD4 white blood cell count as primary outcome. This study demonstrated a positive effect on clinical control of HIV with an increase of the CD4 cell count by 2.0 cells/mm$^3$ (95% CI, 0.1–4.0; $P = .04$). These results were downgraded to moderate quality (grade B) because of imprecision of results.

**Process Outcomes: Compliance.**—Thirteen studies reported compliance with recommendations as a primary outcome.1 Eleven studies reported compliance as the percentage of appropriate tests or reported data from which this statistic could be calculated. In the study by Bates et al25 this was the rate of cancelled tests after a reminder was triggered signaling that the physician was about to order a redundant test. In all other cases, compliance was derived from the number of tests performed after a test was suggested by the CCDSS system. We did not meta-analyze the results because of heterogeneity of the populations studied and the types of intervention. Fifty-four percent (n = 7) reported a positive effect on compliance with recommendations. The scope of the CCDSS tended to influence the effect. Reminders aimed at influencing the laboratory test ordering behavior of 1 or 2 tests tended to have no effect or a smaller effect than reminders aimed at 3 or more conditions (see Figure 4). Computerized clinical decision support systems aimed at laboratory testing may have a positive effect on compliance with recommendations, but this evidence is considered of low quality (grade C) because of important heterogeneity of the results.

**Economic Outcomes.**—Three studies reported economic outcomes, namely Bates et al25 Smith et al21 and Khan et al29 Bates et al25 reported the possible charge savings when implementing a CCDSS that reminds a physician upon ordering a new test that the same test had already been ordered within certain test-specific intervals. They calculated that the 24% decrease in test ordering correlated to a $35 000 annual savings for the clinic. These findings suggest that a direct cost savings of $6.14 per patient per year is possible. Smith et al21 investigated the costs of various strategies to improve laboratory monitoring of medications, including one arm that used a CCDSS within the Kaiser Permanente EHR aimed at primary care physicians. Based on costs for patient identification, chart reviews, resources required to contact patients, laboratory testing, result review of normal and abnormal tests, follow-up visits for abnormal tests, and service cost of letters, a cost for each arm was calculated and compared with the usual care group. They concluded that CCDSSs are probably not cost-effective. These conclusions were made based on the actual expenditures, not on a cost-benefit analysis including the cost of avoided events. Khan et al29 studied the effects of the Vermont Diabetes Information System in ambulatory care on the use of emergency room and hospital resources. They showed that patients cared for with the Vermont Diabetes

---

*References 20, 22–26, 30, 34, 35, 37, 39, 41, 42, 45.*

---

**Abbreviations:** AMSS, Anticoagulation Management Support System; DDMA, Diabetes Disease Management Support System; INR, international normalized ratio; KP EHR, Kaiser Permanente Electronic Health Record; LMR, Longitudinal Medical Record; NR, not reported; PRODIGY, Prescribing Rationally With Decision Support in General Practice Study; RMRS, Regenstrief Medical Record System; UTI, urinary tract infection; VDIS, Vermont Diabetes Information System.

---


* Calculated from number of practices and mean or median amount of whole/full-time-equivalent general practitioners.
Figure 2. Summary of risk of bias of all included studies. The risk of bias assessment of MacLean 2009 was the same as for their 2006 paper. Green indicates low risk of bias, orange indicates moderate risk of bias, and red indicates high risk of bias.

Figure 3. Forest plot of studies (Claes et al.38 Fitzmaurice et al.40 and Poller et al.46,47) reporting adverse (bleeding and thrombotic) effects as outcome for computerized clinical decision support systems (CCDSSs) aimed at anticoagulation testing. The x-axis shows the effect on risk ratio, where results more than 1 favor the intervention group and results less than 1 favor the control group. Heterogeneity: $\chi^2 = 1.26$, df = 2, $P = .53$, $I^2 = 0\%$. Test for overall effect: $Z = .66$ $P = .51$. Abbreviation: M-H, Mantel-Haenszel.
Information System had 11% lower hospital charges, stayed fewer days in hospitals, and had 25% lower emergency room charges than the control patients, resulting in a $2426 (95% CI, 205–4647; \( P = .03 \)) annual savings per patient. Current evidence suggests that CCDSSs may reduce direct and downstream costs of laboratory testing by amounts ranging from $6.14 to $2426 annually per patient.

DISCUSSION

This systematic review summarizes the evidence presented by randomized clinical trials on the effect of CCDSSs on laboratory test ordering of health care professionals in primary care, hospital outpatient care, and hospital inpatient care. There is no evidence that CCDSSs focused on changing laboratory test ordering behavior for diabetes or anticoagulation have an effect on clinical outcomes. Reported clinical outcomes were glycosylated hemoglobin for interventions targeting diabetes and clinical events and TTR for interventions targeting the management of anticoagulation. No study reporting clinical events as an outcome for the evaluation of a CCDSS focused on test ordering for anticoagulation therapy showed a significant effect, including a large multicenter trial conducted by Poller et al.\(^{46,47}\) Although we were unable to pool the results for TTR, only a minority of studies reporting this outcome showed a significant effect. When implemented for the follow-up of HIV testing, Robbins et al.\(^{36}\) found that CCDSSs had a significant effect on improving CD4 cell count. Overall, CCDSSs aimed at improving laboratory test ordering behavior seem to have little or no effect on clinical outcomes. Studies reporting economic outcomes show an improvement on cost-effectiveness, but it must be noted that only studies that demonstrated an effect on clinical outcomes conducted an additional cost-effectiveness evaluation. This could be an important source of publication bias for this outcome measure.

We found an effect on compliance with recommendations in a majority of the studies reporting this process outcome. We defined compliance with recommendations as the percentage of recommended tests that were ordered or the percentage of redundant tests that were cancelled. The effect sizes varied strongly, ranging from the largest effect measured by van Wyk et al.\(^{45}\) (30% increase of appropriate ordering; risk ratio, 2.55; 95% CI, 2.26–2.87) to a negative effect measured by Hetlevik et al.\(^{42,43}\) (risk ratio, 0.9; 95% CI, 0.8–1.01). Bates et al.\(^{25}\) evaluated the effect of a CCDSS that notified when a potentially redundant laboratory test was ordered for 13 different tests. When triggered, the redundancy reminders reduced the percentage of redundant tests performed by half (27% of tests performed in the intervention group compared with 51% in the control group). Van Wijk et al.\(^{23}\) evaluated a CCDSS for the implementation of the Dutch College of General Practitioners guidelines on laboratory testing on laboratory test ordering behavior. This CCDSS included reminders based on 54 different guidelines and proposed laboratory tests

Figure 4. Forest plot summarizing studies reporting compliance as outcome. The x-axis shows the effect on risk ratio, where results more than 1 favor the intervention group and results less than 1 favor the control group. Studies above the line report on a computerized clinical decision support system (CCDSS) targeting 1 or 2 conditions; studies below the line report on a CCDSS targeting 3 or more conditions.
based on a series of indications. When compared with a restricted laboratory test ordering form containing only the 15 most popular tests, the CCDSS reduced the average number of tests from 6.9 to 5.5 per order. A limitation to this study is that no formal evaluation of appropriateness was made. The authors assumed that reducing the number of tests implied a reduction in the number of inappropriate tests. Positive effects were also found by Feldstein et al evaluating the effect of CCDSSs reminding physicians of 10 possible drug-laboratory interactions in Kaiser Permanente primary care practices. That this trend is not absolute is illustrated by the study by Matheny et al studying a similar CCDSS aimed at 14 drug-laboratory interactions within the Longitudinal Medical Record at Brigham & Women’s Hospital. They could not report a significant effect on appropriateness of testing, but an important note to be made is that the baseline rates of overdue testing were already very low before the implementation of the CCDSS, hence leaving very little margin for improvement.

We also found that when the developers of the CCDSS were included in the evaluation of their system, this tended to lead to a higher chance of finding positive effects. Roshanov et al previously pointed out that involving the developers in the evaluation process may be a potential source of bias. Our findings add strength to this conclusion.

**Strengths and Limitations of This Review**

Our review has some important strengths. First, our literature review was very thorough, screening more than 5000 articles and hand searching reference lists of previously conducted systematic reviews for additional citations. Second, we limited the intervention specifically to CCDSSs aimed at changing the laboratory testing behavior of health care professionals, thus assuming that findings may be more generalizable than in previous systematic reviews. Finally, this study provides a comprehensive and clear summary of randomized clinical trials and explores several tendencies in results.

An important limitation to this study is that we were unable to report an overall effect because of various sources of heterogeneity. First, we observed large heterogeneity in settings due to the inclusion of multiple health care settings and countries with different baseline testing behaviors. Second, even though the functionality of all the studied CCDSSs may be very similar, because they are targeted at different conditions, their effectiveness varies largely and we were unable to make generalizable conclusions.

In our protocol we specified that, if possible, we would meta-analyze results using extracted data for patient and process outcomes. When this strategy was not deemed appropriate because of heterogeneity, we chose to summarize the effect by dichotomizing the results as being positive if they showed a significant effect for at least half of the primary outcomes. This strategy for reporting has been used previously but poses limitations to the interpretation of the results. We reported data from studies with process outcomes as a primary outcome and from studies with process outcomes as a secondary outcome if they showed a significant effect on their primary (clinical) outcomes. We refrained from reporting on every secondary outcome because we feared that including positive effects for secondary outcomes from studies that failed to show an effect on their primary outcome could be a potential source of bias. As a result, some studies reporting positive effects for some secondary outcomes were considered ineffective in our review. This may have led to an underestimation of the effect; however, our results appear consistent with other reviews on CCDSSs. Our review was not designed as a health technology assessment; hence, our findings regarding cost-effectiveness must be interpreted with care. We found some evidence suggesting that CCDSSs aimed at improving appropriateness of laboratory testing may be cost-effective. This evidence supports earlier conclusions that CCDSSs are potentially cost-effective, but lacking strong evidence.

Evaluating the trustworthiness of the knowledge base used to design the reminders or recommendations in the CCDSS proved very difficult. Some authors clearly described the guidelines or recommendations used; however, in a large majority of the studies, this was not mentioned. Roshanov et al evaluated multiple factors that may influence the effectiveness of CCDSSs and found that whether or not the advice was evidence-based did not significantly influence their effectiveness; hence, we chose not to exclude any studies on this basis.

Almost all studies on CCDSSs had difficulties in blinding the participants in their design. This poses a problem in studying the effects of CCDSSs in a randomized controlled design and analyzing its results. Simply by knowing that they are being studied, participants in both intervention and control groups tend to improve their clinical practice. This effect is also known as the Hawthorne effect. We identified only one study that was designed to correct for a possible Hawthorne effect by introducing a cluster randomized design on the physician level (with interpractice controls) and additionally randomizing patients within the cluster to the intervention or usual care (intrapractice controls). The interpractice control group was used to evaluate any Hawthorne effect in the intrapractice controls. Albeit Fitzmaurice et al did not see a Hawthorne effect in their control group, it is to date unclear what the influence of this effect could be in trials investigating complex interventions like CCDSSs, especially when process outcomes are being investigated. Additionally, there is a risk that participants in the intervention groups may discuss the information provided by the CCDSS with participants in the control group, resulting in performance bias. That the inability to properly blind the participants and the Hawthorne effect may have a significant effect on results was painfully demonstrated in the study by Holmes et al wherein a single nurse was responsible for a significant improvement in guideline implementation for the whole control group and subsequently cancelled a difference in effect with the intervention group. To correctly observe a true effect of CCDSSs, it remains imperative that researchers set up sound research designs taking into account the complexity of CCDSSs in order to minimize any risk of bias.

**Implications**

Our results are consistent with other reviews on the effect of CCDSSs on diagnostic testing. Roshanov et al concluded that, based on data from randomized controlled trials, CCDSSs can modify test-ordering behavior of health care professionals, but also concluded that generalizable results were impossible because of differences in implementation, system features, system design, and study design. Based on data from controlled studies and time series, Main et al concluded that, when combining effects on primary and secondary outcomes, CCDSSs significantly improve process or practitioner performance outcomes.
CONCLUSIONS

Computerized clinical decision support systems aimed at improving the laboratory test ordering of health care professionals have shown little or no effect on clinical outcomes and may have an effect on some process outcomes; however, data remain sparse and not generalizable. A sound framework defining various system designs, allowing comparison of effects, and avoiding implementation errors, as well as recommendations on study designs and relevant outcomes to avoid evaluation errors, would greatly improve reporting on CCDSSs.

The authors would like to acknowledge and thank Marleen Michels for her aid and contribution to the systematic search strategy.

References


