# Effectiveness of Practices to Support Appropriate Laboratory Test Utilization

## A Laboratory Medicine Best Practices Systematic Review and Meta-Analysis

*Matthew Rubinstein, MS,[1] Robert Hirsch, PhD,[2] Kakali Bandyopadhyay, PhD,[3] Bereneice Madison, PhD,[1] Thomas Taylor, MS,[1] Anne Ranne, PhD,[1] Millie Linville, MS,[3] Keri Donaldson, MD, MSCE,[4] Felicitas Lacbawan, MD, FCAP, FACM,[5] and Nancy Cornish, MD[1]*

From the [1]Centers for Disease Control and Prevention, Atlanta, GA; [2]Carter Consulting, Atlanta, GA; [3]Deloitte Consulting, Atlanta, GA; [4]Pennsylvania State University, State College; and [5]Quest Diagnostics Nichols Institute, San Juan Capistrano, CA.

**Key Words:** Clinical decision-making; Clinical laboratory services; Laboratory diagnosis; Laboratory medicine; Laboratory test utilization; Meta-analysis; Quality assurance; Quality improvement; Systematic review; Test ordering; Utilization management; Utilization review

*Am J Clin Pathol* March 2018;149:197-221

DOI: 10.1093/AJCP/AQX147

## Abstract

***Objectives:*** *To evaluate the effectiveness of practices used to support appropriate clinical laboratory test utilization.*

***Methods:*** *This review followed the Centers for Disease Control and Prevention (CDC) Laboratory Medicine Best Practices A6 cycle method. Eligible studies assessed one of the following practices for effect on outcomes relating to over- or underutilization: computerized provider order entry (CPOE), clinical decision support systems/tools (CDSS/CDST), education, feedback, test review, reflex testing, laboratory test utilization (LTU) teams, and any combination of these practices. Eligible outcomes included intermediate, systems outcomes (eg, number of tests ordered/performed and cost of tests), as well as patient-related outcomes (eg, length of hospital stay, readmission rates, morbidity, and mortality).*

***Results:*** *Eighty-three studies met inclusion criteria. Fifty-one of these studies could be meta-analyzed. Strength of evidence ratings for each practice ranged from high to insufficient.*

***Conclusion:*** *Practice recommendations are made for CPOE (specifically, modifications to existing CPOE), reflex testing, and combined practices. No recommendation for or against could be made for CDSS/CDST, education, feedback, test review, and LTU. Findings from this review serve to inform guidance for future studies.*

Laboratory testing is integral to modern health care as a tool for screening, diagnosis, prognosis, stratification of disease risk, treatment selection, and monitoring of disease progression or treatment response. It is also a guide for hospital admissions and discharges. As rates of test utilization grow, there is increased scrutiny over the appropriateness of testing, for example to reduce potential for diagnostic error.[1,2] Additionally, with increased capitation and restrictive insurance reimbursements, clinical laboratories are under continuous pressure to improve the value and utility of laboratory investigations while operating in relation to these financial restrictions.[3,4] Within this landscape, rates of inappropriate overutilization and underutilization represent an important gap for quality improvement at one of the earliest points of clinical-laboratory interface.[1,5,6] While utilization management approaches have been described in the literature,[6-14] the effectiveness of these interventions in support of appropriate test utilization is unclear, as are current research gaps.

### Quality Gap: Inappropriate Laboratory Test Ordering

Many laboratory test orders are not supported by appropriate-use protocols (ie, organizational guidelines, local consensus guidelines, algorithms, and local administrative directives on utilization) and are unnecessarily duplicative, with variations in test ordering patterns influenced by a number of factors.[4,12,15] A systematic review by

Zhi et al[5] revealed that overall mean rates of inappropriate over- and underutilization of testing were 20.6% and 44.8%, respectively. In this context, a reduction in duplicate test orders and the use of test orders supported by appropriate use protocols qualify as measurable quality gaps.

### Quality Improvement Practices

Eight practices impacting test utilization were determined relevant to this systematic review, and have been characterized in the literature.[7-14,16] These practices, defined in the glossary (Supplemental Table 1; all supplemental material can be found at *American Journal of Clinical Pathology* online), are: computerized provider order entry (CPOE), clinical decision support systems/tools (CDSS/CDST), education, feedback, reflex testing, test review, laboratory test utilization (LTU) teams, and combined practices.

These practice categories represent approaches to systematically manage the utilization of testing, and are generally preanalytical within the total testing process.[17,18] As such, each practice serves to impact the appropriateness of specific testing ordered or performed, a gap generally expressed as overutilization or underutilization of testing within this review's a priori analytical framework and inclusion/exclusion criteria. The source of criteria informing the "appropriateness" of testing is variable, ranging from national recommendations and guidelines to local consensus and administrative protocols. This aspect of variation was not an a priori consideration for analysis and derivation of practice recommendations (ie, not an element of inclusion/exclusion nor an element of data abstraction). However, it is further discussed in the "Conclusions."

## Materials And Methods

This systematic review was guided by the Center for Disease Control (CDC) Laboratory Medicine Best Practices (LMBP) A6 Cycle, a previously validated evidence review and evaluation method for quality improvement in laboratory medicine.[19] Additional resources can be found at https://wwwn.cdc.gov/labbestpractices/.

This systematic review was conducted by a review coordinator, staff trained to apply CDC LMBP methods, and statisticians with expertise in quantitative evidence analysis. The team was advised by a multidisciplinary expert panel consisting of seven members, with subsequent approval or practice recommendations by the LMBP Work Group. Supplemental Table 2 lists the LMBP Work Group members and expert panel members and further describes their roles.

### Ask: Review Questions and Analytic Framework

The question addressed through this systematic review is: "What is the effectiveness of practices used to support appropriate clinical LTU?" This review question was developed in the context of the analytic framework depicted in Figure 1.

Eligibility criteria, relevant population, intervention, comparison, outcome, and setting elements are:

- Population:
  - Personnel legally authorized to order clinical laboratory tests (ie, targeted population for test utilization management interventions)
  - General patient population (no restrictions, including inpatient and outpatient settings, in terms of study eligibility for the systematic review)
- Interventions (quality improvement practice interventions to manage/support appropriate clinical LTU, with quality gaps expressed as inappropriate over- or underutilization of testing): CPOE (CPOE replacement of written test orders and modifications to an existing CPOE), CDSS/CDST, education, feedback, test review, LTU teams, reflex testing, and combined practices.
- Comparison: experimental (eg, randomized controlled trials) and other comparative studies designs (eg, before and after studies) in which the effect of a utilization management intervention was compared to a group lacking it.
- Outcomes:
  - Number of tests (eg, number of test orders, number of tests performed)
  - Costs/charges (eg, cost of test orders, cost of tests performed, cost per diagnosis, overall health care costs)
  - Turnaround time (eg, where reduction in the number of inappropriate tests may not be relevant)
  - Diagnostic yield and diagnostic detection rate
  - Length of hospital stay
  - Other patient-related outcomes (eg, patient satisfaction, patient safety events related to delayed or incorrect diagnosis, adverse drug reactions, readmission rates, morbidity, and mortality)
- Setting:
  - Facility: any health care facility in which laboratory testing is ordered/performed, including reference laboratory testing
  - Targeted testing: any testing identified within eligible studies as being inappropriately utilized (ie, over- or underutilized)

For study inclusion, the intent of this review was not to focus on a single criterion for "inappropriate" utilization nor a single source of that criterion (eg, local consensus guidelines, national guidelines). Rather, focus was cast broadly on the effect of utilization management


DOI: 10.1093/ajcp/aqx147

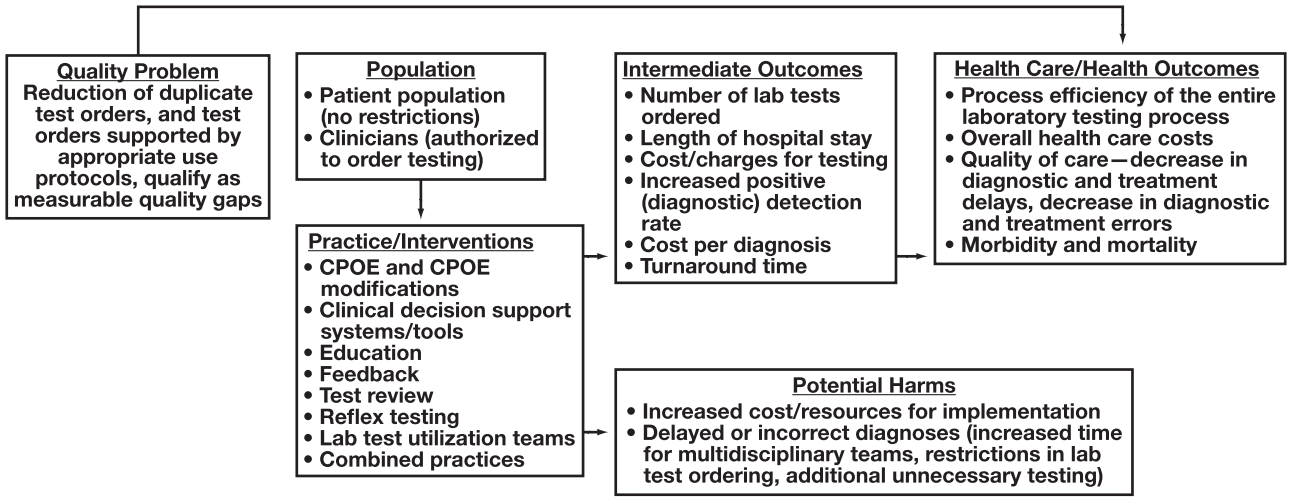


**Figure 1** Analytic framework. CPOE, computerized provider order entry.

approaches. This review (particularly the "Future Research Needs" section) was informed by 2013 and 2014 reviews, which characterized the "landscape" of inappropriate testing, and identified frequent sources of criteria for identification of inappropriate utilization.[5,20]

Finally, several of the outcome types listed above (eg, number of tests, costs of tests, and turnaround time) represent intermediate outcomes and are systems/operational in nature. However, they are proximal indicators of practice effectiveness, such that direct causality in relation to an intervention is more likely. While other outcomes types (eg, length of stay, patient safety events, and morbidity) may be more patient-related outcomes, a concern (discussed in the "Limitations" section) is that they are distal to the intervention and thus influenced by other medical and nonmedical factors during the course of patient care.

### Acquire: Literature Search and Request for Unpublished Studies

With involvement of the expert panel and a CDC librarian, a comprehensive electronic literature search was conducted in seven electronic databases to identify eligible studies in the current evidence base. The initial literature search was conducted April 17, 2014, with two additional searches on September 1, 2015, and January 10, 2016. Further description of the search protocol, as well as the full electronic search strategy for each searched database, are provided in Supplemental Table 3.

### Appraise: Screen and Evaluate Individual Studies

Screening of search results against eligibility criteria was performed by two sets of independent reviewers, with disagreement mediated by consensus discussion or by a third reviewer. The screening process is further described in Supplemental Table 3.

Studies were categorized to specific practice category (ie, CDSS/CDST, CPOE, education, etc) independently by two reviewers, with disagreement mediated by consensus discussion or, if needed, by a third reviewer. Studies were then abstracted and quality appraised using a standard data abstraction form tailored to the topic of this systematic review (Supplemental Table 4). The final data abstraction forms—"evidence summary tables" for each study—represent consensus between two independent abstractions on content and quality appraisal, with a statistician's review of abstracted statistical data and input of qualitative effect size ratings. Use of the data abstraction forms for generation of "evidence summary tables" is further described in Supplemental Table 4.

### Analyze: Data Synthesis and Strength of the Body-of-Evidence (Meta-Analysis Approach)

Two analytic approaches were used in this systematic review: qualitative determinations of overall strength of evidence and quantitative meta-analysis. For the qualitative analysis, groups of studies within a practice category were classified according to the overall strength of their evidence's effectiveness, with ratings of "high," "moderate," "suggestive," or "insufficient". These qualitative ratings take into account the number of studies within a group, their qualitative effect size ratings, and their qualitative quality ratings. Criteria in **Table 1** are the minimum criteria to achieve a particular strength of evidence rating. These criteria are the basis of the body-of-evidence qualitative analyses appearing in the "Results" section and are the primary determinant of the best practice recommendation categorizations appearing in the "Conclusions" section.



DOI: 10.1093/ajcp/aqx147

**Table 1**
**Criteria for Determining Strength of Body-of-Evidence Ratings**[a]

| Strength of Evidence Rating | No. of Studies | Effect Size Rating | Quality Rating |
|---|---|---|---|
| High | ≥3 | Substantial | Good |
| Moderate | 2 | Substantial | Good |
| | or ≥3 | Moderate | Good |
| Suggestive | 1 | Substantial | Good |
| | or 2 | Moderate | Good |
| | or ≥3 | Moderate | Fair |
| Insufficient | Too few | Minimal | Fair |

[a]Adapted from Christenson et al.[19]

Quantitative meta-analysis was conducted on a subset of included studies meeting the following criteria: (1) have a similar outcome (eg, number of tests ordered or performed, or cost of tests), (2) have an intervention satisfying inclusion criteria, (3) have a quality rating of "fair" or "good," and (4) have information necessary to calculate point and interval estimates. In preparation for meta-analysis, data in studies were summarized as either Cohen *d* or as an odds ratio (OR), representing standardized measures of effect for each included study. Continuous data were summarized using Cohen *d*. For continuous data, necessary information included means, standard deviations (or standard errors, confidence interval [CI], exact *P* value, or test statistic), and sample sizes. Nominal data, on the other hand, were summarized using ORs. For nominal data, necessary information included numerators and denominators of proportions.

Meta-analyses were conducted using Comprehensive Meta-Analysis software (version 3) from www.Meta-Analysis.com. All meta-analyses assumed random effects.[21] Unlike the fixed effect approach, the random effects approach does not assume homogeneity of effect across studies, an assumption which is not reasonable for the evidence base analyzed in this review. The $I^2$ statistic, was used to determine consistency (homogeneity) of effects.[22] Most studies had $I^2$ close to 100% (nearly all variation due to heterogeneity), reinforcing the decision to use a random effects approach. Results of meta-analyses were summarized numerically using the summary OR (for number of tests ordered) or Cohen *d* (for cost of tests ordered) and their 95% CI.

Point and interval estimates were summarized graphically using forest plots, appearing in the "Results" section. Study-specific ORs and summary ORs were classified as having minimal (OR < 2.5 or OR > 0.4), moderate (OR ≥ 2.5 or OR ≤ 0.4 and OR < 4.5 or OR > 0.2), or substantial (OR ≥ 4.5 or OR ≤ 0.2) effect sizes. Study-specific Cohen *d* and summary Cohen *d* were classified as having minimal ($|d| < 0.5$), moderate ($|d| \geq 0.5$ and $|d| < 0.8$), or substantial ($|d| \geq 0.8$) effect sizes.[19] These cutoffs were the criteria for qualitative effect size ratings used for the body-of-evidence qualitative analysis described above.

## Results

### Study Selection

A total of 23,231 bibliographic records were identified through seven electronic databases, published between the periods of January 1980 to January 2016. The bibliographic record included published studies, as well as conference abstracts and proceedings. No unpublished studies were successfully obtained for screening. After removing duplicates and non-English articles, a total of 22,207 bibliographic records were identified. Screenings of titles and abstracts independently by two reviewers resulted in 812 studies to be considered for inclusion by full-text review, with elimination of 21,395 studies (including editorials, articles unrelated to the systematic review topic). Subsequent full-text screening resulted in 95 studies determined eligible for inclusion in the systematic review, eliminating 717 studies not meeting inclusion criteria (eligible interventions and/or outcomes not present). Lastly, in accord with the LMBP methodology, 12 of these 95 studies were eliminated from final inclusion following study quality assessment and determination of a "poor" quality rating by two independent data abstractors.

This resulted in 83 studies for inclusion in this systematic review, of which 32 could not be meta-analyzed due to the absence of necessary information. The study selection flow diagram is depicted in **Figure 2**. Full bibliographic information for each study is provided in Supplemental Table 4.

### Study Characteristics and Risk of Bias

While studies with the aim of impacting quality gaps identified as underutilization were eligible for inclusion, the final set of studies that could be qualitatively analyzed and meta-analyzed addressed inappropriate overutilization (additional discussion on this gap in the "Limitations" section). Further, the information among the available evidence base supported statistical determinations of number of tests (representing either number of tests ordered or number of tests performed) as the measure of effect that could be analyzed across a majority of included studies.

The number of tests outcome was calculated as the difference between the number of tests pre- and postintervention, or between the intervention group and control group (for some studies, this was expressed as number of tests *ordered*, for others it was expressed as number of tests *performed*, generalized in this review as "number of tests"), with the assumption made that the testing ordered/performed postintervention was


DOI: 10.1093/ajcp/aqx147

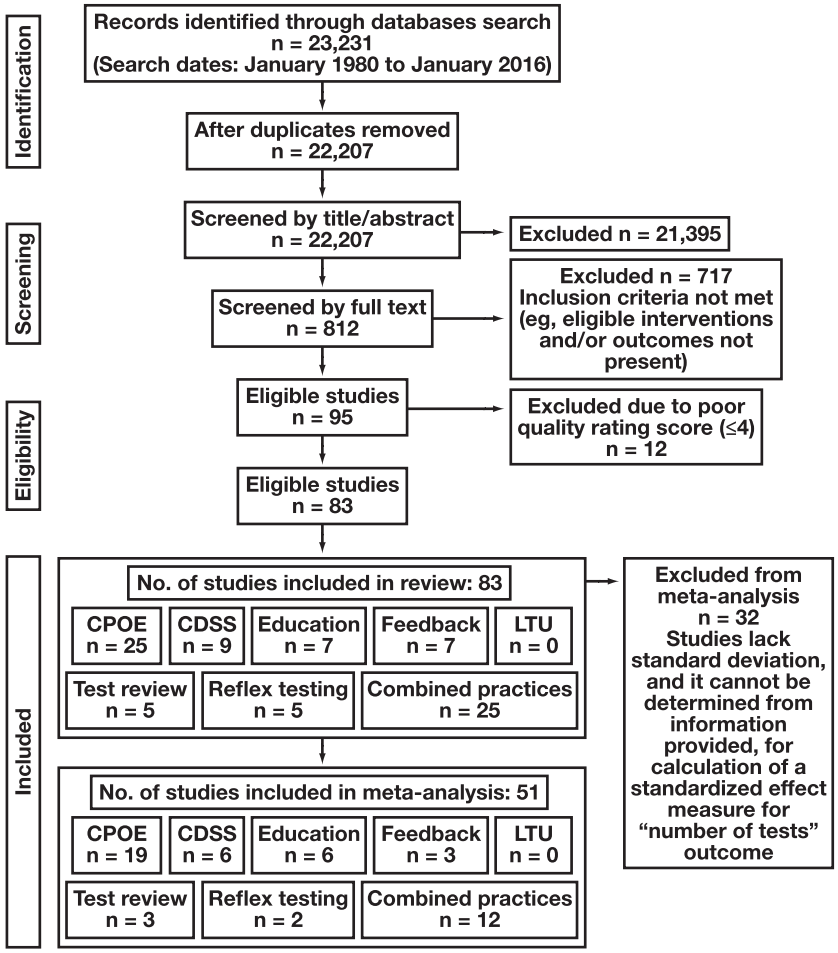


**Figure 2** Study selection flow diagram. CDSS, clinical decision support system; CPOE, computerized provider order entry; LTU, laboratory test utilization.

appropriate. These data permitted calculation in most cases of one of the standardized effect measures—Cohen *d* or OR—for inclusion in qualitative and quantitative analysis summaries (described below). As will be further described in the "Discussion" section, the quality gap addressed by all but one included study was in relation to inappropriate overutilization of testing. Each study variably defined the standard by which the quality gap of inappropriate overutilization was established, with considerable variation across studies as to criteria for determining the presence of inappropriate test utilization through utilization audit, as well as how thoroughly investigators described these criteria in their reporting. This is discussed in more detail in the "Limitations" section and in the "Future Research Need" section.

Data available in the current evidence base also supported determinations of "cost of tests" outcome, serving as this review's secondary outcome measure analyzable across many of the included studies.

To allow for more robust synthesis of eligible studies with number of tests as the outcome, the Cohen *d* statistics (with CIs) were converted to ORs (with CIs).[23] Studies with cost of tests ordered as an outcome were compared using Cohen *d*, since subsequent conversion to an OR was unnecessary.

The available evidence base demonstrated considerable variability in terms of targeted test-ordering clinicians, targeted testing, patient setting, patient inclusion/exclusion criteria, facility type and size, study design, and the specific intervention used within a practice category. The "Applicability and Generalizability" subsection of the "Discussion" section provides details on distribution of characteristics across studies. Additionally, they varied in terms of author affiliation, with 46% of included studies having laboratory professionals in authorship (eg, pathologists, lab consultants, lab managers, and lab directors).

Supplemental Table 4 provides detailed evidence summary tables for the 83 included studies, as well as for



DOI: 10.1093/ajcp/aqx147

the 12 eligible studies excluded for poor quality ratings. Supplemental Table 4 also provides criteria for study quality point deductions. With a possible score of 0 to 10, average quality score for all included studies was 6.7 (standard deviation, ±1.11) for an average "fair" quality rating across all included studies. Supplemental Table 4 lists the most common reasons for quality point deductions. Supplemental Table 5 provides consolidated intervention descriptions, targeted testing descriptions, and targeted clinical staff descriptions for each included study.

### Qualitative and Quantitative Analysis

Results of quantitative analyses and meta-analysis are provided in subsequent sections, grouped by overall practice category (ie, CDSS, CPOE, etc). Subgroup analyses are also provided also for the following three practices: CDSS, CPOE, and combined practices. As described in the "Methods" section, the body-of-evidence qualitative analysis tables are the primary determinant of recommendation categorizations, with criteria for strength of evidence ratings provided in Table 1. Strength of body-of-evidence ratings are based on the outcome measure that could be meta-analyzed across a majority of studies: number of tests. These analyses are additionally summarized in the "Conclusions" section.

The source of appropriate utilization criteria (eg, national guidelines, local consensus, administrative protocols, etc) was not an element of inclusion/exclusion criteria and was not an aspect of how grouping of studies for analysis proceeded. More generally, all but one included study expressed its quality gap as inappropriate overutilization of testing. Studies were grouped, then (for qualitative and quantitative aggregation), by those that assessed a common utilization management practice (eg, education) for a gap expressed generally in the individual paper as inappropriate overutilization of the targeted testing.

### Computerized Provider Order Entry

Twenty-five studies assessing the effectiveness of CPOE-alone practices (ie, not in combination with another practice type) were included in this systematic review. Interventions within the CPOE practice category were subgrouped as redundant test alerts, display of test costs, limiting test availability in the CPOE user interface, and CPOE replacement of written test orders.

#### *Body-of-Evidence Qualitative Analysis*

Twenty-four studies in the CPOE practice category were qualitatively analyzed using the LMBP rating criteria, with results summarized in **Table 2**. They have a "high" rating for the overall strength of evidence of effectiveness in relation to the primary outcome measured: number of tests. Tierney et al (1993) (available in Supplemental Table 2) supported derivation only of the secondary outcome assessed by this review—the cost of test orders—and the study is therefore not included in this qualitative analysis. Twenty-one of these 24 studies assessed the impact of modifications to the CPOE, while three studies (Georgiou et al,[30] Hwang et al,[32] and Westbrook et al[47]) assessed the impact of CPOE systems as replacement for written test orders.

For the 21 studies assessing the impact of modifications to the CPOE, standard effect measures (OR or Cohen *d* with standard error) could not be determined for two of them (Horn et al[31] and Procop et al[42]). Thus, 19 CPOE modification studies were meta-analyzed in relation to the primary outcome measure. Eight of these studies (Bates et al,[26] Bridges et al,[27] Fang et al,[28] Feldman et al,[29] Le et al,[34] Lippi et al,[36] Probst et al,[41] and Waldron et al[46]) additionally supported derivation of the secondary outcome—the cost of test orders—and (along with Tierney et al 1993 [available in Supplemental Table 2]) are further discussed in the "Additional Outcomes Data" subsection of the "Results" section.

An additional five studies were excluded from the review due to poor quality ratings (Manka et al 2013, McArdle et al 2014, Mekhjian et al 2003, Procop et al 2014, and Stair et al 1998 [available in Supplemental Table 2]).

#### *Meta-Analysis*

Nineteen of the 25 studies examining CPOE modifications contained sufficient information to be included in a meta-analysis, with the primary outcome of number of tests. The forest plot in **Figure 3** presents the meta-analysis results.

The summary OR (95% CI) is 0.125 (0.081-0.194) indicating a substantial and statistically significant reduction in the number of tests associated with the intervention. The results are heterogeneous ($I^2$ = 99%, $P$ < .001).

The 19 studies included in the meta-analysis had one of three types of CPOE modification. These are an alert for or a block of redundant tests within a specified time interval, a display of cost of a test at the time of ordering, or a limit of test availability in the CPOE user interface.

Nine of the 19 studies had an alert for, or a block of, tests repeated within a specified time interval (Bansal et al,[24] Bates et al,[25] Bridges et al,[27] Li et al,[35] Lippi et al,[36] Love et al,[37] May et al,[38] Pageler et al,[40] and Waldron et al[46]). The subgroup forest plot for these nine studies is in Supplemental Figure 1. The summary OR (95% CI) for those studies is 0.241


DOI: 10.1093/ajcp/aqx147

**Table 2**
**Body-of-Evidence Qualitative Analysis for Computerized Provider Order Entry Practice**[a]

| Study | Quality Rating | Effect Size Rating |
|---|---|---|
| Bansal et al, 2001[24] | Fair | Minimal |
| Bates et al, 1999[25] | Fair | Moderate[b] |
| Bates et al, 1997[26] | Fair | Minimal |
| Bridges et al, 2014[27] | Good | Moderate[b] |
| Fang et al, 2014[28] | Good | Substantial[b] |
| Feldman et al, 2013[29] | Good | Substantial[b] |
| Georgiou et al, 2011[30] | Fair | Minimal |
| Horn et al, 2014[31] | Fair | Cannot be determined |
| Hwang et al, 2002[32] | Good | Substantial |
| Kahan et al, 2009[33] | Good | Substantial[b] |
| Le et al, 2015[34] | Good | Substantial[b] |
| Li et al, 2014[35] | Fair | Moderate[b] |
| Lippi et al, 2015[36] | Good | Substantial[b] |
| Love et al, 2015[37] | Fair | Minimal[b] |
| May et al, 2006[38] | Good | Minimal[b] |
| Olson et al, 2015[39] | Good | Substantial[b] |
| Pageler et al, 2013[40] | Good | Substantial[b] |
| Probst et al, 2013[41] | Fair | Minimal[b] |
| Procop et al, 2015[42] | Fair | Cannot be determined |
| Shalev et al, 2009[43] | Good | Minimal[b] |
| Solis et al, 2015[44] | Fair | Substantial |
| Vardy et al, 2005[45] | Fair | Minimal |
| Waldron et al, 2014[46] | Good | Substantial[b] |
| Westbrook et al, 2006[47] | Good | Minimal |

[a]Overall strength of evidence of effectiveness rating is "high": 9 studies were good/substantial, 1 study was good/moderate, 1 study was fair/substantial, 2 studies were fair/moderate, 3 studies were good/minimal, 6 studies were fair/minimal, 2 studies were standard effect measure cannot be determined, and 5 studies were excluded.
[b]$P < .05$.

(0.148-0.391), which is a moderate effect that is statistically significant. These articles are heterogeneous ($I^2 = 98.4\%$, $P < .001$).

Three of the 19 studies included in the meta-analysis described the effect of displaying the cost of a test when it was ordered (Bates et al,[26] Fang et al,[28] and Feldman et al[29]). The subgroup forest plot for these studies is in Supplemental Figure 2. The summary OR (95% CI) is 0.028 (<0.001-5.573), which is a substantial effect that is not statistically significant. These studies are heterogeneous ($I^2 = 99.5\%$, $P < .001$), but in the same direction favoring the intervention. Because Bates et al[26] was identified as having significant study design flaws that would result in an underestimate of intervention effect, it was identified as an effect outlier. Summary results for this subgroup are also provided after removing data from this study from meta-analysis. Without Bates et al,[26] the summary OR (95% CI) is 0.0041 (0.0002-0.0827). This is a substantial effect that is statistically significant. The studies are heterogeneous ($I^2 = 88.0\%$, $P = .004$), but in the same direction favoring the intervention.

Seven of the 19 studies included in the meta-analysis addressed the effect of changing test availability on a CPOE interface (Kahan et al,[33] Le et al,[34] Olson et al,[39] Probst et al,[41] Shalev et al,[43] Solis et al,[44] and Vardy et al[45]). The subgroup forest plot for these studies are in Supplemental Figure 3. The summary OR (95% CI) is 0.080 (0.032-0.199), which is a substantial effect that is statistically significant. These studies are heterogeneous ($I^2 > 99.9\%$, $P < .001$).

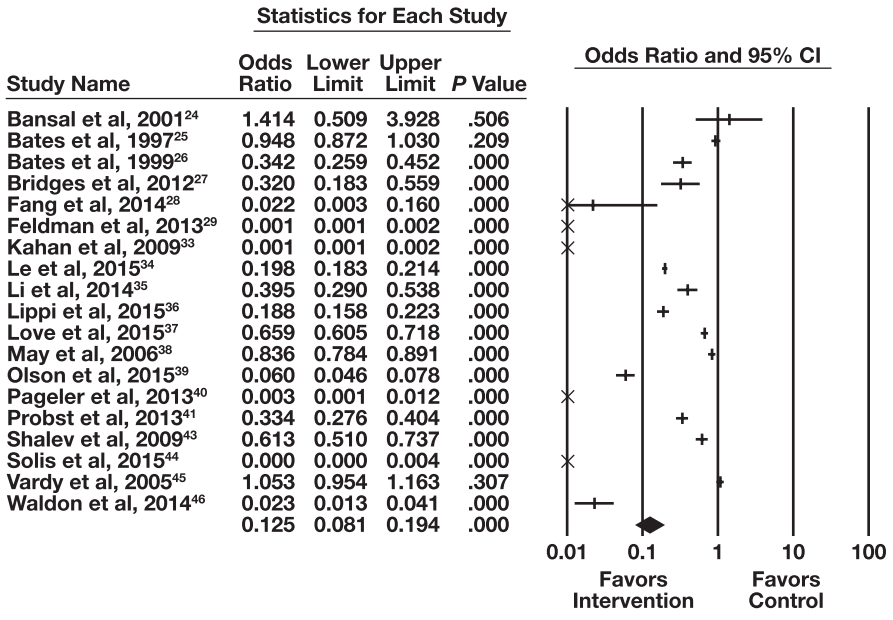


**Figure 3** Forest plot for computerized provider order entry modification studies. CI, confidence interval.



DOI: 10.1093/ajcp/aqx147

### Clinical Decision Support Systems/Tools

Nine studies assessing the effectiveness of CDSS/CDST-alone practices (ie, not in combination with another practice type) were included in this systematic review. Interventions within the CDSS/CDST practice category were subgrouped as reminders of guidelines and proposed testing.

#### *Body-of-Evidence Qualitative Analysis*

Eight studies in the CDSS/CDST practice category were qualitatively analyzed using the LMBP rating criteria, with results summarized in Table 3. They have a "suggestive" rating for the overall strength of the evidence's effectiveness in relation to the primary outcome measure assessed and the number of tests. Tierney et al 1988 (available in Supplemental Table 2) supported derivation only of the secondary outcome measure assessed by this review—the cost of test orders—and is therefore not included in this qualitative analysis.

Among the eight studies assessing the impact of CDSS/CDST, standard effect measures (OR or Cohen *d* with its standard error) could not be determined for two of them (Howell et al[50] and Roukema et al[54]). Thus, six CDSS/CDST practice studies were meta-analyzed in relation to the primary outcome measure assessed. Two of these studies (Nightingale et al[52] and Poley et al[53]) additionally supported derivation of the secondary outcome—the cost of test orders—and (along with Tierney et al 1988 [available in Supplemental Table 2]) are further discussed in the "Additional Outcomes Data" subsection of the "Results" section.

#### *Meta-Analysis*

Six of the nine studies examining CDSS/CDST contained sufficient information to be included in a meta-analysis with the primary outcome of number of tests. The forest plot in Figure 4 presents the meta-analysis results. The summary OR (95% CI) for the CDSS practice is 0.310 (0.141-0.681). This is a moderate effect that is statistically significant. These studies are heterogeneous ($I^2$ = 99.9%, $P$ < .001), but in the same direction favoring the intervention.

The six studies included in the meta-analysis had one of two types of intervention. Three of the studies (Bindels et al,[48] Collins et al,[49] and McKinney et al[51]) had the CDSS/CDST remind the individual ordering a test of guidelines for the utilization of the test for individual patients. The other three studies (Nightingale et al,[52] Poley et al,[53] and vanWijk et al[55]) involved the CDSS/CDST proposing appropriate testing for individual patients.

For the three studies with a reminder of guidelines, the subgroup forest plot is in Supplemental Figure 4. The summary OR (95% CI) for these studies is 0.457 (0.196-1.070). This indicates a minimal effect that is not statistically significant. These studies are heterogeneous ($I^2$ = 95.7%, $P$ < .001), but in the same direction favoring the intervention.

For the three studies with proposed testing for a patient, the subgroup forest plot is in Supplemental Figure 5. The summary OR (95% CI) for these studies is 0.216 (0.069-0.672). This indicates a substantial effect that is statistically significant. These studies are heterogeneous ($I^2$ = 99.9%, $P$ < .001), but in the same direction favoring the intervention.

**Table 3**
**Body-of-Evidence Qualitative Analysis for Clinical Decision Support Systems/Tools Practice**[a]

| Study | Quality Rating | Effect Size Rating |
|---|---|---|
| Bindels et al, 2003[48] | Fair | Minimal |
| Collins et al, 2014[49] | Good | Minimal[b] |
| Howell et al, 2014[50] | Fair | Cannot be determined |
| McKinney et al, 2015[51] | Good | Moderate[b] |
| Nightingale et al, 1994[52] | Fair | Substantial[b] |
| Poley et al, 2007[53] | Good | Minimal[b] |
| Roukema et al, 2008[54] | Fair | Cannot be determined |
| vanWijk et al, 2001[55] | Good | Substantial[b] |

[a]Overall strength of evidence of effectiveness rating is "suggestive": 1 study was good/substantial, 1 study was good/moderate, 1 study was fair/substantial, 2 studies were good/minimal, 1 study was fair/minimal, and 2 studies were standard effect measure cannot be determined.
[b]$P$ < .05.

### Education

Seven studies assessing the effectiveness of education-alone practices (ie, not in combination with another practice type) were included in this systematic review.

#### *Body-of-Evidence Qualitative Analysis*

Seven studies in the education practice category were qualitatively analyzed using the LMBP rating criteria, with results summarized in Table 4. They have a "suggestive" rating for the overall strength of evidence of effectiveness in relation to the primary outcome measure assessed and the number of tests.

For the seven studies assessing the impact of education, a standard effect measure (OR or Cohen *d* with its standard error) could not be determined for one of them (Chonfhaola et al[58]). Thus, six education studies were meta-analyzed in relation to the primary outcome measure. Two of these studies (Baral et al[57] and DellaVolpe et al[60]) additionally supported derivation of the secondary outcome—the cost of test orders—are further discussed in the "Additional Outcomes Data" subsection of the "Results" section.


DOI: 10.1093/ajcp/aqx147

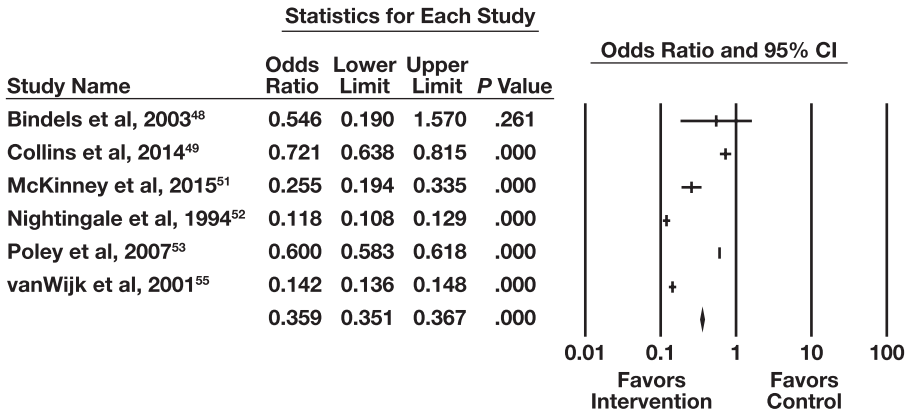


**Figure 4** Forest plot for clinical decision support systems/tools studies. CI, confidence interval.

*Meta-Analysis*

Six of the seven studies examining education contained sufficient information to be included in a meta-analysis with the primary outcome of number of tests. The forest plot in Figure 5 presents the meta-analysis results.

The summary OR (95% CI) are 0.224 (0.127-0.393). This is a moderate effect that is statistically significant. The results are heterogeneous ($I^2$ = 94.8%, $P$ < .001), but in the same direction favoring the intervention.

## Feedback

Seven studies assessing the effectiveness of feedback-alone practices (ie, not in combination with another practice type) were included in this systematic review.

**Table 4**
**Body-of-Evidence Qualitative Analysis for Education Practice[a]**

| Study | Study Quality | Effect Size Rating |
|---|---|---|
| Abio-Valvete et al, 2015[56] | Poor (study excluded) | Cannot be determined |
| Baral et al, 2001[57] | Fair | Substantial[b] |
| Chonfhaola et al, 2013[58] | Fair | Cannot be determined |
| Dawes et al, 2015[59] | Fair | Moderate[b] |
| DellaVolpe et al, 2014[60] | Fair | Minimal[b] |
| Eisenberg, 1977[61] | Good | Substantial[b] |
| Gardezi, 2015[62] | Fair | Minimal[b] |
| Kotecha et al, 2015[63] | Poor (study excluded) | Cannot be determined |
| Thakkar et al, 2015[64] | Good | Minimal[b] |

[a]Overall strength of evidence of effectiveness rating is "suggestive": 1 study was good/substantial, 1 study was fair/substantial, 1 study was fair/moderate, 1 study was good/minimal, 2 studies were fair/minimal, 1 study was standard effect measure cannot be determined, and 2 studies were excluded.
[b]$P$ < .05.

*Body-of-Evidence Qualitative Analysis*

Seven studies in the feedback practice category were qualitatively analyzed using the LMBP rating criteria, with results summarized in Table 5. They have a "suggestive" rating for the overall strength of evidence of effectiveness in relation to the primary outcome measure assessed and the number of tests.

For the seven studies assessing the impact of feedback, standard effect measures (OR or Cohen $d$ with its standard error) could not be determined for four of them (Gama et al,[67] Verstappen et al,[69] Verstappen et al,[70] and Winkens et al[71]). Thus, three feedback studies were meta-analyzed in relation to the primary outcome.

*Meta-Analysis*

Three of the seven studies examining feedback provided sufficient information to be included in the meta-analysis, with the primary outcome of number of tests. The forest plot in Figure 6 presents the meta-analysis results.

The summary OR (95% CI) is 0.116 (0.003-4.599). This indicates a substantial effect that is not statistically significant. The results are heterogeneous ($I^2$ = 98.9%, $P$ < .001), but in the same direction favoring the intervention.

## Reflex Testing

Five studies assessing the effectiveness of reflex testing-alone practices (ie, not in combination with another practice type) were included in this systematic review.

*Body-of-Evidence Qualitative Analysis*

Five studies in the reflex testing practice category were qualitatively analyzed using the LMBP rating criteria, with results summarized in Table 6. They have a



DOI: 10.1093/ajcp/aqx147

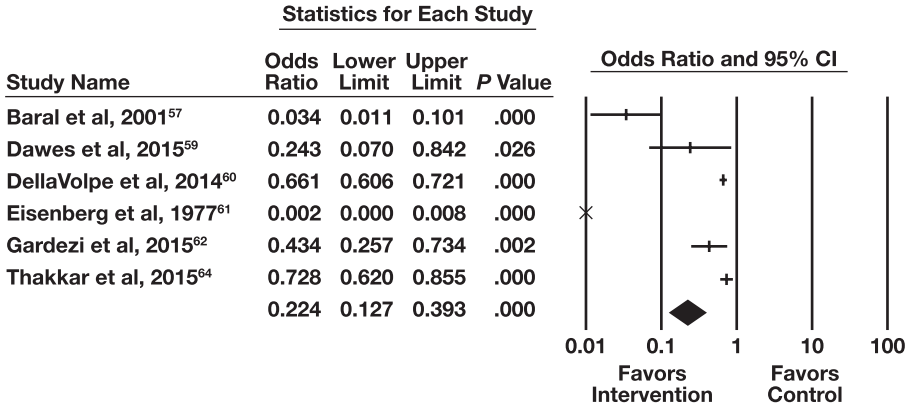


**Figure 5** Forest plot for education studies. CI, confidence interval.

"moderate" rating for the overall strength of evidence of effectiveness in relation to the primary outcome measure assessed and the number of tests.

For the five studies assessing the impact of reflex testing, standard effect measures (OR or Cohen *d* with its standard error) could not be determined for three of them (Bonaguri et al,[73] Tampoia et al,[76] and Wu et al[78]). Thus, two reflex testing studies were meta-analyzed in relation to the primary outcome measure.

*Meta-Analysis*

Two of the five eligible studies had sufficient information to be included in the meta-analysis, with the primary outcome measure of number of tests. The forest plot in Figure 7 presents the meta-analysis results.

The summary OR (95% CI) is 0.008 (0.003-0.022). This indicates a substantial effect that is statistically significant. The results are homogeneous ($I^2$ < 0.001%, $P$ = .513).

### Test Review

Five studies assessing the effectiveness of test review-alone practices (ie, not in combination with another practice type) were included in this systematic review.

**Table 5**
**Body-of-Evidence Qualitative Analysis for Feedback Practice**[a]

| Study | Quality Rating | Effect Size Rating |
|---|---|---|
| Baker et al, 2003[65] | Good | Minimal |
| Bunting et al, 2004[66] | Good | Substantial[b] |
| Gama et al, 1992[67] | Fair | Cannot be determined |
| Miyakis et al, 2006[68] | Good | Minimal[b] |
| Verstappen et al, 2004[69] | Fair | Cannot be determined |
| Verstappen et al, 2004[70] | Fair | Cannot be determined |
| Winkens et al, 1992[71] | Fair | Cannot be determined |

[a]Overall strength of evidence of effectiveness rating is "suggestive": 1 study was good/substantial, 2 studies were good/minimal, and 4 studies were standard effect measure cannot be determined.
[b]$P$ < .05.

*Body-of-Evidence Qualitative Analysis*

Five studies in the test review practice category were qualitatively analyzed using the LMBP rating criteria, with results summarized in Table 7. They have an "insufficient" rating for the overall strength of the evidence's effectiveness in relation to the primary outcome measure assessed and the number of tests.

For the five studies assessing the impact of test review, standard effect measures (OR or Cohen *d* with its standard error) could not be determined for two of them (Chu et al[81] and Miller et al[84]). Thus, three test review studies were meta-analyzed in relation to the primary outcome measure.

*Meta-Analysis*

Three of the five eligible studies in this practice category had sufficient information to be included in the meta-analysis, with the primary outcome of number of tests. The forest plot in Figure 8 presents the meta-analysis results.

The summary OR (95% CI) is 0.388 (0.314-0.480). This indicates a moderate effect that is statistically significant. The results are homogeneous ($I^2$ < 0.001%, $P$ = .851).

### Laboratory Test Utilization Team

No studies were found for inclusion assessing the effectiveness of an LTU team as the only practice assessed, in isolation. LTU was observed only in combination with other practices (eg, education and LTU). The combined practices section provides reference to a comparative plot assessing the impact of combined practices with, and those without, an LTU component.

### Combined Practices

Twenty-five studies assessing the effectiveness of combinations of practices were included in this systematic review. While not all possible combinations


DOI: 10.1093/ajcp/aqx147

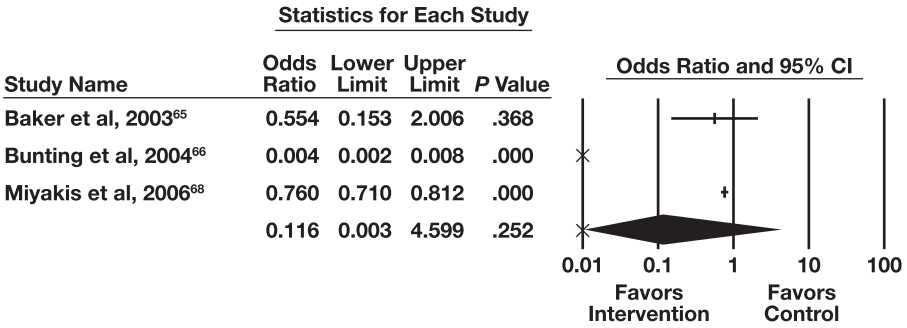


**Figure 6** Forest plot for feedback studies. CI, confidence interval.

of practices were observed in the available evidence base, **Table 8** provides detail on the combinations that were observed.

*Body-of-Evidence Qualitative Analysis*

Twenty-five studies in the combined practices category were qualitatively analyzed using the LMBP rating criteria, with results summarized in Table 8. They have a "moderate" rating for the overall strength of evidence of effectiveness in relation to the primary outcome measure assessed: number of tests.

For the 25 studies assessing the impact of combined practices, standard effect measures (OR or Cohen *d* with its standard error) could not be determined for 13 of them (Gilmour et al,[89] Hutton et al,[91] Larochelle et al,[94] Lum,[95] MacPherson et al,[96] Riley et al,[100] Roggeman et al,[101] Rosenbloom et al,[102] Spiegel et al,[104] Vegting et al,[107] Vidyarthi et al,[108] Wang et al,[109] and Warren[110]). Thus, 12 combined practices studies were meta-analyzed in relation to the primary outcome measure.

**Table 6**
**Body-of-Evidence Qualitative Analysis for Reflex Testing Practice**[a]

| Study | Quality Rating | Effect Size Rating |
|---|---|---|
| Baird et al, 2009[72] | Good | Substantial[b] |
| Bonaguri et al, 2011[73] | Fair | Cannot be determined |
| Cernich et al, 2014[74] | Poor (study excluded) | Cannot be determined |
| Froom et al, 2012[75] | Good | Substantial[b] |
| Tampoia et al, 2007[76] | Fair | Cannot be determined |
| VanWalraven et al, 2002[77] | Poor (study excluded) | Cannot be determined |
| Wu et al, 1999[78] | Fair | Cannot be determined |

[a]Overall strength of evidence of effectiveness rating is "moderate": 2 studies were good/substantial, 3 studies were standard effect measure cannot be determined, and 2 studies were excluded.
[b]$P < .05$.

*Meta-Analysis*

Twelve of the 25 studies examining combined practices had sufficient information to be included in the meta-analysis, with the primary outcome of number of tests. The forest plot in **Figure 9** presents the meta-analysis results.

The summary OR (95% CI) is 0.411 (0.149-1.131). This indicates a minimal effect that is not statistically significant. These studies are heterogeneous ($I^2 > 99.9\%$, $P < .001$), but in the same direction favoring the intervention.

Three additional forest plots and three comparison plots are now provided. These serve to indicate the effect of select groupings of combined practice studies (inclusion of education and feedback, inclusion of CDSS and/or CPOE modifications, and inclusion of LTU), and to compare their effects to single practice studies.

Of the 12 combined practices studies in the meta-analysis, eight included both education and feedback. The subgroup forest plot for those studies is in Supplemental Figure 6. The summary OR (95% CI) is 0.292 (0.097-0.880). This indicates a moderate effect that is statistically significant. These studies are heterogeneous ($I^2 = 99.9\%$, $P < .001$), but in the same direction favoring the intervention.

These eight studies with an education and feedback component were compared to the studies in the education practice (studies assessing the effect of education alone) and to studies in the feedback practice (feedback alone). That comparison is illustrated in Supplemental Figure 7. There is no statistically significant difference between the OR for the combined practices studies containing both and education and feedback component, and the education alone studies ($P = .775$) or the feedback alone studies ($P = .847$).

Of the 12 studies in the combined practice meta-analysis, three included a CPOE modification and/or CDSS/CDST component. The forest plot for these studies is in Supplemental Figure 8. The summary OR (95% CI) is 0.750 (0.552-1.018). This indicates a minimal effect that is not statistically significant. These studies are homogeneous ($I^2 < 0.001\%$, $P = .546$).



DOI: 10.1093/ajcp/aqx147

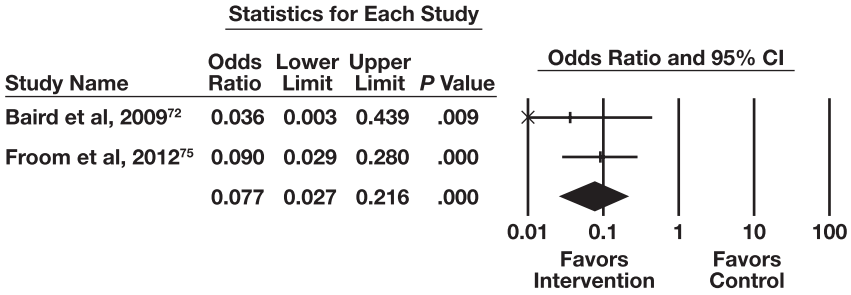


**Figure 7** Forest plot for reflex testing studies. CI, confidence interval.

These three studies were compared to studies in the CPOE practice (CPOE alone) and to studies in the CDSS/CDST practice (CDSS alone). That comparison is illustrated in Supplemental Figure 9, with CDSS and/or CPOE combinations referred to in the forest plot as "health IT." There is no statistically significant difference between the ORs for the health IT studies and the CDSS alone studies ($P = .225$). The difference between the ORs for the combined practice studies and the CPOE alone studies, however, is statistically significant ($P < .001$).

Of the 12 studies in the combined practice meta-analysis, three included an LTU component. The forest plot for these studies is in Supplemental Figure 10. The summary OR (95% CI) is 0.412 (0.555-3.046). This indicates a moderate effect that is statistically significant. These studies are heterogeneous ($I^2 > 99.9\%$, $P < .001$), but in the same direction favoring the intervention.

These three were compared to the nine combined practices studies that did not include LTU. That comparison is illustrated in Supplemental Figure 11. There is no statistically significant difference between the ORs for studies that include LTU and those that do not include LTU ($P = .953$).

**Table 7**
**Body-of-Evidence Qualitative Analysis for Test Review Practice[a]**

| Study | Quality Rating | Effect Size Rating |
|---|---|---|
| Aesif et al, 2015[79] | Fair | Minimal[b] |
| Barazzoni et al, 2002[80] | Fair | Moderate[b] |
| Chu et al, 2013[81] | Fair | Cannot be determined |
| Dickerson et al, 2014[82] | Fair | Minimal[b] |
| Dolazel et al, 2015[83] | Poor (study excluded) | Cannot be determined |
| Miller et al, 2014[84] | Fair | Cannot be determined |

[a]Overall strength of evidence of effectiveness rating is "insufficient": 1 study was fair/moderate, 2 studies were fair/minimal, 2 studies were standard effect measure cannot be determined, and 1 study was excluded.
[b]$P < .05$.

### Additional Outcomes Data

Data available in the current evidence base supported determinations of cost of tests outcome, this review's secondary outcome measure. Meta-analysis forest-plots, as well as the strength of evidence ratings, for cost data are provided in Supplemental Figures 12 to 15. Sixteen of 83 studies included in this review could be meta-analyzed for the secondary outcome, cost of tests. Fourteen of these 16 studies were also meta-analyzed for the number of tests outcome. Tierney et al 1988 and Tierney et al 1993 were meta-analyzed only for cost of tests. Cost outcome definitions for these 16 studies are provided in Supplemental Table 6. Limitations of the cost of tests outcome measure are provided in the "Limitations" section.

Supplemental Table 6 conveys other outcomes-related data present in the eligible evidence base and conveys *P* values (when reported) for the 30 studies that could not be meta-analyzed in relation to the review's primary or secondary outcome measure. Of these 30 studies, 19 report *P* values; of these 19 studies, 18 indicated a statistically significant favorable impact of the utilization management practice intervention.

### Applicability and Feasibility Data

Two pie charts and three cross-attribute tables support more detailed discussion in the "Applicability and Generalizability" subsection of the "Discussion" section. **Figure 10** and **Figure 11** indicate the distribution of facility setting and practice category across all included studies, while **Table 9**, **Table 10**, and **Table 11** provide additionally detail on the landscape of two study-level attributes—facility setting and patient setting—relative to practice categories (CPOE, CDSS, etc). Definitions for facility setting types are provided in the glossary (Supplemental Table 1). For a consolidated list of interventions (including practice composition of "combined practice" studies), targeted testing, and targeted clinical staff, readers are referred to Supplemental Table 5. Finally, **Table 12** provides additional information as to distribution of study-level characteristics within the assessed evidence base.


DOI: 10.1093/ajcp/aqx147

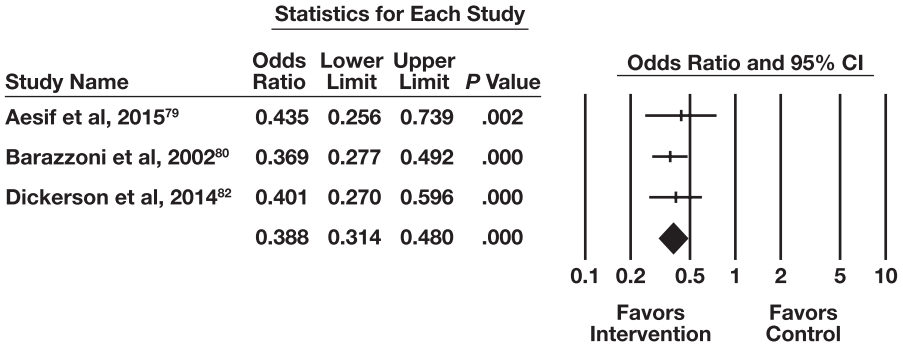


**Figure 8** Forest plot for test review studies. CI, confidence interval.

## Discussion

### Additional Benefits and Economic Evaluation

Several subsections of this section serve to further inform implementation decisions, while indicating limitations of the current evidence base. Frameworks guiding decision-making exist, such as GRADE (Grading of Recommendations Assessment, Development, and Evaluation Working Group) Evidence to Decision frameworks, and may be applied by health care decision-makers when choosing to adopt (or adapt) recommendations in new contexts.[112]

To this end, this first subsection refers readers to the data provided in the "Additional Outcomes Data" of the "Results" section, suggesting benefit beyond this review's primary outcome ("number of tests"). These data suggests statistically significant favorable impact (via *P* values) for studies that could not be meta-analyzed. While the patient-related outcomes provided by some of the

**Table 8**
**Body-of-Evidence Qualitative Analysis for Combined Practice[a]**

| Study | Practice Combination | Quality Rating | Effect Size Rating |
|---|---|---|---|
| Bareford et al, 1990[85] | Education, feedback | Poor (study excluded) | Cannot be determined |
| Baricchi et al, 2012[86] | Education, LTU | Fair | Minimal[b] |
| Calderon-Margalit et al, 2005[87] | Education, feedback, LTU | Good | Substantial[b] |
| Dowling et al, 1989[88] | Education, feedback | Good | Minimal[b] |
| Gilmour et al, 2015[89] | Education, reflex | Fair | Cannot be determined |
| Isofina et al, 2013[90] | CDSS, education | Poor (study excluded) | Cannot be determined |
| Hutton et al, 2009[91] | CPOE modification, education, LTU | Fair | Cannot be determined |
| Janssens et al, 2015[92] | CPOE modification, LTU | Good | Minimal |
| Kroenke et al, 1987[93] | Education, feedback | Good | Minimal[b] |
| Larochelle et al, 2014[94] | CPOE modification, CDSS, education | Good | Cannot be determined |
| Lum, 2006[95] | Education, test review, LTU | Fair | Cannot be determined |
| MacPherson et al, 2005[96] | Education, LTU | Fair | Cannot be determined |
| McNicoll et al, 2015[97] | Education, feedback | Good | Moderate[b] |
| Minerowicz et al, 2015[98] | Education, feedback | Good | Substantial[b] |
| Newman et al, 2015[99] | Education, feedback | Good | Moderate |
| Riley et al, 2015[100] | CPOE modification, test review | Fair | Cannot be determined |
| Roggeman et al, 2014[101] | CDSS, education | Fair | Cannot be determined |
| Rosenbloom et al, 2005[102] | CPOE modification, CDSS | Fair | Cannot be determined |
| Samuelson et al, 2015[103] | CDSS, reflex | Good | Minimal |
| Spiegel et al, 1989[104] | Feedback, LTU | Fair | Cannot be determined |
| Thomas et al, 2006[105] | Education, feedback | Fair | Minimal[b] |
| Tomlin et al, 2011[106] | Education, feedback | Good | Minimal[b] |
| Vegting et al, 2012[107] | CPOE modification, eduaction, feedback | Fair | Cannot be determined |
| Vidyarthi et al, 2015[108] | Education, feedback | Good | Cannot be determined |
| Wang et al, 2002[109] | CDSS, education, LTU | Good | Cannot be determined |
| Warren, 2013[110] | CDSS, LTU | Fair | Cannot be determined |
| White et al, 2013[111] | CPOE modification, CDSS | Good | Minimal |

CDSS, clinical decision support systems; CPOE, computerized provider order entry; LTU, laboratory test utilization.
[a]Overall strength of evidence of effectiveness rating is "moderate": 2 studies were good/substantial, 2 studies were good/moderate, 6 studies were good/minimal, 2 studies were fair/minimal, 13 studies were standard effect measure cannot be determined, and 2 studies were excluded.
[b]$P < .05$.



DOI: 10.1093/ajcp/aqx147

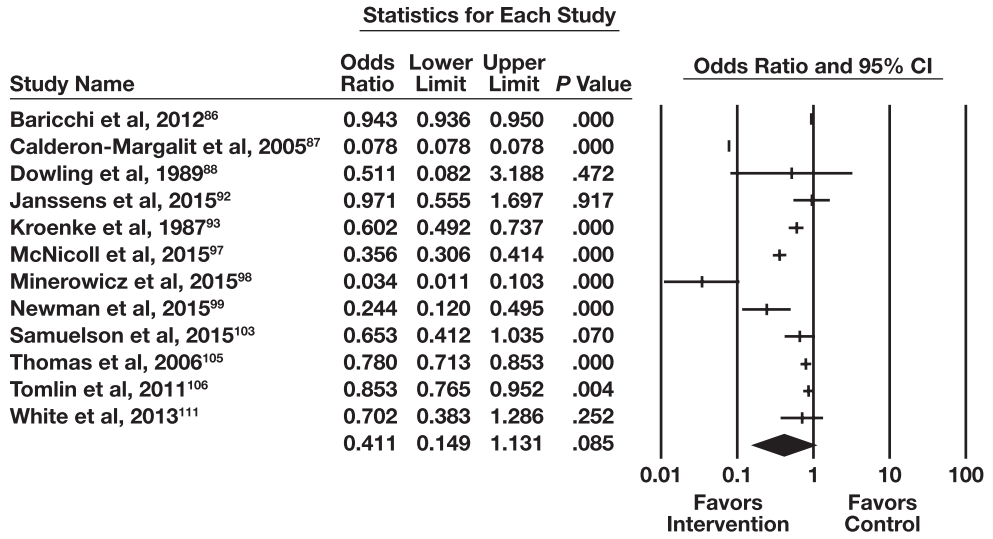


**Figure 9** Forest plot for combined practice studies. CI, confidence interval.

studies are distal outcomes, impacted by many factors during the course of care, the following pattern arises: the rates of the patient-important outcomes assessed (eg, morbidity, mortality, and length of stay) were not significantly different pre- or postintervention, therefore interventions apparently were not associated with adverse impact on these outcomes. Additional discussion on limitations associated of the patient-related outcome measures encountered in the current evidence base appears in the "Limitations" section.

Additionally, while concerns associated with the economic evaluation supported by the current evidence base are also noted in the "Limitations" section, what the meta-analyzed "costs of tests"

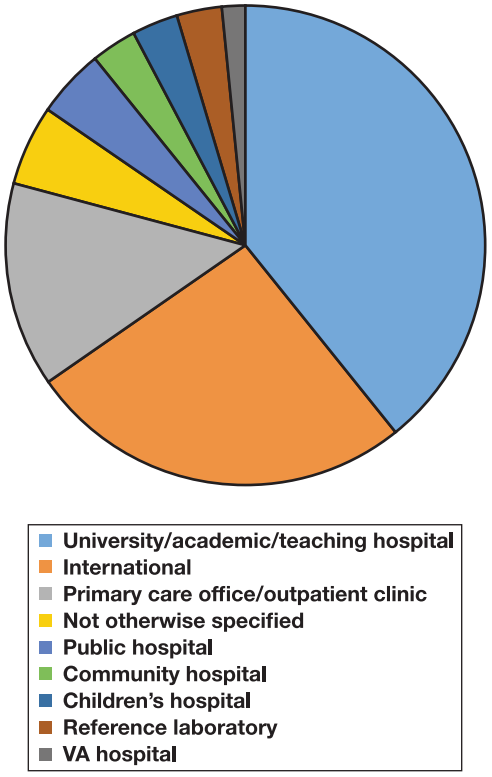


**Figure 10** Pie chart for facility setting across all included studies. VA, Veterans Affairs.

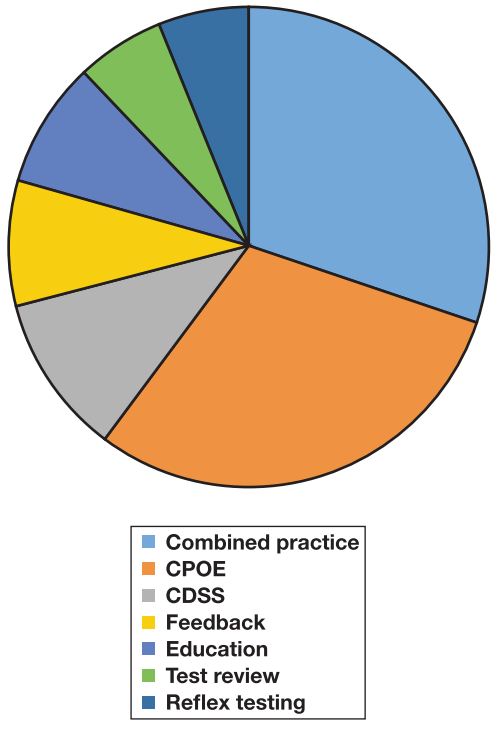


**Figure 11** Pie chart for practice category across all included studies. CDSS, clinical decision support systems; CPOE, computerized provider order entry.


DOI: 10.1093/ajcp/aqx147

**Table 9**
**Practice to Facility Setting Cross-Attribute Table**

| | Practice Category | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Facility Setting** | **CDSS** | **CPOE** | **Combined** | **Education** | **Feedback** | **Reflex Testing** | **Test Review** | **Total[a]** |
| University/academic/teaching hospital | 3 | 17 | 19 | 5 | 1 | 4 | 3 | 52 |
| Children's hospital | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 4 |
| Community hospital | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 4 |
| Public hospital | 1 | 3 | 2 | 0 | 0 | 0 | 0 | 6 |
| VA hospital | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 2 |
| Primary care office/outpatient clinic | 4 | 4 | 4 | 0 | 5 | 1 | 0 | 18 |
| Reference laboratory | 0 | 0 | 0 | 0 | 2 | 1 | 1 | 4 |
| International (outside US) | 5 | 8 | 9 | 1 | 6 | 3 | 2 | 34 |
| Not otherwise specified | 1 | 1 | 2 | 1 | 1 | 1 | 0 | 7 |
| Total[a] | 17 | 37 | 36 | 8 | 15 | 11 | 7 | 131 |

CDSS, clinical decision support systems; CPOE, computerized provider order entry; VA, Veterans Affairs.
[a]The totals may (and do) include the same study more than once if the study authors reported more than one setting.

outcome does provide is complementary indication of the proximal effects of these interventions, expressed as cost-savings.

### Applicability and Generalizability

From the pie charts in the "Applicability and Feasibility Data" subsection of the "Results" section, the following patterns emerge: (1) academic/university/teaching as well as primary care/outpatient clinic facility settings are strongly represented, while community hospitals, children's hospitals, and public hospitals (including Veterans Affairs [VA] hospitals) are less represented, and (2) education, feedback, reflex testing, and test review interventions are less represented as single-practice interventions compared to CPOE and CDSS. From the cross-attribute tables, observable in Table 9 is that within university/academic/teaching hospitals CPOE and combined practice interventions are more strongly represented relative to other intervention types. Because CPOE interventions most commonly appeared as modifications to an existing CPOE system (see "Results" section), this finding is perhaps reflective of increased resource availability in such settings in support of improvement efforts, a point perhaps also true of combined practice interventions (which may require more resources to plan and implement relative to single-practice interventions). A pattern observable in Table 10 is that most interventions occurred in hospital inpatient settings, perhaps

**Table 10**
**Practice to Patient Setting Cross-Attribute Table**

| | Practice Category | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Patient Setting** | **CDSS** | **CPOE** | **Combined** | **Education** | **Feedback** | **Reflex Testing** | **Test Review** | **Total[a]** |
| Emergency department | 4 | 3 | 1 | 2 | 0 | 1 | 2 | 13 |
| Hospital inpatient and outpatient | 4 | 4 | 7 | 2 | 2 | 3 | 2 | 24 |
| Hospital inpatient only | 3 | 23 | 16 | 3 | 2 | 3 | 2 | 52 |
| Hospital outpatient only | 0 | 1 | 5 | 1 | 0 | 2 | 0 | 9 |
| Primary care/clinic outpatient | 6 | 6 | 7 | 0 | 11 | 2 | 1 | 33 |
| Total[a] | 17 | 37 | 36 | 8 | 15 | 11 | 7 | 131 |

CDSS, clinical decision support systems; CPOE, computerized provider order entry.
[a]The totals may (and do) include the same study more than once if the study authors reported more than one setting.



DOI: 10.1093/ajcp/aqx147

**Table 11**
**Facility Setting to Patient Setting Cross-Attribute Table**

| | Patient Setting | | | | | |
|---|---|---|---|---|---|---|
| **Facility Setting** | **Emergency Department** | **Hospital Inpatient and Outpatient** | **Hospital Inpatient Only** | **Hospital Outpatient Only** | **Primary Care/ Outpatient Clinic** | **Total[a]** |
| University/academic/ teaching hospital | 4 | 12 | 32 | 4 | 0 | 52 |
| Children's hospital | 1 | 0 | 3 | 0 | 0 | 4 |
| Community hospital | 0 | 2 | 1 | 1 | 0 | 4 |
| Public hospital | 2 | 1 | 3 | 0 | 0 | 6 |
| VA hospital | 1 | 0 | 1 | 0 | 0 | 2 |
| Primary care office/outpatient clinic | 0 | 0 | 0 | 0 | 18 | 18 |
| Reference laboratory | 0 | 0 | 1 | 0 | 3 | 4 |
| International (outside US) | 4 | 6 | 9 | 3 | 12 | 34 |
| Not otherwise specified | 1 | 3 | 2 | 1 | 0 | 7 |
| Total[a] | 13 | 24 | 52 | 9 | 33 | 131 |

VA, Veterans Affairs.
[a]The totals may (and do) include the same study more than once if the study authors reported more than one setting.

**Table 12**
**Distribution of Attributes Across All Included Studies**

| Attribute | Value | Percentage |
|---|---|---|
| Article type | Abstract only | 7.2% (6/83) |
| | Full article | 92.8% (77/83) |
| Article year | ≥2006 | 68.7% (57/83) |
| | <2006 | 31.3% (26/83) |
| Country | US | 59% (49/83) |
| | Non-US | 41% (34/83) |
| Study design | Postintervention only | 8.4% (7/83) |
| | Retrospective before and after | 8.4% (7/83) |
| | Retrospective before and after, with concurrent control | 4.8% (4/83) |
| | Prospective postintervention group, with historical control | 13.3% (11/83) |
| | Before and after without concurrent control | 36.2% (30/83) |
| | Before and after with concurrent control | 2.4% (2/83) |
| | Other quasi-experimental | 15.7% (13/83) |
| | randomized trial | 10.8% (9/83) |
| Facility size | ≥300 beds | 37.4% (31/83) |
| | <300 beds | 3.6% (3/83) |
| | Unclear/not reported | 59% (49/83) |
| Funding | Funded | 33.7% (28/83) |
| | Not funded | 13.3% (11/83) |
| | Unclear/not reported | 53% (44/83) |
| Investigator background | Pathology/laboratory only | 9.6% (8/83) |
| | Nonpathology/ laboratory only | 45.8% (38/83) |
| | Both | 36.2% (30/83) |
| | Unclear/not reported | 8.4% (7/83) |

reflecting that (1) most testing occurs in hospital labs (a point supported by basic Centers for Medicare and Medicaid Services data), and (2) inpatients may experience increased testing in terms of scope or frequency.

Finally, a pattern emerging from Table 11 is that there is greater focus on primary care/outpatient clinic patient settings among studies outside the US relative to other patient setting types, possibly a reflection of intervention planning occurring in different reimbursement environments (eg, national health care systems) than in the US.

Further, a few studies were identified from the meta-analysis forest plots as representing effect outliers, such that the desired effect of the study was not only "substantial," but was also stronger relative to other studies quantitatively combined within the same practice category. Characteristics of these studies may be generalizable. Three CPOE modification studies were effect outliers: Pageler et al[40] (a "good" quality study) exhibited strong administrative support for the intervention; Waldron et al[46] (a "good" quality study) utilized a hard stop (while hard stops may not be appropriate in many cases, it may be appropriate for some types of esoteric testing, for example); and Solis et al[44] (a "fair" quality study) sent multiple emails and reminders to test ordering clinicians, informing about the change to the CPOE system. One education study was an effect outlier: Eisenberg[61] (a "good" quality study) exhibited an education intervention of considerable duration.

Finally, two combined practice studies were effect outliers. Calderon-Margalit et al[87] (a "good" quality study, combining education, feedback, and LTU) exhibited strong administrative restrictions and control within the intervention. Minerowicz et al[98] (a "good" quality study, combining education and feedback) exhibited


DOI: 10.1093/ajcp/aqx147

continued repetition of the feedback component. These findings are consistent with existing claims that the impact of an intervention may benefits by strength (eg, repetition, substantial administrative support, etc) of intervention components,[11,113,114] although whether this was a direct cause of these two studies having stronger effect is unclear, as many organizational and process variables may impact magnitude of effect.

### Feasibility of Implementation

The practices evaluated in this systematic review are generally feasible for implementation in most health care settings, though ease of implementation may vary from setting to setting. Drivers of implementation feasibility may include organizational and management structures, the business environment and the laboratory's business process needs (including financial targets and budget considerations), the patient population being served through the provision of laboratory testing, and staff competencies. Implementation of any new practice in a hospital setting may therefore encounter challenges due to budgets, programming resources, training needs, and commitment/follow up from key stakeholders. Ultimately, the successful implementation of practices depends on demonstrating meaningful impact on the quality of care and patient outcomes, as well as potential cost effectiveness, in a way that complements administrative and clinical objectives. Examples of implementation challenges encountered within the included evidence base are listed in Supplemental Table 7.

### Associated Harms

Some of the practices evaluated in this review may have unintended impacts on patient care or practitioner diagnostic workflows. In addition, a concern for implementation of LTU teams and test review resulting is the possible delay of timely patient care causing patient dissatisfaction and potential increase in hospital stay. Further, the ability to display test cost, definitions of test cost, and capacity to keep this information up to date, may vary from setting to setting, and in relation to local practice standards. A comprehensive LTU approach should evaluate the merit of planned test utilization practice interventions and evaluate the impact of any change in laboratory test ordering process or procedures, including the possibility of unintended systems and/or patient outcomes, prior to implementation. Some of the examples of implementation challenges listed in Supplemental Table 7 may also inform potential associated harms.

## Conclusion

### Practice Recommendations

Recommendations are categorized as "recommend," "not recommended," and "no recommendation for or against due to insufficient evidence." Recommendation categorizations in this review are a function of the current available evidence base, and of the LMBP methodology, including a priori analysis criteria (eg, selected effect measure rating cutoffs, the LMBP quality assessment tool, and the LMBP strength of body of evidence matrix). Recommendations are made for single practice categories, as well as for the combined practice category. The approach for recommendation categorization is described in the "Materials and Methods" section, with criteria indicated in Table 1.

For practices with a "no recommendation for or against" categorization, this finding does not rule out the potential current value of these practices; rather, it indicates a need for additional studies evaluating the effect of these practices, as well as a need for greater access to relevant unpublished studies. Despite there being no firm "best practice" recommendations for most practices, the data do suggest the practices have the potential to promote appropriate test utilization, such that additional research should be pursued.

Given the current evidence base, and the data available to support the LMBP qualitative analyses and the meta-analyses, recommendations arising from analyses are made in relation to evidence-based assessing practices targeting quality issues that are locally determined (ie, within the health care setting of the individual study) to represent inappropriate test "overutilization" and are made in relation to this review's primary outcome measure, "number of tests." Recommendations arising from this systematic review do not serve to endorse specific appropriate use protocols guiding individual investigators in auditing appropriateness of test utilization (ie, organizational guidelines, local consensus guidelines, algorithms, pathways, appropriateness criteria, and local administrative directives on utilization). Rather, recommendations arising from this systematic review relate to the means (CDSS, CPOE, education, feedback, etc) of supporting/promoting appropriate test utilization; in other words, recommendations relate to utilization management practices.

In instances where health care organizations choose to implement a practice (or a combination of practices) with the goal of test utilization management, we recommend the quality and applicability of available guidelines (local and national) and protocols guiding utilization



DOI: 10.1093/ajcp/aqx147

appropriateness be carefully assessed and validated in the local setting, to further evaluate potential patient harms relative to benefits. Decision to implement should include review by the institution's medical executive committee (or equivalent), involve a use-case (eg, practice impact modeling) within the institution to assess impact before implementation, and involve feedback informing continuous quality improvement.

Practice recommendations are summarized in Table 13, with additional detail provided in the remainder of this section.

### Recommendation for Computerized Provider Order Entry Practices

Use of CPOE is recommended as a best practice to support appropriate clinical LTU.

The overall strength of the evidence's effectiveness of the practice for utilization management is rated as high. The evidence base analyzed to arrive at this recommendation examined modifications to existing CPOE systems (not CPOE replacement of written test orders). The pooled effect size rating (OR = 0.125, 95% CI = 0.081-0.194, $P < .001$) for 19 meta-analyzed studies is substantial for the number of tests outcome. Effects across studies were not consistent ($I^2 = 99.3$ %, $P < .001$); however, all studies demonstrated a favorable effect for the number of tests outcome.

**Table 13**
**Summary of Practice Recommendations**

| Practice Category | Practice Recommendation |
|---|---|
| CPOE | Use of CPOE is recommended as a best practice to support appropriate clinical LTU |
| CDSS/CDST | No recommendation for or against due to insufficient evidence is made for CDSS/CDST as a best practice to support appropriate clinical LTU |
| Education | No recommendation for or against due to insufficient evidence is made for education as a best practice to support appropriate clinical LTU |
| Feedback | No recommendation for or against due to insufficient evidence is made for feedback as a best practice to support appropriate clinical LTU |
| Reflex testing | Use of reflex testing practices is recommended as a best practice to support appropriate clinical LTU |
| Test review | No recommendation for or against due to insufficient evidence is made for test review as a best practice to support appropriate clinical LTU |
| LTU team | No recommendation for or against due to insufficient evidence is made for LTU team as a best practice to support appropriate clinical LTU |
| Combined practices | Use of combined practices is a recommended as a best practice to support appropriate clinical LTU |

CDSS/CDST, clinical decision support systems/tools; CPOE, computerized provider order entry; LTU, laboratory test utilization.

### Recommendation for Clinical Decision Support Systems/Tools Practices

No recommendation for or against due to insufficient evidence is made for CDSS/CDST as a best practice to support appropriate clinical LTU.

The overall strength of evidence of effectiveness of the practice for utilization management is rated as suggestive. The pooled effect size rating (OR = 0.310, 95% CI = 0.141-0.681, $P = .004$) for six meta-analyzed studies is moderate for the number of tests outcome. Effects across studies were not consistent ($I^2 = 99.9\%$, $P < .001$); however, all studies demonstrated a favorable effect for the number of tests outcome.

### Recommendation for Education Practices

No recommendation for or against due to insufficient evidence is made for education as a best practice to support appropriate clinical LTU.

The overall strength of the evidence's effectiveness of the practice for utilization management is rated as suggestive. The pooled effect size rating (OR = 0.224, 95% CI = 0.127-0.393, $P < .001$) for six meta-analyzed studies is moderate for the number of tests outcome. Effects across studies were not consistent ($I^2 = 94.8\%$, $P < .001$); however, all studies demonstrated a favorable effect for the number of tests outcome.

### Recommendation for Feedback Practices

No recommendation for or against due to insufficient evidence is made for feedback as a best practice to support appropriate clinical LTU.

The overall strength of evidence of effectiveness of the practice for utilization management is rated as suggestive. The pooled effect size rating (OR = 0.116, 95% CI = 0.003-4.599, $P = .252$) for three meta-analyzed studies is substantial for the number of tests outcome. Effects across studies were not consistent ($I^2 = 98.9\%$, $P < .001$), however all studies demonstrated a favorable effect for the number of tests outcome.

### Recommendation for Reflex Testing Practices

Use of reflex testing practices is recommended as a best practice to support appropriate clinical LTU.

The overall strength of evidence of effectiveness of the practice for utilization management is rated as moderate. The pooled effect size rating (OR = 0.008, 95%


DOI: 10.1093/ajcp/aqx147

CI = 0.003-0.022, $P < .001$) for two meta-analyzed studies is substantial for number of tests outcome. Effects across studies were consistent ($I^2 < 0.001\%$, $P = .513$), with all studies demonstrating a favorable effect for the number of tests outcome.

### Recommendation for Test Review Practices

No recommendation for or against due to insufficient evidence is made for test review as a best practice to support appropriate clinical LTU.

The overall strength of evidence of effectiveness of the practice for utilization management is rated as insufficient. The pooled effect size rating (OR = 0.388, 95% CI = 0.314-0.480, $P < .001$) for three meta-analyzed studies is moderate for number of tests outcome. Effects across studies were consistent ($I^2 < 0.001\%$, $P = .851$), with all studies demonstrating a favorable effect for the number of tests outcome.

### Recommendation for Laboratory Test Utilization Team

No recommendation for or against due to insufficient evidence is made for LTU teams as a best practice to support appropriate clinical LTU.

No studies were found for inclusion assessing the effectiveness of an LTU team as the only practice assessed in isolation. LTU was observed only in combination with other practices (eg, education and LTU; refer to Table 8). No recommendation for or against due to insufficient evidence is made for LTU—apart from combination with other practices—as a best practice to support appropriate clinical LTU. Subgroup analyses for practice interventions with an LTU component appear in the "Results" section for combined practices.

### Recommendation for Combined Practices

Use of combined practices is recommended as a best practice to support appropriate clinical LTU.

The overall strength of the evidence's effectiveness of the practice for utilization management is rated as moderate. The evidence base analyzed to arrive at recommendation examined was not inclusive of all possible practice combinations (combinations observed are indicated in Table 8). The pooled effect size rating (OR = 0.411, 95% CI = 0.149-1.131, $P = .085$) for 12 meta-analyzed studies is minimal for number of tests outcome. Effects across studies were not consistent ($I^2 > 99.9\%$, $P < .001$); however, all studies demonstrated a favorable effect for the number of tests outcome.

### Limitations

An important limitation of this study was the inability to obtain unpublished data from relevant quality improvement/research efforts. This may have limited this systematic review's ability to (1) achieve greater strength of body-of-evidence ratings for specific practices, (2) evaluate practices in more diverse settings, and (3) better avoid potential for publication bias.

Next, this systematic review highlights the presence in the current evidence base of the limited number of "good" quality studies, largely due to incomplete reporting. In addition to the data collection form discussed in the next section (and provided in Supplemental Table 8), investigators are referred to a centralized repository of available reporting standards at the Enhancing the Quality and Transparency of Health Research Network's website (http://www.equator-network.org/).

Third, a current trend in laboratory practices suggests the important role of LTU teams and somewhat relatedly diagnostic management teams (refer to the Supplemental Table 1 glossary for definitions of LTU teams and diagnostic management teams). However, a limited number of studies were available to determine effectiveness of LTU teams, as this practice invariably appeared in the included evidence base in combination with another utilization management practice.

Fourth, these studies shared the common aim of introducing utilization management practices in order to support rational use of diagnostic resources by clinicians ordering testing, and reduce variability in test ordering behaviors. However, there was considerable variation across studies as to criteria for determining the presence of inappropriate test utilization through utilization audit, as well as how thoroughly investigators described this criteria (and its source), along with efforts to validate criteria in the local setting. This review did not seek to document the source of criteria used nor assess the validity of the criteria used. Rather, it focused on utilization management approaches in local settings where inappropriate test utilization was determined to be occurring. Future systematic review updates may incorporate a component to assess the quality/validity of inappropriate utilization criteria utilized within a study.

Fifth, the primary outcome measure supported by the current evidence base (such that an outcome could be meta-analyzed across a majority of studies) was number of tests, which has limitations as an outcome. While it is a reflection of resource utilization, and is a measure proximal to the interventions assessed, it neglects to include information reflecting the proportion of testing that was concordant with appropriate utilization criteria. Furthermore, there is not a formal assessment on whether



DOI: 10.1093/ajcp/aqx147

this reduction resulted from elimination of only inappropriate testing or also included testing that would have been appropriate.

Sixth, the analysis of cost outcomes supported by the current evidence base has limitations. While "costs of tests" may complement the "number of tests" outcomes, it was variably defined and derived by investigators (see Supplemental Table 6), and often represented an ad hoc analysis within the studies. As another expression of resource use, it supports proximal cost-minimization or cost-consequence analyses, neglecting cost-effectiveness. However, as is, it is assumed to be an important measure to decision-makers (a recent study estimated in vitro diagnostic testing to represent 2.3% of all health care spending in the US, or about US$73 billion annually),[4,115] while providing a point of reference for future economic evaluations. While detailed discussion on health care payment models, health care economic evidence, and evaluation approaches (eg, to more clearly depict possible health economic benefits) is beyond the scope of this review, investigators are encouraged to review available resources.[4,115–117]

Seventh, convincingly establishing the impact of test utilization practices on patient-related outcomes, especially those as distal as patient length of stay, morbidity, and mortality, is challenging within the primary evidence base, given such outcomes are influenced by many factors in health care.[2,118] Nevertheless, as supported by the current evidence base, the observation that these interventions are apparently not associated with adverse impact on these outcomes is an important one. While establishing causality, through study designs such as randomized controlled trials, is often unrealistic in such investigations, improved incorporation of "big data" (ie, large data sets and analytics from health information technology/systems, including medical data warehouses) may support more robust outcomes associations, as well as more robust economic evaluations in the context of test utilization management practices.[119]

Alternatively, investigators may focus on surrogate or intermediate patient outcomes more proximal to the interventions; for example, outcomes or metrics relating to effect on patient management.[120–125] While such measures have limitations of their own, they may include diagnostic yield, time to treatment, etc, as may better support claims that use of such interventions will lead to improved patient health outcomes. Further, increased use of intermediate outcomes, such as rates of guideline-concordant testing among the primary evidence base, would also benefit the evidence base, as would other discrepant analyses demonstrated, for example, by Larochelle et al[94] and Le et al.[34] Additional discussion on laboratory-related outcomes can be found in the literature.[118,126]

Eighth, for the combined practice category, all possible combinations of practices were obviously not observed in the eligible evidence base, impacting generalizability of combined practice recommendations. Further, while the intensity/frequency of individual components within a combined practice was often not explicitly clear, in general there appeared to be a pattern that frequency/intensity of practice components may have been less than that of individual practice studies, potentially contributing to lower effect sizes among the combined practice studies. This does not rule out the potential increased value of combining practices to impact test ordering practices but is a reflection of the current evidence base and the way in which combined practices were planned and/or implemented. A prior systematic review suggests benefit in combining practices (expressed in the review as "multidimensional interventions") to impact test ordering behaviors.[12] Further, it is reasonable to assume there is some overlap of practices otherwise depicted by investigators as using a single practice (eg, CDSS alerts might be assumed to have an educational influence, while not explicitly/reproducibly involving an educational intervention).

Finally, there was only one study (Roukema et al[54]) that attempted to impact a utilization quality gap expressed as underutilization of testing, reflecting a substantial gap in the evidence given that overall mean rates of inappropriate underutilization appear to exceed those of inappropriate overutilization (20.6% and 44.8%, respectively).[5,16] Relatedly, no studies were found for inclusion where the intervention directly targeted patients, eg, through text or email reminders of necessary testing.

### Future Research Needs

In guiding future quality improvement studies, investigators may consider the following definition of appropriate test utilization, which is informed by several resources: utilization that is consistent with current subject-specific expert knowledge or evidence-based standards for usage, matching the patient with the correct test(s) at the right time and in the correct order, and performed with expectation of informing patient-management decisions for the benefit of health outcomes, in a cost-effective manner.[16,20,120,124,127-133] Further, it has defined criteria for identifying occurrences of inappropriate utilization through test utilization audits (criteria may, for example, relate to testing frequency and timing, test choice in relation to alternative tests, clinical indications for tests, or probability a test result will provide clinically actionable information). Inappropriate utilization may


DOI: 10.1093/ajcp/aqx147

involve underutilization, overutilization, or misuse (eg, misinterpretation) of testing in relation to specific patient populations and clinical settings.

This emphasizes criteria-based assessments during utilization audits, as well as the link from testing to clinical decision-making and patient-relevant outcomes, with consideration of cost-effectiveness. This definition should be assessed in light of the fact that "appropriateness" in testing remains a complex, multidimensional, multiperspective concept.[127,128] Additionally, it should be assessed with consideration that some element of subjectivity may be unavoidable in the provision of testing for a specific patient in a specific context.[128]

Specific needs for future research as suggested by this review include (1) assessment of test utilization management practice intervention in additional settings/contexts (ie, the underrepresented settings/contexts indicated in the "Applicability and Generalizability" section: community hospitals, children's hospitals, and public hospitals including VA hospitals); (2) better description of the source of criteria applied in tests utilization audits for determinations of inappropriate test utilization, as well as description of efforts to validate the criteria in the local setting, as well as better description of the test utilization audit data collection process (and how it reliably applied the criteria)[5,20,132]; (3) incorporation of patient-related outcomes (surrogate or health outcomes) affected by inappropriate test utilization; (4) inclusion of cost data as permits more robust economic evaluations, such as cost-effectiveness, allowing for comparative analysis of an intervention in terms of both cost and effect; (5) assessment of alternate utilization management interventions, head-to-head, within individual studies; (6) assessment of practice intervention not examined in this review, for example diagnostic management teams (defined and distinguished from LTU teams in the glossary); and (7) assessment of quality gaps representing inappropriate underutilization of testing.

Additionally, while laboratory professionals were frequently involved as investigators and authors within the included evidence base (46% of included studies), we recommend more robust representation of laboratory professionals on teams investigating the impact of practices to manage appropriate test utilization. Recent articles provide guidance on the role of the laboratory at this point of clinical-laboratory interface, as well as on relevant lab-related performance measures laboratory professionals may monitor, since labs should be "continuously engaged in auditing, monitoring, and improving the appropriateness of test req uests…."[6,16,115,131,134,135] This should include efforts to further evolve how appropriate test utilization is benchmarked and monitored though lab-related performance measures or indicators (ideally standardized or "harmonized" across laboratories and health care organizations),[134,136] as may involve increased use of health information technology, web technology, and new approaches to integrating and analyzing available, relevant data.[6,119]

Available in Supplemental Table 8 is a data collection form, as can assist in preparing existing data or performing prospective studies for publication, or for submission to the CDC LMBP initiative as unpublished studies.

*Corresponding author: Matthew Rubinstein, MS, Centers for Disease Control and Prevention, MS-G23, 1600 Clifton Rd, Atlanta, GA 30333; ksr3@cdc.gov.*

*Funding source: Work conducted by Kakali Bandyopadhyay and Millie Linville was supported by CDC funding for the Laboratory Medicine Best Practices Initiative to Westat-Deloitte under contract (HHSD2002013M53968B/2002013F57569). Work conducted by Robert Hirsch was supported under contract to Carter Consulting (200-2010-37203).*

*Acknowledgments: We thank Jacqueline Goolsby (DLS, CDC), Elizabeth Leibach (DLS, CDC), Paramjit Sandhu (DLS, CDC), Sonya Strider (DLS, CDC), Joanna Taliano (CDC Library), the LMBP Workgroup, and the LMBP test utilization expert panel.*

*Disclaimer: The findings and conclusions in this article are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.*

## References

1. Balogh E, Miller BT, Ball J; Institute of Medicine (U.S.). *Committee on Diagnostic Error in Health Care: Improving Diagnosis in Health Care*. Washington, DC: The National Academies Press; 2015. xxvii, 444.
2. Hallworth MJ, Epner PL, Ebert C, et al. IFCC Task Force on the Impact of Laboratory Medicine on Clinical Management and Outcomes. Current evidence and future perspectives on the effective practice of patient-centered laboratory medicine. *Clin Chem*. 2015;61:589-599.
3. Plebani M. Towards a new paradigm in laboratory medicine: the five rights. *Clin Chem Lab Med*. 2016;54:1881-1891.
4. Rohr UP, Binder C, Dieterle T, et al. The value of in vitro diagnostic testing in medical practice: a status report. *Plos One*. 2016;11:e0149856.
5. Zhi M, Ding EL, Theisen-Toupal J, et al. The landscape of inappropriate laboratory testing: a 15-year meta-analysis. *Plos One*. 2013;8:e78962.
6. Plebani M, Panteghini M. Promoting clinical and laboratory interaction by harmonization. *Clin Chim Acta*. 2014;432:15-21.
7. Baird G. The laboratory test utilization management toolbox. *Biochem Med (Zagreb)*. 2014;24:223-234.
8. Janssens PM. Managing the demand for laboratory testing: options and opportunities. *Clin Chim Acta*. 2010;411:1596-1602.
9. Huck A, Lewandrowski K. Utilization management in the clinical laboratory: an introduction and overview of the literature. *Clin Chim Acta*. 2014;427:111-117.





DOI: 10.1093/ajcp/aqx147

10. Plebani M, Zaninotto M, Faggian D. Utilization management: a European perspective. *Clin Chim Acta*. 2014;427:137-141.
11. Yeh DD. A clinician's perspective on laboratory utilization management. *Clin Chim Acta*. 2014;427:145-50.
12. Solomon DH, Hashimoto H, Daltroy L, et al. Techniques to improve physicians' use of diagnostic tests: a new conceptual framework. *JAMA*. 1998;280:2020-2027.
13. Fryer AA, Smellie WS. Managing demand for laboratory tests: a laboratory toolkit. *J Clin Pathol*. 2013;66:62-72.
14. Smellie WS. Demand management and test request rationalization. *Ann Clin Biochem*. 2012;49:323-336.
15. Barth JH. Clinical quality indicators in laboratory medicine. *Ann Clin Biochem*. 2012;49(pt 1):9-16.
16. Ferraro S, Panteghini M. The role of laboratory in ensuring appropriate test requests. *Clin Biochem*. 2017;50:555-561.
17. Plebani M. Exploring the iceberg of errors in laboratory medicine. *Clin Chim Acta*. 2009;404:16-23.
18. Barr JT, Silver S. The total testing process and its implications for laboratory administration and education. *Clin Lab Manage Rev*. 1994;8:526-542.
19. Christenson RH, Snyder SR, Shaw CS, et al. Laboratory medicine best practices: systematic evidence review and evaluation methods for quality improvement. *Clin Chem*. 2011;57:816-825.
20. Hauser RG, Shirts BH. Do we now know what inappropriate laboratory utilization is? An expanded systematic review of laboratory clinical audits. *Am J Clin Pathol*. 2014;141:774-783.
21. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ*. 2011;342:d549.
22. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557-560.
23. Borenstein M. *Introduction to Meta-Analysis*. Chichester, UK: John Wiley & Sons; 2009:xxviii, 421.
24. Bansal P, Aronsky D, Aronsky D, et al. A computer based intervention on the appropriate use of arterial blood gas. *Proc AMIA Symp*. 2001:32-6.
25. Bates DW, Kuperman GJ, Rittenberg E, et al. A randomized trial of a computer-based intervention to reduce utilization of redundant laboratory tests. *Am J Med*. 1999;106:144-150.
26. Bates DW, Kuperman GJ, Jha A, et al. Does the computerized display of charges affect inpatient ancillary test utilization? *Arch Intern Med*. 1997;157:2501-2508.
27. Bridges SA, Papa L, Norris AE, et al. Duplicated laboratory tests: evaluation of a computerized alert intervention abstract. *J Healthc Qual*. 2014;36:46-53.
28. Fang DZ, Sran G, Gessner D, et al. Cost and turn-around time display decreases inpatient ordering of reference laboratory tests: a time series. *BMJ Qual Saf*. 2014;23:994-1000.
29. Feldman LS, Shihab HM, Thiemann D, et al. Impact of providing fee data on laboratory test ordering: a controlled clinical trial. *Jama Intern Med*. 2013;173:903-908.
30. Georgiou A, Lang S, Rosenfeld D, et al. The use of computerized provider order entry to improve the effectiveness and efficiency of coagulation testing. *Arch Pathol Lab Med*. 2011;135:495-498.
31. Horn DM, Koplan KE, Senese MD, et al. The impact of cost displays on primary care physician laboratory test ordering. *J Gen Intern Med*. 2014;29:708-14.
32. Hwang JI, Park HA, Bakken S. Impact of a physician's order entry (POE) system on physicians' ordering patterns and patient length of stay. *Int J Med Inform*. 2002;65:213-223.
33. Kahan NR, Waitman DA, Vardy DA. Curtailing laboratory test ordering in a managed care setting through redesign of a computerized order form. *Am J Manag Care*. 2009;15:173-6.
34. Le RD, Kosowsky JM, Landman AB, et al. Clinical and financial impact of removing creatine kinase-MB from the routine testing menu in the emergency setting. *Am J Emerg Med*. 2015;33:72-75.
35. Li L, Georgiou A, Vecellio E, et al. What is the effect of electronic pathology ordering on test re-ordering patterns for paediatric patients? *Stud Health Technol Inform*. 2014;204:74-79.
36. Lippi G, Brambilla M, Bonelli P, et al. Effectiveness of a computerized alert system based on re-testing intervals for limiting the inappropriateness of laboratory test requests. *Clin Biochem*. 2015;48:1174-1176.
37. Love SA, McKinney ZJ, Sandoval Y, et al. Electronic medical record-based performance improvement project to document and reduce excessive cardiac troponin testing. *Clin Chem*. 2015;61:498-504.
38. May TA, Clancy M, Critchfield J, et al. Reducing unnecessary inpatient laboratory testing in a teaching hospital. *Am J Clin Pathol*. 2006;126:200-206.
39. Olson J, Hollenbeak C, Donaldson K, et al. Default settings of computerized physician order entry system order sets drive ordering habits. *J Pathol Inform*. 2015;6:16.
40. Pageler NM, Franzon D, Longhurst CA, et al. Embedding time-limited laboratory orders within computerized provider order entry reduces laboratory utilization. *Pediatr Crit Care Med*. 2013;14:413-419.
41. Probst CA, Shaffer VA, Chan YR. The effect of defaults in an electronic health record on laboratory test ordering practices for pediatric patients. *Health Psychol*. 2013;32:995-1002.
42. Procop GW, Keating C, Stagno P, et al. Reducing duplicate testing: a comparison of two clinical decision support tools. *Am J Clin Pathol*. 2015;143:623-626.
43. Shalev V, Chodick G, Heymann AD. Format change of a laboratory test order form affects physician behavior. *Int J Med Inform*. 2009;78:639-644.
44. Solis DLE, Rabrich J, Van Leer P. Improving physician order practices and cost savings by changing the electronic medical records. *Ann Emerg Med*. 2015;66(4 suppl 1):S34.
45. Vardy DA, Simon T, Limoni Y, et al. The impact of structured laboratory routines in computerized medical records in a primary care service setting. *J Med Syst*. 2005;29:619-626.
46. Waldron JL, Ford C, Dobie D, et al. An automated minimum retest interval rejection rule reduces repeat CRP workload and expenditure, and influences clinician-requesting behaviour. *J Clin Pathol*. 2014;67:731-733.
47. Westbrook JI, Georgiou A, Dimos A, et al. Computerised pathology test order entry reduces laboratory turnaround times and influences tests ordered by hospital clinicians: a controlled before and after study. *J Clin Pathol*. 2006;59:533-536.
48. Bindels R, Hasman A, Kester AD, et al. The efficacy of an automated feedback system for general practitioners. *Inform Prim Care*. 2003;11:69-74.
49. Collins RA, Triulzi DJ, Waters JH, et al. Evaluation of real-time clinical decision support systems for platelet and cryoprecipitate orders. *Am J Clin Pathol*. 2014;141:78-84.



DOI: 10.1093/ajcp/aqx147

50. Howell LP, MacDonald S, Jones J, et al. Can automated alerts within computerized physician order entry improve compliance with laboratory practice guidelines for ordering pap tests? *J Pathol Inform*. 2014;5:37.
51. McKinney ZJ, Peters JM, Gorlin JB, et al. Improving red blood cell orders, utilization, and management with point-of-care clinical decision support. *Transfusion*. 2015;55:2086-2094.
52. Nightingale PG, Peters M, Mutimer D, et al. Effects of a computerised protocol management system on ordering of clinical tests. *Qual Health Care*. 1994;3:23-28.
53. Poley MJ, Edelenbos K, Mosseveld M, et al. Cost consequences of implementing an electronic decision support system for ordering laboratory tests in primary care: evidence from a controlled prospective study in the Netherlands. *Clin Chem*. 2007;53:213-9.
54. Roukema J, Steyerberg EW, van der Lei J, et al. Randomized trial of a clinical decision support system: impact on the management of children with fever without apparent source. *J Am Med Inform Assoc*. 2008;15:107-113.
55. van Wijk MA, van der Lei J, Mosseveld M, et al. Assessment of decision support for blood test ordering in primary care: a randomized trial. *Ann Intern Med*. 2001;134:274-281.
56. Abio-Calvete MDLO, Botón Contreras ME, et al. Demand management of analytical request in erythropathogy section of a tertiary hospital. *Haematologica*. 2015;100:581-582.
57. Baral N, Koner BC, Lamsal M, et al. Thyroid function testing in eastern Nepal and the impact of CME on subsequent requests. *Trop Doct*. 2001;31:155-157.
58. Chonfhaola ACM, O'Riain M, Murray M. Demand management in a university laboratory in the west of Ireland: the financial benefit of distribution of local guidelines for vitamin B12 and folate testing. *Haematologica*. 2013;98:216.
59. Dawes L, Miller L, Subramoney M, et al. The impact of education and audit on compliance with a clinical guideline for fetal fibronectin testing and the management of threatened preterm labour at National Women's Health, Auckland City Hospital. *BJOG Int J Obstet Gynaecol*. 2015;122:237-8.
60. DellaVolpe JD, Chakraborti C, Cerreta K, et al. Effects of implementing a protocol for arterial blood gas use on ordering practices and diagnostic yield. *Healthc (Amst)*. 2014;2:130-135.
61. Eisenberg JM. An educational program to modify laboratory use by house staff. *J Med Educ*. 1977;52:578-581.
62. Gardezi SA. Troponin: think before you request one. *BMJ Qual Improv Rep*. 2015;4.
63. Kotecha N, Cardasis J, Narayanswami G, et al. Reducing unnecessary lab tests in the MICU by incorporating a guideline in daily ICU team rounds. *Am J Respir Crit Care Med*. 2015;191:no pagination.
64. Thakkar RN, Kim D, Knight AM, et al. Impact of an educational intervention on the frequency of daily blood test orders for hospitalized patients. *Am J Clin Pathol*. 2015;143:393-397.
65. Baker R, Falconer Smith J, Lambert PC. Randomised controlled trial of the effectiveness of feedback in improving test ordering in general practice. *Scand J Prim Health Care*. 2003;21:219-223.
66. Bunting PS, Van Walraven C. Effect of a controlled feedback intervention on laboratory test ordering by community physicians. *Clin Chem*. 2004;50:321-326.
67. Gama R, Nightingale PG, Broughton PM, et al. Modifying the request behaviour of clinicians. *J Clin Pathol*. 1992;45:248-249.
68. Miyakis S, Karamanof G, Liontos M, et al. Factors contributing to inappropriate ordering of tests in an academic medical department and the effect of an educational feedback strategy. *Postgrad Med J*. 2006;82:823-829.
69. Verstappen WH, van der Weijden T, Dubois WI, et al. Improving test ordering in primary care: the added value of a small-group quality improvement strategy compared with classic feedback only. *Ann Fam Med*. 2004;2:569-575.
70. Verstappen WH, van der Weijden T, ter Riet G, et al. Block design allowed for control of the hawthorne effect in a randomized controlled trial of test ordering. *J Clin Epidemiol*. 2004;57:1119-1123.
71. Winkens RA, Pop P, Grol RP, et al. Effect of feedback on test ordering behaviour of general practitioners. *BMJ*. 1992;304:1093-1096.
72. Baird GS, Rainey PM, Wener M, et al. Reducing routine ionized calcium measurement. *Clin Chem*. 2009;55:533-540.
73. Bonaguri C, Melegari A, Ballabio A, et al. Italian multicentre study for application of a diagnostic algorithm in autoantibody testing for autoimmune rheumatic disease: conclusive results. *Autoimmun Rev*. 2011;11:1-5.
74. Cernich JT, Hamilton M, Mitre N. Increased use of lab algorithm to reduce cost of screening for hypothyroidism in pediatric type 1 diabetes. *Diabetes*. 2014;63:A322-A323.
75. Froom P, Barak M. Cessation of dipstick urinalysis reflex testing and physician ordering behavior. *Am J Clin Pathol*. 2012;137:486-489.
76. Tampoia M, Brescia V, Fontana A, et al. Application of a combined protocol for rational request and utilization of antibody assays improves clinical diagnostic efficacy in autoimmune rheumatic disease. *Arch Pathol Lab Med*. 2007;131:112-116.
77. Van Walraven C, Goel V. The effect of a hepatitis serology testing algorithm on laboratory utilization. *J Eval Clin Pract*. 2002;8:327-32.
78. Wu AH, Ghani F, Prigent F, et al. Reflex testing II: evaluation of an algorithm for use of cardiac markers in the assessment of emergency department patients with chest pain. *Clin Chim Acta*. 1999;288:97-109.
79. Aesif SW, Parenti DM, Lesky L, et al. A cost-effective interdisciplinary approach to microbiologic send-out test use. *Arch Pathol Lab Med*. 2015;139:194-198.
80. Barazzoni F, Grilli R, Amicosante AM, et al. Impact of end user involvement in implementing guidelines on routine pre-operative tests. *Int J Qual Health Care*. 2002;14:321-327.
81. Chu KH, Wagholikar AS, Greenslade JH, et al. Sustained reductions in emergency department laboratory test orders: impact of a simple intervention. *Postgrad Med J*. 2013;89:566-571.
82. Dickerson JA, Cole B, Conta JH, et al. Improving the value of costly genetic reference laboratory testing with active utilization management. *Arch Pathol Lab Med*. 2014;138:110-113.
83. Dolezal A, Krasowski MD. Paraneoplastic autoantibody panel ordering as a model for laboratory test utilization analysis and intervention. *Lab Invest*. 2015;95:498A.
84. Miller CE, Krautscheid P, Baldwin EE, et al. Genetic counselor review of genetic test orders in a reference laboratory reduces unnecessary testing. *Am J Med Genet A*. 2014;164A:1094-1101.
85. Bareford D, Hayling A. Inappropriate use of laboratory services: long term combined approach to modify request patterns. *BMJ*. 1990;301:1305-1307.






DOI: 10.1093/ajcp/aqx147

86. Baricchi R, Zini M, Nibali MG, et al. Using pathology-specific laboratory profiles in clinical pathology to reduce inappropriate test requesting: two completed audit cycles. *BMC Health Serv Res*. 2012;12:187.
87. Calderon-Margalit R, Mor-Yosef S, Mayer M, et al. An administrative intervention to improve the utilization of laboratory tests within a university hospital. *Int J Qual Health Care*. 2005;17:243-248.
88. Dowling PT, Alfonsi G, Brown MI, et al. An education program to reduce unnecessary laboratory tests by residents. *Acad Med*. 1989;64:410-412.
89. Gilmour J, Weisman A, Orlov S, Vecchiarelli J, et al. Reducing unnecessary free thyroid hormone testing at an academic ambulatory hospital: a quality improvement (QI) initiative. *Thyroid*. 2015;25:A219-A20.
90. Iosfina I, Merkeley H, Cessford T, Geller G, et al. Implementation of an on-demand strategy for routine blood testing in ICU patients. *Am J Respir Crit Care Med*. 2013;187.
91. Hutton HD, Drummond HS, Fryer AA. The rise and fall of C-reactive protein: managing demand within clinical biochemistry. *Ann Clin Biochem*. 2009;46:155-158.
92. Janssens PM, Staring W, Winkelman K, et al. Active intervention in hospital test request panels pays. *Clin Chem Lab Med*. 2015;53:731-742.
93. Kroenke K, Hanley JF, Copley JB, et al. Improving house staff ordering of three common laboratory tests: reductions in test ordering need not result in underutilization. *Med Care*. 1987;25:928-935.
94. Larochelle MR, Knight AM, Pantle H, et al. Reducing excess cardiac biomarker testing at an academic medical center. *J Gen Intern Med*. 2014;29:1468-1474.
95. Lum G. Evaluation of a protocol to control utilization of B-type natriuretic peptide testing. *Am J Clin Pathol*. 2006;126:190-194.
96. MacPherson RD, Reeve SA, Stewart TV, et al. Effective strategy to guide pathology test ordering in surgical patients. *Anz J Surg*. 2005;75:138-143.
97. McNicoll L, Butterfield K, Caliendo A, et al. Vampire medicine: do we need all those blood tests? *J Am Geriatr Soc*. 2015;63:S213.
98. Minerowicz C, Abel N, Hunter K, et al. Impact of weekly feedback on test ordering patterns. *Am J Manag Care*. 2015;21:763-768.
99. Newman DB, Siontis KC, Chandrasekaran K, et al. Intervention to reduce inappropriate ionized calcium ordering practices: a quality-improvement project. *Perm J*. 2015;19:49-51.
100. Riley JD, Procop GW, Kottke-Marchant K, et al. Improving molecular genetic test utilization through order restriction, test review, and guidance. *J Mol Diagn*. 2015;17:225-229.
101. Roggeman N, Aronson C, Netols C. Review of red blood cell (RBC) irradiation requests results in decreased utilization. *Transfusion*. 2014;54:246A-247A.
102. Rosenbloom ST, Chiu KW, Byrne DW, et al. Interventions to regulate ordering of serum magnesium levels: report of an unintended consequence of decision support. *J Am Med Inform Assoc*. 2005;12:546-553.
103. Samuelson BT, Glynn E, Holmes M, et al. Use of a computer-based provider order entry (CPOE) intervention to optimize laboratory testing in patients with suspected heparin-induced thrombocytopenia. *Thromb Res*. 2015;136:928-931.
104. Spiegel JS, Shapiro MF, Berman B, et al. Changing physician test ordering in a university hospital: an intervention of physician participation, explicit criteria, and feedback. *Arch Intern Med*. 1989;149:549-553.
105. Thomas RE, Croal B, Ramsay C. Reducing inappropriate laboratory testing through provider education and feedback. *J Clin Outcomes Manage*. 2006;13:472-477.
106. Tomlin A, Dovey S, Gauld R, et al. Better use of primary care laboratory services following interventions to "market" clinical guidelines in New Zealand: a controlled before-and-after study. *BMJ Qual Saf*. 2011;20:282-290.
107. Vegting IL, van Beneden M, Kramer MH, et al. How to save costs by reducing unnecessary testing: lean thinking in clinical practice. *Eur J Intern Med*. 2012;23:70-75.
108. Vidyarthi AR, Hamill T, Green AL, et al. Changing resident test ordering behavior: a multilevel intervention to decrease laboratory utilization at an academic medical center. *Am J Med Qual*. 2015;30:81-87.
109. Wang TJ, Mort EA, Nordberg P, et al. A utilization management intervention to reduce unnecessary testing in the coronary care unit. *Arch Intern Med*. 2002;162:1885-1890.
110. Warren JS. Laboratory test utilization program: structure and impact in a large academic medical center. *Am J Clin Pathol*. 2013;139:289-97.
111. White P, Kenton K. Use of electronic medical record-based tools to improve compliance with cervical cancer screening guidelines: effect of an educational intervention on physicians' practice patterns. *J Low Genit Tract Dis*. 2013;17:175-181.
112. Alonso-Coello P, Schunemann HJ, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices, 1: introduction. *BMJ*. 2016;353:i2016.
113. Bosse HM, Mohr J, Buss B, et al. The benefit of repetitive skills training and frequency of expert feedback in the early acquisition of procedural skills. *BMC Med Educ*. 2015;15:22.
114. Moulaert V, Verwijnen MG, Rikers R, et al. The effects of deliberate practice in undergraduate medical education. *Med Educ*. 2004;38:1044-1052.
115. Crawford JM, Shotorbani K, Sharma G, et al. Improving American healthcare through "clinical lab 2.0": a Project Santa Fe report. *Acad Pathol*. 2017;4:2374289517701067.
116. Brunetti M, Pregno S, Schünemann H, et al. Economic evidence in decision-making process in laboratory medicine. *Clin Chem Lab Med*. 2011;49:617-621.
117. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)--explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. *Value Health*. 2013;16:231-250.
118. Bruns DE. Laboratory-related outcomes in healthcare. *Clin Chem*. 2001;47:1547-1552.
119. Shirts BH, Jackson BR, Baird GS, et al. Clinical laboratory analytics: challenges and promise for an emerging discipline. *J Pathol Inform*. 2015;6:9.
120. Thornbury JR, Fryback DG. Technology assessment: an American view. *Eur J Radiol*. 1992;14:147-156.
121. Staub LP, Lord SJ, Simes RJ, et al. Using patient management as a surrogate for patient health outcomes in diagnostic test evaluation. *bmc Med Res Methodol*. 2012;12:12.



DOI: 10.1093/ajcp/aqx147

122. Staub LP, Dyer S, Lord SJ, et al. Linking the evidence: intermediate outcomes in medical test assessments. *Int J Technol Assess Health Care*. 2012;28:52-58.
123. Pearl WS. A hierarchical outcomes approach to test assessment. *Ann Emerg Med*. 1999;33:77-84.
124. Epner PL, Gans JE, Graber ML. When diagnostic testing leads to harm: a new outcomes-based approach for laboratory medicine. *BMJ Qual Saf*. 2013;22(suppl 2):ii6-ii10.
125. Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, et al. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ*. 2012;344:e686.
126. Horvath AR, Pewsner D. Systematic reviews in laboratory medicine: principles, processes and practical considerations. *Clin Chim Acta*. 2004;342:23-39.
127. Barth JH. Selecting clinical quality indicators for laboratory medicine. *Ann Clin Biochem*. 2012;49:257-261.
128. Cappelletti P. Appropriateness of diagnostics tests. *Int J Lab Hematol*. 2016;38(suppl 1):91-99.
129. Alonso-Cerezo MC, Martín JS, García Montes MA, et al. Appropriate utilization of clinical laboratory tests. *Clin Chem Lab Med*. 2009;47:1461-1465.
130. Graber ML, Trowbridge R, Myers JS, et al. The next organizational challenge: finding and addressing diagnostic error. *Jt Comm J Qual Patient Saf*. 2014;40:102-110.
131. Melanson SE. Establishing benchmarks and metrics for utilization management. *Clin Chim Acta*. 2014;427:127-130.
132. van Walraven C, Naylor CD. Do we know what inappropriate laboratory utilization is? A systematic review of laboratory clinical audits. *JAMA*. 1998;280:550-558.
133. Price CP. Application of the principles of evidence-based medicine to laboratory medicine. *Clin Chim Acta*. 2003;333:147-154.
134. Sciacovelli L, Panteghini M, Lippi G, et al. Defining a roadmap for harmonizing quality indicators in laboratory medicine: a consensus statement on behalf of the IFCC working group "laboratory error and patient safety" and EFLM task and finish group "performance specifications for the extra-analytical phases". *Clin Chem Lab Med*. 2017;55:1478-1488.
135. Shahangian S, Snyder SR. Laboratory medicine quality indicators: a review of the literature. *Am J Clin Pathol*. 2009;131:418-431.
136. Plebani M. Harmonization in laboratory medicine: the complete picture. *Clin Chem Lab Med*. 2013;51:741-751.
137. Ogrinc G, Mooney SE, Estrada C, et al. The SQUIRE (Standards for QUality Improvement Reporting Excellence) guidelines for quality improvement reporting: explanation and elaboration. *Qual Saf Health Care*. 2008;17(suppl 1):i13-32.
138. Harris AD, McGregor JC, Perencevich EN, et al. The use and interpretation of quasi-experimental studies in medical informatics. *J Am Med Inform Assoc*. 2006;13:16-23.





DOI: 10.1093/ajcp/aqx147