Department of Veterans Affairs
Health Services Research & Development Service | Evidence-based Synthesis Program

QUERI

Electronic Health Record-based Interventions for Reducing Inappropriate Imaging in the Clinical Setting: A Systematic Review of the Evidence

January 2015

Prepared for:
Department of Veterans Affairs
Veterans Health Administration
Quality Enhancement Research Initiative
Health Services Research & Development Service
Washington, DC 20420

Prepared by:
Evidence-based Synthesis Program (ESP) Center
West Los Angeles VA Medical Center
Los Angeles, CA
Paul G. Shekelle, MD, PhD, Director

Investigators:

Principal Investigator:
Paul G. Shekelle, MD, PhD

Contributing Investigators:
Caroline Lubick Goldzweig, MD, MS
Greg Orshansky, MD
Neil M. Paige, MD, MSHS
Brett A. Ewing, MS

Research Associates:
Isomi M. Miake-Lye, BA
Jessica M. Beroes, BS

VA HEALTH CARE Defining EXCELLENCE in the 21st Century

PREFACE

Quality Enhancement Research Initiative's (QUERI) Evidence-based Synthesis Program (ESP) was established to provide timely and accurate syntheses of targeted healthcare topics of particular importance to Veterans Affairs (VA) clinicians, managers and policymakers as they work to improve the health and healthcare of Veterans. The ESP disseminates these reports throughout the VA, and some evidence syntheses inform the clinical guidelines of large professional organizations.

QUERI provides funding for four ESP Centers and each Center has an active university affiliation. The ESP Centers generate evidence syntheses on important clinical practice topics, and these reports help:

- develop clinical policies informed by evidence;
- guide the implementation of effective services to improve patient outcomes and to support VA clinical practice guidelines and performance measures; and
- set the direction for future research to address gaps in clinical knowledge.

In 2009, the ESP Coordinating Center was created to expand the capacity of HSR&D Central Office and the four ESP sites by developing and maintaining program processes. In addition, the Center established a Steering Committee comprised of QUERI field-based investigators, VA Patient Care Services, Office of Quality and Performance, and Veterans Integrated Service Networks (VISN) Clinical Management Officers. The Steering Committee provides program oversight, guides strategic planning, coordinates dissemination activities, and develops collaborations with VA leadership to identify new ESP topics of importance to Veterans and the VA healthcare system.

Comments on this evidence report are welcome and can be sent to Nicole Floyd, ESP Coordinating Center Program Manager, at Nicole.Floyd@va.gov.

Recommended citation: Goldzweig C L, Orshansky G, Paige N M, Ewing B A, Miake-Lye I M, Beroes J M, Shekelle P G. Electronic Health Record-Based Interventions for Reducing Inappropriate Imaging in the Clinical Setting: A Systematic Review of the Evidence. VA ESP Project #05-226; 2014.

This report is based on research conducted by the Evidence-based Synthesis Program (ESP) Center located at the **West Los Angeles VA Medical Center, Los Angeles, CA,** funded by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, Quality Enhancement Research Initiative. The findings and conclusions in this document are those of the author(s) who are responsible for its contents; the findings and conclusions do not necessarily represent the views of the Department of Veterans Affairs or the United States government. Therefore, no statement in this article should be construed as an official position of the Department of Veterans Affairs. No investigators have any affiliations or financial involvement (*eg*, employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties) that conflict with material presented in the report.

TABLE OF CONTENTS

TABLES

FIGURES

EVIDENCE REPORT

INTRODUCTION

There are widespread concerns within Veterans Affairs (VA) healthcare and in non-VA US healthcare that the costs of healthcare are rising at unsustainable rates. One driver of cost is the increasing use of radiology imaging procedures, particularly advanced imaging techniques such as computed tomography (CT) scanning, magnetic resonance imaging (MRI), and CT angiography. Advances in imaging capabilities allow physicians to image ever-finer areas within the body, and the ease with which many of these tests are ordered has led to dramatic increases in the rates of use of many tests. For example, the use of CT scans in the Emergency Department (ED) grew by 330% in the 12 years from 1996 through 2007, at a time when the rate of ED visits grew by only 11%.[3] Similarly, other investigators reported a 3-fold increase in the likelihood of getting a CT scan or MRI during an ED visit between 1998 and 2007.[4]

These dramatic increases in utilization have led to increased scrutiny regarding the clinical value of these imaging studies. In some cases, strong evidence exists that the imaging studies provide no value, or even harm patients. For example, a meta-analysis of early lumbar imaging for patients with acute low back pain included 5 randomized controlled trials (RCTs) where patients were randomized to receive or not receive early imaging in the form of a plain film, a CT, or an MRI. At 3 months, there was no improvement in pain or function among patients who had received imaging.[5] In other cases, strong professional opinion considers certain tests to be of little value, mainly because alternate tests are preferred or the probability of an abnormal image is exceedingly remote.

There is widespread agreement that more appropriate use of certain imaging tests could both improve quality and save costs. When the American Board of Internal Medicine Foundation asked physician specialty groups to identify procedures or tests that they judged were of little value, imaging tests were frequently identified, such as the use of CT scans for minor head injury in the ED (American College of Emergency Physicians), imaging studies in patients with nonspecific low back pain (American College of Physicians), imaging for uncomplicated headache (American College of Radiology), CT angiography for patients with low clinical probability of pulmonary embolus and a negative D-dimer assay (American College of Chest Physicians), and stress cardiac imaging in patients without high-risk markers for coronary artery disease (American College of Cardiology).

The recognition that more appropriate use of imaging could improve quality and reduce costs has led to the development of interventions to encourage more appropriate radiology utilization. Some of these interventions have made use of the clinical decision support capabilities of electronic health records. VA has been a leader in the use of electronic health records and clinical decision support, and VA leadership therefore requested a review of published studies assessing the effect of electronic health record (EHR)-based interventions to improve the appropriateness of imaging.

METHODS

TOPIC DEVELOPMENT

This report was developed based on a nomination from the VHA Choosing Wisely Workgroup, including Dr. David Atkins, Director of Health Services Research and Development, Dr. Charles Anderson, Chief Consultant for Diagnostic Services, and Sherrill Snuggs, Utilization Officer in the Office of Patient Care Services, in an effort to implement some of the segments of the "Choosing Wisely" campaign. In 2011, the American Board of Internal Medicine, working with Consumer Reports, organized the "Choosing Wisely" campaign in conjunction with 9 other leading medical professional groups, including the American College of Physicians, the leading medical organization dealing with adult medical care and highly relevant to Veterans' care.[6] These organizations were asked to identify 5 tests or treatments that in their professional opinions, based on published guidelines, were often used inappropriately. Leading this list were several imaging studies including studies such as stress cardiac imaging in asymptomatic patients, routine preoperative chest x-rays in asymptomatic patients, and routine imaging in the absence of "red flag symptoms" for low back pain.

In recognition of the risks and costs associated with inappropriate imaging, VA leadership has requested an evidence synthesis which evaluates studied methods for reducing inappropriate imaging that center around the EHR.

The final key questions are:

Key Question 1: What is the effectiveness of EHR-based interventions in reducing unnecessary or inappropriate imaging?

Key Question 2: Do EHR-based interventions vary in results by system?

Key Question 3: What are the harms or potential harms associated with EHR-based interventions used to reduce inappropriate imaging?

The PROSPERO registration number is CRD42014009557.

SEARCH STRATEGY

Our search strategy had 4 components:

The primary component was a search of an existing database developed from 4 prior, broad-based reviews of health information technology (IT).[7-10] These 4 reviews were done using similar search strategies (covering the period 1995-2013) and inclusion/exclusion criteria, and were designed to identify all published hypothesis-testing studies of clinical health IT. Hypothesis-testing studies included randomized trials, controlled before-and-after studies, time series studies, and pre-post studies. These studies were further classified according to the health IT functionality, including clinical decision support (CDS), computerized provider order entry, patient care reminders, e-prescribing, health information exchange, etcetera. In the most recent review, summary data were presented which showed that of 1057 health IT studies in the

database, 417 were classified as CDS. These 417 titles and abstracts were searched for studies eligible for this review (*eg*, CDS aimed at reducing inappropriate radiology use).

The second component of our search was reference mining of 3 potentially relevant systematic reviews:

1. The impact of computerized provider order entry systems on medical imaging services: a systematic review.[11]

   This review included 8 studies classified as ordering of medical imaging examinations by the provider.

2. Computerized clinical decision support systems for chronic disease management: a decision-maker-researcher partnership systematic review.[12]

   This review included 36 studies. It also contained a list of criteria for characterizing CDS systems that included whether it was integrated with the electronic health record (EHR) or computerized physician order entry (CPOE), whether the CDS was automated through the EHR, whether it was pilot tested, whether it gave feedback at the time of care, whether the CDS suggested procedures, and whether the authors of the article were also the developers of the CDS. We adapted this list for our own use.

3. Systematic review of clinical decision support interventions with potential for inpatient cost reduction.[13]

   This review identified 3 studies of radiology interventions.

Our third component was a targeted search of PubMed from 2011 to 9/10/2014, looking specifically at decision support for imaging, along with 2 searches constructed to identify articles related to 2 key references in Web of Science and PubMed (see Appendix B for this search strategy).

Our fourth component was to reference mine all included articles.

STUDY SELECTION

All reference titles and abstracts were screened in duplicate. If either reviewer selected a title or abstract, it was included for further review. Full text articles were then reviewed in duplicate, with all discrepancies discussed with the group. References were selected based on the following inclusion criteria:

Participants: Adult population. Studies aimed only at children were excluded. Studies with mixed populations were included.

Intervention: EHR-based interventions for reducing imaging for diagnostic purposes (as opposed to screening) considered inappropriate or unnecessary based on clinical guidelines. This meant that studies seeking to increase the use of radiographic imaging like mammography for breast cancer screening were excluded. Studies of systems running on personal digital assistants were

excluded. Studies of web-based interventions or computerized, stand-alone systems that we judged could be easily incorporated into the EHR were included.

Comparator (study design): Usual care.

Outcome: Rates of imaging procedures judged as unnecessary based on existing clinical guidelines. Studies that reported on changes in appropriateness (as opposed to decrease in appropriateness) were also included. Studies that targeted the use of imaging procedures stated as being overused and then reporting only utilization data were included. Utilization outcomes were considered separate from appropriateness outcomes. A table of all outcomes included as "appropriateness" is in Appendix C.

Timing: All times.

Setting: Ambulatory, hospital, and emergency department settings.

DATA ABSTRACTION

Data were extracted by 2 reviewers, and discrepancies were reconciled with the group. Articles had data abstracted on study design, time period, setting, imaging modality, intervention, comparison, sample size, target of intervention, findings, IT design, data entry for intervention, and implementation characteristics. See Appendix D for the full list of data abstracted from each article.

QUALITY ASSESSMENT

We assessed the quality of studies by their design and the degree to which they reported information about intervention and implementation characteristics (see Table 1).

Table 1. Criteria Used to Assess the Quality of Included Studies

IT Design
Is it integrated with CPOE?
Does it give real time feedback at point of care?
Does the CDS suggest a recommended course of action?
Intervention Classification
Data Entry Source
Is it automated through EHR (*eg*, only uses data already being entered for clinical care)?
Does clinical staff enter data specifically for intervention?
Implementation Characteristics
Was it pilot tested or used an iterative process of development/ implementation?
Was there any user training/ clinician education?
Are the authors also the developers and part of the user group for the CDS?
Was there use of audit-and-feedback (or other internal incentive)?
Are there any other implementation components not already discussed?

DATA SYNTHESIS

We constructed evidence tables showing the study characteristics and results for all included studies. We used as the primary outcome the effect of the intervention on the appropriateness outcome. As a secondary outcome, we used the effect on the utilization. For studies presenting count data or for which a count could be calculated (from a percentage), an odds ratio and associated standard error were calculated. For comparability, the log odds ratios and their standard errors were converted into Cohen's d effect sizes.[14] For studies presenting means and measures of variation, Cohen's d effect sizes were calculated directly. For each study, we used the difference between the pre-intervention period and the post-intervention period, or the difference between the time series projection of performance in the absence of the intervention compared to actual performance during the intervention, or the difference between providers randomized to the intervention or control, as appropriate to the study design and the available data. Results were converted to effect sizes for the analysis. Random effects meta-analyses were conducted using the H-K variance estimator.

After collecting data on the interventions, implementations, and settings but prior to extraction of outcomes data, we developed 4 hypotheses regarding effectiveness of the intervention, one in each category of intervention characteristics, settings, implementation, and target:

1. Interventions will vary in their effectiveness according to the following rank order: (a) interventions that present only information, (b) interventions that include a pop-up or reminder that the selected radiographic examination does not meet current guidelines, (c) interventions that require an active override for providers to continue to order a radiographic examination not supported by guidelines (*ie*, "soft stop"), or (d) interventions that forbid providers from ordering a radiographic examination not supported by guidelines unless/until consultation with a peer or expert (*ie*, "hard stop").

2. Interventions will be more effective in settings that are integrated networks of care (*eg*, VA, Kaiser) than in other settings.

3. Interventions will be more effective if they include other implementation components, such as audit and feedback, academic detailing, etcetera.

4. Interventions may vary by the radiographic modality they target.

RATING THE BODY OF EVIDENCE

The evidence was assessed using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) criteria, which uses the domains of study design limitations, inconsistency, indirectness, and imprecision in results.[15] The GRADE Working Group classified the quality of evidence across outcomes according to the following criteria:

- High = Further research is very unlikely to change our confidence on the estimate of effect.

- Moderate = Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

- Low = Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
- Very Low = Any estimate of effect is very uncertain.

PEER REVIEW

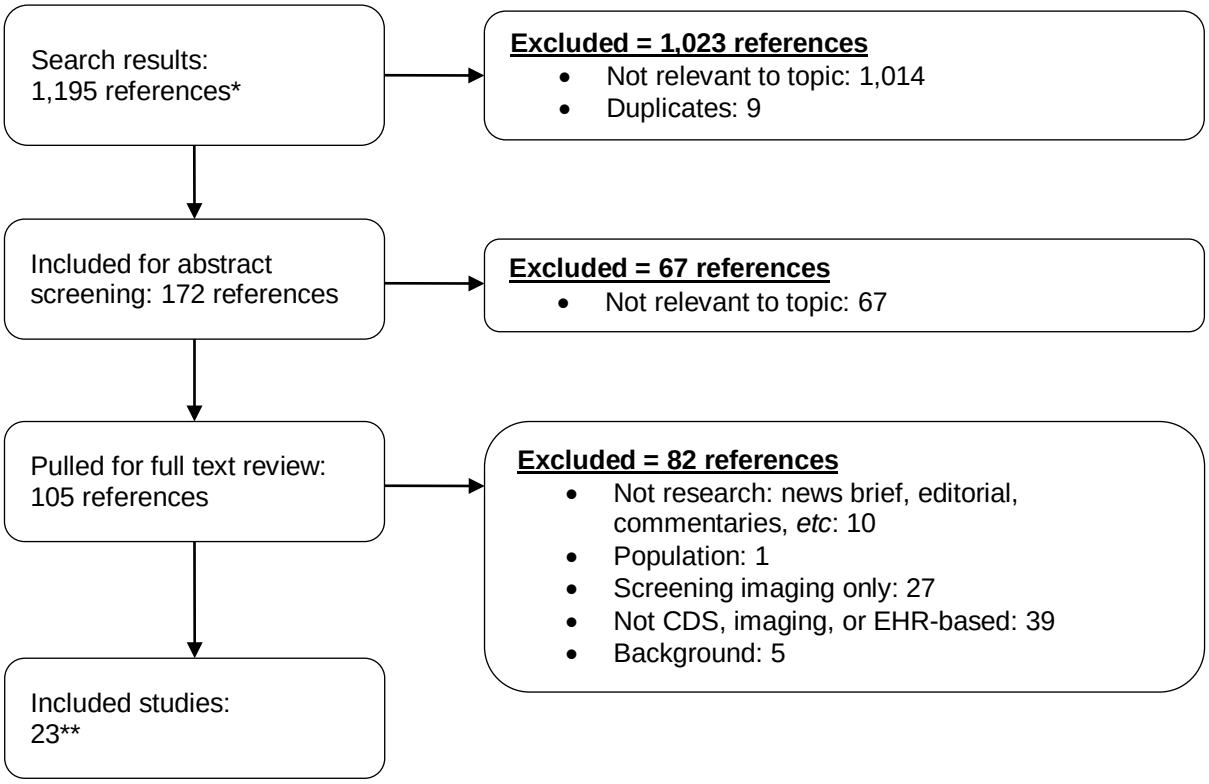
A draft version of this report was reviewed by 2 technical experts and 2 members of VA clinical leadership. Reviewer comments were addressed and our responses were incorporated into the final report. The complete set of comments and responses can be found in Appendix E.

RESULTS

LITERATURE FLOW

From all sources, we retrieved 1,195 titles. From these, we identified 172 titles as being potentially relevant. After reviewing these 172 abstracts, we identified 105 titles for full text review. Of these, we rejected a total of 82 articles, with 10 rejected for their study design (*eg*, a commentary, editorial, review, *etc*), 27 rejected as being about radiology imaging for screening, one rejected for being in a child-aged population, and 39 rejected as not being EHR-based or not about CDS or not about imaging (see Figure 1 for literature flow details).

Figure 1. Literature Flow Chart



* Results from prior systematic reviews (N=152), the existing database (N=226), the update searches (N=793), and articles identified during reference mining (N=24) were deduplicated to reach this number.
** Manuscript reference list includes additional references cited for background and methods plus websites relevant to key questions.

Of the 23 articles included, 3 were randomized trials,[16-18] 7 were time series studies,[19-25] and 13 were pre-post studies.[26-38] Seven studies collected data prior to the year 2000,[16-18,20,21,27,36] and 7 studies included data collection within the past 5 years (2009 or later).[19,22,29-31,33,38] Ten interventions targeted what was sometimes called "high cost imaging," which usually included CT and MRI and occasionally nuclear medicine tests as well.[16,19,22,24,25,31,33-35,37,38] Four interventions targeted pulmonary CT angiography,[23,28,32,38] Two studies targeted chest x-ray,[17,30] 4 interventions targeted multiple radiologic investigations,[20,26,29,36] and 3 studies had other radiologic targets.[18,21,27]

Studies were mostly single institution implementations, and at US academic medical centers. Thirteen studies, or just over half of all studies, were from the "HIT leaders," meaning US academic centers who have a long history of development, implementation, evaluation, and publication of health IT.[16-18,22-25,31,33,36-38] Ten of these studies (35%) came from institutions affiliated with Harvard.[16,18,22-25,31,33,37,38] Three studies came from integrated healthcare delivery organizations, in particular Kaiser,[21] VA,[32] and Virginia Mason.[19]

In 5 studies, the intervention consisted of simply the display of information, such as the cost of tests, relevant guidelines, or an appropriateness rating for the requested radiology examination for that indication.[16,17,21,29,36] Nine studies displayed patient-specific information about whether or not the requested study was consistent with existing guidelines (or something similar).[18,23,25,27,28,30,31,35,38] Four studies included what we characterized as a "soft stop," meaning for radiology orders that the CDS rated as inconsistent with guidelines or inappropriate, the provider needed to enter a reason why the CDS advice was being over-ridden.[20,24,26,34] Five studies included a "hard stop," meaning providers were prevented from ordering radiology examinations the CDS classified as inappropriate without getting approval from some external person, usually a radiologist or senior clinician.[19,22,32,33,37]

Almost all interventions were integrated with CPOE, gave real-time feedback, and a recommended course of action. Only one study specifically indicated that the intervention was developed iteratively or pilot tested,[18] about one-third of studies reported clinician training was part of the implementation process,[19,23,26-28] 5 studies reported including audit-and-feedback as part of the implementation,[19,22,24,26,33] and 7 studies reported on other implementation characteristics, barriers, and facilitators (including having a culture of evidence-based medicine, a phased implementation, salaried physicians, and a risk contract with the payor of care).[19,22,25,26,28,33,35] The Evidence Tables present details of all 23 included studies (see Appendix A).

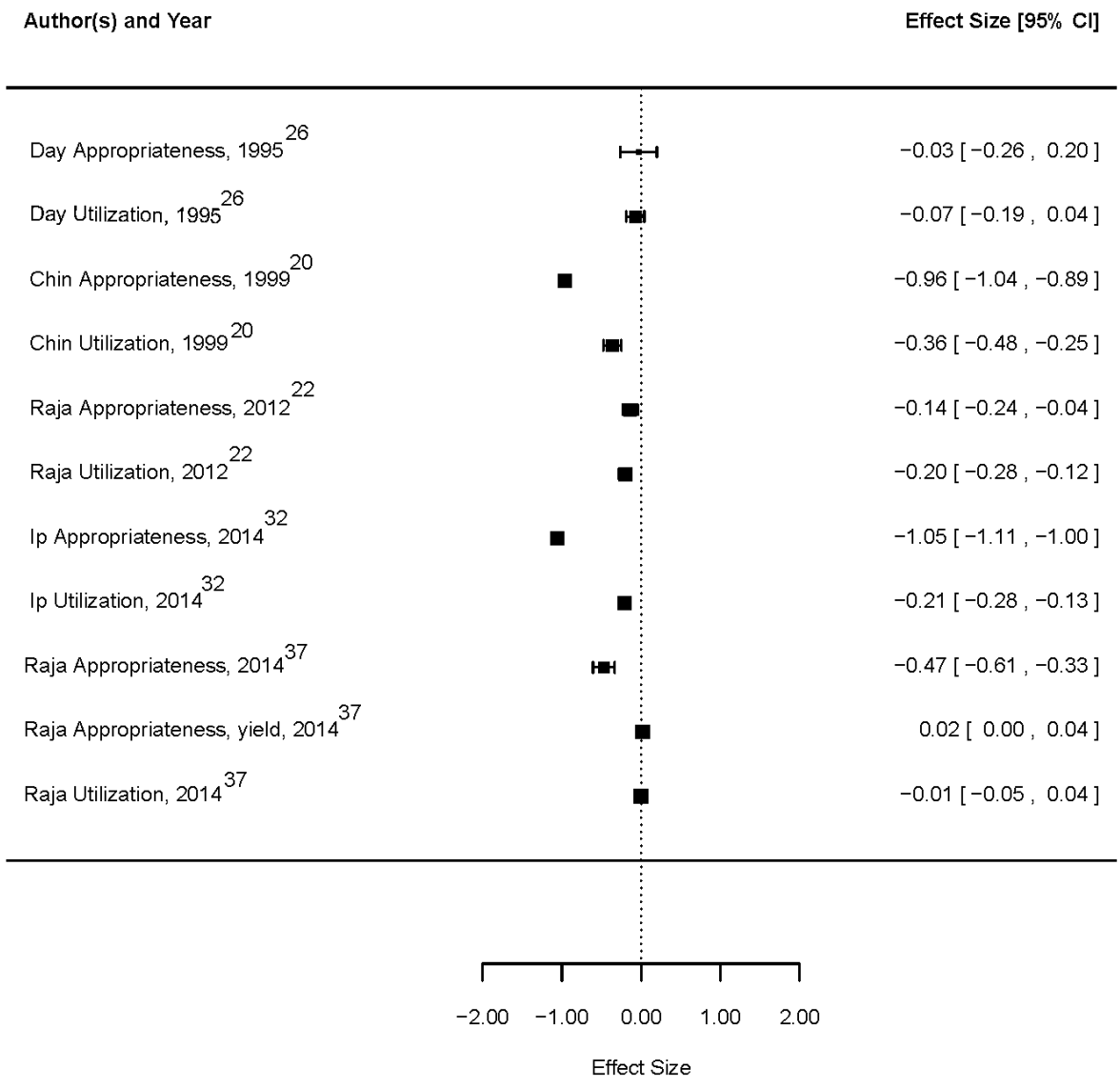
Two studies did not present data sufficient to include in our quantitative analysis, one because it did not present comparative data without the intervention[26] and one because the outcome was an aggregate measure of many tests, and data specific to the radiology targets were not presented.[36] Of the remaining 21 studies, 13 reported an appropriateness outcome and 13 reported a utilization outcome, while 5 studies reported both. See Table 2 for which outcomes were reported in which studies.

Table 2. Articles with Type of Outcome Presented

Author, Year	Appropriateness	Utilization
Bates, 1997[16]		X
Blackmore, 2011[19]		X
Carton, 2002[20]	X	
Chin, 1999[21]	X	X
Day, 1995[27]	X	X
Dresher, 2011[28]	X	
Durand, 2013[29]		X
Flamm, 2013[30]	X	
Gupta, 2014[31]	X	
Harpole, 1997[18]		X
Ip, 2014[33]	X	X
Ip, 2013[22]		X
Raja, 2012[23]	X	X
Raja, 2014[38]	X	X
Rosenthal, 2006[24]	X	
Sanders, 2001[34]		X
Sistrom, 2009[25]		X
Solberg, 2010[35]	X	
Soo Hoo, 2011[32]	X	
Tierney, 1988[17]		X
Vartanians, 2010[37]	X	

Figure 2 presents the results from the 5 studies reporting both appropriateness and utilization outcomes. While the 2 outcomes are directionally consistent within study, in 3 studies the effect size for appropriateness was much greater than for utilization. In general, this is to be expected. Most radiology examinations are not ordered for inappropriate reasons, and thus an intervention targeted at reducing inappropriate test ordering will have a larger effect on appropriateness than it will on utilization. For example, if a particular radiological procedure is ordered appropriately 70% of the time and inappropriately 30% of the time among 100 consecutive orders, then an intervention that successfully reduces inappropriate use by 50% will result in only 15 inappropriate test orders, while utilization will decrease only 15% (from 100 test orders to 85 test orders, the latter consisting of 70 appropriate orders and 15 inappropriate orders). Thus, we expect that utilization will be less affected than appropriateness for these interventions. The study by Raja and colleagues published in 2014 reported both at an appropriateness outcome, as in the proportion of radiologic tests ordered that met a national standard and the patient positive yield of radiologic examinations. We included both here for comparison.

Figure 2. Studies Reporting Both Appropriateness and Utilization Outcomes



ES=effect size, CI=confidence interval

KEY QUESTION 1: What is the effectiveness of EHR-based interventions in reducing unnecessary or inappropriate imaging?

KEY QUESTION 2: Do EHR-based interventions vary in results by system?

We combine our presentation of results for these 2 key questions, since they are interrelated.

Thirteen studies contributed to each pooled analysis, one pooled analysis for appropriateness and one pooled analysis for utilization. Five studies contributed data to both. Our primary outcome was the effect on appropriateness.

Figure 3 displays the results for individual studies reporting appropriateness. Nine of the 13 studies reported statistically significant benefits of the intervention, 2 reported a benefit that was not statistically significant, and 2 studies reported no effect. The random effects pooled estimate from all 13 studies was an effect size of 0.48 (95% CI: -0.71, -0.25). This equates to a "moderate" sized effect, according to a conventional classification.[39] Substantial heterogeneity is present, as indicated by the I^2 statistic of 99.5%, and the visual inspection of the plot. Neither Begg's nor Egger's test indicated the presence of publication bias (p=0.951, p=0.339, respectively). A sensitivity analysis that included the appropriateness outcome instead of the yield outcome for Raja (2014) yielded little difference in the pooled estimate (0.52, 95% CI -0.73, -0.31).[38]

As a clinical example of what constitutes a moderate-sized effect, consider the results of the study by Gupta and colleagues, which reported an effect size of 0.67 associated with implementation of the CDS intervention. In clinical terms, this meant that before CDS implementation the percentage of head CT examinations ordered for appropriate reasons in the ER for patients with mild traumatic brain injury was 49%. After implementation of the CDS, this rate increased to 76.5%.[31] As a second example, the study by Rosenthal and colleagues reported an effect size of 0.63 associated with implementation of a CDS, giving appropriateness scores for a number of radiologic procedures. Before the intervention, 6% of procedures were judged as being of low utility; after the intervention this value dropped to 2%.[24]

Figure 3. Results for Appropriateness from Individual Studies

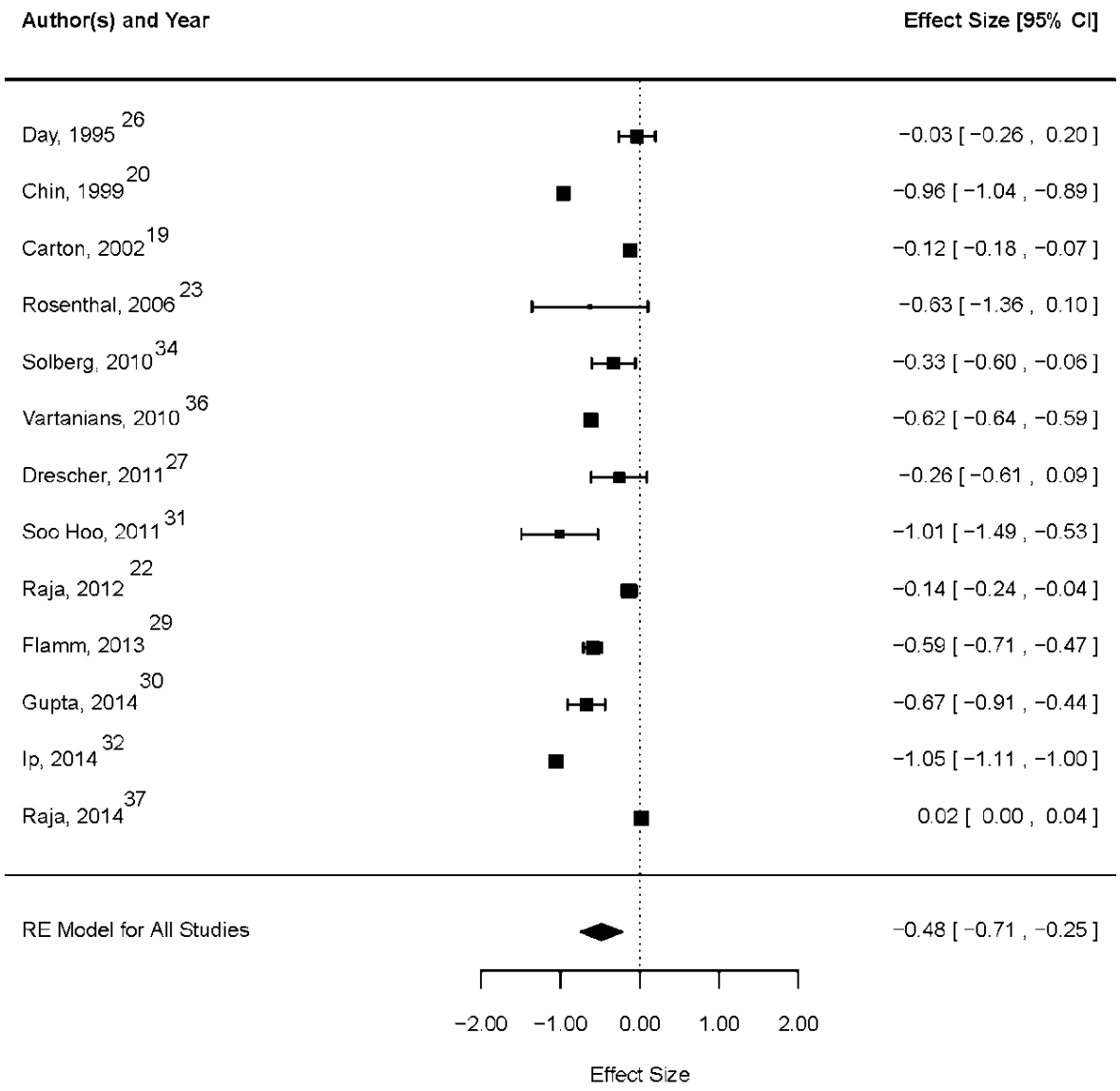
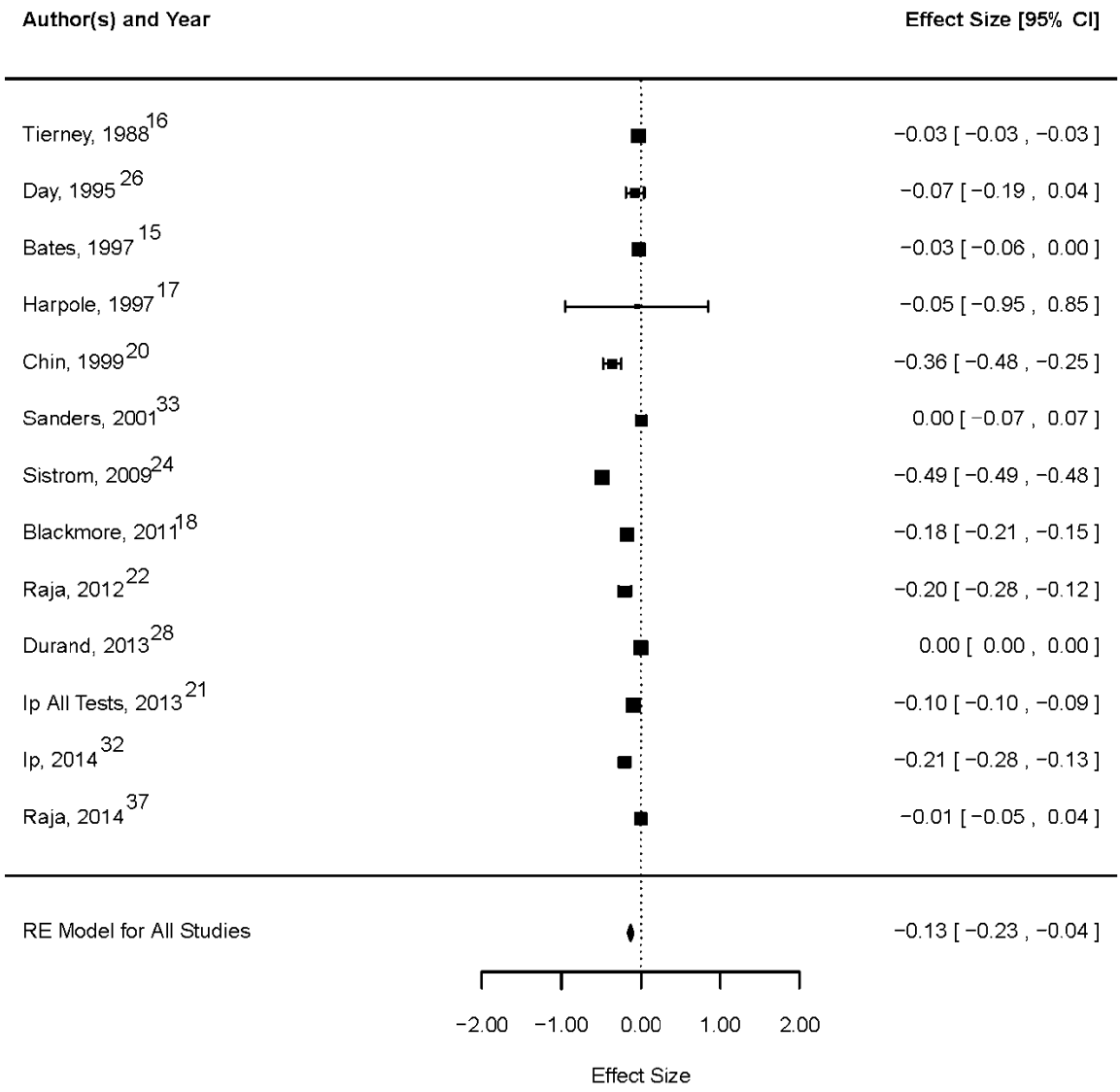


Figure 4 displays the results for the 13 studies reporting utilization outcomes. Six studies reported statistically significant benefits of the intervention, and 7 studies reported essentially no effect. The random effects pooled estimate from all 13 studies was an effect size of 0.13 (95% CI: -0.23, -0.04). This equates to a "small" sized effect, according to a conventional classification. Despite a narrower range of effect sizes in these 13 studies (10 of the 13 studies cluster between zero and an effect size of 0.21) compared to the range of effect sizes for the appropriateness outcomes, like the appropriateness studies substantial heterogeneity is present, as indicated by the I2 statistic of 100%. Neither Begg's nor Egger's test indicated the presence of publication bias (p=0.392, p=0.259, respectively).

Figure 4. Results for Utilization from Individual Studies



Effect of Intervention Characteristics, Setting, Implementation, and Target on Effectiveness

We explored 4 hypotheses regarding effectiveness, one each for characteristics of the intervention, the setting (integrated care delivery versus other settings), the implementation process (the use of audit and feedback was the only implementation characteristic with sufficient data to support a stratified analysis), and the radiologic target of the intervention. We had insufficient studies to support robust pooled estimates of individual strata or to support multivariable analyses. Furthermore, since appropriateness was our primary outcome we present stratified results only for this outcome. Nevertheless, some patterns are apparent.

Figure 5 presents the results of the appropriateness outcomes, stratified by intervention characteristics and settings. “A” interventions provided information only; in the included study it presented internally developed guidelines any time an upper GI series order was placed.[21] “B”

interventions presented information on appropriateness or guidelines specifically tailored to the individual patient, often as a pop-up or alert. Some of these interventions also recommended alternative interventions, but did not include any barrier for the clinician to order the test. "C" interventions in general were similar to "B" interventions, but required the ordering clinician to justify with free text why they were overriding the decision support recommendation that a study was inappropriate. We called this a "soft stop." "D" interventions included what we called a "hard stop," meaning the intervention prevented the clinician from ordering a test contrary to the CDS determination of inappropriateness, until additional discussion with or permission obtained from another clinician or radiologist.

All of the "D" studies reported moderate-to-large effects on appropriateness, while the "B" and the "C" studies reported more variable and generally lesser effects. The one "A" study, by Chin and colleagues,[21] reported a large effect. This study displayed locally developed guidelines for appropriate upper gastro intestinal (UGI) ordering at the time an order was placed. This study was conducted in an integrated care delivery setting with a high baseline rate of inappropriate use (45%), which may partly explain these strikingly successful findings.

Regarding setting, the 2 studies conducted in integrated care settings all reported large effects on appropriateness. Studies conducted in other settings found much smaller effects, while studies at 10 US institutions that are leaders in health IT reported heterogeneous results, although 4 of the 6 studies of appropriateness at the health IT leaders reported statistically significant benefits of the intervention.

Figure 5. Appropriateness Results Stratified by Intervention Characteristics and by Setting

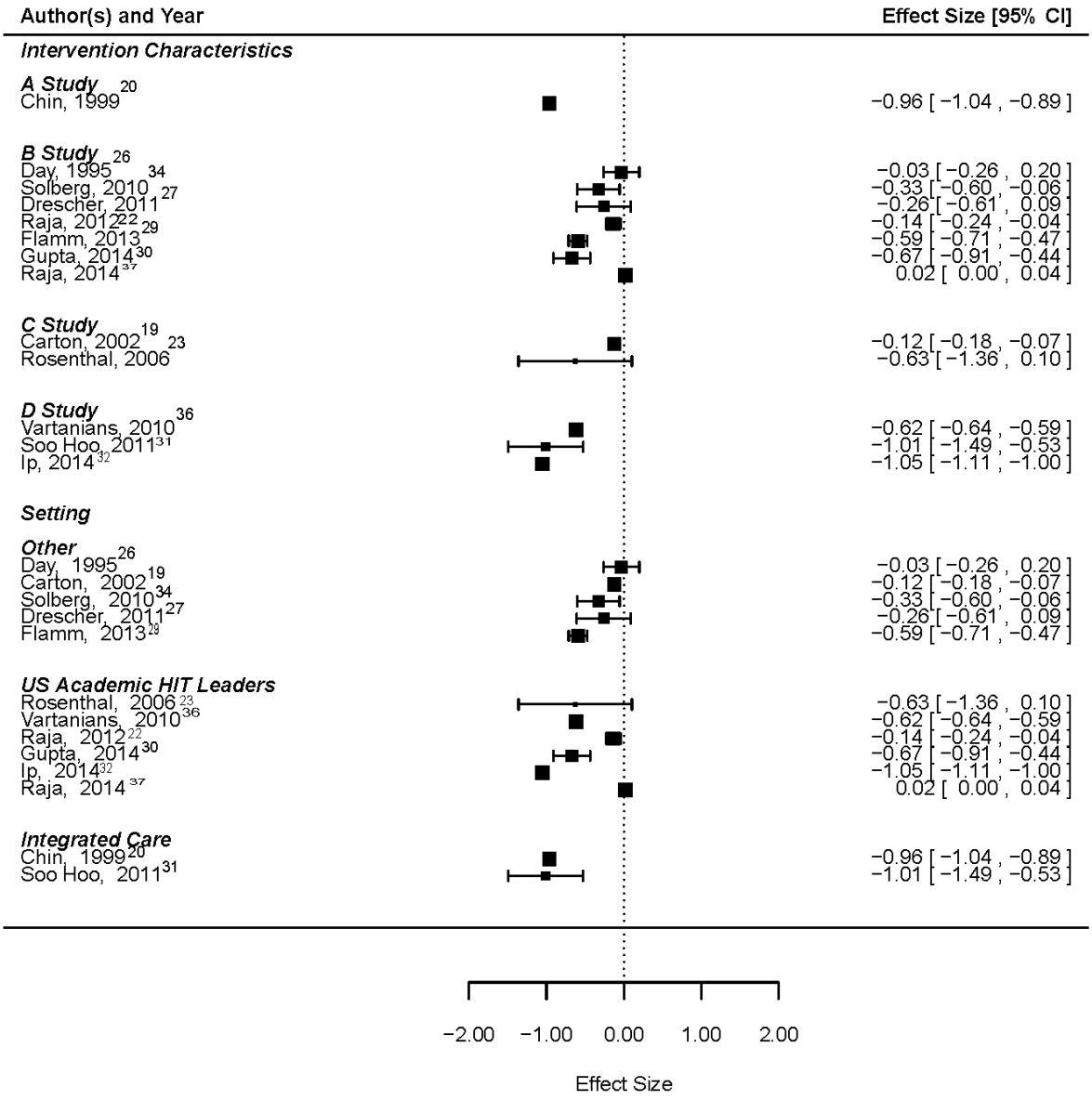
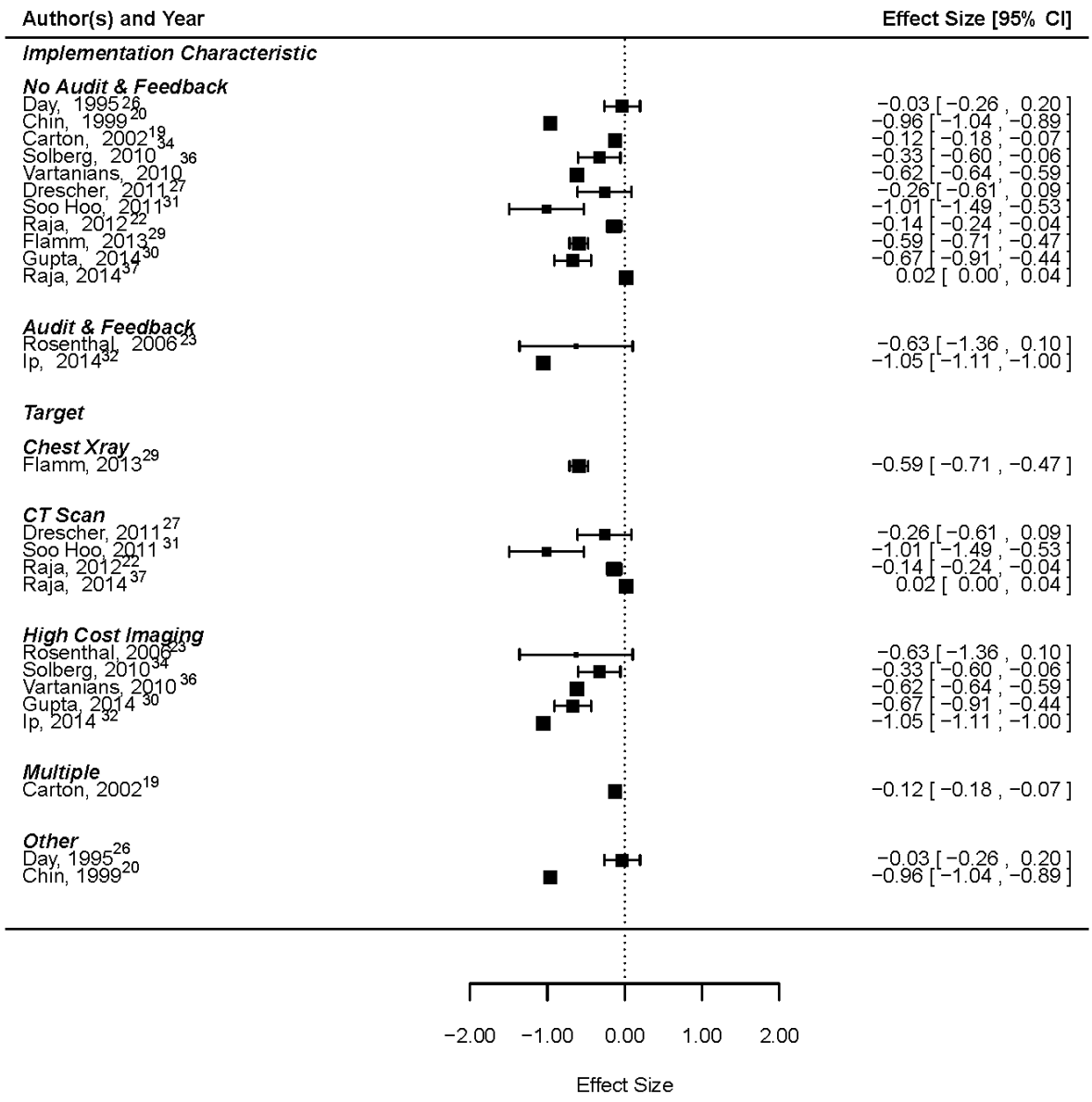


Figure 6 presents the results of the appropriateness, respectively, stratified by whether or not the publication reported that audit-and-feedback was part of the implementation process and by the target of the intervention. There did not seem to be clear patterns, although both studies using audit-and-feedback reported moderate-to-large effects. In the study by Rosenthal and colleagues, "physicians, who [ordered] more than a few examinations with low utility scores [were] contacted by one of their senior clinical leaders and counseled about appropriate ordering."[24] In the study by Ip and colleagues, audit-and-feedback was accomplished by sending to primary care providers quarterly practice pattern variation reports depicting the number of back MRIs ordered per the number of low back pain related visits.[33]

No pattern of differential effectiveness is as clearly apparent as in the preceding figures, although in a study of one intervention a statistically significant effect was seen for CT scans and nuclear radiology tests but not MRI,[22] suggesting the possibility of a differential effect of interventions based on their target.

Figure 6. Appropriateness Results Stratified by Implementation Characteristics and by Target



Two studies which otherwise met our inclusion criteria could not be included in the pooled analysis because they did not present sufficient data. In one study a rural family medicine clinic in Canada made computerized decision support available to any interested physician. The majority of physicians were infrequent users of the system, which was voluntary. Among 904 diagnostic imaging studies ordered using the system, clinical guidelines applied to 58%. Of these, 29% were identified as inappropriate and an alternative diagnostic strategy suggested. Of these, physicians followed the suggestion in 25% of cases. This study could not be included in our pooled analysis because it did not present data from the pre-intervention period.[26] The second study was a time series study of displaying the costs for a large number of outpatient diagnostic tests (urinalysis, complete blood count, serum electrolytes, *etc*).[36] Three radiologic tests relevant to this review were included: chest x-ray, head CT scan, and head MRI. This study could not be included in our pooled analysis because it did not report results separately for the individual tests. The effect of the intervention on the aggregate of all these diagnostic tests was a statistically significant 14% reduction in utilization. This significant finding is in contrast to 3 other studies of an intervention that presented cost data, all of which reported no effect on utilization.

VA Studies

We identified one study conducted in a VA setting.[32] This was a single site, pre-post study at an urban, academically affiliated VA. Locally developed guidelines were developed for the appropriate use of CT angiography for suspected pulonary embolus. These were then incorporated in the CPOE fuction of the VA electronic health record (CPRS version 1.0.27). The order entry menu screen included the Wells criteria point score. The ordering physician was required to complete the checklist to generate a Wells score. If the score was 4 or less, the physician was required to get a D-dimer test. If the D-dimer test was above a certain threshold, the order was allowed to proceed without any additional barriers. If the D-dimer was below that threshold, the ordering physician was required to consult with the on-call chest radiology attending for approval (a “hard stop”). In the 2 years prior to the intevention, the yield of positive CT angiography studies was 3.1%. This increased to 16.5% after inplementation of the intervention. There was no assessment of possible patient harms or physician reactions to the use of the intervention.

Summary of Findings and Quality of Evidence for Key Question 1 and 2

Twenty-one studies provide moderate quality evidence that EHR-based interventions can reduce inappropriate test ordering by a moderate amount, and reduce overall utilization by a small amount. Low quality evidence supports that interventions that include a “hard stop,” preventing ordering clinicians from overriding a decision support determination that a test is inappropriate, and implementation in an integrated care delivery setting, are associated with greater effectiveness. Audit-and-feedback may be a useful implementation tool, but data are too sparse to draw conclusions. We judged the quality of evidence regarding appropriateness and utilization as moderate, due to heterogeneity in the results. We judged the quality of evidence regarding the characteristics as low, due to the sparseness of the data and indirect nature of the comparisons. That is, these characteristics have not been tested as a priori hypotheses for differential effectiveness within the same study.

KEY QUESTION 3: What are the harms or potential harms associated with EHR-based interventions used to reduce inappropriate imaging?

Four studies reported on harms associated with their interventions.[18,26,28,30] One study evaluating a decision support tool to reduce unnecessary pre-operative testing found that with the intervention there was an increase in the percent of pre-operative chest x-rays inappropriately not ordered.[30] Prior to the intervention 1.9% of patients did not get a chest x-ray when indicated compared to 9.3% after the intervention. The clinical impact of this is not known. Another study of a decision support tool to reduce abdominal kidney, ureter, bladder (KUB) x-rays identified 12 KUB studies out of a total of 255 performed against the advice of the tool where there were positive findings.[18] Of these 12, six KUB studies were felt to have significantly influenced patient outcomes, making it unclear whether following the locally developed guidance could have endangered the patient. The 2 other studies reported on qualitative information from physician surveys which primarily identified lack of interest in using the decision support tools because of time constraints and perceived inefficiencies.[26,28]

Summary of Findings

There are few data on the potential harms of decision support tools to reduce inappropriate radiology test ordering. Future studies should evaluate for harms – particularly investigating whether guidelines when applied in practice provide unanticipated results, or when there are issues related to workflow, efficiency, or provider dissatisfaction that could impact a decision support tool's effectiveness. For example in a study of CDS to prevent drug-drug interactions, the use of a "hard stop" intervention – while effective in changing prescribing – resulted in delays in treatment for 4 patients, resulting in preventative stopping of the study by the Institutional Review Board.[1] Another study, excluded from our review because it assessed a pediatric population, surveyed physicians and found that most felt the CDS was "a nuisance" and "not relevant to the complex or high risk patients they had to treat."[2] This highlights the need for assessment of harms and unintended effects in every evaluation.

Quality of Evidence for Key Question 3

We judged the quality of evidence for harms as very low, meaning any estimate is uncertain.

SUMMARY AND DISCUSSION

SUMMARY OF EVIDENCE BY KEY QUESTION

Key Question 1: What is the effectiveness of EHR-based interventions in reducing unnecessary or inappropriate imaging?

Key Question 2: Do EHR-based interventions vary in results by system?

Twenty-one studies provide moderate quality evidence that EHR-based interventions can reduce inappropriate test ordering by a moderate amount, and reduce overall utilization by a small amount. Low-quality evidence supports that interventions that include a "hard stop," preventing ordering clinicians from overriding a decision support determination that a test is inappropriate, the use of audit-and-feedback as part of the implementation, and that are conducted in an integrated care delivery setting, are associated with greater effectiveness.

Key Question 3

There are few data on the potential harms of decision support tools to reduce inappropriate radiology test ordering. Future studies should evaluate for harms – particularly investigating whether guidelines when applied in practice provide unanticipated results, or when there are issues related to workflow, efficiency, or provider dissatisfaction that could impact a decision support tool's effectiveness. For example, in a study of CDS to prevent drug-drug interactions, the use of a "hard stop" intervention – while effective in changing prescribing – resulted in delays in treatment for 4 patients, resulting in preventative stopping of the study by the Institutional Review Board.[1] Another study, excluded from our review because it assessed a pediatric population, surveyed physicians and found that most felt the CDS was "a nuisance" and "not relevant to the complex or high risk patients they had to treat."[2] This highlights the need for assessment of harms and unintended effects in every evaluation.

LIMITATIONS

Publication Bias

The most important limitation to this review is the likely existence of publication bias. Although we did not detect any statistical evidence of publication bias, there surely must be more implementations of EHR-based interventions to improve appropriateness of radiology test ordering then the 22 published studies we found. Our expectation is that many such interventions are done and never formally evaluated or published. How the results of these implementations may differ from the published studies is unknown, but we expect both effective and ineffective implementations have likely occurred and not been published. This lack of publication is a major impediment toward more rapid learning of how health IT can best be implemented.

Study Quality

In common with many other areas of health IT, we found key information on context and implementation to be lacking in published studies. For example, only one study reported on pilot testing of the intervention, and only about one-third of studies reported on clinician training as part of the implementation. This may perpetuate the belief that these kinds of health IT interventions can be developed separate from the workflow of practicing clinicians, and then

simply “turned on” with the expectation that clinicians will know how to use the intervention and use it correctly. The dearth of reporting of possible harms is another key limitation. Every intervention has to be expected to potentially cause harms and these need to be explicitly measured.

Heterogeneity

Heterogeneity in effectiveness is a prominent finding of our review. Our stratified analyses explained some of this heterogeneity, but the greater portion remains unexplained. It has been postulated that the majority of heterogeneity in health IT evaluations is due to details of the context and implementation that go unreported in published study. We expect the same to be true here. Heterogeneity is a primary contributor to low quality of evidence assessments.

Applicability of Findings to the VA Population

Only 3 studies were performed in integrated care delivery settings, and only one study was performed in VA. However, there is evidence suggesting that interventions implemented in integrated care delivery settings may be more effective than in other settings, indicating VA may realize benefits equal to or greater than the average benefit reported here.

RESEARCH GAPS/FUTURE RESEARCH

We identified the following research gaps:

Direct comparisons are needed of different intervention characteristics. We found suggestive evidence that interventions with a “hard stop” are more effective than other interventions, but to prove this hypothesis requires testing the 2 methods head-to-head in the same study. This should be easy to do, since randomization can occur at the provider level, and consist of the CDS with and without the hard stop.

More research is needed on possible harms. Harms of a CDS intervention with a “hard stop” have been reported in other clinical situations. An explicit assessment of harms should be incorporated into every study of interventions.

About half of our included studies collected data more than 7 years ago, and information technology and attitudes about the use of information technology change over time, so data from more recent time periods would be helpful.

One study reported differential effectiveness by target, and this should be assessed in future studies.

Like all health IT evaluations, more information about context and implementation is needed.

CONCLUSIONS

Computerized decision support integrated with the electronic health record can reduce inappropriate use of diagnostic radiology testing by a moderate amount. The use of a “hard stop” as part of the intervention and use in an integrated care delivery setting may all increase effectiveness. There are few data on the potential harms of decision support tools to reduce inappropriate radiology test ordering. Future studies should evaluate for harms.

REFERENCES

1. Strom BL, Schinnar R, Aberra F, et al. Unintended effects of a computerized physician order entry nearly hard-stop alert to prevent a drug interaction: a randomized controlled trial. *Archives of Internal Medicine.* 2010;170(17):1578-1583.

2. Bowen S, Johnson K, Reed MH, Zhang LP, Curry L. The Effect Of Incorporating Guidelines Into a Computerized Order Entry System for Diagnostic Imaging. *Journal of the American College of Radiology.* Apr 2011;8(4):251-258.

3. Kocher KE, Meurer WJ, Fazel R, Scott PA, Krumholz HM, Nallamothu BK. National trends in use of computed tomography in the emergency department. *Ann. Emerg. Med.* 2011;58(5):452-462. e453.

4. Korley FK, Pham JC, Kirsch TD. Use of advanced radiology during visits to US emergency departments for injury-related conditions, 1998-2007. *JAMA.* 2010;304(13):1465-1471.

5. Chou R, Fu R, Carrino JA, Deyo RA. Imaging strategies for low-back pain: systematic review and meta-analysis. *The Lancet.* 2009;373(9662):463-472.

6. Cassel CK, Guest JA. Choosing wisely: helping physicians and patients make smart decisions about their care. *JAMA.* 2012;307(17):1801-1802.

7. Chaudhry B, Wang J, Wu S, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med.* May 16 2006;144(10):742-752.

8. Buntin MB, Burke MF, Hoaglin MC, Blumenthal D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. *Health Affairs.* 2011;30(3):464-471.

9. Goldzweig C, Orshansky G, Paige N, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. *Annals of Internal Medicine.* 2013;159(10):677.

10. Jones SS, Rudin RS, Perry T, Shekelle PG. Health information technology: an updated systematic review with a focus on meaningful use. *Annals of Internal Medicine.* 2014;160(1):48-54-54.

11. Georgiou A, Prgomet M, Markewycz A, Adams E, Westbrook JI. The impact of computerized provider order entry systems on medical-imaging services: a systematic review. *Journal of the American Medical Informatics Association.* 2011;18(3):335-340.

12. Roshanov PS, Misra S, Gerstein HC, et al. Computerized clinical decision support systems for chronic disease management: a decision-maker-researcher partnership systematic review. *Implement Sci.* 2011;6(1):92.

13. Fillmore CL, Bray BE, Kawamoto K. Systematic review of clinical decision support interventions with potential for inpatient cost reduction. *BMC Med. Inform. Decis. Mak.* Dec 2013;13.

14. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Chapter 7. *Introduction to meta-analysis*: John Wiley & Sons; 2011.

15. Grading of Recommendations Assessment DaEGWG. GRADE working group. 2013; http://www.gradeworkinggroup.org/index.htm. Accessed 9/5/2014.

16. Bates DW, Kuperman GJ, Jha A, et al. Does the computerized display of charges affect inpatient ancillary test utilization? *Arch Intern Med.* Nov 24 1997;157(21):2501-2508.

17. Tierney WM, McDonald CJ, Hui SL, Martin DK. Computer predictions of abnormal test results: effects on outpatient testing. *JAMA.* 1988;259(8):1194-1198.

18. Harpole LH, Khorasani R, Fiskio J, Kuperman GJ, Bates DW. Automated evidence-based critiquing of orders for abdominal radiographs: impact on utilization and appropriateness. *J Am Med Inform Assoc.* Nov-Dec 1997;4(6):511-521.

19. Blackmore CC, Mecklenburg RS, Kaplan GS. Effectiveness of clinical decision support in controlling inappropriate imaging. *J Am Coll Radiol.* Jan 2011;8(1):19-25.

20. Carton M, Auvert B, Guerini H, et al. Assessment of radiological referral practice and effect of computer-based guidelines on radiological requests in two emergency departments. *Clinical radiology.* 2002;57(2):123-128.

21. Chin HL, Wallace P. Embedding guidelines into direct physician order entry: simple methods, powerful results. Paper presented at: Proceedings of the AMIA Symposium1999.

22. Ip IK, Schneider L, Seltzer S, et al. Impact of provider-led, technology-enabled radiology management program on imaging. *Am J Med.* Aug 2013;126(8):687-692.

23. Raja AS, Ip IK, Prevedello LM, et al. Effect of computerized clinical decision support on the use and yield of CT pulmonary angiography in the emergency department. *Radiology.* Feb 2012;262(2):468-474.

24. Rosenthal DI, Weilburg JB, Schultz T, et al. Radiology order entry with decision support: initial clinical experience. *Journal of the American College of Radiology.* 2006;3(10):799-806.

25. Sistrom CL, Dang PA, Weilburg JB, Dreyer KJ, Rosenthal DI, Thrall JH. Effect of Computerized Order Entry with Integrated Decision Support on the Growth of Outpatient Procedure Volumes: Seven-year Time Series Analysis 1. *Radiology.* 2009;251(1):147-155.

26. Curry L, Reed MH. Electronic decision support for diagnostic imaging in a primary care setting. *Journal of the American Medical Informatics Association.* May 2011;18(3):267-270.

27. Day F, Hoang LP, Ouk S, Nagda S, Schriger DL. The impact of a guideline-driven computer charting system on the emergency care of patients with acute low back pain. *Proc Annu Symp Comput Appl Med Care.* 1995:576-580.

28. Drescher FS, Chandrika S, Weir ID, et al. Effectiveness and acceptability of a computerized decision support system using modified Wells criteria for evaluation of suspected pulmonary embolism. *Ann. Emerg. Med.* 2011;57(6):613-621.

29. Durand DJ, Feldman LS, Lewin JS, Brotman DJ. Provider Cost Transparency Alone Has No Impact on Inpatient Imaging Utilization. *Journal of the American College of Radiology.* Feb 2013;10(2):108-113.

30. Flamm M, Fritsch G, Hysek M, et al. Quality improvement in preoperative assessment by implementation of an electronic decision support tool. *J Am Med Inform Assoc.* Jun 2013;20(e1):e91-96.

31. Gupta A, Ip IK, Raja AS, Andruchow JE, Sodickson A, Khorasani R. Effect of clinical decision support on documented guideline adherence for head CT in emergency department patients with mild traumatic brain injury. *J Am Med Inform Assoc.* Feb 17 2014.

32. Hoo GW, Wu CC, Vazirani S, Li Z, Barack BM. Does a clinical decision rule using D-dimer level improve the yield of pulmonary CT angiography? *AJR Am J Roentgenol.* May 2011;196(5):1059-1064.

33. Ip IK, Gershanik EF, Schneider LI, et al. Impact of IT-Enabled Intervention on MRI Use for Back Pain. *Am J Med.* Feb 7 2014.

34. Sanders DL, Miller RA. The effects on clinician ordering patterns of a computerized decision support system for neuroradiology imaging studies. *Proc AMIA Symp.* 2001:583-587.

35. Solberg LI, Wei F, Butler JC, Palattao KJ, Vinz CA, Marshall MA. Effects of electronic decision support on high-tech diagnostic imaging orders and patients. *Am J Manag Care.* Feb 2010;16(2):102-106.

36. Tierney WM, Miller ME, McDonald CJ. The effect on test ordering of informing physicians of the charges for outpatient diagnostic tests. *N Engl J Med.* May 24 1990;322(21):1499-1504.

37. Vartanians VM, Sistrom CL, Weilburg JB, Rosenthal DI, Thrall JH. Increasing the Appropriateness of Outpatient Imaging: Effects of a Barrier to Ordering Low-Yield Examinations 1. *Radiology.* 2010;255(3):842-849.

38. Raja AS, Gupta A, Ip IK, Mills AM, Khorasani R. The use of decision support to measure documented adherence to a national imaging quality measure. *Acad Radiol.* Mar 2014;21(3):378-383.

39. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Assoc, Inc; 1988.